A Critical Missing Element: Critical Thinking at Rwanda's Public Universities and the Implications for Higher Education Reform

Rebecca Schendel

Submitted for examination for the degree of PhD in
Education & International Development
July 2013
Institute of Education, London

Abstract

In the years since the genocide, the Government of Rwanda has contributed significant financial resources to the re-establishment and development of its public higher education sector. This investment has largely been justified in terms of the contribution of university graduates to the country's vision of becoming a service-based knowledge economy, capable of reducing its reliance on foreign aid and technical assistance. Implicit in this vision for the future is an assumption that a university education will help students to improve in their ability to think critically about problems and to use evidence when making decisions.

This study empirically investigated this assumption by administering a version of the Collegiate Learning Assessment – a performance-task-based test of critical thinking, adapted for use in Rwanda – to a random sample of 220 students enrolled at three of Rwanda's most prestigious public institutions. Assessment results were supplemented with in-depth case studies at two of the institutions involved in the study. Results of the study suggest that Rwandan students are not significantly improving in their critical thinking ability during their time at university. Critical thinking ability in Rwanda seems to be largely influenced by the academic experiences provided within university Faculties, as the use of innovative classroom practices appears to have a positive impact on the cultivation of critical thinking skills. However, results indicate that such practices cannot be assumed, as faculty motivation and understanding of pedagogical innovations can significantly affect their effective implementation.

The international community has largely focused its higher education reform efforts on improvements in institutional efficiency, but the results of this study indicate that student learning outcomes cannot be ignored. Without similar support for initiatives that seek to improve pedagogy, regional revitalisation efforts are unlikely to have a substantial effect on development objectives.

Word Count Declaration

I confirm that the word length of

1) the thesis, including footnotes, is 97,466 words,

2) the bibliography is 11,001 words,

and,

3) the appendices is 24,915 words.

Signed ..... Date .....
Candidate

Acknowledgements

This study would truly not have been possible without the support, collaboration and, at times, hard work of numerous friends, family members and colleagues in Rwanda, London and the U.S.

I am particularly grateful to colleagues at NUR, KIST and SFB for allowing me to enter your communities and learn about your worlds. Above all, my thanks go to Dr. Jolly Mazimhaka, Dr. Reid Whitlock, Dr. Bonfils Safari and Dr. Chika Ezeanya for all of your support throughout the process. To the study participants at all three universities, I offer my most heartfelt appreciation. This study would never have happened without your assistance.

I am deeply indebted to the Ministry of Education and the Higher Education Council in Kigali for supporting this research and for ensuring that the information is passed along to those who will find it useful. I am also profoundly thankful to Michael and to Charles at Generation Rwanda for facilitating my research funding and making this study possible.

A seemingly endless supply of volunteers helped at every stage of this project - completing translations, reviewing assessment drafts, scoring assessments, and reading proposals. I am particularly grateful to the students from Generation Rwanda, who offered to field-test and pilot all of the tools used in the study, and to Lambert and Abdul for being ever willing to help with all of the logistical challenges.

I am deeply grateful to my supervisors, Andy Tolmie and Moses Oketch, for your belief in me and for all of your assistance along the way.

To my superstar research team – Clarisse, Emile, Patricie, Bienvenu, Paulin, Aimé, Eric, Annet and Jean Baptiste – I thank you from the bottom of my heart for all of your hard work and professionalism at every stage.

And, most of all, my love and thanks go to Drew for inspiring me and making all of this possible, to Amy and Emily for endless hours of emotional support, and to my parents, who not only instilled the value of critical thinking in me at a young age but supported me in every sense at every juncture in this journey. Thank you for always being the foundation.

Table of Contents

List of Tables

List of Figures

List of Acronyms

AAU	Association of African Universities
ANOVA	Analysis of Variance
AVU	African Virtual University
CAAP	Collegiate Assessment of Academic Proficiency
CAE	Council for Aid to Education
CAT	Continuous Assessment Test
CCTST	California Critical Thinking Skills Test
CLA	Collegiate Learning Assessment
DFID	Department for International Development (Great Britain)
DHS	Demographic Health Survey
DIF	Differential Item Functioning
EAC	East African Community
FAED	Faculty of Architecture & Environmental Design, KIST
FAS	Faculty of Applied Sciences, NUR; Faculty of Applied Sciences, KIST
FEM	Faculty of Economics & Management, NUR
FOE	Faculty of Engineering, KIST
FOS	Faculty of Science, NUR
GATS	General Agreement on Trade in Services
HEC	Higher Education Council (Rwanda)
HERANA	Higher Education Research and Advocacy Network in Africa
IMF	International Monetary Fund
IT	Information Technology
KHI	Kigali Health Institute
KIE	Kigali Institute of Education
KIST	Kigali Institute of Science & Technology
MANOVA	Multivariate Analysis of Variance
MAPP	Measure of Academic Proficiency & Progress
MCAR	Missing Completely at Random
MDGs	Millennium Development Goals
MINEDUC	Ministry of Education, Rwanda
M.P.	Member of Parliament
NHEC	National Higher Education Council (Rwanda)
NSSE	National Survey of Student Engagement
NUR	National University of Rwanda
OECD	Organisation for Economic Co-operation & Development
PCA	Principal Components Analysis
PHEA	Partnership for Higher Education in Africa
PISA	Programme for International Student Assessment
RPF	Rwandan Patriotic Front
SAT	Scholastic Aptitude Test
SES	Socio-Economic Status
SFB	School of Finance and Banking
U.K.	United Kingdom
UNESCO	United Nations Educational, Scientific and Cultural Organisation
UNR	Université Nationale du Rwanda
U.S.	United States
USAID	United States Agency for International Development
ZPD	Zone of Proximal Development

Chapter 1: Research Background

Since the independence struggles of the 1950s and 1960s, universities have played a key symbolic function across Africa[1]. In the final years of the colonial era, the demand for higher education was inextricably linked with nationalist platforms. Universities were seen as critical components of the nation-building process, as they represented the power of education to unify a population (Assié-Lumumba & CODESRIA, 2006) and symbolised the potential for equality with the West (Coleman & Court, 1993). Alongside the national airline, postal service and banking system, the national university was perceived to be an important symbol of legal and institutional independence from the colonisers (Ajayi *et al*, 1996; Lulat, 2005).

Despite this early emphasis on the importance of higher education, contemporary African universities have a dubious international reputation. Narratives of declining quality dominate the literature, presenting a portrait of a sector that has struggled for years to maintain standards in the face of overwhelming challenges. However, with the rise of the global 'knowledge economy', higher education has assumed a position of renewed importance across the continent, prompting the international community to focus some of its attention on revitalising the region's universities.

This study aims to contribute to this ongoing revitalisation debate by suggesting a new methodology for assessing academic quality at Africa's universities. In this introductory chapter, a rationale is presented for a study of this nature. The chapter begins with a discussion of how current reform efforts are attempting to address problems of quality in the region and concludes with an examination of how such a study might start to fill existing gaps in the reform agenda.

1.1 Narratives of Declining Quality

The body of literature addressing the issue of quality in African higher education generally focuses on historical explanations for why universities have declined in prominence and effectiveness in the years since independence. Scholars in this domain tend to emphasise one of two major themes within this historical narrative: the impact of

[1] Although Africa is a diverse continent, much of the existing literature on institutional quality discusses African higher education in the aggregate. Reform efforts have also tended to be largely uniform across the continent, despite diversity in national contexts. As a result, the term 'African higher education' has been used when framing the wider debates in the field. The use of this term is not intended to imply that the higher education systems in all African countries have experienced identical political histories or economic circumstances or that contemporary African institutions face identical barriers to quality.

political influences on the functioning of universities, and the financial crisis that crippled many university budgets in the 1970s and 1980s. Both perspectives offer important insights into the entrenched challenges facing contemporary higher education systems in the region.

1.1.1 Political influences

The political perspective tends to discuss African higher education in terms of an internal "crisis of identity" that began in the early 1970s (Goma & Tembo, 1984). Despite an unprecedented amount of aid and technical assistance throughout the 1960s, most African nations remained 'underdeveloped' a decade after independence. New theories emerged to explain the apparent lack of impact, many of which concentrated on the need to refocus development assistance on 'basic needs' and extreme inequalities within many African contexts. Against this backdrop, higher education was increasingly portrayed as disconnected from the general population, focused entirely on urban life, dependent on Western tradition and burdened by a "tendency to foster careerism, aloofness and ivory-towerism in [its] graduates" (Kajubi, 1984, p. 27). In the years following independence, African universities were predominantly elite institutions, focused on training a small percentage of the population for leadership positions. However, governments across the region viewed higher education as a public good and therefore allocated significant amounts of public spending to university scholarships and maintenance of public institutions. By the early 1970s, there was a growing sense of disenchantment with higher education, as hopes that "[the university] would rapidly transform the developing world" appeared not to have materialised (Ajayi, in International Association of Universities, 1979, p. 19). Although such criticisms were largely premature, given the relative youth of African higher education systems (ibid.), dissatisfaction with the high cost of higher education spread throughout the region, prompting debates about the role of higher education in development.

In response to the increasing discontent, the newly formed Association of African Universities (AAU) convened a conference in Accra in 1972 with a stated goal of redefining the role of the university in contemporary African society. In a move for self-preservation, those attending the conference agreed that they needed to reorganise the higher education sector in order to respond to growing concerns from their governments and fellow citizens. The overarching message that emerged in Accra was that higher education in Africa needed to move away from Western tradition and establish a new role that would be articulated by and for Africans (Yesufu & Association of African Universities, 1973). The Western ideal of universities as institutions focused on the "advancement of knowledge" was cast aside, and a new mandate that universities should work to "change

society quickly" emerged (Wandira, 1977, p. 45). The new university ideal gained traction in the 1970s as a way for Africa to move away from the colonial roots of its higher education system, creating something new and more relevant for the continent. In this new role, universities in Africa would increasingly be referred to as "developmental" universities, given their focus on the development goals of their surrounding communities. This move away from traditional notions of the university as an elite institution, focused on liberal notions of teaching 'knowledge for knowledge's sake', to the concept of a "developmental" university, focused explicitly on assisting government to reduce poverty in the region, had two critical impacts on the higher education sector, both with long-term implications.

First, universities responded to criticisms of elitism by dramatically expanding access. Between 1975 and 1995, African universities experienced a 290% increase in enrolment (Shabani, 1995, p. 174). The surge in enrolments continues to the present day, as social demand for upward mobility via university education has increased with rising enrolment. As of 2009, universities in Sub-Saharan Africa enrolled approximately four million students (World Bank, 2009, p. 46). This represents a nine-fold increase in enrolment since the mid-1970s.

Second, the emphasis on the importance of service to the nation as the primary mission of developmental universities resulted in increased government interference in higher education affairs. Although many of the nationalists who led their countries to independence in the 1960s were university-trained, the influence of intellectuals on politics had waned by the early 1970s. Many of the governments that came to power in the 1970s assumed authority through military, rather than democratic, means, and, as a result, many regimes fundamentally distrusted universities and feared the potential for academics to disagree with government policy (Zeleza, 2004). Conflicts between universities and the state increased, resulting, in some instances, in violence towards students and high-profile intellectuals[2]. The transition towards a "developmental university" model also encouraged many governments to demand increased financial accountability (Maliyamkono *et al*, 1982) and, in some instances, to take direct control of universities in order to make them "more responsive to development needs" (Ajayi *et al*, 1996, p. 95). As a result, university autonomy was often significantly curtailed. In reaction, many prominent African academics opted to emigrate to universities in the U.S and Europe. This so-called 'brain drain' exacerbated declining standards at many

[2] One extreme example was the 1972 kidnapping and disappearance of the Vice-Chancellor of Makerere University in Uganda, following his refusal to grant an honorary doctorate to one of Idi Amin's political appointees (Ajayi *et al*, 1996).

universities, as those who replaced the emigrating faculty were often less experienced (ibid.). Furthermore, those that opted to stay often chose to work closely with government, leading to an increasing dominance of politics over the functions of the university (Mkandawire, 2005).

1.1.2 Financial pressures

The impact of this internal "crisis" was further compounded by serious financial pressures during the 1980s and 1990s. The impact of Structural Adjustment on national budgets and the simultaneous trend away from international assistance for higher education had severe ramifications on the operations of universities across the region.

Throughout the 1960s and 1970s, international aid and philanthropic organizations were active supporters of African higher education. Widespread support for growth models of development (Rostow, 1960) and the prevalence of human capital theory (Schultz, 1961) played a key role in positioning education, including higher education, as a vital component of international aid to Africa (Eisemon & Kourouma, 1994). However, by the early 1980s, some of the most powerful development organisations, specifically the World Bank, began to publicly question the link between higher education and development in Africa[3]. Years of development investment appeared to be yielding little economic return. At the same time, increasingly conservative Western governments were beginning to exert pressure on development agencies to justify their interventions by demonstrating results. In response to this internal and external pressure, education economists at the World Bank were tasked with analysing the returns to investment on the Bank's education programmes. In 1986, George Psacharopoulos *et al* published a seminal paper for the Bank (Psacharopoulos *et al*, 1986), which argued that the private rate of return on investment in higher education vastly outweighed any social rate of return. The Psacharopoulos report claimed that, although individuals benefited significantly from enrolment in higher education, a corresponding benefit to *society* could not be identified. It therefore advocated shifting the burden of pay for higher education from the government to the individual. Furthermore, the paper argued that the social rate of return on investment in primary education was double that of higher education (ibid.). The Psacharopoulos analysis implied that higher education could not be linked directly to economic development in Africa, suggesting instead that international organisations and governments should prioritise the funding of primary education in order to maximise return on investment.

[3] The influence of the World Bank is particularly significant, as it is the largest multilateral funder of education in the region (Brock-Utne, 2000).

More recent analysis indicates that the Psacharopoulos calculations were seriously flawed. Bennell (1996), in particular, has highlighted a number of significant methodological issues with the Psacharopoulos paper, including the use of a limited sample[4], the reliance on historical data[5], and the lack of comparability within the sample[6]. Bennell has also argued that many of the assumptions underlying the calculations were largely inapplicable to the African context (ibid.). For example, the Psacharopoulos paper ignored any earnings generated by the rural and informal sectors of the economy, despite their prominence in the region, while also disregarding the impact of the non-competitive nature of many entrenched wage-paying professions, such as public sector jobs, on differential earnings. Positive externalities, such as tax revenues or increased earnings resulting from improved health outcomes, were also not included in the calculations (Birdsall, 1996; Bloom, Canning & Chan, 2006).

Despite these analytical shortcomings, the results of the Bank's rate of return analysis prompted a near universal withdrawal of international aid to African universities[7]. As a result, governments across Africa suddenly assumed almost the entire financial burden of their higher education institutions. At the same time, the global economic context triggered a wave of new policies aimed at protecting international loan agencies, such as the World Bank and the International Monetary Fund, from the impact of countries defaulting on their loans. In an attempt to prevent default, the loan-giving agencies developed a new type of loan for developing countries, known as Structural Adjustment Loans, which attached conditions to any future financial lending (Clark, 2006). As a result of the conditions attached to their loans, many African governments found themselves in the difficult position of experiencing simultaneously rising budgets and declining external investment. This resulted in many governments cutting their own national budgets for higher education.

By the early 1990s, national and international investment in African higher education had been drastically reduced. However, student enrolment across the region continued to expand, as African families increasingly viewed university education as a

---

[4] The Psacharopoulos paper used data from 18 countries, which were taken as representative of all 46 countries in sub-Saharan Africa (Bennell, 1996, p. 184).

[5] Of the 18 studies included in the report, 15 were based on data that were at least 10 years old (ibid., p. 190)

[6] For example, some of the studies included in the analysis were adjusted for particular variables, such as school drop-out rates and socio-economic background, while others were not (ibid., p. 196).

[7] There were exceptions to this general rule, particularly Sweden, Norway and the Netherlands, who continued to support higher education despite international trends (Boeren, 2005; Gmelin, King & McGrath, 2001; Lindow, 2011). Some of the major American philanthropic organisations, particularly the Ford and Rockefeller Foundations, also supported African higher education throughout this period (King, 2009).

means of social mobility (Lebeau, 2008). Reductions in funding prevented universities from supporting rising student enrolments with a similar expansion in faculty members, leaving most institutions inadequately staffed (Salmi, Verspoor & International Association of Universities (IAU), 1994). As university budgets were squeezed, real wages of academic members of staff declined to what many considered to be unacceptably low levels (Court & Kinyanjui, 1986), forcing many faculty members to take supplementary jobs in order to make ends meet. This left them with less time in the classroom and fewer opportunities to work on other traditional academic pursuits, such as research or mentoring new recruits to the profession (Lim, 1999). Others opted to leave academia altogether, seeking more highly paid positions as consultants to government or international organisations (Holm, 2012). Many of the faculty members who remained at universities in the region were less qualified than their forebears, due to a lack of possibilities for training and mentorship, and less committed to their institutions, due to conflicting pressures on their time from supplementary employment (Tettey & Partnership for Higher Education in Africa, 2009). Furthermore, very few new faculty members were in the "pipeline", due to a lack of postgraduate training programs and a decline in the prestige of the profession (ibid.). When combined with the pressures of ever-increasing student enrolment, the consequences of this human resource challenge were far-reaching and catastrophic.

Declining resources also significantly curtailed the ability of universities to manage or improve their infrastructure. By the early 1990s, most universities were housed in crumbling buildings designed to accommodate far fewer students. This caused short-term problems, in that students could not physically fit into lecture halls, dormitories or libraries on campus, and also contributed to long-term degradation, as the pressure of overpopulation accelerated deterioration (Coombe, Ford Foundation & Rockefeller Foundation, 1991). Universities were also poorly equipped with basic educational materials, such as textbooks, laboratory equipment and computers (Salmi, Verspoor & International Association of Universities (IAU), 1994).

Against this backdrop, universal primary education was adopted as the central education priority for the international development agenda, following the World Conference on Education for All in 1990 and the adoption of the Millennium Development Goals (MDGs) in 2000. With international organisations focused almost exclusively on primary education efforts, higher education continued to struggle with limited funding from both international and domestic sources.

1.2 Reforms and Revitalisation

Despite the continued prioritisation of funding for primary education, the rhetoric of the international community has gradually shifted back to one of support for higher education in recent years. With increasing frequency, policymakers – including representatives of the Bank itself – are arguing that universities do have a critical role to play in Africa's development (Association for the Development of Education in Africa Working Group on Higher Education & Association of African Universities, 2004; Collins, 2011; World Bank, 2009). The explanation for this dramatic shift can be found in the contemporary context of international higher education. Changes in the conceptualisation of the role of universities in Africa is linked to an overall *international* shift in the rhetoric surrounding higher education, which in turn is linked to the rapid proliferation of technology and the impact of globalisation on international economic structures.

The economy of the 21st century is generally referred to as the 'knowledge economy', due to the fact that the most powerful economies around the world are now based fundamentally on innovation and the production of new knowledge, rather than the production of agricultural goods or manufactured items. Knowledge is supplanting physical capital as the most lucrative source of present and future wealth around the world, meaning that investment in human capital is increasingly seen as having the potential to yield higher returns than investment in the production of goods (Serageldin, 2000). These economic changes have obvious implications for educational institutions, and countries increasingly view higher education, in particular, as vital for national competitiveness (Ashcroft & Rayner, 2011). Within the African context, the growing international consensus around the idea that universities can and should be "engines of development" (Castells, 1994) has refocused attention on the implications of the region's struggling higher education sector (Moja, 2004; Task Force on Higher Education and Society, World Bank & UNESCO, 2000; World Bank, 2002).

As a result of this altered stance towards higher education, the international lending community has encouraged a wave of reforms in recent years, aimed at revitalising Africa's universities. Critically, given a general shift in the international development community away from a welfare state approach and towards market-based solutions to global poverty (Clark, 2006), the reforms have all centred on a fundamentally neoliberal vision for higher education. This emphasis has largely focused reform efforts on the economic challenges facing higher education institutions in Africa (Association for the Development of Education in Africa Working Group on Higher Education & Association of African Universities, 2004). Although there have been a few projects aimed at increasing research productivity (Dias, 1992; Shabani, 1995; Vogel, 2012) and improving

infrastructure (Lindow, 2011), revitalisation has focused primarily on three areas: improvements in strategic planning and management of institutions; differentiation within higher education systems; and changing modes of delivery of instruction (Maassen, Pinheiro & Cloete, 2007).

1.2.1 Institutional reforms

African universities have long been criticised for their inefficiency (Psacharopoulos *et al*, 1986; Teferra & Altbach, 2004; World Bank, 1988; World Bank, 2010). Reforms in recent years have therefore focused on improving the efficiency of institutions by streamlining budgets, creating mission statements and implementing strategic plans. Institutions have been encouraged to eliminate excess non-academic staff, improve university governance and alter the structure of courses and programmes to allow the maximum number of graduates to complete their studies in the minimum time required (European University Association, 2010; UNESCO, Association of African Universities & Priority Africa Programme, 1993).

The reform agenda has also demanded that institutions mitigate their financial difficulties by experimenting with alternative funding strategies. Although international agencies have altered their stance towards universities in recent years, they have not changed their funding arrangements. International aid and national education budgets continue to focus on primary education, as a result of international pressure to meet the targets set by the MDGs. Universities are therefore required to diversify their financing in order to stay afloat (Association of African Universities, Donors to African Education/Working Group on Higher Education & National University of Lesotho, 1995). Many have hearkened back to the work of Psacharopoulos *et al* (1986), arguing that the financial burden of higher education should fall on the beneficiaries themselves, given the high private rate of return on investment in higher education. A number of institutions have therefore adopted fee-paying structures that place the onus on the individual student to pay for their education, either upfront or in the long term via student loan schemes (Teferra & Altbach, 2004). As tuition fees are largely insufficient for covering rising costs (ibid.), institutions have also had to devise additional methods of generating income. One frequent practice is the provision of part-time evening courses to members of the public who would not normally gain admission but are willing to pay for training (Mamdani, 2007). Another is for faculty members to offer their services as consultants to government or other industries in the region (Mamdani, 2007; Mohamedbhai, 2008).

Many institutional level reforms have been implemented through 'partnership' models, a development model that has grown substantially in popularity since 2000 (Fraser, 2009). This approach has been particularly apparent in the projects sponsored by

USAID (the United States Agency for International Development) and in the work of the Partnership for Higher Education in Africa (PHEA), a project founded in 2000 that pooled resources from the seven major American philanthropic foundations[8] to assist in the revitalisation of Africa's universities. Partnership models of reform generally pair an African institution with an institution in the West[9]. Most tend to focus on the development of research capacity through the provision of short-term training courses, access to institutional resources and support for postgraduate study for faculty members. Partnership projects have also worked to improve technological infrastructure, such as internet bandwidth, and to strengthen regional research consortia on the continent.

1.2.2 System-wide reforms

In addition to institutional reforms, international organisations have emphasised the potential benefit of more diversified higher education systems. The 'market' for higher education around the world is expanding to include a growing diversity of tertiary education institutions, including technical and vocational schools, distance learning providers and private universities (Altbach, 1998). As a result of the neoliberal approach of the development industry, market-driven competition has been lauded by many as the best solution to problems of quality and access at African universities. The World Bank, for example, argues that increasingly diversified higher education systems lead to the alleviation of enrolment pressure on public universities and an increase in competition between providers, which results in higher quality across the sector (Task Force on Higher Education and Society, World Bank & UNESCO, 2000; UNESCO, 2009; World Bank, 2009; World Bank & Salmi, 1994). The Bank also advocates for the introduction of "niche" institutions, such as schools of business or institutes of technology, which prioritise a particular subset of programmes. Niche institutions are seen to improve system efficiency, as institutions focused on a small number of programmes can limit their need for a wide array of educational materials, while also reducing overlap within the sector (Association for the Development of Education in Africa Working Group on Higher Education & Association of African Universities, 2004).

Advocates of sector differentiation have also encouraged a dramatic increase in the number of private institutions across the continent (Fehnel, 2003; Ng'ethe *et al*, 2008; Teferra & Knight, 2008). Many see privatisation as a potentially welcome trend for the region, given that private institutions can remain relatively autonomous from government

---

[8] The Carnegie, Ford, MacArthur, Rockefeller, Hewlett, Mellon, and Kresge Foundations

[9] The United Kingdom has also sponsored a number of partnership programmes, such as the Higher Education Links Programme (Stephens, 2009). Since 2000, these partnerships have not required a U.K. partner (King, 2009).

budgets and can react more flexibly to 'market demand'. As private institutions often have access to higher levels of funding than public institutions, advocates for privatisation also argue that private institutions are better equipped to provide higher salaries and better infrastructure than public institutions.

1.2.3 A focus on new modes of delivery

The final component of the current reform agenda is the promotion of new, cost-efficient modes of delivery of education. Specifically, distance learning has been promoted as a solution to problems of over-enrolment at institutions across the continent (Fehnel, 2003; Task Force on Higher Education and Society, World Bank & UNESCO, 2000; Teferra & Knight, 2008; World Bank, 2009). In 1997, the World Bank established the African Virtual University (AVU) with the stated goal of increasing access to higher education in Africa (Teferra & Knight, 2008). In addition to providing instruction, the AVU advocates for pre-existing institutions to expand their enrolments through distance learning programmes.

1.2.4 Evidence of impact

When taken as a whole, the reform efforts of the past decade have yielded mixed results for Africa's universities. There is no doubt that some of the reforms have had a positive impact (Lindow, 2011). Strategic planning has taken effect, and a growing proportion of universities have adopted mission statements and more streamlined budgets. An emphasis on differentiation has encouraged the rapid expansion of private institutions across the continent, which has allowed more students to access higher education. The proliferation of technology across the region has also generated some positive results for both students and faculty. Students are increasingly able to use the internet for research, which has eased some of the challenges caused by under-resourced libraries. Technology has also helped to mitigate the isolation felt by many academics at African institutions, as email and the internet provide access to international journals and the potential for connecting with academics elsewhere in the world (Teferra & Knight, 2008).

However, the reforms of the past decade have not yet resolved many of the most fundamental challenges facing African higher education. Although contemporary higher education systems do tend to include a wide diversity of institutions (Ng'ethe *et al*, 2008), the increase in institutional variety has not yet yielded the results advocated by supporters of differentiation. First, the addition of new institutions has not solved the problem of over-enrolment. As much of the demand for higher education is driven by a desire for social mobility, the proliferation of new institutions has simply fuelled a further

increase in the number of students hoping to access higher education (Lebeau, 2008). The simultaneous move towards cost-sharing arrangements has also exacerbated inequality of access, as most of the growth has been at private institutions where students must pay substantial fees in order to gain access (Oketch, 2003). Second, the introduction of new institutions has not led to true differentiation of the higher education system. There are very few policies governing the expansion of the higher education sector in most African countries. As a result, new institutions tend to copy existing models, rather than establishing innovative approaches. The pressure of the market has also resulted in increasingly blurry lines between different types of institutions. Most African higher education systems tend to be binary, meaning that they have two main types of institution: universities and polytechnics. However, as institutions have transitioned towards a reliance on student fees, there is an increasing tendency for mission "drift", meaning that universities and polytechnics have started to respond to student demand by introducing academic programmes typically outside the purview of their institutional type (Ng'ethe *et al*, 2008). The same holds true for "niche" institutions, many of which have responded to market demand by introducing new academic programmes outside their institutional missions.

In fact, there is evidence that current reform efforts may actually be contributing to *declining* academic quality in the region (Assié-Lumumba & CODESRIA, 2006; Association for the Development of Education in Africa Working Group on Higher Education & Association of African Universities, 2004; Brock-Utne, 2000; Court, 1995; Lebeau, 2008; Mamdani, 2007). The financial prowess of some of the new private institutions has allowed them to recruit faculty members from the public institutions. Given the general shortage in faculty numbers in the region, recruited faculty members tend to simply take on their new positions *in addition to* their existing contracts. This contributes to further over-commitment of faculty members and reduces standards of quality across the sector (Teferra & Altbach, 2004). The lack of any regulatory structure in some contexts has also allowed the proliferation of private providers of a very low academic standard (Materu, 2007). At the same time, the emphasis on cost-effectiveness and efficiency has resulted in an increasingly narrow curricular focus at both private and public institutions, reducing the concept of higher education to one of 'training' for specific technical careers. Cost-sharing arrangements have tended to move the orientation of universities away from student learning and towards commercial interests, leading to an increased focus on consultancies and short courses while distracting from the central task of encouraging student learning within undergraduate programmes. The focus on new modes of delivery has also presented a number of new challenges to institutional quality,

as African students are generally ill-prepared for distance education, given their unfamiliarity with technology or self-guided instruction (Kapur & Crowley, 2008). Infrastructural issues, such as the persistent lack of electricity and broadband connectivity, remain significant barriers to the successful delivery of online programmes, particularly to those from less advantaged backgrounds (Amutabi & Oketch, 2003; Klees, 2002), and there is limited understanding within the region regarding how best to regulate the quality of online education (Materu, 2007).

1.3 A Fundamental Gap in the Reform Agenda

Most alarmingly, despite a decade of reforms, university graduates continue to struggle with unemployment, while expatriates remain the candidates of preference for technical and leadership positions across the continent (Ashcroft & Rayner, 2011). Although this imbalance is likely due to a myriad of factors, one crucial reason appears to be the low capacity of many local university graduates. As a recent study by the Higher Education Research and Advocacy Network in Africa (HERANA) articulated, "African universities tend to be highly efficient in producing large numbers of undergraduates, but are far less successful in delivering quality instruction" (as quoted in Lindow, 2011, p. 43).

The current reform agenda is not yet having a positive effect on the issue of graduate capacity. This is likely due to the fact that most international organisations define quality in terms of efficiency. Although the issue of quality assurance is now central to many national and regional higher education strategies (e.g. Inter-University Council for East Africa, 2010), the focus remains largely on inputs and outputs. Any discussion of how to improve quality, as defined in terms of student learning outcomes, remains largely missing from the discourse[10]. In the early 1990s, there was some recognition of the need for improved pedagogy at many African universities (Matiru, Zentralstelle für Erziehung Wissenschaft und Dokumentation & Deutsche Stiftung für internationale Entwicklung, 1991; Nwaboku & UNESCO, 1996; UNESCO, Association of African Universities & Priority Africa Programme, 1993; UNESCO Regional Office for Education in Africa, 1992; van den Bor & Shute, 1991). However, the prominence of pedagogy as a reform objective was short-lived. Although UNESCO and the German Development Fund launched a programme aimed at improving teaching methods in African universities, it only lasted a few years and had very little sustained impact (World Bank, 2009). In recent years, teaching and learning have hardly been mentioned in the reform literature.

[10] The only significant exception to this can be found in South Africa, where the education sector policy is based on Spady's philosophy of Outcomes-based Education (1994).

The lack of focus on student learning outcomes is a significant gap in the current revitalisation agenda. As Ajayi *et al* wrote over fifteen years ago, "it has to be appreciated that if the training of the graduates is poor in quality and substance, their contribution to society will be inferior and a stumbling block to real national development and progress" (1996, p. 205). The perceived low quality of university graduates remains an equally critical development issue today. However, determining how to work with institutions to improve the academic quality of their programmes requires some understanding of current realities, and many within the international development community maintain that measurement of student learning outcomes is an all but unachievable task (e.g. World Bank, 2009). The field of educational research offers a fresh perspective on this challenge, given the range of potential methodologies that have been used to investigate academic quality at universities elsewhere in the world[11]. The vast majority of such methods were developed initially for use in other contexts, so any uncritical application is likely to yield contextually erroneous results. However, if used thoughtfully and with appropriate adaptations and modifications, they offer new ways of thinking about the revitalisation debate in Africa.

This study represents a first attempt to use one such method to inform higher education reform efforts in the region. The study considers one indicator of academic quality – the improvement of critical thinking skills – within the confines of one national higher education system. By analysing the factors that appear to help or hinder the development of critical thinking at Rwanda's universities, the study investigates a number of new areas for potential reform, highlighting the crucial importance of pedagogy and considering the vital question of cultural context. The study therefore offers a new perspective on the revitalisation debate by expanding current perceptions of how academic quality might be assessed and improved at Africa's universities.

The study is presented in the seven subsequent chapters. In Chapter 2, a rationale is presented for the selection of Rwanda as the study context. In Chapter 3, the concept of critical thinking is explored in some detail, and a conceptual framework is provided as a guide to the study. Chapter 4 is dedicated to a review of the study methodology. Chapter 5 focuses on the procedure used to select and adapt the study's central assessment tool. In Chapters 6 and 7, results of the data collection and analysis are presented and explored.

[11] As the question of academic quality has been the focus of debates regarding lower levels of education in recent years, there is a growing body of evidence that has relied on educational research methods to assess learning in primary and secondary schools in the region. However, such methods have not yet been used to inform reform efforts in the higher education sector.

The conclusion in Chapter 8 draws out implications of the findings for higher education reform, both within Rwanda and for the wider region.

Chapter 2: Higher Education in Rwanda

This chapter outlines the rationale for situating this study in the East African country of Rwanda by articulating the particular characteristics of the setting that make it an ideal location for an investigation of academic quality at the university level. The contemporary higher education landscape in Rwanda is then described in further detail. The chapter concludes with a discussion of why the study is focused on critical thinking as its central indicator of academic quality.

2.1 Rwanda as a "Crucial Case"

There were two main drivers for selecting Rwanda as the study context. First, the author lived in Rwanda between 2008 and 2010, working as the director of programmes for a local university scholarship initiative. Situating the study in Rwanda, therefore, carried the benefit of pre-existing knowledge of Rwanda's educational landscape, culture and politics. This personal connection to Rwanda also assisted with access to critical 'gatekeepers' at both the national and the institutional level. Second, and perhaps more importantly, Rwanda can be considered a "crucial case" in the African higher education landscape (Eckstein, 1975), as its tertiary education sector is both typical and atypical for the region. This section outlines the characteristics of Rwanda's higher education sector that qualify it as a particularly appropriate location for a study of this nature.

2.1.1 A brief history of higher education in Rwanda

Unlike many countries in Africa, Rwanda's history as a defined nation state did not begin with the arrival of colonialism[12]. Rather, the majority of the land currently defined as Rwanda existed as a feudal kingdom, unified under the control of a strong central *mwami* (or king), long before the emergence of a colonial state. The *mwami* was an all-powerful figure who controlled his subjects through both psychological and military domination (Prunier, 1995). This existing feudal structure was manipulated by the

[12] It is important to acknowledge that the question of how history has been constructed and retold since independence is a particularly contentious issue in Rwanda (Freedman *et al*, 2011). History as an academic subject is believed to have fuelled much of the ethnic conflict of the early 1990s (McLean Hilker, 2011; Walker-Keleher, 2006), and the various available histories of the country offer conflicting views and explanations that remain highly political and controversial within Rwanda today (see, for example, Mamdani, 2001; Pottier, 2002; Prunier, 1995; Uvin, 1998). The summary presented here is therefore not intended to be a definitive account of Rwandan history. Rather, the brief outline of historical events has been included in order to give some basic context to the study.

Belgians during the colonial era in order to maintain control over the population (Mamdani, 2001).

Education in the colonial era was highly selective and restricted to the upper echelon of society (Mamdani, 2001), a structure that carried ethnic, as well as economic, implications (McLean Hilker, 2011). When the Belgians arrived in Rwanda, they opted to recruit landowners from the Tutsi ethnic group to assist them in their rule over the Rwandan people. This political structure was rationalised in ethnic terms. The Belgians highlighted the 'superiority' of the Tutsi people, describing them as an "unusually intelligent" African tribe (Prunier, 1995, p. 6). Educational institutions helped to maintain the political structure, as schools were used to train Tutsi children to assume positions of leadership in the Belgian colonial structure. Children from the other ethnic groups in Rwanda – the Hutu and the Twa – were given no option to attend school.

In the 1950s, the Belgians reversed their strategy and began promoting the Hutu majority as the true leaders of Rwanda. It was in the context of this sudden support for popular rule that Rwanda experienced its first wave of ethnic violence. In 1959, a violent revolt against the Tutsi elite drove much of the Tutsi population over the border into neighbouring Uganda. By the early 1960s, the balance of power had shifted towards the Hutu majority. In 1962, Rwanda was granted independence from Belgium, and the first Hutu regime, under President Kayibanda, took control of the country.

The Belgian model of colonial educational provision largely ignored higher education (Lulat, 2005). The first higher education institution in Rwanda was the *Grand Séminaire de Nyakibanda*, established in 1936 by the Roman Catholic Church to train Rwandans for the priesthood. No other institutions of higher education were established during the colonial era, so any Rwandan interested in pursuing a higher degree had to attempt to do so in Belgium or at one of the universities in neighbouring Zaire. In 1963, the new Government of Rwanda joined other newly formed African nations in creating a national university for the new republic by founding the *Université Nationale du Rwanda* (UNR in French; NUR in English) under a joint agreement with the Dominican Fathers of Quebec (World Bank, 2004). Over the subsequent three decades, the higher education sector expanded gradually in Rwanda. After President Habyirimana took power in 1973, educational institutions formalised a system of access along ethnic lines, establishing quotas that explicitly privileged Hutu children over Tutsis at all levels of education (Hayman, 2005; Obura, UNESCO & International Institute for Educational Planning, 2003).

By the early 1990s, tensions between Rwanda's ethnic groups reached a tragic boiling point. Between April and July 1994, approximately 800,000 Tutsis and moderate

Hutus were killed at the hands of local militias across Rwanda[13]. An army of Rwandan refugees from south-western Uganda, known as the Rwandan Patriotic Front (RPF), ultimately stopped the killing and assumed control of the capital, taking over the Rwandan government and establishing a new regime based on the principles of unity and reconciliation for a new Rwanda[14], but peace came too late to prevent the complete devastation of the nation. In addition to the unimaginable loss of life, the country had lost the majority of its infrastructure. Government buildings, businesses, agriculture, transportation systems and schools had all been destroyed. The higher education sector did not escape the destruction. A Ministry of Education report from 1994 graphically outlines the situation at NUR at the end of the genocide: "All the infrastructure of the NUR is in a deplorable state: doors smashed, windows broken, files scattered in the corridors. Vehicles, scientific and laboratory equipment, office furniture, classrooms, and staff and student residences, have all been looted" (MINEPRISEC/MINESUPRES 1994, p. 18, as quoted and translated in Obura, UNESCO & International Institute for Educational Planning, 2003, p. 115). Similar devastation at other institutions of higher education, along with a substantial loss of faculty members and staff, forced the effective closure of the sector.

2.1.2 A unique higher education context

The events of the genocide are crucial for understanding why Rwanda's higher education sector is unique in the region. First, the complete devastation of human resources and infrastructure necessitated a rebuilding of all of the nation's institutions. Although tragic in every other sense, the genocide did allow the Rwandan government to create an entirely new higher education sector starting in the late 1990s (Mazimhaka & Daniel, 2003). As outlined in the introductory chapter, one of the main challenges facing contemporary higher education in Africa is the historical legacy of failed policies and declining respect for the region's universities. The genocide altered this reality for Rwanda, as the RPF government had the opportunity to start afresh with its higher education system. Rwanda is therefore an interesting case in which to consider the question of academic quality, as the higher education system in Rwanda is relatively young and unburdened by past challenges.

Second, the genocide played a critical role in the articulation of contemporary higher education policy in Rwanda. In the months following the war, educational reform

[13] For more detailed discussion of the events of the genocide, see Prunier (1995), Uvin (1998), Gourevitch (1998) and Mamdani (2001).

[14] The RPF has remained in power as the leading political party in Rwanda until the present day. Paul Kagame, former commander of the RPF army, has served as President of the Republic since 2003.

was identified as one of the first priorities in the rebuilding of the country (Mugisha, 2010; Obura, UNESCO & International Institute for Educational Planning, 2003). The dramatic loss of human life during the genocide was devastating for a country with limited resources aside from its population. The RPF government therefore prioritised the re-establishment and expansion of higher education, arguing that it had a crucial role to play in replacing the country's lost human resources (Mazimpaka, Daniel & Association of African Universities, 2000; Obura, UNESCO & International Institute for Educational Planning, 2003).

The events of the genocide also profoundly influenced the relationship between the Rwandan government and the international community. Rwanda is a tiny country with no natural resources and one of the highest population densities on the African continent. As the majority of Rwandans are rural agricultural workers with no regular income stream, the government is substantially limited in its ability to raise domestic revenue through taxes. International assistance has therefore been a critical component of national rebuilding since 1994. Foreign assistance funds just under 50% of Rwanda's national budget (Hayman, 2007, p. 373). Much of this aid is delivered in the form of direct budget support, but the country also relies heavily on technical assistance from expatriates (Hayman, 2009b). Despite this reliance on international assistance, the legacy of the genocide has allowed the Rwandan government to maintain some control over its policy objectives, even in instances where national priorities have not fallen in line with the priorities of the donor community (ibid.). In many ways, the government's success at carving out a policy space for its own agenda can be attributed to international feelings of guilt about the events of 1994, given the lack of foreign intervention during the genocide and the support that the international community gave to the Habyirimana regime in the early 1990s (Uvin, 1998). The legacy of the genocide also complicates the international community's relationship with President Kagame and his government. The Rwandan government is simultaneously seen as an authoritarian regime that allows increasingly little space for political opposition and as a 'development darling' that espouses a clear and progressive approach to development throughout its policies (Hayman, 2009b). Although the social and economic success of many of Rwanda's recent policies have clearly played a role in the international conceptualisation of Rwanda as a progressive developmental state, the legacy of the genocide has played an equally important role in the international community's acceptance of some of the more controversial political dimensions of the RPF's rule in Rwanda.

It is this legacy that has allowed Rwanda to focus on higher education, despite international pressure to prioritise primary education. Since the late 1990s, Rwanda has

spent a higher proportion of its education budget on higher education than almost any other country in sub-Saharan Africa (World Bank, 2004)[15]. As discussed in the first chapter, international support for the MDGs has generally resulted in African governments directing their national education budgets almost exclusively towards the funding of primary education. However, the particular relationship between the Rwandan government and the international community in the years since the genocide has allowed Rwanda to maintain high levels of financial support for its universities (Hayman, 2007). In the years directly following the genocide, this support resulted in the rapid re-establishment and subsequent expansion of the higher education sector (Mazimhaka & Daniel, 2003). NUR reopened in January 1995, a mere six months after the end of the war. In 1997, the Ministry of Higher Education merged with the Ministry of Primary and Secondary Education to form one coordinated Ministry of Education (MINEDUC), and the new Ministry spearheaded the establishment of a number of new public institutions, including the Kigali Health Institute (KHI) in 1996, the Kigali Institute of Science, Technology and Management (KIST) in 1997[16], and the Kigali Institute of Education (KIE) in 1999. In 2007, an independent government agency, the Higher Education Council (HEC), was established to regulate Rwanda's growing higher education sector (Official Gazette of the Republic of Rwanda, 2006)[17].

It has now been nearly 20 years since the genocide, and higher education remains a focus of Rwanda's education strategy. However, the rationale for supporting higher education has gradually moved away from the post-genocide rhetoric of replacing human capital. Since 2000, support for the expansion of the higher education sector has instead been justified in terms of the sector's crucial role in *Vision 2020*, Rwanda's national development strategy. Chief amongst the priorities articulated in *Vision 2020* is the development of Rwanda's human capacity in order to transform the country into a knowledge-based economy, capable of competing on the international economic stage (Republic of Rwanda Ministry of Finance and Economic Planning, 2000). The strategy elaborates the need for a skilled workforce, trained in disciplines such as information technology (IT), engineering and management and capable of confronting the myriad problems facing the country (ibid.). Similarities in the size, resources and political circumstances of Rwanda and the East Asian country of Singapore inspired the RPF

[15] In 2008, 26% of Rwanda's education budget was allocated to higher education (MINEDUC, 2008).

[16] In 2006, the Faculty of Management, initially housed within KIST, was transferred to the newly established School of Finance & Banking (SFB). KIST then dropped Management from its name, transitioning to become the Kigali Institute of Science & Technology.

[17] The HEC was initially called the National Higher Education Council (NHEC).

government to base *Vision 2020* largely on Singapore's economic development strategy. Following Singapore's lead, the Government of Rwanda positioned higher education as a central component of its national development strategy (Murenzi & Hughes, 2006; Tikly & Great Britain Department for International Development, 2003). A highly educated workforce is seen as crucial for economic growth and the attraction of foreign investment, both vital outcomes necessary for Rwanda to be able to wean itself off its current dependence on foreign aid (Chu, 2009). Rather than focusing exclusively on replacing human resources, the government's support for higher education is now couched in terms of these wider developmental goals for the country (Hayman, 2007; Palmer *et al*, 2007).

In light of its importance for national development, the government has expanded its support for higher education beyond the funding of institutions. Since 1998, the government has also funded one of the only student assistance schemes in the region (World Bank, 2009). The student bursary programme was designed to increase access to higher education, while also addressing historical biases within the sector. As discussed earlier in the chapter, ethnicity was the most important criteria for admission to university in pre-genocide Rwanda. Today, admission to university is based on a National Examination, which all students must complete in their selected subject at the end of secondary school (Mugisha, 2010). The new system has been welcomed as an important step towards unity and reconciliation in the country, as university admission is now based on academic merit, rather than student background (Kamali, 2006). The National Examination is also used as the qualifying examination for the student bursary programme. Under the scheme, the highest-scoring students in each subject are offered a government scholarship to study a particular discipline at one of the public universities. Bursary recipients are provided with tuition and fees, in addition to a modest monthly stipend (World Bank, 2004)[18]. Depending on their family's income level, students may receive the bursary as a need-based grant, or they may be responsible for paying back a portion of their bursary upon obtaining employment.

2.1.3 Regional similarities

Rwanda's higher education sector has a substantial advantage over other systems in the region, given the high levels of rhetorical and financial support that it receives from the government. However, despite such assistance, Rwandan universities face many of the same challenges as other institutions in the region.

As elsewhere on the continent, student enrolment continues to rise rapidly each year. In 2011, there were 73,674 students enrolled in universities across Rwanda

[18] Higher income students do not receive the monthly stipend.

(MINEDUC, 2012)[19]. Although still only 4.8% of the eligible population (World Bank, 2011a)[20], student enrolments have risen between 15 and 25% a year since 1995 (Hayman, 2005; MINEDUC, 2012). During this same period, the sector has expanded from a single institution to a system comprising 31 institutions (MINEDUC, 2012). Much of this increase is due to growth in the private sector. In 1996, there were two private institutions in Rwanda. Today, private institutions account for roughly half of the country's higher education enrolment (ibid.)[21]. However, as in other countries in the region, expansion of the private sector has not solved the problem of over-enrolment. Rwanda's overall population is growing at a rate of 3% per year (World Bank, 2013). This burgeoning population, coupled with the expansion of universal access to basic education across the country, is exerting significant demographic pressure on the tertiary education system.

Rwanda's institutions also suffer from an acute lack of qualified faculty members. The problem of faculty qualifications is an issue throughout the region, but the genocide dramatically exacerbated the problem in Rwanda, as many academics died during the war. NUR, for instance, was left with only 18% of its staff at the end of the genocide (Obura, UNESCO & International Institute for Educational Planning, 2003, p. 49). Although institutions have attempted to bolster their faculty ranks by recruiting professors from the diaspora and hiring expatriate lecturers, faculty recruitment and retention remain major issues for Rwanda's universities. In 2011, the system officially employed 2,583 faculty members (MINEDUC, 2012). However, this number double-counts any faculty member holding part-time positions at multiple institutions. As elsewhere in Africa, Rwandan faculty members, particularly at the public institutions, often opt to augment their small salaries by accepting part-time positions at other institutions or consultancies with domestic and international organisations (World Bank, 2004). The actual number of faculty members is therefore likely to be much lower than this number suggests. The turnover rate of faculty members is also a serious issue, particularly for the public universities. In 2010 alone, 23% of the faculty members at Rwanda's public universities opted to leave the sector for jobs elsewhere (Sindayigaya, 2010).

Universities in Rwanda also remain under-resourced in terms of technology, laboratory equipment and libraries. At the public universities, for example, the student to computer ratio is approximately 10:1 (Sindayigaya, 2010). Broadband access remains an issue for many students, and libraries continue to be sorely outdated and over-subscribed.

---

[19] Statistics from 2012 were not yet available at the time of writing.

[20] Enrolment in higher education in Africa as a whole is 6% of the eligible population (World Bank, 2011b).

[21] 35,772 students were enrolled in private institutions in 2011 (MINEDUC, 2012).

Reform efforts, spearheaded by the international community, have been implemented in Rwanda in much the same way as in other countries in the region. Although such interventions cannot accurately be described as 'revitalisation' efforts in the Rwandan context, given the relative youth of the higher education system, they reflect the general trend towards neoliberal solutions to institutional concerns discussed in Chapter 1. Britain's Department for International Development (DFID), Rwanda's largest bilateral donor (World Bank, 2004), has focused its efforts on governance and management within MINEDUC, while other organisations, such as USAID and the World Bank, have concentrated on differentiation of the higher education system (Maassen, Pinheiro & Cloete, 2007). The Netherlands Programme for Post-Secondary Education has provided support for university libraries, and Japan, China and Korea have provided aid and technical assistance for infrastructural development. The most recent *Education Sector Strategic Plan* (MINEDUC, 2010) outlines a number of priorities for the sector which also reflect international trends in higher education revitalisation, such as encouragement of greater privatisation, expansion of cost-sharing policies (such as tuition fees and income generation), investment in distance learning, and strengthening of university management. As in other contexts in the region, none of the reforms have focused on improving student learning outcomes at Rwanda's universities.

Although representative of the region in many ways, Rwanda's institutions of higher learning are unique, as they have had the advantage of substantial levels of government support. The reconstruction of the sector in the years following the genocide has also allowed Rwanda's universities to move away from many of the entrenched barriers to success evident in other African contexts. New institutions have been carefully developed with an eye to national needs and a concern for institutional quality, while Rwanda's progress towards other development outcomes, such as improvements in public health, the establishment of supportive structures for establishing small businesses and high levels of public security, has created a strong "enabling environment" for university graduates to capitalise on their education (Palmer *et al*, 2007).

Rwanda's public institutions have benefited in particular from the government's support. The public sector receives the vast majority of government funding for higher education. The government's merit-based bursaries are also reserved for use at public institutions, so those students finishing secondary school with the highest examination scores are overwhelmingly likely to attend public universities. As a result, the public institutions are widely considered to be the most prestigious in the country. This also makes the public sector more representative of the wider university student population,

as private institutions are generally populated by wealthier students who are able to pay the required tuition fees.

The high level of financial support, high calibre of entering students and representative nature of the student population make Rwanda's public universities an ideal location for a study of academic quality in the region. Given the likely positive effect of high financial support on academic quality, an investigation of student learning outcomes within Rwanda's public higher education sector offers an important perspective, as any concerns identified are likely to be indicative of wider trends in the region.

2.2 The Contemporary Context

Today, Rwanda's higher education system consists of 31 institutions. Of these 31, 17 are public (seven universities and 10 colleges or polytechnics) (Sindayigaya, 2010). The *National Qualifications Framework for Higher Education* governs the public sector, forming the basis of regular four-year *Higher Education Strategic Plans* developed by the Ministry of Education. Rwanda's public institutions are semi-autonomous, meaning that they are allowed to make their own curricular and administrative decisions but are obligated to follow the Ministry's strategic guidelines (Republic of Rwanda National Council for Higher Education, 2009). Private institutions must be accredited by the HEC in order to operate in Rwanda, but they do not receive government funding, nor do they report directly to the Ministry.

In the past five years, Rwanda's higher education system has undergone a number of dramatic and fundamental changes. As many of the students currently enrolled at universities in Rwanda have experienced these changes during their university careers, they are a key component of the contemporary context of the study.

The first and most fundamental change was the adoption of a new language policy in 2008 (MINEDUC, 2010). Given its colonial legacy, Rwanda had a Francophone education system throughout most of its history. However, much of the diaspora, including President Kagame and other officials in the government, grew up in Anglophone contexts, notably Uganda. As a result, a significant proportion of the current Rwandan population does not speak French. For the first fourteen years following the genocide, the government dealt with this challenge by assuming a bilingual education policy, in which public schools were required to offer classes in both English and French (Obura, UNESCO & International Institute for Educational Planning, 2003). However, in 2008, the Government of Rwanda announced that, starting in 2010, English would be the only official language of instruction at the secondary and tertiary level. The government rationalised the change in language policy in economic terms, arguing that Rwanda could

integrate more effectively into the East African Community (EAC) and the Commonwealth[22] if its educated population spoke English (McLean Hilker, 2011). It is also likely that the change in language policy had a political dimension, as diplomatic relations between France and the RPF regime are strained, due to French support for the Habyirimana regime during the genocide (Samuelson & Freedman, 2010). As a result of the new language policy, all universities are now required to teach exclusively in English. This has had obvious ramifications for both students and faculty members. Some faculty members unable to teach in English have left university employment, while students who grew up speaking French in primary and secondary school are now pursuing their university studies entirely in English.

Another set of changes was implemented in order to better align Rwanda's higher education system with other nations in the EAC. In 2008, Rwanda's public universities adopted the modular system of instruction used in Kenya, Uganda and Tanzania (Gahutu, 2010). In 2011, Rwanda's higher education system also altered its calendar from a January to October academic year to a September to June academic year, in line with the rest of the countries in the EAC and the Commonwealth.

There have also been recent changes in the government's financial support for higher education. In 2011, the government drastically reduced the number of available student bursaries. The bursary benefits also changed, as fees considered 'additional', such as dissertation and graduation fees, were removed from the bursary scheme. At the same time, the government reduced its financial support to public institutions of higher education from 24% of the total education budget in 2009-2010 to a projected 14% by 2014-2015 (MINEDUC, 2010). As a result of the cuts, many institutions were required to raise tuition fees in the 2011 academic year. These changes have not yet reduced enrolment numbers, but it is likely that they have skewed access towards wealthier students who can afford to pay higher tuition fees. Although no public reason was given for the sudden decrease in financial support for higher education, there is some speculation that the international donor community succeeded in pressuring the government to redirect some of the funding allocated to higher education to primary education in order to achieve the MDGs by 2015 (Hayman, 2005; Hayman, 2007; McLean Hilker, 2011; Schweisfurth, 2006). This seems like a plausible explanation, as Rwanda has lost some of its leverage with the international community in recent years, given allegations of Rwandan support for rebels in the eastern Democratic Republic of Congo.

[22] Rwanda was accepted into the Commonwealth in 2009.

2.3 Critical Thinking as an Indicator of Academic Quality

'Academic quality' is a complex and multi-faceted concept. As a result, there are myriad ways that one might choose to assess academic quality in a particular context. International rankings, such as those published by *Times Higher Education* and *U.S. News & World Report,* tend to measure academic quality at universities by using a range of input and output indicators, such as research outputs, student-to-faculty ratios and graduation rates. National governments are liable to collect similar data when monitoring the performance of their public institutions. Output and input indicators are appropriate data points to collect when conceptualising institutional quality in terms of efficiency. However, this study aimed to assess academic quality in terms of student learning outcomes. It was therefore vital to identify a learning objective that would be seen as a valid indicator of academic quality in the Rwandan context.

Critical thinking was selected as this central indicator for a number of reasons. First, critical thinking has gained traction in recent years as a crucial learning outcome for university students in the context of the knowledge economy. Manpower forecasting models have long been rejected as a strategy for higher education policy, given the speed of change within contemporary industries (Fulton, Gordon & Williams, 1980). As governments can no longer know which technical skills will be necessary in the future, structuring a curriculum around current workforce needs is widely viewed as a problematic method for developing a competitive workforce. Instead, there is growing international consensus that higher education should encourage more general abilities, such as critical thinking, that can support university graduates throughout their careers, regardless of future changes in the economy. Critical thinking is also seen to be a mandatory skill for the information age, as easy access to an overwhelming quantity of information requires individuals to have the ability to make determinations about the quality and reliability of a wide range of evidence (Guile, 2006; Peck, 2012).

Although African scholars have long argued that the cultivation of critical thinking skills is a priority for higher education in the region (Ajayi *et al*, 1996; Assié-Lumumba & CODESRIA, 2006; Cloete & Centre for Higher Education Transformation (South Africa), 1997; Crossman, 2004; Mosha, 1986; Musaazi, 2011; Nyerere, 1968; Wiredu, 1980), this perspective has rarely been discussed by national governments or the international community. However, in the context of the global knowledge economy, governments across the region are beginning to recognise that an emphasis on technical training is likely to disadvantage university graduates by restricting them to certain technical positions, rather than cultivating the more general cognitive skills necessary for interacting with information and developing solutions to problems (Fonlon, 2009;

Mamdani, 2007). This growing recognition of the importance of such skills has influenced domestic higher education policies across the region. International organisations have also started to argue that critical thinking is a crucial component of higher education for development in Africa (Africa-UK Engineering for Development Partnership, 2012; Association for the Development of Education in Africa Working Group on Higher Education & Association of African Universities, 2004; Brannelly, Lewis & Ndaruhutse, 2011; Task Force on Higher Education and Society, World Bank & UNESCO, 2000; UNESCO, 2009), reflecting a growing consensus around critical thinking as a key learning outcome of university education in the region.

Likely as a result of these international and regional trends, critical thinking has been articulated as an explicit learning objective of the higher education system in Rwanda. In a 2011 speech to the NUR student body, President Paul Kagame summarised his vision of how universities can and should be contributing to Rwanda's national development:

> At the core of every university's mission is the pursuit of new knowledge, seeking new ways of doing things to meet present and future challenges…Let me state that it is the people who carry ideas and skills and are not afraid to make discoveries that will shape the future – whether of an institution, such as this one, a nation, like ours, or the world. This is the educated society our universities should be helping to create – one that is less reliant on government for all the answers but one that will become the driving force and a strong partner of modernisation. And to adequately take on this responsibility requires that you adopt a new mentality that permits you to turn ideas and knowledge into tools of transformation. (Kagame, 2011)

Rwanda's recent policy documents reflect a similar emphasis on higher education as a means of empowerment and independence from outside assistance, indicating an assumption that students will graduate from university with the ability to find their own solutions to entrenched problems in society. Critical thinking is listed as a component of MINEDUC's mission and is included as one of the "high-level objectives" for education in the national *Economic Development and Poverty Reduction Strategy* (MINEDUC, 2010). Critical thinking is also described in the *Education Sector Strategic Plan* as a necessary "life skill", required to "face the challenges of health and nutrition, HIV/AIDS, family planning, gender awareness, social inclusion and the promotion of peace, unity and reconciliation" (MINEDUC, 2010, p. 2). Critical thinking for decision-making is also clearly viewed as a crucial component of workforce development in Rwanda, as "generic cognitive skills", such as "analysis, evaluation and critical skills", are mentioned as central learning outcomes in the *National Qualifications Framework for Higher Education Institutions* (MINEDUC, 2007).

International organisations working in Rwanda have added their voices to the government rhetoric, indicating widespread consensus around the importance of critical thinking for Rwanda's development. In a 2003 DFID study, for example, stakeholders across Rwanda highlighted critical thinking as one of the most important skills for integration into the global knowledge economy (Tikly & Great Britain Department for International Development, 2003). Other nations pursuing a "high skills strategy" of economic development – including Singapore, the model for Rwanda's development strategy – have also highlighted the importance of critical thinking for innovation and entrepreneurship, two of the core objectives of Rwanda's development strategy (Brown, Green & Lauder, 2001).

2.4 Previous Assessments of Critical Thinking in Rwanda

Despite the importance of critical thinking for Rwanda's development strategy, there has never been an attempt to empirically assess the acquisition of critical thinking skills at Rwanda's universities. Although a few studies have investigated the role of critical thinking in the secondary school curriculum in Rwanda (Freedman *et al*, 2011; McLean Hilker, 2011; Rutayisire, Kabano & Rubagiza, 2004; Walker-Keleher, 2006), no analysis has ever been conducted at the university level.

However, there are indications that universities may have cause for concern in this area. Highly skilled workers continue to be recruited from outside Rwanda to fill technical and leadership positions across the country (Hayman, 2005; Palmer *et al*, 2007), despite the high cost of expatriate employees and the government's goal of self-reliance. Recent evaluations have found that the transition from tertiary education to the workplace is poor, as many university graduates have trouble accessing the labour market within the first few years after graduation (Educational Consultants India Limited, 2001; Zoyem, 2010). One likely explanation for these trends is that Rwanda's university graduates may not be sufficiently qualified to fill many of the posts required by the national development strategy. Two recent studies have investigated this apparent skills gap and have concluded that graduates seem to struggle with a pronounced lack of problem solving ability. In 2010, the Government of Rwanda commissioned a *National Skills Survey* to assess current skills gaps in the workforce, as identified by employers in various sectors across the country. Problem solving emerged as one of the most significant concerns, particularly at managerial and professional levels (Republic of Rwanda National Council for Higher Education, 2011). A recent study of engineering capacity also indicated that graduates of Rwanda's engineering programmes were poorly equipped to confront complex problems and were therefore having trouble finding employment, despite a

continued need for engineers in many industries (Africa-UK Engineering for Development Partnership, 2012). Problem solving is not necessarily the same as critical thinking. However, the two terms are used interchangeably in much of the Rwandan policy literature. Although never explicitly defined, the term 'critical thinking' is generally used to imply the use of evidence to make decisions and solve complex problems. The apparent lack of problem solving capacity in the workforce identified by these studies, therefore, suggests that universities may not yet be succeeding in achieving their stated goal of cultivating critical thinking skills in their students.

Singapore, Rwanda's development model, has struggled with a similar lack of critical thinking skills in its workforce (Brown, Green & Lauder, 2001; Tikly & Great Britain Department for International Development, 2003). Within the Singaporean context, this has largely been attributed to the country's culture of conformism and respect for authority. Similar concerns could be raised in Rwanda. There is a clear preference for consensus politics in contemporary Rwanda, likely due to a widespread belief within the country that the genocide was sparked by international pressure to adopt multi-party democracy in the early 1990s (Hayman, 2009a; Mutebi, Stone & Thin, 2003). Scholars have also argued that respect for authority is a fundamental aspect of Rwandan culture (e.g. Prunier, 1995), suggesting that critique may not be encouraged in Rwandan society. However, the emphasis on critical thinking in Rwandan education policy indicates high-level support for the cultivation of critical thinking skills as an important learning outcome of a university education in Rwanda.

The development of critical thinking skills is clearly valued as an important objective of the higher education sector in Rwanda. However, there is anecdotal evidence that students may not be acquiring such skills by the time they graduate from university. Despite these concerns, there has never before been an attempt to empirically assess the ability of Rwandan students to demonstrate critical thinking skills, nor has there been any analysis of how universities in Rwanda may be helping or hindering the development of such skills in their student populations. It was this gap that this study aimed to fill.

Before beginning any empirical work, it was necessary to build a conceptual framework that could be used as a guide to the design and implementation of the study. The next chapter outlines the elements of the proposed framework, clarifying the author's positioning within ongoing debates about critical thinking and exploring the relevant findings from prior research focused on the development of critical thinking at the university level.

Chapter 3: Towards a Conceptual Framework

Despite widespread consensus around its importance as an educational objective, critical thinking remains one of the most debated constructs in the field. There is little agreement over the definition of the term, and there are questions about how the construct should be conceived, taught and assessed. In order to investigate the acquisition of critical thinking skills in the Rwandan context, it was therefore necessary to build a conceptual framework that could guide the study by defining the central construct in light of these ongoing debates and justifying the selection of variables to consider during analysis. The conceptual framework was intended to "specialise" the study within the theoretical landscape (Dowling & Brown, 2010), providing a guide for the research questions and a means of linking the study findings to previous work in the field.

This chapter outlines the steps involved in the creation of the study's underlying conceptual framework. The chapter begins with an analysis of a number of possible definitions of critical thinking. The chapter then proceeds to a discussion of how critical thinking can be assessed and an examination of the evidence that critical thinking can be improved at the university level. The findings from prior research into critical thinking are then analysed in some detail in order to clarify which individual and institutional factors seem to have an influence on the development of critical thinking at university. The chapter concludes with a brief discussion of the role of culture in the cultivation of critical thinking skills.

3.1 Defining the Central Construct

The first priority in the elaboration of the conceptual framework was the selection of an underlying definition and theory of critical thinking to guide the study. In this section, a number of possible definitions of critical thinking are examined in light of some of the fundamental debates in the field. The section concludes with a discussion of the central theory of critical thinking development chosen to underpin the study.

3.1.1 Possible definitions

Most scholars agree that the first philosopher to write about the different kinds of thinking that could be fostered through education was John Dewey. In *How We Think* (1933 (renewed 1960)), Dewey set out to articulate the difference between what he called "reflective thinking" and other types of thinking, defining reflective thinking as the "active,

persistent, and careful consideration of any belief or supposed form of knowledge in the light of the grounds that support it and the further conclusions to which it tends" (p.9). In Dewey's view, reflective thinking requires a state of doubt, in which a thinker realises the need for more information, and a subsequent act of searching for reliable material to resolve the doubt. The nature of the problem under question, therefore, defines the final outcome, and the outcome controls the process of thinking. For Dewey, in order to be "genuinely thoughtful", one must be "willing to sustain and protract that state of doubt which is the stimulus to thorough inquiry, so as not to accept an idea or make positive assertion of a belief until justifying reasons have been found" (ibid., p. 16). In other words, an individual involved in reflective thinking needs to be willing to submit his or her ideas to a process of critique.

The concept of differentiated types of thinking was further advanced in the 1950s, largely due to the emergence of developmental models of cognition and the attempted application of such models to planning within schools. Bloom's *Taxonomy of Educational Objectives* (1956) is probably the best-known example of this attempt to translate psychological theories into a schematic for use in education. Bloom's Taxonomy outlines a progression of cognitive skills from "lower-order" skills, including "knowledge", or the memorisation of facts, to "higher-order" skills, including analysis, evaluation and creation of new ideas, stipulating that students need to master lower level processes before moving on to higher levels. Bloom and his colleagues made no explicit reference to Dewey, but there is certainly a connection between Dewey's conceptualisation of reflective thinking and the "higher-order" thinking skills articulated in the Taxonomy. As the Taxonomy was highly influential, particularly in the U.S., the "higher-order" thinking skills it referenced – analysis, synthesis and evaluation – became commonly accepted learning objectives for students in the later stages of schooling. The presentation of a linear progression of cognitive skills in the Taxonomy suggested that students should automatically gain "higher-order" thinking skills during secondary and tertiary education.

By the early 1960s, the concept of differentiated types of thinking was quite well established. However, 'critical thinking' did not yet feature in the discourse. In 1962, Robert Ennis published "A Concept of Critical Thinking" in the *Harvard Educational Review* . In this short article, Ennis advanced the idea of "critical thinking" as a unique type of thinking based on logical reasoning and argumentation. Over the next few years, Ennis would clarify his definition of a critical thinker as being somebody who is proficient at judging: "whether a statement follows from the premises, whether something is an assumption, whether an observation statement is reliable, whether a simple generalisation is warranted, whether a hypothesis is warranted, whether a theory is

warranted, whether an argument depends on an ambiguity, whether a statement is over vague or over specific, and whether an alleged authority is reliable" (Ennis, 1964, p. 599). This definition betrays Ennis' roots in the field of philosophy and his belief in the role of logic as an important societal value. Although Ennis did not explicitly link his concept of critical thinking to developmental theories of the time, he seems to have perceived critical thinking as a desired end-goal of the education process, implying that logic should be included in the school curriculum as a means of cultivating the development of critical thinking skills in the student population.

In the early 1980s, Ennis revised his initial definition of critical thinking by incorporating decision-making processes in 'real-world' situations, moving away from an exclusive focus on critical thinking as an academic exercise based on logical argument and analysis of hypotheses. In 1985, Ennis wrote that critical thinkers possess four fundamental sets of abilities: "clarity-related abilities…inference-related abilities, abilities related to establishing a sound basis for inference, and abilities involved in going about decision making in an orderly and useful way, often called problem solving" (Ennis, 1985, p. 48). In other words, critical thinking is "reflective and reasonable thinking that is focused on deciding what to believe or do" (ibid., p. 45).

Ennis' definition inspired both support and debate. Some theorists incorporated the central components of his definition into their own definitions of critical thinking. Sternberg, for instance, defines critical thinking as "mental processes, strategies, and representations people use to solve problems, make decisions and learn new concepts" (Sternberg, 1985, p. 46). Other contemporaries of Ennis argued that his definition was too narrow to encapsulate the full concept of critical thinking. Lipman (1988), for example, claimed that Ennis' definition was too focused on the outcomes of the thinking process, rather than the various components of the process itself, stressing that creativity and empathy are also critical components of decision making. To Lipman, the context of a decision is key, as are the criteria used when considering alternatives. Lipman also argued that critical thinking must be "self-correcting", thereby linking the concept of critical thinking to Flavell's (1976) concept of "metacognition" – the process by which a person consciously thinks about the cognitive strategy he or she is using to make a decision. In Lipman's view, therefore, critical thinking is "skilful, responsible thinking that facilitates good judgment because it (1) relies upon criteria, 2) is self-correcting, and 3) is sensitive to context" (Lipman, 1988, p. 39).

In addition to defining critical thinking, Ennis, Lipman and others involved in the emerging 'Critical Thinking Movement' in the U.S. began to argue that universities needed to explicitly *teach* critical thinking skills as an essential component of an undergraduate

education in order for students to acquire the ability to think critically throughout their lives. In their view, universities could not assume that students would gain such skills simply as a result of enrolment at university. In response to this argument, in 1990, the state of California mandated that all public universities must include critical thinking as an explicit learning outcome of a university education. Anticipating that such a mandate would require an underlying definition in order for success to be assessed, the state's Department of Education asked the American Philosophical Association to assemble an expert panel to articulate a consensus definition of critical thinking. The panel's definition was as follows:

> "We understand critical thinking to be purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation, and inference, as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which that judgment is based." (Facione, 1990, p. 2)

As this definition was conceived largely by philosophers, it is perhaps unsurprising that the report reflects a return to the conceptualisation of critical thinking as synonymous with formal logic and argumentation. Other definitions from the same period also focus on the analysis and evaluation of arguments, including Parker and Moore's definition of critical thinking as "the careful and deliberate determination of whether to accept, reject, or suspend judgment about a claim" (1989, p. 3) and Scriven and Fisher's definition of critical thinking as the "skilled and active interpretation and evaluation of observations and communications, information and argumentation" (1997, p. 21).

In 1994, Dewey's concept of "reflective thinking" returned to the discourse through King and Kitchener's Model of Reflective Judgment, a model of intellectual development in late adolescence and adulthood (King & Kitchener, 1994). King and Kitchener argue that reflective thinking is not the same as critical thinking. However, their differentiation reflects the assumptions of the period, as they define critical thinking as being synonymous with logic. In their theory, King and Kitchener postulate that critical thinking is used to solve "well-defined problems", while reflective thinking is necessary to solve "ill-structured" problems that have no "correct solution" and "no way to prove definitively that a proposed solution is correct" (ibid., p. 6). Such problems, they argue, must be evaluated in "light of existing information…that may be incomplete and unverifiable" (ibid.) and, as such, require an epistemological understanding of the unverifiable nature of knowledge, not simple familiarity with the standards of logical reasoning.

Contemporary definitions of critical thinking incorporate elements of all of these definitional strands, emphasising the role of judgment and logic, as well as creativity and the ability to propose solutions to ill-structured problems. In his most recent definition of

critical thinking, even Ennis has written that, "Critical thinking is focused, skilled, active, reasonable thinking, incorporating the identification, clarification, and due consideration of the situation, relevant background information, reasons, evidence, and alternatives in deciding what to believe or do" (Ennis, 2009, p. 82), clarifying that, in this definition, "alternatives" imply both alternatives suggested by external parties and alternatives generated by the thinker him- or herself. Others have attempted to combine past definitions by listing specific skills as key components of critical thinking, while also linking the act of critical thinking to real-world applications that assume the existence of ill-structured problems. Definitions in this vein include:

> The ability to "process and utilise new information ... reason objectively and draw objective conclusions from various types of data; evaluate new ideas and techniques efficiently; become more objective about beliefs, attitudes, and values; evaluate arguments and claims critically; and make reasonable decisions in the face of imperfect information" (Pascarella & Terenzini, 2005, p. 155)
>
> [A willingness and ability to] "use … cognitive powers of analysis, interpretation, inference, evaluation, explanation, and self-monitoring meta-cognition to make purposeful judgments about what to believe or what to do" (Giancarlo-Gittens, 2009, p. 19)
>
> "…considering an issue from multiple perspectives, critically examining evidence (and attending to information that may run counter to or disconfirm initial ideas), valuing claims that are backed by appropriate and adequate evidence, reasoning objectively and dispassionately, and arriving at informed judgments and decisions" (Chun, 2010, p. 23)

To some extent, variation in the precise wording of the definitions outlined in this section is a result of differences between the disciplines involved in discussions of critical thinking. Psychologists, for example, tend to conceptualise critical thinking differently from sociologists or philosophers. However, even within the same discipline, the multi-faceted nature of the concept inspires a number of theoretical debates. Theorists disagree over the connections between critical thinking and problem solving (Johnson, 1992), critical thinking and intelligence (Halpern, 1996; Sternberg & Berg, 1992), critical thinking and creativity (Bailin *et al*, 1999b; Halpern, 1996; Lipman, 2003), and critical thinking and emotional intelligence (Moon, 2008). The question of criteria and standards is another ongoing debate, as is the philosophical argument around whether it is the definition of critical *thinking* or critical *thinkers* that is most important (ibid.). Of the myriad debates surrounding the concept of critical thinking, two are particularly contentious within the current literature: whether or not critical thinking can be considered a skill (or set of skills), and whether critical thinking is a general or a domain-specific phenomenon.

3.1.2 Fundamental debates

3.1.2.1 Critical thinking as a skill

Bailin *et al* (1999a) have argued that there is a tendency to organise education in terms of content, skills and attitudes. Within this typology, critical thinking is generally classified as a "skill". However, this classification has led to significant disagreement within the literature.

On one side of the debate are those theorists who conceptualise critical thinking as an ability – or set of abilities – that can be cultivated, improved and applied to various situations. Some articulate critical thinking as a pre-conceived sequence of skills that can be applied to any situation requiring critical thought (Bell, 2005; Cottrell, 2003). Others emphasise the individual components of critical thinking without defining a particular sequence for their application (Ennis, 2009; Kneale, 1999). Another group specifically argues against the idea of a sequence, contending that critical thinking is an array of abilities that can be used in differing ways whenever a thinker is confronted with a particular situation. In this conceptualisation, the selection of which thinking strategy to use in a given situation is part of the critical thinking process. Halpern (1996), for example, argues that there are skills that students can be taught to recognise and apply and that, if recognised and applied appropriately to different scenarios, the use of such skills makes the students more effective thinkers.

The rhetoric in support of 'graduate skills' tends to assume critical thinking is one overarching ability that can and should be developed through education in order to prepare graduates for their professional careers. In the context of the knowledge economy, many argue that the education system needs to focus on fostering strategies and processes for lifelong learning, rather than teaching specific content (Star & Hammer, 2008). Critical thinking is often included as one such strategy that should be encouraged. Others contend that the acquisition of critical thinking skills makes graduates more flexible employees (Assiter, 1995). Regardless of the level of specificity in the definition, all such arguments assume that critical thinking is an ability or set of abilities that can be learned by students in preparation for their participation in the workforce.

On the other side are those who disagree with reducing critical thinking to a set of pre-conceived skills. For the theorists in this camp, prescribing certain steps ignores the importance of context and the role of individual disposition (Barnett & Society for Research into Higher Education, 1997; Brown, 1998; Moseley, 2005). Bailin *et al* (1999a), for example, have interpreted the promotion of a skills approach as encouraging rote application of individual processes, rather than reasoned judgment within a situation. The

same authors have argued against the idea of critical thinking as a uniform skills-based process of thinking, given differences in cognition between individuals.

A significant component of the debate is whether or not critical thinking requires certain dispositions or "habits of mind" (Case, 2009). In his early work on critical thinking, Ennis acknowledged that critical thinking involved both *skills* (such as inference, analysis and evaluation) and the *disposition* to use such skills (Ennis, 1985). Giancarlo-Gittens has defined the disposition towards critical thinking as "a person's internal motivation to think critically when faced with problems to solve, ideas to evaluate, or decisions to make" (Giancarlo-Gittens, 2009, p. 20). She argues that it is a person's disposition that dictates whether or not he or she will apply the critical thinking skills that he or she possesses when confronted with a challenge or problem. During the drafting of the California report in 1990, the panellists could not agree whether disposition should be an integral part of their definition of critical thinking (Facione, 1990). Most of the theorists writing in the past two decades have acknowledged the role of disposition in the use of critical thinking skills (Case, 2009; Halpern, 1996; Lipman, 2003; Lun, Fischer & Ward, 2010). Some have even gone so far as to argue that critical thinking is more about disposition than ability (Brown, 1998). Others do not explicitly mention disposition in their definitions but imply that disposition is part of critical thinking by assuming the use of critical thinking skills in real-world situations (Chun, 2010; Pascarella & Terenzini, 2005). The question of disposition is therefore central to the skills debate, as some argue that critical thinking can never be viewed purely as a skill, given that a disposition to use critical thinking is a prerequisite for the action of thinking critically.

In many respects, the skills debate is the result of diverse interpretations of the concept of a 'skill'. There is no question that education policy in many national contexts has moved towards an increasingly functionalist view of university education that privileges skills development and preparation for the workforce (Crossley & Watson, 2003). Many who reject this movement tend to have a similarly negative reaction to defining critical thinking as a skill. However, others view the development of "high skills" (Brown, Green & Lauder, 2001), including critical thinking, as a process far more nuanced than the simple acquisition of competencies (Tikly & Great Britain Department for International Development, 2003). In the latter conceptualisation, the development of skills is not a dogmatic process of learning concrete steps for rote application. Rather, skills are viewed as complex strategies that individuals can choose to use, depending on circumstances and prior knowledge of a situation. In many ways, this conceptualisation brings the two sides of the debate together, as it retains the focus on individual

competencies that can be acquired and developed while also acknowledging the importance of disposition in the use of such skills in real-world situations.

3.1.2.2 Critical thinking: generic or domain-specific

The skills debate has its roots in a deeper area of disagreement surrounding critical thinking: the debate as to whether critical thinking is generic or domain-specific. Those who perceive critical thinking as differing fundamentally between disciplines or domains tend to reject the conceptualisation of critical thinking as a skill (or set of skills). In contrast, those who see critical thinking as a range of general thinking strategies are more likely to describe the strategies as skills. A discussion of the generalist-specifist debate is therefore fundamental for understanding the roots of the contemporary skills debate.

As the first definitions of critical thinking were based in the tradition of logical reasoning, the origins of the concept were framed in generalist terms. In his early definitions of critical thinking, Ennis (1962; 1964) outlined a set of generic reasoning strategies that he claimed could be applied to any argument. Dewey's description of reflective thinking suggests that he was also a generalist, as he argued that "the various ways in which men … think can be told and can be described in their general features" (Dewey, 1933 (renewed 1960), p. 3). Since these early definitions, many theorists have continued in the generalist vein, arguing that critical thinking is a general construct that can be applied to any number of situations. Scriven (2009), for example, has postulated that some aspects of critical thinking are common to all fields. Ryan (1992) examined a number of field-specific arguments in scientific theory and determined that they all exhibited certain generic reasoning strategies. In his seminal work on the generalisability of critical thinking, Norris argues that critical thinking is generalisable for four reasons: 1) "the idea of 'thinking-in-general' makes sense philosophically", 2) "some commonality exists in the thinking required from field to field and subject to subject", 3) "critical thinking provides a 'fund of resources' for dealing effectively with multiple different fields or subjects," and 4) "the ability to think critically as learned in one field or subject has a positive influence on thinking critically in other fields" (1992, p. 1).

Others disagree with the concept of general cognitive traits or generic standards of reasoning or judgment. Paul (1987) asserts that any thinking involving judgment relies on context-specific criteria. McPeck (1981) agrees that critical thinking is context-specific, arguing that reflective scepticism can only be applied to a problem area by using the standards of that domain, while Bereiter and Scardamalia (1998) refute the possibility of general levels of understanding that can be applied across disciplines, arguing that deep

understanding can only be achieved within a particular field or context. More recently, Moore (2004; 2011) has argued that modes of argumentation and judgment differ by field. Expanding beyond judgment criteria, some have argued that cognition itself is always context-bound, as cognitive processes depend on both the subject matter and the authenticity of the task to which they are applied (Anderson & Bloom, 2001).

In addition to more conceptual philosophical arguments about the nature of critical thinking, the generalist-specifist debate has featured two particular areas of disagreement. First, theories of expertise have divided the field. Despite early arguments that humans rely on general *schemata* (previously learned and specialised techniques) to help forge links between novel situations and familiar problems (Newell & Simon, 1972), more recent empirical assessments of expert knowledge have indicated that experts tend to use domain-specific techniques to organise knowledge (Chi, Glaser & Farr, 1988). Metacognitive strategies, used extensively by experts, have also been found to be strongest when learned through specific content areas (ibid.). Some, such as Moseley (2005), have argued that certain skills, such as logic and argumentation, can be applied in such a way as to help a learner gain expertise in any area. However, theories of expertise have been used most frequently to support specifist arguments.

Second, the debate has revolved around the concept of transferability. Transferability is, in many respects, the central goal of education. As students are eventually expected to apply their learning to experiences outside the classroom, the primary objective of education is to provide them with knowledge and skills that can be transferred from the classroom to the outside world. However, empirical studies have not always provided support for the existence of transfer between domains. The first studies of transfer were conducted by Thorndike and Woodworth in 1901. Their investigation, focused on whether the study of Latin and other 'difficult' subjects had a general impact on student attention spans or facility with learning, found little evidence of impact across disciplines (Thorndike & Woodworth, 1901). Follow up studies established that similarities between the learning environment and the transfer setting can increase the likelihood of transfer occurring (Thorndike, 1913). More recent studies have reinforced the hypothesis that "near" transfer (transfer between similar circumstances) is more common than "far" transfer (Drew & Staff and Educational Development Association, 1998), suggesting that transfer is rare between disciplines. Those in the specifist tradition have argued that the lack of empirical support for transfer reinforces their argument that critical thinking is not a general concept (Brown, 1998; Hyland & Johnson, 1998; McPeck, 1992).

Theorists in the situated cognition tradition (Lave, 1988; Rogoff, 2003) have taken these arguments one step further by arguing that the basic concept of transfer is flawed, given that cognition is variable within individuals depending on the particulars of a given context. In her landmark study of mathematical ability (1988), Lave argued that experimental studies of transfer were invalid because they removed the participants from their natural habitats, thereby fundamentally changing the cognitive situation. After watching participants use mathematical ability in everyday situations, Lave concluded that there was no evidence of transfer from classroom exercises to real-world applications, suggesting that cognition is an entirely socially constructed phenomenon that changes from context to context.

Although it is clear that transfer does not always occur, generalists maintain that it is still reasonable to believe that transfer *can* occur. First, there is some evidence that studies of transfer have not always considered all of the possible types of transfer. Salomon and Perkins' (1989) theory of transfer outlines two fundamentally different types of transfer – "low road" transfer, or the "spontaneous, automatic transfer of highly practiced skills, with little need of reflective thinking", and "high road" transfer, or "explicit conscious formulation of abstraction in one situation that allows making a connection to another" (p. 118). As the two kinds of transfer are fostered in different ways, it is likely that educational environments may facilitate one kind of transfer more than another. It is also likely that studies of transfer have focused only on one type of transfer, providing a feasible explanation for varying results across studies purporting to investigate the same phenomenon. Second, although transfer *should* take place between classrooms and the 'real world', most generalists acknowledge that there is no *guarantee* that transfer will occur without explicitly planning for it and building strategies for transfer into the curriculum (Kuhn, 2005). Research has demonstrated that there are educational practices that seem to foster transfer better than others, such as explicitly connecting lessons with past or future applications (Cowan, 1994) and building in opportunities to practice in a variety of situations, including simulations similar to actual future applications (Perkins & Salomon, 1994). Metacognition has also been found to positively affect transfer across contexts, suggesting that the fostering of metacognitive strategies should assist with transfer (Mayer & Wittrock, 1996). These data suggest that Lave's arguments against transfer may be too extreme. Although there is no question that humans are affected by their social environments, cognition clearly differs systematically between individuals, regardless of context. The observation of a lack of transfer between classroom practices and every-day circumstances is likely to indicate a failure of education to foster transfer, rather than proof of the impossibility of transfer between domains.

Although many identify explicitly with one side or the other of the generalist-specifist divide, others have reached a more nuanced position, combining aspects of both traditions. Mayer and Wittrock (1996) have demonstrated that transfer seems to depend on both a deep understanding of content knowledge and a familiarity with more general skills and processes. Kuhn (1999), building on this argument, has suggested that critical thinking skills must first be situated within individual subject matter but must also be defined more generally, so that students can recognise them in different domains. In his study of university students, Entwistle (1997) determined that general skills were used in most disciplines but that certain disciplines emphasised certain skills more than others, lending support for the relevance of both within-discipline acculturation and across-discipline transferability. Most theorists who support a more nuanced view of the generalist-specifist debate agree that there are certain general standards that can be applied to reasoning within any disciplinary context (Siegel, 1992) and general concepts, such as the determination of the credibility of sources, that are relevant to all domains (Blair, 1992). Lipman (2003) agrees, arguing that general criteria do exist across disciplines, even while individual subjects organise knowledge and arguments in different ways. Some theorists have actually argued that there is a danger in the specifist argument, in that it implies a return to rigid disciplinary education at the expense of interdisciplinary work (Davies, 2006).

3.1.3 The development of critical thinking

In addition to debates over the nature of critical thinking, there has long been discussion around whether or not critical thinking can be improved – and, if it can, how and why such improvement might occur. These debates have their origins in more general theories of cognitive development. Two perspectives dominate the contemporary literature on cognition: the constructivist perspective, based largely on the work of Piaget, and the situated cognition perspective, inspired by the work of Vygotsky.

Piaget revolutionised the field of psychology by arguing that individuals actively participate in their own intellectual development. Rejecting the claims of the behaviourist school, which had argued that individuals develop through a series of learned behavioural responses to external stimuli (Skinner, 1954), Piaget (1975) theorised that cognition develops through a process of "equilibration", in which individuals actively seek to balance their pre-existing knowledge with any new information that contradicts it. In response to such "conflicts" with new external stimuli, individuals must construct their own understanding of how new information can be synchronised with their pre-existing cognitive schema. The construction of new schema that incorporate the influence of the new stimulus is what is often referred to colloquially as 'learning'. However, Piaget

argued, learning cannot be assumed, as individuals can also compensate for new knowledge by adapting it to fit their pre-existing schema. The process of compensation does not result in any cognitive development, as there is no change to the pre-existing cognitive structure. Piaget also contributed the concept of "cognitive readiness" to the field (Moseley, 2005). According to Piaget, an individual's level of "readiness" to respond to stimuli impacts their capacity to construct new cognitive understanding, with lower levels of readiness typically resulting in compensation, rather than construction. Although he disagreed with most of the fundamental tenets of behaviourism, Piaget agreed with the conceptualisation of cognition as a developmental process in which individuals gradually progress through a series of phases, demonstrating increasingly sophisticated cognitive ability. However, unlike the behaviourists, who viewed complex thinking as a combination of individual behaviours (Skinner, 1954), Piaget saw cognition as a system of interrelated cognitive processes.

Vygotsky's socio-cultural theory (1978) contradicted the idea of an invariant cognitive progression. Although he agreed that individuals develop cognitively over time, he suggested that developmental trajectories differ depending on historical or cultural context. One of the central arguments of Vygotsky's theory is that the stimulus-response interaction in the more advanced stages of cognitive development is always mediated by activities and tools which are historically and socially specific (Kozulin, 1998). Although he agreed with Piaget's theory that development occurs as the result of conflicts, Vygotsky argued that conflicts occur when individuals fail to find adequate solutions to novel problems (Vygotsky, Cole & Luria, 1978). Such conflicts are then overcome through interactions with others (ibid.). Vygotsky suggested that individuals exhibit a range of ability, depending on the amount of guidance they may have received from more skilful tutors or peers. He called this the Zone of Proximal Development (ZPD). Those acting without the benefit of any guidance will tend to exhibit lower ability, while those who have worked with a more experienced partner are likely to demonstrate higher ability. Vygotsky defined the ZPD as "the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers" (ibid., p. 86).

Most theories of intellectual development draw inspiration from at least one of these central traditions within cognitive psychology. Theories of epistemological development have also contributed to contemporary understandings of cognition. Perry's landmark study of undergraduates (1970) introduced the idea that, in addition to developing cognitive skills, students develop an increasingly sophisticated understanding

of the nature of knowledge itself. Perry's scheme of epistemological development included nine "positions", which were arranged in a linear progression, ranging from "Basic Duality" (in which students see every problem as having a right and a wrong answer and therefore seek the 'right' answers from authorities) to "Commitment" (in which students are able to affirm their identities among multiple perspectives and responsibilities). Perry observed that students in his sample moved from duality through "Multiplicity" (in which they perceived the existence of multiple perspectives but tended to value them equally) to "Relativism" (at which point they began to understand that perspectives should be supported by reasons and that individuals should understand the reasons behind their own perspectives). Some students in his sample proceeded to an even more sophisticated stage (that of Commitment) where they had experienced committing to an individual perspective or opinion, perceived the implications of doing so, and realised that the construction of individual opinions would be an ongoing activity throughout their life. In addition to outlining the progression itself, Perry highlighted three alternatives to progression: "temporising" (in which a student delays in a position and explicitly hesitates before taking the next step), "escape" (in which a student detaches from the progression altogether) and "retreat" (in which a student entrenches himself in one of the earlier positions in the progression). Although Perry's study was limited in its generalisability, as it focused only on male students at Harvard University, his scheme highlighted the importance of epistemology as a critical component of cognitive development.

Perry's study also inspired many other studies of epistemological development. Belenky *et al* (1986) built on Perry's model by examining differences in the development trajectories of males and females. Baxter Magolda (1992) offered a simplified model, proposing four progressive domains of knowing and reasoning ("Absolute Knowing", "Transitional Knowing", "Independent Knowing", and "Contextual Knowing"). In a study of 70 undergraduates, she found that, although all of the students progressed through the same domains, male and female students did so in slightly different ways. King and Kitchener's Model of Reflective Judgment (1994) must also be situated in this paradigm. Their model is also developmental, consisting of seven stages, each characterised by distinct assumptions about knowledge and how it is acquired. Although they identify seven qualitatively different stages of epistemological development, King and Kitchener organise their seven stages in three overarching categories ("Pre-reflective Thinking", "Quasi-reflective Thinking", and "Reflective Thinking"). The overarching classifications reflect their observation that individuals tend to demonstrate a range of ability within a developmental category, depending on the support they receive from tutors and peers

(Kitchener & Fischer, 1990). Similar to Vygotsky's theory of the ZPD, King and Kitchener (1994) argue that individuals are likely to demonstrate "functional" ability within their developmental category in the absence of structured support (p. 28). However, with support, it is more likely that they will demonstrate "optimal" ability within their stage. Although recognising the fluidity of individual movement through the stages identified in their model, King and Kitchener do maintain that reflective thinking is a developmental process. The evidence from their extensive longitudinal and cross-sectional studies suggests a developmental trajectory, as participants were unable to comprehend the epistemological assumptions underpinning the higher stages in the model without progressing through the lower stages first (ibid.).

All of these theories of epistemological development propose similar end-points and suggest a developmental progression from simplistic, dualistic thinking to nuanced, reflective thinking. They also rest on the assumption that development occurs as the result of an epistemological conflict when a learner is suddenly confronted with the need to see knowledge in a different way. However, there are differences in the specific mid-points and the particularities of the trajectories proposed by the various theories. Some, such as Perry's, suggest a linear progression through the stages, while others, such as King and Kitchener's, propose a more fluid forward motion, allowing for flexibility within stages.

The study of metacognition (Flavell, 1976) has also influenced understandings of intellectual development. King and Kitchener (1994) found that problem solving strategies were related to the stages of epistemological development, as they observed that the metacognitive strategies used by their study participants developed in complexity as they gained an increasingly sophisticated epistemological understanding. Kuhn (1995) also argues that metacognition follows a developmental trajectory, as she has observed that the frequency of use of particular metacognitive strategies appears to change gradually over time, with individuals eventually relinquishing the use of inadequate strategies in favour of more effectives ones. In fact, Kuhn has observed that the selection of which metacognitive strategy to use in a particular situation becomes a strategy in itself, once an individual has had sufficient experience to recognise the need to make such metacognitive choices (ibid.).

Theories of cognition, metacognition and epistemological development have all contributed to contemporary understandings of how critical thinking develops. However, most theories of critical thinking focus exclusively on the cognitive elements. In contrast, Kuhn (1999) has proposed a coherent and convincing developmental theory of critical thinking that incorporates all of these underlying constructs. In her theory, Kuhn argues

that critical thinking requires cognitive strategies (such as abstraction and the ability to differentiate theory from evidence), metacognitive strategies that allow for control over the thinking process, and a sophisticated epistemological worldview that recognises the uncertainty of knowledge. Without all three of these "meta-knowing competencies", individuals are unlikely to be able to solve ill-structured problems (ibid., p. 23). As cognition, metacognition and epistemology have all been found to follow developmental trajectories, Kuhn suggests that critical thinking should also be viewed in developmental terms. However, her research indicates that many of these meta-knowing competencies appear to remain incompletely developed in adults – or at least are not identifiable in their everyday behaviour – suggesting that developmental progression cannot be assumed (Kuhn, 1991). Unlike those in the situated cognition field, Kuhn does not attribute these findings to a lack of transferability between domains. Rather, she argues that individuals will only be able to use meta-knowing competencies in real-world situations if they have had sufficient opportunity to use and refine them *and if they value their use*. Kuhn postulates that this requires an advanced stage of epistemological development, as it is only once individuals see knowledge as something that can be evaluated that they value the use of critical thinking as a way to assess claims and contradictory evidence. In other words, Kuhn's theory implies that the disposition to think critically can only emerge in later stages of intellectual development when an individual has reached a sufficient level of epistemological understanding (Kuhn, 2005).

3.1.4 Positioning the study

As this study aimed to assess critical thinking as an indicator of academic quality at Rwandan universities, it was necessary to identify which of the various conceptualisations of critical thinking outlined in this section would be most relevant for use in the study context. This final section therefore considers the available options in light of Rwanda's education strategy.

The use of the term 'critical thinking' in the Rwandan policy literature implies a conceptualisation of critical thinking as a general phenomenon that can be fostered within a particular discipline and then applied to a multitude of possible situations outside of the classroom. This conceptualisation of critical thinking resonates with Kuhn's theory of critical thinking development. Her research has indicated that general "meta-knowing competencies" are applicable to ill-structured problems across domains, provided they are sufficiently developed through practice in individual content areas. Furthermore, she suggests that individuals must develop a number of general cognitive and metacognitive skills – and must reach a sufficiently advanced level of epistemological understanding – before they are capable of solving ill-structured problems in the 'real world'. Given both

its contextual relevance and its conceptual cogency, Kuhn's theory was adopted as the underlying model of critical thinking development guiding the study.

Although she assumes that critical thinking is a general phenomenon, Kuhn does not suggest that the term refers to one overarching skill. Rather, her theory implies that 'critical thinking' is most appropriately used as an umbrella term describing an individual's ability and disposition to use a number of individual cognitive skills – including the analysis and use of information, the evaluation of arguments, and the reliance on evidence – when making decisions or proposing new solutions to ill-structured problems. Of the available definitions of critical thinking, this conceptualisation adheres most closely to the definitions of critical thinking articulated by Pascarella and Terenzini (2005) and Chun (2010). As Pascarella and Terenzini's definition includes developmental components, such as "becom[ing] more objective", Chun's definition was judged to be the more feasible definition to assess. Chun's definition was therefore selected as the underlying definition for use in this study.

3.2 Assessing Critical Thinking

The next question to address during the creation of the conceptual framework was whether it would be theoretically possible to assess such a conceptualisation of critical thinking. This section focuses on debates around the general feasibility of assessing the construct. A more detailed exploration of the validity of specific assessment formats for use in the study context is presented as part of the discussion of the assessment selection process in Chapter 5.

Disagreements over whether or not critical thinking can be assessed largely follow the same lines as other debates within the field. Some theorists argue that critical thinking can never be validly assessed. For example, those in the situated cognition tradition see the concept of cognitive assessment as anathema, given that they believe the testing situation itself influences the cognitive process (Lave, 1988). From this perspective, any assessment of critical thinking would simply reflect an individual's cognitive processes in the particular context of the assessment, so there would be little purpose to such an exercise, as assessment results would have no relation to the individual's cognitive behaviour in any other circumstance (as discussed in Pellegrino *et al*, 2001). Others within the specifist tradition believe that assessment is feasible but do not view general tests of critical thinking ability as valid, given that such assessments presume the existence of general critical thinking skills. Most generalists believe that critical thinking can be assessed. However, even within this group, there is substantial disagreement over the validity of different kinds of assessment tools.

Evaluation of an assessment's validity relies on three types of validity: *content validity*, which justifies the selection of particular items in an assessment as a valid sample of a student's understanding of the domain under consideration; *criterion validity*, which justifies the positioning of an assessment within the domain under consideration by analysing correlations between results of the assessment and results of other validated assessments of related abilities or outcomes; and *construct validity*, which justifies the positioning of an assessment within the domain under consideration by explicitly linking its structure and scoring methodology to the theoretical construct itself (Kane, 2006). Of these three, content and construct validity are considered of primary importance, as they are directly related to underlying theoretical understandings of the phenomenon in question. As a result, much of the discussion about the validity of individual critical thinking assessments focuses on these two areas.

Most disagreements over the construct validity of critical thinking assessments are the result of differences of opinion regarding the definition of critical thinking. Given the diversity of definitions and conceptualisations of critical thinking discussed in the previous section, it would seem likely that existing critical thinking assessments would reflect a wide range of underlying definitions. Although very little empirical work has considered this question, a recent study of three critical thinking assessments did conclude that the three tests appeared to rely on different definitions of critical thinking (Steedle, Kugelmass & Nemeth, 2010). Such diversity leads to debate, as those who disagree with the underlying definition of critical thinking used in a particular assessment are unlikely to perceive the assessment as having strong construct validity. Even when there is agreement about definitions, disagreements can arise over the construct validity of scoring methods, given the lack of consensus around which standards should be used when assessing whether an individual's thinking is sufficiently 'critical' or not.

There is also debate as to the construct validity of different assessment formats. Critical thinking is generally assessed via multiple-choice assessments, open-ended essay tests or performance-task-based assessments. Depending on an individual's conceptualisation of critical thinking, certain assessment types are seen to be valid, while others are not. Those who view critical thinking as consisting of a number of skills, for example, would tend to support assessments developed psychometrically, in which individuals are rank-ordered based on their ability to demonstrate particular critical thinking skills (Gardner & Clark, 1992), while those who view critical thinking as a holistic ability would be more likely to support assessments that rate respondents based on some measure of overarching ability (Klein *et al*, 2005). Debates around whether critical thinking is an ability, a disposition or both have also affected the determination of

assessment validity (Case, 2009). Early assessments of critical thinking mostly followed a multiple-choice format, in which respondents were asked to demonstrate critical thinking skills through the selection of particular responses. This format was criticised by many for ignoring the role of disposition, as multiple-choice questions do not provide any insight into the likelihood that an individual would choose to use critical thinking skills in a real-world situation (Ennis, 2009). A number of assessments have since been developed which purport to assess both critical thinking ability and the disposition to use such skills (Facione, 1990). Those who view disposition as a crucial component of critical thinking are more likely to support the use of these instruments.

Concerns around content validity, meanwhile, tend to focus on the particular content of individual assessment questions. As discussed earlier in the chapter, those taking a more nuanced view of the generalist-specifist debate concede that different disciplines tend to organise the assessment of knowledge in different ways. From this perspective, differential experience with individual disciplines or domains could result in differential performance on a given critical thinking assessment, depending on the familiarity of respondents with assessment questions based in particular areas of academic study.

Even where there is agreement around the validity of an instrument, disagreement can arise over its use. An assessment may only be valid for use in certain contexts, depending on the familiarity of an intended population with the content of the assessment questions. The validity of an assessment also depends upon the intended use of the instrument, as assessment formats may be valid for some purposes and not others. For example, an assessment format demonstrating low inter-rater reliability would be an invalid choice for a summative high-stakes assessment of critical thinking ability. However, it might be entirely valid to use the same testing format formatively in a classroom setting (Stiggins, 1987).

Most theorists who view critical thinking in generalist terms believe that it is feasible to assess critical thinking. As this study relies on a generalist conceptual framework, it was assumed that it would be possible to assess the critical thinking ability of university students in Rwanda. However, the inextricable relationship between an assessment's validity and the intended context and purpose raised questions about how critical thinking might validly be assessed in the study. The selection of an assessment tool that would be valid for use in Rwanda therefore emerged as a crucial component of the study design.

3.3 Improving Critical Thinking at University

Although the evidence suggests that critical thinking develops over time, it does not necessarily follow that education can impact the developmental progression. The potential role played by a university education in the development of critical thinking ability was therefore another important question to consider when constructing the conceptual framework. This section summarises the available evidence regarding how enrolment in university may influence the development of student critical thinking ability.

3.3.1 Models of university impact

Before considering the available evidence, it is useful to review some of the models of university impact that have been proposed by theorists of higher education. Perhaps the most widely used model of university impact is Astin's Input-Environment-Outcome model (1970). Astin's model suggests that learning outcomes can be seen as a function of inputs (the demographic characteristics, family backgrounds, pre-university academic and social experiences of incoming students) and the university environment. In the model, inputs are seen to shape outcomes both directly and indirectly, as they can both have a direct impact on an outcome and shape the way in which a student interacts with his or her university environment.

Most subsequent models have built on Astin's general framework. Tinto's Theory of Student Departure (1975), for example, seeks to explain student withdrawal from university by considering student inputs and interactions with the university environment. Tinto added an additional feature to the basic impact model framework by including student interactions with the external environment outside university as a critical variable in determining outcomes. Pascarella's General Model for Assessing Change (1985) posits that change in university occurs as a result of five variables: Student Background/Precollege Traits (Input), Structural/Organisational Characteristics of Institutions (Environment), Institutional Environment (Environment), Interactions with Agents of Socialisation (Environment) and Quality of Student Effort (interaction between Input and Environment).

All three models are quite similar, in that they give weight to both individual student characteristics and the influence of the university environment. Furthermore, they allow for student agency, by suggesting that students interact with their environments in a variety of ways. Assessing the impact of university attendance on a given outcome is a complex endeavour. All three theorists concede that there is always a risk of confounding effects, such as maturation or practice effects, that can impact outcomes, either positively or negatively. They also acknowledge the difficulty of proving causal impact. If change is detected, it is very difficult to prove that the change occurred as

a direct result of university attendance. If change is not determined, it can be even more difficult to prove that the lack of change is the result of a failure on the part of the institution. It could be that similar students not enrolled at university would *decline* in their abilities during a certain period, implying that university does have an impact, even without a demonstrable change in outcomes (Pascarella & Terenzini, 2005). Despite these imperfections, input-environment-outcome models tend to be the most widely used models of university impact.

3.3.2 Improving critical thinking at university

A number of empirical studies have attempted to assess whether or not universities can improve the critical thinking ability of their students. It is difficult to directly compare the results of these studies, as they use different measures of critical thinking and varying research designs. However, the evidence suggests that critical thinking can be improved as a result of university education, although university attendance does not appear to guarantee such improvement.

In their 1991 and 2005 meta-analyses of university impact in the United States[23], Pascarella and Terenzini concluded that students do appear to improve in their critical thinking abilities during university. In 1991, they determined that the weight of the evidence showed fourth-year students to be better abstract reasoners, more skilled at using evidence to address ill-structured problems, and more able to demonstrate greater intellectual flexibility and more sophisticated abstract frameworks than their first-year counterparts. In 2005, they reached the same conclusion, although they determined that the weight of evidence was less conclusive than it had been in 1991, indicating either that their initial analysis was overly optimistic or that students were less likely to improve in the 1990s and early 2000s than had previously been the case. Studies referenced in the 2005 analysis (e.g. Facione & Facione, 1997; Hagedorn *et al*, 1999; Mines *et al*, 1990) found statistically significant improvements on multiple measures of critical thinking, as well as similar constructs, such as reflective judgment (Baxter Magolda, 1990; King & Kitchener, 1994). Two studies also indicated a modest improvement in the disposition to think critically during university, as assessed via the California Critical Thinking Disposition Inventory (Facione & Facione, 1997; Giancarlo & Facione, 2001). Gains were identifiable whether or not studies controlled for maturation or pre-university characteristics and regardless of the study timeframe (i.e. gains in different studies were observed after one year, three years and four years of university).

[23] Technically, neither volume is an official meta-analysis, as the studies involved adopted diverse methods of inquiry. Rather, Pascarella and Terenzini reviewed all available studies and used 'weight of the evidence' criterion to determine responses to their research questions.

Such results have been confirmed within individual institutions (e.g. Hatcher, 2009) and in other national contexts. In a recent study of more than 3,500 students in Colombia, Saavedra and Saavedra (2011) found that fourth-year university students demonstrated significantly higher critical thinking ability (as measured via the Graduate Skills Assessment) than first-year students. Although the study used a cross-sectional design, the results are persuasive, as the researchers took great pains to ensure the comparability of the samples. The effects of individual-level characteristics, incoming ability and college fixed effects were controlled, and sensitivity analysis was used to assess selection bias[24]. As gains were not equivalent across institutions, Saavedra and Saavedra conjectured that some institutions have a greater impact on the development of critical thinking skills than others. This finding also allowed them to reject the hypothesis that gains were the result of maturation effects.

When considered together, these studies suggest that critical thinking can be improved through university education. However, results from other investigations demonstrate a lack of improvement, implying that critical thinking does not automatically improve as a result of university attendance. In the U.S., two large-scale studies have recently considered whether or not students improve in their critical thinking ability during university. In a longitudinal study of 2,500 American undergraduates enrolled at 24 institutions, Arum and Roksa (2011) observed that students did not generally improve in their critical thinking ability – as measured by the Collegiate Learning Assessment (CLA) – during their first two years at university. In a similar study of over 2,000 students enrolled at 17 liberal arts institutions, the Wabash Study of Liberal Arts Institutions reported that students only improved by an average of 1% on the Collegiate Assessment of Academic Proficiency Critical Thinking Test (Blaich & Wise, 2010). The results of the two studies are quite similar, suggesting some validity in the findings. However, it is important to acknowledge that Arum and Roksa's results have not been universally accepted. The lack of improvement claimed in their study relies on their assessment that 55% of the sample demonstrated a statistically significant improvement in test scores, when the difference in test scores was compared to the standard error of difference. It does not, however, necessarily follow that 45% of the sample did *not* significantly improve, as Arum and Roksa have claimed. It simply indicates that a statistically significant gain could not be identified for these students (Astin, 2011). As there is no consensus around how much students can be expected to improve during university, the selection of a 95% confidence

[24] Sensitivity analysis is a method of quantifying uncertainty in a mathematical model by systematically assessing the importance of various input variables in determining a given outcome (Saltelli *et al*, 2008).

interval must be viewed as largely arbitrary, casting some doubt over Arum and Roksa's claims (Pascarella *et al*, 2011). Furthermore, gain scores at the individual level are notoriously unreliable, especially when the assessment is open-ended, as was the case with the instrument they used (ibid.).

Despite these shortcomings, the results of the two studies do seem to indicate that improvement in critical thinking does not always occur as a result of university attendance. This hypothesis has been supported by studies in other contexts. Pithers and Soden (1999) found no significant improvement in critical thinking ability amongst university students in Scotland and Australia (as assessed via the Smith Whetton Critical Reasoning Test). Phan's (2011) longitudinal analysis of over 200 students in Hong Kong actually demonstrated a *decline* in the use of critical thinking during university[25].

On balance, the evidence seems to suggest that a university education can improve critical thinking ability but that such improvement cannot be assumed. This presents a question as to whether improvement depends on the individual student, the institution or both. The next section will address this issue by analysing the individual and institutional factors that appear to be associated with improvements in critical thinking at the university level.

3.4 Factors Associated with Improvement in Critical Thinking

The literature on gains in critical thinking tends to follow Astin's model of impact by focusing on the respective influence of inputs and the university environment. The study of inputs generally concentrates on two areas: the characteristics of incoming students, and the characteristics of faculty members[26]. Analysis of institutional effects, meanwhile, tends to focus on both structural and environmental characteristics. This section has been organised along similar lines, with first inputs and then institutional characteristics being analysed in terms of their respective impact on the development of critical thinking ability. As with overall studies of university impact, it is challenging to aggregate studies of critical thinking, given the diversity in instruments and research designs used in different studies. However, the weight of the evidence does indicate that

[25] It is important to note that Phan's study relied on the Reflective Thinking Questionnaire (Kember *et al*, 2000), a measure which captures critical thinking disposition, rather than ability. Phan did not identify a decline in student *ability* to demonstrate critical thinking skills. Rather, his study suggests that the likelihood of using critical thinking seems to decline university.

[26] Strictly speaking, the only inputs in Astin's model are student characteristics. Faculty characteristics are considered to be an element of the university environment. However, as there is a substantial body of recent literature that considers the influence of the university environment on incoming faculty characteristics, it seemed more theoretically appropriate to classify faculty characteristics as an input, rather than as a static aspect of the university environment.

there are a number of individual and institutional characteristics that seem to have a substantial impact on the development of critical thinking skills at the university level[27].

In addition to directly influencing critical thinking ability, the evidence suggests that a number of individual and institutional characteristics also *indirectly* influence critical thinking ability by directly affecting levels of student engagement. Although studies differ in their analysis of the impact of various individual and institutional characteristics, there is considerable consensus around the theory that student engagement with the learning process is a fundamental condition for improving critical thinking. In fact, studies indicate that the degree of a student's engagement with his or her university education may be the single most important factor influencing the development of critical thinking ability (Astin, 1984; Astin, 1993b; Carini, Kuh & Klein, 2006; Kuh & the Documenting Effective Educational Practice Project, 2005). Factors that have been found to positively influence student engagement have therefore also been included in this discussion, as they are likely to have an indirect influence on improvements in critical thinking.

3.4.1 The influence of student and faculty inputs on critical thinking

3.4.1.1 Student inputs

Astin (1991) outlines seven characteristics of incoming university students that typically have an impact on university outcomes: fixed characteristics (such as demographic background and type/location of secondary school attended); educational background; incoming cognitive ability; incoming aspirations, expectations, values and attitudes; incoming behavioural patterns; initial choice of academic subject; and initial choices regarding university experiences (i.e. living on or off campus, opting to work during university, or enrolling full- or part-time). Subsequent theories of higher education have adopted these broad categories while also adding additional details. Sociological research has suggested that family background should also be considered as a relevant background characteristic. Theorists of learning orientations have also indicated the relevance of incoming conceptions of learning and incoming student motivations (Entwistle, 1997). Much of the literature on student inputs has its roots in Coleman's landmark study of the individual characteristics found to affect achievement in schools (1966). Higher education theory has expanded on Coleman's contributions by considering

[27] Most of the research into gains in critical thinking at the university level has been conducted in the United States. This is likely to be an indication of the prominence of the Critical Thinking Movement in the U.S. in the 1980s and 1990s. Although every attempt has been made to include evidence from other contexts, the review of the evidence included here is therefore admittedly biased toward the American context.

both the impact of the characteristics themselves and the interactions between incoming characteristics and the choices students make during their time at university.

Any study of critical thinking at the university level assumes that students arrive at university with differing levels of incoming critical thinking ability. Those that enter with high levels of critical thinking ability are more likely to finish university with high levels of ability, as there is little evidence that critical thinking ability declines as a result of university attendance. Furthermore, research suggests that critical thinking may be generative, as students with higher incoming levels of ability seem to progress faster than others in their critical thinking development during university (Arum & Roksa, 2011; Phan, 2011). Studies linking student inputs and critical thinking therefore tend to consider the impact of individual student characteristics on both incoming cognitive ability and gains in critical thinking during university.

Five student input variables have received particular attention in the literature: demographics; educational background; family background; incoming motivations; and initial choice of academic subject.

A. Background characteristics

The first three variables (demographics, educational background and family background) can be classified as 'background student characteristics', as none can change as the result of enrolment at university. Studies investigating the impact of such characteristics on critical thinking ability have indicated that certain background characteristics seem to have more of an influence than others on both incoming critical thinking and gains in critical thinking during university.

Although few studies have attempted to assess it directly, the aspect of a student's background that appears to have the most significant impact on critical thinking ability is his or her socio-economic status. Socio-economic status (SES) is a particularly complex variable to assess. Studies attempting to assess the influence of SES on learning outcomes do not generally attempt to quantify SES by using measures of family income or expenditure, likely because students rarely have an accurate understanding of their family's financial situation. Instead, studies tend to rely on proxy measures of SES. In the American context, race is often included as a crude proxy for SES, as racial minorities in the U.S. are overwhelmingly likely to be of lower socio-economic status than their white counterparts. Race is a highly confounded variable, so any conclusions based on an analysis of racial differences are likely to mask the impact of a number of intervening variables that may have a significant influence on an outcome in their own right. As a

result, researchers also tend to collect additional information about a student's family (e.g. parental educational level or occupation) and secondary school, in order to assemble a more complete picture of an individual student's background.

Studies have indicated that there is a significant difference in both incoming critical thinking ability and gains in critical thinking ability between white and minority students in the U.S. (Arum & Roksa, 2011; Shavelson, 2010; Terenzini *et al*, 1994). In the aggregate, minority students seem to be more likely to enter university with lower levels of critical thinking ability and to show less significant gains in critical thinking over the course of their university careers.

Arum and Roksa (2011) also identified a strong correlation between incoming critical thinking ability and secondary school characteristics. This is unsurprising, as it is generally assumed that prior schooling should have an impact on the incoming cognitive ability of students. Secondary school characteristics were not, however, found to have a significant impact on gains in critical thinking during university.

Furthermore, Arum and Roksa (2011) found that students from families without any university graduates demonstrated lower incoming critical thinking ability than those from university-educated families. Such differences seemed to increase over time, with students from families without any university graduates demonstrating smaller gains in critical thinking during university than those from university-educated families. Other features of family background seemed to have a less systematic effect. The language spoken at home was found to have no impact on either incoming critical thinking ability or growth over time. Parental occupation was also found not to have a significant effect on critical thinking. This last finding was corroborated by Terenzini *et al* (1995).

Although none of these variables can be taken as a direct proxy of SES, all are likely to be correlated with SES, at least in the American context. Low-income students in the U.S. are more likely to be racial minorities and to come from families with minimal levels of education. They are also more likely to attend low-quality primary and secondary schools, which are less likely to foster the development of critical thinking skills in their students. The weight of the evidence, therefore, suggests that SES has a significant impact on both incoming critical thinking ability and the potential for a student to improve in their critical thinking ability during university.

The other background characteristic that is often included in studies of critical thinking is gender. Most studies indicate that there is no significant correlation between gender and incoming critical thinking ability (Arum & Roksa, 2011; Facione, 1991). However, growth studies have proven to be less conclusive. In a recent longitudinal study,

Arum and Roksa (2011) found no difference in gains between male and female students. Facione (1991), on the other hand, did identify significant gender-based differences in gains in his study. When considering gain studies in the aggregate, Pascarella and Terenzini (2005) found that, on balance, studies indicate that male students demonstrate smaller gains in critical thinking skills and larger gains in metacognitive skills than female students. In one recent study, no significant relationship was found between gender and raw assessment scores, but when scores were adjusted for incoming critical thinking, women were found to perform better than men (Shavelson, 2010). However, another study using the same instrument found that gender was not significantly correlated with critical thinking ability, even once scores were adjusted for incoming ability (Council for Aid to Education, 2008). There is some evidence that student reasoning patterns may differ along gender lines (Baxter Magolda, 1992) and that gender may also play a role in how students explain their cognitive gains during university (Kuh, 1995). The overall weight of the evidence, however, suggests that gender has little systematic impact on critical thinking ability.

B. Incoming attitudes and motivations

In addition to fixed background characteristics, students enter university with pre-existing motivations and attitudes towards university study. Unlike fixed characteristics, however, attitudes and motivations can change as the result of interaction with the university environment (Entwistle, 1997). This change can be either positive or negative, as university experiences have been found to both increase and decrease student motivation.

Incoming student attitudes and motivations have been found to have a particularly significant impact on student approaches to learning. Marton and Saljo (1976) were the first to advance the theory that student approaches to learning are a critical component of learning outcomes. Marton and Saljo proposed two qualitatively different approaches: a "surface" approach, in which students are motivated simply by progressing to the next stage and, therefore, apply the minimum amount of effort in order to progress, and a "deep" approach, in which students are motivated by a desire to learn and understand and, therefore, engage meaningfully and appropriately with the task at hand. In 1979, Ramsden (1979) identified a third category: a "strategic" approach, in which students are motivated explicitly by the desire for higher marks and, therefore, strategically select activities that can help them to achieve their goal. Nicholls (1983) also defined student orientations in terms of "task orientation", in which students are focused on understanding or mastery, and "ego orientation", in which students are focused on

outperforming other students and appearing competent to others. The study of learning orientations suggests that learning is dependent on what students intend to gain from the learning process (Entwistle, 1997). Research in this domain also indicates that approaches to learning are influenced by pedagogy. While good teaching has been found to increase student motivation, leading to "deeper" approaches, institutional and teaching norms also appear to have the potential to encourage increasingly "surface" approaches (Biggs, 2001; Entwistle, 1997).

There are very few studies considering the impact of motivation on critical thinking ability. However, the evidence suggests that approaches to learning influence levels of student engagement. It is therefore highly likely that incoming student motivations and approaches to learning have an indirect impact on the development of critical thinking skills[28].

C. Selection of academic subject

Academic subject is often considered to be an environmental variable, as there is some evidence that students enrolled in different disciplines may have qualitatively different academic experiences (Ramsden, 1997). However, in contexts where students are free to select their subject of study, academic subject must also be considered as an input, as there may be a correlation between critical thinking ability and the motivation to select a particular subject of study. For example, in their study, Arum and Roksa (2011) concluded that social science/humanities and science/math students demonstrated the largest gains in critical thinking ability, while those studying business, education and social work demonstrated the smallest gains. Although this could be a function of the student experience in different disciplines, they also identified a correlation between student background characteristics and chosen academic field. Students studying social science/humanities and science/math subjects were more likely to come from more advantaged social backgrounds and to demonstrate higher incoming ability. This raises a question as to whether identifiable correlations between academic discipline and critical thinking ability reflect systematic differences in academic experiences between disciplines or systematic differences in the incoming characteristics and/or motivations of student who select particular areas of study. The evidence to date is inconclusive on this point.

[28] Pascarella and Terenzini (2005) noted that there are very few studies considering the connection between demographic background and incoming motivation. It is, therefore, possible that fixed background characteristics may also have an indirect impact on critical thinking as a result of a correlation between student backgrounds and incoming attitudes and motivations.

3.4.1.2 Faculty inputs

Analysis of the influence of faculty characteristics on student learning outcomes tends to focus on the impact of faculty attitudes and teaching orientations[29]. Studies have found that faculty behaviour and pedagogical choices are typically influenced both by disciplinary and personal background and by beliefs about the purpose of education (Braxton, Olsen & Simmons, 1998; Stark *et al*, 1990). Kember and Gow (1994) postulated that there are two main teaching orientations amongst university faculty members: a "knowledge transmission" orientation and a "learning facilitation" orientation. In subsequent analysis at one institution, they found that university departments with a greater learning facilitation orientation were more likely to have students demonstrating deep approaches to learning than those with a knowledge transmission orientation. In a more recent article, Entwistle and Peterson (2004) have argued that the evidence overwhelmingly supports the theory that instructor perceptions of teaching have a significant impact on student learning outcomes.

In her study of four "extreme case" institutions in the U.S., Tsui (2001) found similar results in terms of the influence of faculty attitudes on student critical thinking ability. She identified three faculty attitudes which seemed to have a particularly strong positive effect: instructor belief in the potential for students to improve their critical thinking ability; enthusiasm for teaching and willingness to give extra effort to teaching practices; and a perception of teaching as being a process of mutual learning, rather than the transmission of facts. Although instructors tend to teach the way they were taught (Haas & Keeley, 1998), Tsui found that the culture of an institution can have a profound impact on such faculty attitudes, regardless of background, suggesting that incoming faculty attitudes can be modified as a result of institutional context.

3.4.2 The influence of the institutional environment on critical thinking

3.4.2.1 Institutional structure

Studies linking critical thinking ability with institutional characteristics tend to focus on elements of the university environment. However, a handful of studies have also considered the relationship between an institution's structure and the critical thinking ability of its students.

---

[29] Although studies also assume that teaching ability is likely to have an impact on student development, the difficulty of constructing a useful measure of holistic 'teaching ability' typically results in studies focusing on assessment of pedagogical practices and student outcomes as proxies of teaching ability.

Pascarella and Terenzini (2005) concluded that institutional structure has very little influence on cognitive development. They did identify one study that found that students enrolled at larger institutions demonstrated less significant gains in cognitive ability than those enrolled at smaller institutions (Dey, 1991). However, they could find no other studies corroborating these results. Kuh (1995) observed that students enrolled at liberal arts colleges demonstrated greater gains in critical thinking ability than those enrolled at other types of institutions in the U.S, but he did not control for student background characteristics in his study. As it is likely that selection of the liberal arts experience may be correlated with other factors that may influence critical thinking ability, such as student motivation, concrete conclusions cannot be drawn from these findings. Although Arum and Roksa (2011) found pronounced differences in critical thinking ability across institutions in their sample after controlling for individual-level factors, they could identify no systematic relationship between critical thinking ability and any structural characteristics of institutions in the sample.

There is some evidence that institutional type may have an impact in other contexts. Saavedra and Saavedra (2011), for example, found that students attending private universities in Colombia demonstrated greater gains in critical thinking ability than students attending public universities, even after controlling for incoming ability, socio-economic status and area of study. However, the results of this study have not yet been corroborated elsewhere.

There also appears to be little relationship between a university's structure and the learning environment it provides for its students (Kuh & the Documenting Effective Educational Practice Project, 2005). Even a university's selectivity, generally assumed to have a profound effect on the learning environment, does not appear to be correlated with the implementation of educational practices found to influence student engagement (Pascarella *et al*, 2006).

Although institutions clearly have an impact on the development of critical thinking skills, the evidence suggests that the source of differences between institutions is not their size, type, structure or selectivity. Rather, differences in critical thinking ability between institutions appear to be the result of differences in the institutional environment.

3.4.2.2 The institutional environment

The preponderance of the literature on the development of critical thinking skills at university focuses on the influence of particular characteristics of the university experience. Although there is a vast range of characteristics that could be considered, the

literature considering the relationship between university experiences and critical thinking tends to concentrate on three domains: academic experiences; interactions between members of the university community; and extracurricular activities. There is also some literature that considers the impact of institutional culture on the development of critical thinking skills. Although the literature generally seeks to determine which characteristics may have a positive impact on critical thinking ability, it is important to acknowledge that university experiences may also have a negative impact on student abilities. Entwistle (1997) has argued that the student experience can be structured in such a way as to actively discourage students from deep approaches to learning. Such an effect would also tend to have a negative impact on the cognitive development of students.

A. Academic experiences

a) Field of study

There is conflicting evidence regarding the impact of academic discipline on critical thinking ability. In his 1991 study of critical thinking, Facione found significant differences between the post-test scores of students enrolled in different academic fields. Lehman and Nisbett (1990) observed dissimilarities in the reasoning skills of students enrolled in different fields of study, while Palmer and Marra (2004) identified differences in epistemology across domains, arguing that the variation was likely to be linked to differences in the academic experiences between disciplines. Ramsden (1997) has claimed that teaching methods do vary across disciplines, lending some support for Palmer and Marra's hypothesis.

However, other studies have found no systematic differences between students enrolled in different disciplines. Schommer and Walker (1995) could not identify any epistemological differences between academic domains, while Terenzini *et al* (1995) found critical thinking skills were largely unaffected by academic field. In fact, they observed that any discrepancies between disciplines tended to disappear when pre-university differences were taken into account. Arum and Roksa (2011) found a similar effect in their sample.

As discussed in Section 3.4.1, variation in critical thinking ability along disciplinary lines may be the result of pre-university characteristics related to the selection of academic field, rather than any systematic variation in the qualitative experiences of different academic disciplines. Pascarella and Terenzini (2005) could find little consistent connection between field of study and critical thinking in their review of existing studies. In fact, the only differences that could be identified tended to be attributable to pre-university abilities and motivation or to differences in departmental culture. Moon (2008)

concurs with Pascarella and Terenzini's conclusion, arguing that it is difficult to generalise about the impact of discipline on critical thinking, given that many other variables – such as pedagogy, curriculum, teacher ability, interactions between teachers and students, and incoming student ability – appear to have a much more significant effect and are not necessarily correlated with particular fields of study.

b) The university curriculum

Results are more conclusive regarding the impact of the university curriculum on critical thinking ability. In fact, the evidence indicates that the curriculum can have a highly significant impact on gains in critical thinking ability.

The most important characteristic appears to be the level of integration and inter-relatedness between courses within a curriculum. A multitude of studies have demonstrated that the cultivation of clear connections between individual courses can have a significant impact on the development of critical thinking skills (Kember & Leung, 2005; Schilling, 1991; Terenzini *et al*, 1995; Wright, 1992). Although some argue that interdisciplinarity has a significant effect (Everett & Zinser, 1998; Wright, 1992), others disagree, arguing that a high level of connection between the courses is more important than diversity in subject matter. There is also evidence that critical thinking is most likely to develop when all of the courses in the curriculum reinforce the use of such skills. Stand-alone courses that encourage the use of critical thinking skills may actually generate resentment amongst students, as such courses are likely to demand significantly higher levels of cognitive engagement than more traditional course formats (Entwistle & Peterson, 2004).

In addition to inter-relatedness between courses, the *sequence* of courses appears to be a significant factor (Terenzini *et al*, 1995). Explicitly incorporating critical thinking into the curriculum in a progressive manner appears to have a particularly profound effect on student outcomes (Hatcher, 2009). By gradually exposing students to the use of critical thinking in different circumstances – and by progressively expecting a more advanced demonstration of such skills – universities can help to "scaffold" the cognitive development of their students throughout their time on campus. "Scaffolding" refers to the process of "'controlling' those elements of [a] task that are initially beyond a learner's capacity, thus permitting him to concentrate upon and complete only those elements that are within his range of competence" (Wood, Bruner & Ross, 1976, p. 90). As theorised by Vygotsky in his description of the ZPD, scaffolding appears to encourage optimal performance within a student's developmental stage, thereby supporting the development of "higher-level" cognitive strategies, such as critical thinking (King & Kitchener, 1994;

Rosenshine & Meister, 1992). The evidence suggests that the typical course sequence, in which first-year students are enrolled in large lecture-style introductory courses, can be detrimental to critical thinking development, as it is often too late to wait for later "advanced" courses to demand the use of critical thinking skills in coursework and assignments (Lonka & Ahola, 1995; Tsui, 2002). In contrast, by gradually exposing students to the use of critical thinking, progressive curricula appear to support students in the development of their critical thinking skills.

In addition to the inter-connectedness and sequence of courses in the curriculum, the perceived relevance of course content seems to make a difference. In particular, students appear to demonstrate deeper approaches to learning if they can see the relevance of both the content of their coursework (Entwistle, 1997) and the particular classroom practices and activities chosen by their instructors (Kuhn, 2005).

c) An appropriate level of challenge

Another component of the academic experience found to be fundamentally important in the development of critical thinking ability is the provision of an appropriate level of academic challenge. Studies agree that critical thinking is most likely to develop if students are exposed to high levels of academic challenge while also having access to sufficient academic support (Blaich & Wise, 2010; Hammond & Gibbons, 2001; Moon, 2008). Appropriate levels of challenge have also been found to have a positive impact on approaches to learning (Entwistle, 1997; Ramsden, 1997).

As discussed earlier in the chapter, cognitive development appears to occur as a result of "conflicts" when students are confronted with new information that challenges their pre-existing understanding. In order to stimulate such conflicts, students must be exposed to new ideas and new situations (King & Kitchener, 1994). Studies have indicated that certain kinds of coursework are particularly effective at providing a sufficient level of academic challenge to stimulate the development of critical thinking skills. For instance, requiring students to complete a substantial amount of reading and writing seems to have a positive effect on critical thinking (Tsui, 2002), likely because reading exposes students to new perspectives while writing requires them to develop their own arguments. Challenging class discussions and independent research projects also seem to stimulate the development of critical thinking skills (ibid.), as they require students to actively construct their own understanding of course content and provide an opportunity for students to practice metacognitive skills, such as determining what is unknown about a given question or situation (Bransford, Brown & Cocking, 1999).

However, challenge in itself is not necessarily conducive to cognitive growth. In fact, without simultaneous provision of an appropriate level of support, students are likely to retreat from the moments of conflict that appear to be necessary for intellectual development (Baxter Magolda, 1999; Bransford, 1979; Perry & Harvard University Bureau of Study Counsel, 1970). As discussed in the previous section, it appears to be helpful to "scaffold" the overall student learning experience by gradually requiring the use of more sophisticated cognitive skills in the curriculum. Such gradual exposure is likely to limit the tendency to disengage with more challenging assignments. Research also suggests that instructors can provide scaffolding within their individual courses by carefully balancing the level of familiarity with a particular teaching method with the level of familiarity with particular course content. For example, in a series of experiments related to the use of debate in classrooms, Kuhn (2005) observed that students were more likely to engage with more sophisticated arguments if allowed to debate a familiar subject, such as music, than when they were required to debate more 'academic' subjects.

For both curricular and course-based support, the construction of an appropriate "scaffold" requires an understanding of students' prior learning (Rosenshine & Meister, 1992). Studies have shown that pre-existing misconceptions can significantly limit student learning, even after extensive instruction (e.g. Dunbar, Fugelsang & Stein, 2007). As a result of the entrenched nature of previously-held beliefs, teachers are most effective when they can gauge their students' prior knowledge and present new learning in such a way as to allow students to make connections between the new content and their pre-existing understanding (Bransford, Brown & Cocking, 1999; Kuh & the Documenting Effective Educational Practice Project, 2005). Appropriate scaffolding for critical thinking also requires an understanding of a student's level of epistemological understanding. Support that is likely to be useful for students demonstrating low levels of epistemological understanding is unlikely to be very effective for students at higher stages of intellectual development (King & Kitchener, 1994). Cognitive load theory provides a theoretical explanation for why prior knowledge has such an influence on the development of critical thinking skills. The theory suggests that working memory has a finite capacity, meaning that its two primary functions – short-term storage and cognitive processing – must constantly vie for the same cognitive space (Kalyuga, 2006). As a result, the amount of prior knowledge that a student possesses within a given subject area is likely to affect his or her ability to think critically within it, as familiarity with content frees space in working memory for more sophisticated cognitive processes (ibid.). Prior knowledge also appears to have a substantial effect on metacognition, as the entrenched nature of pre-existing

beliefs can make it more difficult for an individual to un-learn an ineffective but frequently used cognitive strategy than to learn a new, more effective, one (Kuhn, 1999).

d) Active and collaborative learning

The particular teaching methodologies used in university classrooms also appear to have a significant impact on the development of critical thinking skills. The use of so-called 'active' teaching methodologies has consistently been found to have an impact on the cognitive development of students. Dewey (1933 (renewed 1960)) was the first to argue that project-based learning was the best way to stimulate reflective thinking. A number of studies have since suggested that Dewey's assertions were correct, as they have demonstrated the positive influence of experiential, problem-based and inductive pedagogical techniques on the development of critical thinking ability (Kuhn, 2005; Osborne, 2010; Renner & Lawson, 1975; Renner & Paske, 1977). Authentic learning experiences that approximate real-world scenarios seem to be particularly helpful for increasing the likelihood of transfer, as the classroom learning environment more closely approximates the situations in which students will one day be asked to apply their skills (Bransford, Brown & Cocking, 1999; Kuhn, 2005). The use of ill-structured problems in class also appears to be a useful strategy for helping students to develop their epistemological understanding and practice their metacognitive skills (Kitchener & Fischer, 1990). In addition to the use of active teaching methods, studies have found that collaborative learning techniques, such as the incorporation of class discussions and group projects into course curricula, can encourage the development of critical thinking skills (Johnson, Johnson & Smith, 1996; Karabenick & Collins-Eaglin, 1996; McKeachie *et al*, 1986; Skon, Johnson & Johnson, 1981; Tsui, 2002).

However, not all active and collaborative methods yield gains in critical thinking (Arum and Roksa, 2011; Blaich and Wise, 2010). Some of the specific instructional techniques found to improve critical thinking ability include class presentations, group projects and class discussions (Tsui, 2002). The crucial dimensions of such activities for cognitive development are the level of active engagement with the subject matter and the amount of interaction with various viewpoints and perspectives.

Class discussions that require students to defend their own positions on controversial issues, for instance, have been found to positively influence the development of critical thinking skills (Paul, 1981). Similarly, although traditional lecturing methods have been found to be ineffective for encouraging critical thinking (Tsui, 2002), lecturing that requires a substantial level of internal or external student dialogue appears to have a potentially positive impact on intellectual development, given the high level of interaction

and substantial amount of exposure to alternate perspectives inherent in such a method (Baxter Magolda, 1999). Interactive lectures and discussions can also provide an opportunity for instructors to model the standards or criteria used within a particular discipline when assessing claims or evaluating evidence (Baxter Magolda, 1999; Osborne, 2010). In contrast, class discussions in which instructors test for student recall or allow students to simply ask clarifying questions about the course content do not seem to have the same effect on cognitive development.

Group projects require a similar level of student participation and engagement with other perspectives in order to influence critical thinking ability (Kember & Leung, 2005; Smith, 1977; Smith, 1981; Terenzini *et al*, 1995). Much like class discussions, group projects can be beneficial for the development of critical thinking ability when they require students to explain or defend a position or when they force students to contend with perspectives different from their own (Rosenshine & Meister, 1992). Group projects that focus on consensus, on the other hand, are unlikely to impact critical thinking. A lack of diversity in the opinions or understandings of group members can also limit the benefit of collaborative learning for cognitive development (Howe, Tolmie & Rodgers, 1992).

The overall weight of the evidence suggests that active and collaborative learning methods can have a significant impact on the cultivation of critical thinking skills. However, such activities must require students to engage with multiple viewpoints and develop their own perspectives on complex issues in order to result in such cognitive gains.

e) Assessment and feedback

Studies have also demonstrated strong links between assessment practices and critical thinking. The format of examinations appears to play an important role in cognitive development. There is wide consensus around the superiority of essay-based examinations for assessing student understanding (Entwistle & Entwistle, 1997). However, an open-ended format in itself does not appear to be sufficient for ensuring student engagement. Essay questions must challenge students to demonstrate their own understanding of the course content by applying their knowledge to new situations in order to truly influence cognitive development. In contrast, essay questions that can be answered by regurgitating lecture notes do not appear to stimulate a high level of challenge, as students are not required to demonstrate any deep understanding of the subject area (ibid.).

As with other elements of the curriculum discussed in this section, synthesising examination questions are only likely to stimulate cognitive development if students are

provided with a sufficient level of support to enable them to succeed. For instance, students tend to demonstrate higher ability on synthesising examination questions if they have the opportunity to practice such techniques in class and through other course assignments (Entwistle & Entwistle, 1997; Kember & Leung, 2005; Tsui, 2002). The provision of feedback – both during the term and in relation to examination performance – also appears to play a crucial scaffolding function, as it helps students to understand their current level of ability in relation to the outcomes that they are striving to reach (Bransford, Brown & Cocking, 1999; Kuh & the Documenting Effective Educational Practice Project, 2005; Pellegrino *et al*, 2001; Rosenshine & Meister, 1992; Wiggins, 1998; Wood, Bruner & Ross, 1976).

Assessment practices also seem to have a strong influence on student learning orientations. Students demonstrating a surface approach to learning are likely to select studying strategies that require a minimum amount of effort (Entwistle & Entwistle, 1997). For instance, students with such a learning orientation typically opt to memorise lecture notes or review answers to examinations given in previous years, rather than testing their deeper understanding of the course content (ibid.). However, the regular use of varied and open-ended assessment questions that require students to apply their understanding of the subject matter to new situations is likely to stimulate students to engage more deeply with the course material.

B. Interactions with faculty and peers

a) Faculty-student interactions

Studies have also examined how interactions with faculty members impact gains in critical thinking ability. Pascarella and Terenzini (2005) reference a number of studies suggesting correlations between self-reported gains in critical thinking ability and the frequency of faculty-student interaction outside the classroom. Smith (1977; 1981) observed a significant correlation between faculty encouragement of students and gains in critical thinking ability, a finding corroborated by Terenzini *et al* (1995), while Pithers and Soden (2000) found that negative interactions with faculty members seemed to prohibit the development of student critical thinking skills. Entwistle (1997) has also observed that relationships between faculty and students can have a significant effect, either positive or negative, on student motivation.

However, other studies have found little systematic connection between faculty-student relationships and critical thinking ability. Arum and Roksa (2011) could identify no significant effect, while the Wabash study (Blaich & Wise, 2010) found that the only element of faculty-student interaction that seemed to have an influence on critical

thinking was "good teaching". Interactions outside of the classroom seemed to have little impact.

The weight of the evidence suggests that the most significant influence of faculty-student interactions on critical thinking ability is the faculty-student dynamic within the classroom. Instructor behaviour, in terms of attitudes towards students, classroom practices and provision of feedback, can have a marked influence on the cognitive development of students. However, interactions between faculty members and students outside of the class, although potentially important for other aspects of student development, do not appear to have a systematic effect on the development of critical thinking skills.

b) Relationships with peers

Studies have also considered the impact of peer relationships on critical thinking ability. Kuh (1995) has actually argued that gains in cognitive complexity are more attributable to peer interactions than to academic experiences. In their review, Pascarella and Terenzini (2005) did identify a number of studies that indicated a relationship between peer interactions and cognitive growth (e.g. Astin, 1993a; Terenzini *et al*, 1995; Twale & Sanders, 1999). However, other studies have found little correlation between peer interactions and critical thinking ability (Arum and Roksa, 2011; Blaich and Wise, 2010).

As with collaborative instructional techniques, these discrepancies are likely to be attributable to qualitative differences between different types of peer interaction. Much like class discussions and group projects, peer interactions are only likely to have a significant impact on cognitive growth if they allow students to engage with multiple viewpoints and perspectives. Other types of peer interactions are unlikely to have any systematic effect on critical thinking.

C. Extracurricular activities

A handful of studies have also considered the impact of extracurricular activities on critical thinking ability. Some have demonstrated that involvement in extracurricular activities can have a positive impact on critical thinking (Inman & Pascarella, 1998; Kuh, 1995; Terenzini *et al*, 1995). However, in their review, Pascarella and Terenzini (2005) could identify no systematic correlation between extracurricular involvement and critical thinking across institutional contexts. This is likely due to the same effects discussed in relation to collaborative learning and peer interactions. Extracurricular activities are only

likely to affect the development of critical thinking skills if they encourage exposure to a range of new ideas and perspectives.

D. Institutional Culture

In addition to considering the impact of particular aspects of the university experience on cognitive development, a few studies have examined how an institution's culture may influence the development of critical thinking skills. Kuh (1995) has argued that an institution's "ethos" is significantly correlated with gains in cognitive complexity, while Terenzini *et al* (1994) found that the stronger an institution's "scholarly emphasis", the greater the evidence of cognitive growth. As both "institutional ethos" and "scholarly emphasis" can be interpreted in myriad ways, the most conclusive evidence regarding the influence of institutional culture is Tsui's (2000) analysis of the particular aspects of university culture that appear to impact the development of critical thinking skills. Tsui qualitatively investigated the role of culture at four institutions: two with demonstrably high levels of critical thinking ability and two with low levels of critical thinking ability, as assessed via a national quantitative assessment. She identified three elements of institutional culture as being particularly significant for the cultivation of critical thinking skills. First, an institution's epistemological orientation appeared to play a significant role. Tsui observed that the overall emphasis of the curriculum at the institutions with high levels of student critical thinking ability was on exploring what was not known, rather than transmitting what was known. As a result of their epistemological orientation, the institutions were more likely to use active teaching methods, such as class discussions about ill-structured problems, and to assess student learning via open-ended, synthesising assignments. Second, she identified a positive correlation between high critical thinking ability and an institution's emphasis on student responsibility and self-reflection. Finally, institutions with high levels of student critical thinking ability appeared to actively encourage high levels of social and political awareness within the student population.

Tsui (2001) also found that institutional culture could significantly influence faculty attitudes. In her study, she identified a strong correlation between job satisfaction and the likelihood of faculty members to demonstrate the attitudes found to positively influence critical thinking. Institutional culture and policies, meanwhile, appeared to have a significant impact on job satisfaction. Collegiality within departments and exchange around pedagogical practices emerged as positive influences on faculty morale, as did strong institutional leadership, while incentive structures were found to contribute substantially to faculty job satisfaction. The provision of regular pedagogical training was also found to have a significant impact on faculty attitudes and classroom behaviour. The

results of Tsui's study suggest that institutional culture can have an indirect effect on the cultivation of critical thinking skills through its direct impact on faculty attitudes.

3.4.3 Student engagement and critical thinking

As discussed earlier in the chapter, the literature suggests that student engagement is the single most important factor influencing the development of critical thinking ability (Astin, 1984; Astin, 1993b; Carini, Kuh & Klein, 2006; Kuh & the Documenting Effective Educational Practice Project, 2005). Incoming attitudes and motivations clearly affect student engagement levels. However, research has indicated that institutions can also play an important role. Perhaps unsurprisingly, many of the experiences found to increase student engagement are the same as those found to have a significant impact on critical thinking ability.

Much like the literature on critical thinking, studies have demonstrated that institutions can encourage student engagement by implementing active and collaborative learning methods, providing an appropriate level of academic challenge and incorporating feedback to assist student learning (Carini, Kuh & Klein, 2006; Kuh & the Documenting Effective Educational Practice Project, 2005). Enriching educational experiences, such as internships, field visits and culminating final-year projects, also seem to have a significant impact on student engagement (Kuh & the Documenting Effective Educational Practice Project, 2005). An understanding of prior learning, a supportive campus environment and an institutional culture focused on student learning also appear to be important (ibid.).

In recognition of the importance of student engagement for cognitive growth, institutions in the U.S. created a survey, known as the National Survey of Student Engagement (NSSE), to assess the implementation of the effective educational practices found to encourage student engagement. The NSSE collects information about five clusters of indicators: Level of Academic Challenge, Active and Collaborative Learning, Student Interactions with Faculty Members, Enriching Educational Experiences, and Supportive Campus Environment. Students at participating institutions complete the survey, documenting their experiences in relation to the five clusters, and aggregate scores are used to determine institutional progress on the various indicators. In a recent study, Pascarella *et al* (2010) considered the effectiveness of the NSSE benchmarks as predictors of educational outcomes, controlling for incoming ability. Overall, they determined that the NSSE indicators were stronger predictors of educational outcomes than the indicators currently included in international rankings of institutions. Furthermore, they found that Level of Academic Challenge, Active and Collaborative Learning, Enriching Educational Experiences, and Supportive Campus Environment were all significantly correlated with gains in reasoning and problem solving. The only scale

with no effect on cognitive development was the Student/Faculty Interaction scale, likely because the NSSE considers the frequency of Student/Faculty Interactions, rather than the qualitative dimensions of such interactions.

It is unclear if the results of the Pascarella *et al* study imply that the NSSE benchmarks *directly* increase the critical thinking ability of students or if the benchmarks increase student engagement, thereby *indirectly* leading to cognitive growth. Given the similarities between factors influencing cognitive development and student engagement, it is perhaps most likely that institutions implementing many of the best practices in the NSSE simultaneously increase student engagement and create the conditions that encourage the development of critical thinking skills. Regardless of the precise process, it is clear that many of the indicators included in the NSSE have a significant influence on the development of critical thinking skills at university.

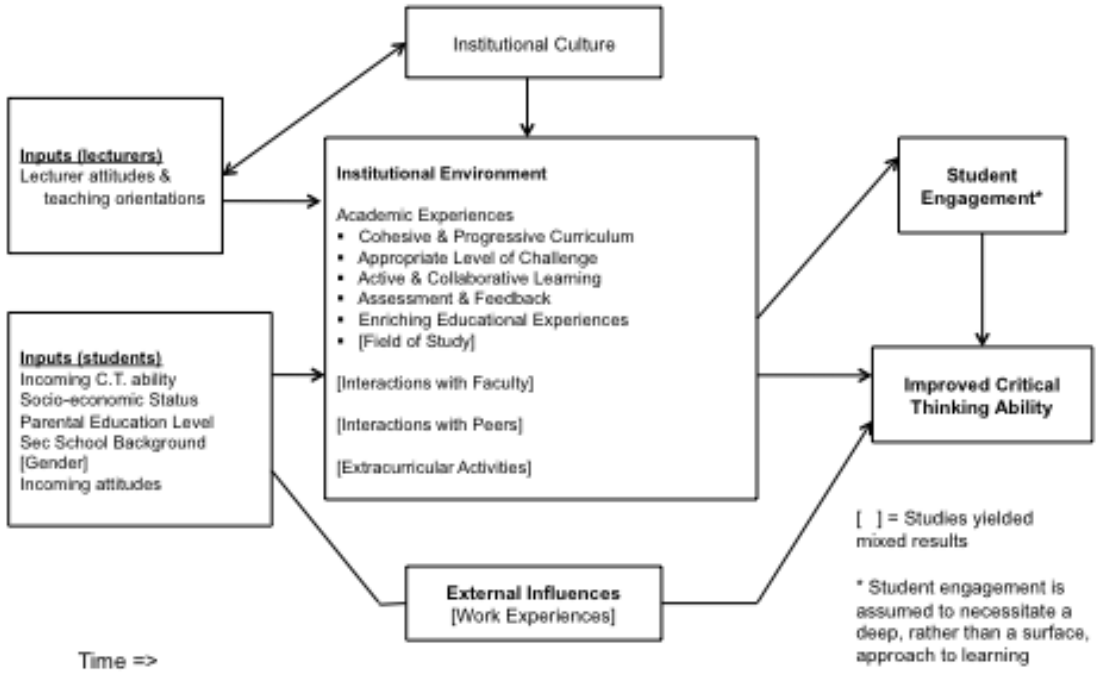
3.4.4 The influence of external experiences on critical thinking

Before concluding this section, it is important to acknowledge that university students also have experiences away from their institution, which may have a positive or negative impact on the development of their critical thinking skills. Although such external experiences may be highly relevant, very few studies have explicitly examined correlations between critical thinking ability and external influences. The only area of inquiry attracting significant interest to date has been the potential impact of work experiences on cognitive development. The findings in this vein have so far been largely inconclusive (as discussed in Pascarella & Terenzini, 2005). Although some theorists have found that work experiences can help with intellectual development (Baxter Magolda, 1992), others have argued that work experiences off campus can encourage surface approaches to learning, as students have limited time for academic work when burdened by work responsibilities (Aronowitz, 2000). Given the diversity of possible work experiences and the confounding influence of other factors that may require students to work during university, no systematic correlation can be assumed between work experiences and the development of critical thinking skills. For some students, work experiences may help with learning outcomes, while others may find that work responsibilities have a detrimental effect on their development (Kuh, 1995).

3.5 Conceptual Framework

Findings from the literature reviewed in the previous sections were aggregated into one overarching conceptual framework (presented below as Figure 3.1). The framework, based loosely on Astin's Input-Environment-Outcome model (1970), outlines the inputs, institutional characteristics and external influences that appear to affect the development of critical thinking skills at the university level.

Figure 3.1: Conceptual Framework



Arrows in the diagram denote the direction of influence. Most arrows move forward, indicating, for example, the influence of inputs on the university environment. Double-headed arrows are used to demarcate factors that can both act as an influence and be influenced by other factors. For example, faculty attitudes are likely to have an impact on institutional culture, but studies also suggest that culture can influence faculty attitudes. Lines without arrows represent movement through time without any direct influence of one factor on another.

All of the factors discussed in this chapter have been included in the framework, with two exceptions. First, as students in Rwanda do not generally have the opportunity to select their field of study at the university level[30], field of study was not included as an

[30] As a condition of their government bursary, students are generally assigned to a particular department based on their National Examination score. Public institutions in Rwanda also have

input variable. Second, race/ethnicity was explicitly excluded, as the investigation of ethnic differences in post-genocide Rwanda is a highly inappropriate and unwelcome practice.

3.6 Investigating Critical Thinking in Rwanda

Although a useful summary of the existing evidence linking individual and institutional factors to the development of critical thinking at universities, the conceptual framework could only be used as a guide to the current study if gauged to be potentially applicable to the Rwandan context. The simple application of Western theories and concepts to non-Western contexts is a problematic practice, given differences in the underlying assumptions and behaviours that operate in different cultures (Crossley & Watson, 2003). It was therefore important to consider the relevance of the conceptual framework for use in the study context. The final section of this chapter considers the potential threats to the framework's applicability in Rwanda and outlines the rationale for proceeding with using it as a guide to the study's design and implementation.

3.6.1 Culture and critical thinking

Although critical thinking is referenced as an important learning outcome in many different national contexts, there is a lack of consensus around whether or not critical thinking is a universal construct. In fact, there is a substantial literature that argues that critical thinking is a uniquely Western approach to reasoning. The debate has generally focused on differences between Western and Eastern models of cognition, likely due to the rising numbers of Asian students studying abroad in Western universities. There has been little discussion of how critical thinking is or is not manifested in African contexts. However, the central arguments in the Western/Eastern debate are instructive for this study, as they highlight the major concerns in assuming similarities in cognitive development across cultures.

Barbara Rogoff (2003) has written extensively on the role of culture in cognitive development. Firmly grounded in the situated cognition tradition, Rogoff argues that cognitive development can only be understood in light of cultural practices, as cognitive functions develop in different ways depending on cultural circumstances. She also asserts that cultural practices change over time and that individuals are often concurrent members of multiple cultural groups, meaning that cognitive development can never be assumed to progress in a similar way across – or even within – cultures. In her study of

---

strict enrolment caps on many undergraduate programmes, meaning that even fee-paying students may not be given their first or second choice of academic subject.

cognition in various cultural contexts, Rogoff highlights myriad cultural practices that can influence cognitive development, including the relative importance of independence versus interdependence – or cooperation versus competition – within a society and the norms surrounding discipline and child rearing. Cross-cultural psychologists have observed differences between cultures in terms of the willingness of students to distinguish themselves from others (Philips, 1972), the interpretations of problems that need to be solved and the proper methods for solving them (Goodnow, 1976), and the definition of what constitutes 'intelligence' (Wober, 1972). There are also cultural norms surrounding the appropriateness of challenging authority figures, leading to significant differences in interactions between instructors and students in various cultural contexts (Harkness & Super, 1982; Omokhodion, 1989).

A number of theorists have argued that such cultural differences lead to fundamental dissimilarities in cognition. Nisbett *et al* (2001), for example, have argued that epistemology is strongly influenced by culture, suggesting that there are fundamental differences between Western and Asian systems of thought. Norenzayan *et al* (2002) claim that Western and Asian students differ in their preference for using formal or intuitive reasoning, suggesting that such differences are likely to be the result of dissimilarities in cultural norms and pedagogical practices. In her analysis of the critical thinking ability of Asian students in the U.K., Durkin (2008) observed that Asian students were less likely to demonstrate criticality in their academic work, not because of a lack of ability but because of an aversion to critique, which they tended to see as a confrontational, and therefore offensive, practice. Culture has also been found to impact what individuals remember, an important element of cognition (Deregowski, 1970).

However, not all scholars agree that there are fundamental differences in cognition between cultures. Chan and Yan (2008) have argued that, although *preferences* regarding the use of reasoning are culturally based, reasoning itself is a human attribute that cannot differ across cultures. In their view, the preference to use a particular type of reasoning is entirely dependent on the scenario in question. As cultural expectations are connected to the particulars of specific scenarios, preferences would tend to differ between cultures. However, they argue, people are equally likely to form logical judgments, regardless of their culture of origin, depending on their level of education and their exposure to the use of reasoning to address problems. Ryan and Louie (2008) present a similar argument, suggesting that it is problematic to conflate Asian students' disposition to use critical thinking with their critical thinking ability. As disposition relates to behaviour, they argue, it is likely that culture influences disposition. However, it does not follow that differences in ability fall along cultural lines. Studies of problem solving in different cultures support

such a differentiation between ability and disposition, as different cultural groups have been found to use the same cognitive functions but to manifest such functions in different ways depending on context (Cole, 1990).

According to Kuhn (1999), critical thinking disposition can only be identified once an individual has reached a sufficiently advanced level of epistemological understanding. However, some theorists have argued that epistemology itself is inherently cultural, as the value placed on different kinds of knowledge and their use may differ between cultural contexts (Nisbett *et al*, 2001). Such a perspective suggests that the ability to evaluate competing claims or contradictory evidence may not be viewed as necessary or important in some cultural contexts. Others disagree with this perspective. Baxter Magolda (1999), for instance, has argued that, much like the differences she observed between the developmental patterns of male and female students, it is the patterns of epistemological development, rather than the stages themselves, that are likely to differ between cultural contexts. Her theory of epistemological development postulates that male and female students progress through the same four stages of intellectual development in fundamentally different ways. For example, she observed that most female students were likely to make the transition from Absolute Knowing (where students believe that knowledge exists and is either right or wrong) to Transitional Knowing (where students admit that uncertainty is possible in particular circumstances) if they had the benefit of a supportive peer community in which they could experiment with expressing their own views in a safe environment. Male students, in contrast, seemed to be more likely to make the transition if they were challenged by their instructors to exchange views with others and seek logical explanations for perceived uncertainty. Although the end-point was the same, the pathway therefore differed for male and female students. Baxter Magolda has argued that culture is likely to affect epistemological development in the same way, with individuals from different cultures progressing through the same stages of epistemological sophistication but along different pathways depending on cultural preferences and behaviours.

Although the disposition to use critical thinking skills is likely to be affected by cultural influences, there is little evidence that critical thinking ability is a culturally specific construct. However, there is no doubt that education is a cultural process. As a result, it is highly likely that the fostering of critical thinking skills through education differs between cultures. Culture can both help and hinder cognitive development. For example, in the U.S., students are taught to value *what* they think more than the reasoning behind their thinking, which can be a significant barrier to the teaching of critical thinking skills in American classrooms (Kuhn, 2005). Education can also foster differences in

dispositions, as instructors tend to model the use of critical thinking skills in different situations. Classroom practices can, therefore, encourage or discourage the use of critical thinking skills in various domains and contexts (Chan & Yan, 2008; Okagaki & Sternberg, 1990). Cultural norms can also influence classroom behaviour, which in turn can affect the development of critical thinking skills. As discussed earlier in the chapter, active participation in classroom discussions or small group projects can have a significant influence on critical thinking. However, cultural norms may prevent students from actively participating in such activities, thereby limiting the influence of such practices in certain contexts.

Cultural norms also appear to influence how students approach learning (Richardson, 1994), as does linguistic ability. Kember and Gow (1990) have found that the effort required to learn in a foreign language is likely to encourage surface approaches to learning. Low levels of linguistic ability may also impact student participation in educational activities, as students tend to withdraw and avoid engaging in dialogue if they are not confident in their communication skills (Shizha, 2012).

Although largely focused on Asian cultures, the themes outlined in this section are relevant for any discussion of critical thinking in Rwanda. Rwandan students do not study in their native language of Kinyarwanda. The effects of learning in a foreign language are therefore likely to influence the cognitive development of Rwandan students. Scholars have also written extensively about the expectation of obedience to authority within Rwandan society (Pottier, 2002; Prunier, 1995; Tikly & Great Britain Department for International Development, 2003). As authoritarianism has been found to override the potential to exercise individual cognitive ability in other African contexts (Wiredu, 1980), it is possible that Rwanda's authoritarian political structure influences the disposition of Rwandan students to use critical thinking outside the classroom. Some scholars have argued that, despite the diversity of African cultures, there are commonalities in African "ways of thinking", such as the privileging of collective decision-making over individual analysis and the reliance on intuitive, rather than analytical, reasoning (Ngara, 2012; Senghor, 1962). Such an argument implies that critical thinking would not be valued in many African societies. Although traditional Rwandan culture certainly shares some of these features, the inclusion of critical thinking as an important learning objective in recent policy documents suggests that reasoning and problem solving may have risen in prominence in contemporary Rwandan culture. Cultures are changing rapidly around the world, given the influences of globalisation and migration (Crossley & Watson, 2003; Yang, 1988). Rwanda has experienced a particularly dramatic transformation in the years since

the genocide, given the repatriation of many members of the diaspora from a multitude of cultural contexts (Tikly & Great Britain Department for International Development, 2003). It is likely that such social transformation has affected – and will continue to affect – the perceived value of critical thinking, both within the Rwandan education system and in wider society.

As critical thinking has been advocated as a particular goal of a university education in Rwanda, a study of how such skills are being fostered was deemed to be an appropriate and relevant area of inquiry. However, given the likely influence of culture on classroom norms and the disposition to use critical thinking skills outside the learning environment, it seemed probable that culture might affect the applicability of the study's conceptual framework to the Rwandan context. In particular, it seemed likely that critical thinking might be affected by additional inputs or that certain aspects of the university environment found to positively influence critical thinking in other contexts might not have the same effect in Rwanda. Given the lack of directly applicable contextual evidence and its reliance on findings from different cultural contexts, the conceptual framework was retained as the best available guide for the study. However, the potential for differences between the Rwandan context and the contexts of the studies included in the framework remained a central consideration throughout the study's implementation.

3.6.2 Assessing critical thinking in Rwanda

The influence of culture was also an important consideration in the determination of how critical thinking might be assessed in the study context. Rogoff (2003) has argued persuasively against the use of assessments across different cultural contexts. As assessments always include cultural nuances and references, an assessment developed in one context cannot validly be used in another, unless the content has been tested and found to be equally comprehensible to the new population. Otherwise, assessment results reflect both levels of cultural understanding and levels of ability on the central construct, thereby undermining the construct validity of the instrument.

Despite such concerns, a number of recent studies have used Western instruments to assess the critical thinking ability of non-Western students (Bataineh & Zghoul, 2006; Lombard & Grosser, 2004; Lombard & Grosser, 2008; Osman & Githua, 2009; Saavedra & Saavedra, 2011). Although this is a questionable practice in any study of critical thinking in a non-Western cultural context, it is particularly problematic when used in cross-cultural comparison. In their recent studies of critical thinking ability in South Africa, for example, Lombard and Grosser (2004; 2008) compared the results of South African students on an American test of critical thinking ability with a similar sample of students in the U.S. Unsurprisingly, the South African students achieved lower scores than their

American counterparts. However, the results have very little practical meaning, as the poor scores of South African students may have had more to do with their cultural background than their critical thinking ability.

In addition to problems caused by the cultural nuances of test questions, assessments developed in other cultural contexts often pose a linguistic challenge to test takers. The relationship between thought and language has been discussed extensively in the literature. Comprehension and communication can both influence thought, indicating an intimate relationship between linguistic and cognitive ability (Greenfield & Bruner, 1969; Halpern, 1996). Scholars have largely refuted the idea that one's language determines the limits of one's thoughts (Cole & Scribner, 1974). However, there is an obvious connection between linguistic ability and the demonstration of cognitive practices, given that cognition can never be observed directly and must always be mediated through some form of communication.

The importance of language is particularly relevant when assessing people in a language other than their mother tongue. A number of recent empirical studies have investigated the demonstrated critical thinking ability of Western and Asian students. As the studies have relied on the use of English assessment tools, any observed differences have been attributable to the cognitive demands of using a second language, rather than any actual difference in critical thinking ability (Floyd, 2011; Lun, Fischer & Ward, 2010). Takano and Noda (1993) observed that students demonstrate a temporary decline in thinking ability when processing a foreign language. This phenomenon has been referred to as the "L2 cognitive deficit" (Long & Harding-Esch, 1977). Thinking in a second language is generally agreed to impose limits on working memory (Just & Carpenter, 1992) and on the ability to process information (Clahsen & Felser, 2006). Cognitive load theory suggests that only limited amounts of information can be stored and processed in working memory, implying that the use of a second language would have a detrimental effect on performance on any cognitive task (Plass, Moreno & Brünken, 2010). It is also likely to affect the speed of work, as students have to spend more time on comprehension when working in a second language (ibid.). The implication of these findings is that the use of a foreign language is likely to limit demonstrated critical thinking ability, meaning that the results of an assessment completed in a second language would tend to indicate lower ability than an assessment completed in one's mother tongue.

It is clear that the validity of a study is significantly improved if it relies on an assessment tool that has been created for or adapted to the particular study context. Generally, the more an assessment has "indigenous validity" (Irwin, Klein & Townsend, 1982) within a population, the more likely it is that the assessment will actually sample

from the intended domain under consideration (Cole & Scribner, 1974). Much as students appear to be less able to demonstrate critical thinking ability in unfamiliar subject areas, the assessment of critical thinking through culturally unfamiliar testing scenarios is likely to substantially underestimate a respondent's optimal critical thinking ability. In contrast, the incorporation of familiar tools and scenarios can improve performance on assessments, as students are better able to understand the content of test questions (Serpell, 2007), reducing cognitive load and allowing for the use of more complex cognitive strategies (Lun, Fischer & Ward, 2010).

As discussed earlier in the chapter, any assessment of critical thinking relies on the selection of a valid assessment tool for a particular context and purpose. The cultural dimension of the current study added an extra layer of complexity to the assessment selection process, as no assessment tool had been tested for use in Rwanda prior to the implementation of the study. The question of how to validly assess critical thinking in the Rwandan context, therefore, emerged as a fundamental component of the study design.

A clearly articulated conceptual framework forms the basis of any effective research design, as there should always be a clear theoretical link between a study's objectives, research questions and methodology (Miles & Huberman, 1994; Robson, 2011). The conceptual framework presented in this chapter served as such a foundation for this study, guiding the implementation of data collection and the selection of variables to be considered during analysis. Details of the chosen study design are outlined in the next chapter.

Chapter 4: Research Methodology

This chapter moves the discussion from the theoretical to the empirical domain, as it presents the methodology chosen to investigate the theoretical understandings of critical thinking outlined in Chapter 3. The study objectives and central research questions will be delineated, and the overarching methodological approach will be specified. The chapter closes with a discussion of the major ethical considerations of the study design.

4.1 Study Objectives & Research Questions

The logic of the conceptual framework suggested two primary objectives for the study. First, the study aimed to determine whether or not students in Rwanda are improving in their critical thinking ability during university. Second, it aspired to identify the factors that seem to help or hinder the acquisition of critical thinking skills at Rwanda's universities. These objectives were re-articulated as a series of research questions to guide the study design:

1) How can critical thinking be appropriately assessed in the Rwandan context?
2) Is there evidence that Rwandan students are improving in their critical thinking ability during their time at university?
3) What individual and institutional factors appear to be associated with student critical thinking ability in Rwanda?
4) How are Rwandan universities helping (or hindering) the development of critical thinking skills in their students?

4.2 Overarching Research Design

Researchers in other contexts have approached similar topics through a variety of different methods. In their study of American universities, Arum and Roksa (2011) considered a comparable set of research questions by analysing the results of a large national dataset over a two-year period. Participants in the Arum and Roksa study took a critical thinking assessment at two different points in their university career. In addition, they completed a version of the NSSE and a survey of their motivations and attitudes. The data was analysed quantitatively to determine the impact of individual and institutional factors on gains in critical thinking during university. Saavedra and Saavedra (2011) also

took an entirely quantitative approach in their study of critical thinking in Colombia, analysing correlations in a large national dataset between scores on a critical thinking assessment and various individual and institutional variables. Unlike Arum and Roksa, their study did not follow a longitudinal structure. Rather, they used a cross-sectional approach, comparing first-year and fourth-year scores as a proxy for growth. Tsui (1998) approached similar questions in a different manner, relying on qualitative analysis of institutional environments. Using the results of a national quantitative study of critical thinking, she first analysed the correlation between a number of institutional factors and self-reported gains in critical thinking. Determining that institutional selectivity was positively correlated with self-reported gains, she opted to complete case studies at four "extreme case" institutions (one high selectivity/low critical thinking, one high selectivity/high critical thinking, one low selectivity/high critical thinking and one low selectivity/low critical thinking). At each institution, Tsui conducted focus groups, individual interviews and observations, using the resulting data to determine which institutional factors seemed to be associated with high gains in critical thinking.

Although fundamentally different in their approach, all three studies provide rich insights into the factors that influence the development of critical thinking skills in university students. As there are both quantifiable and non-quantifiable factors that impact the development of critical thinking skills, it is logical that both quantitative and qualitative approaches would be relevant to the study of critical thinking. However, none of these prior investigations offered a relevant research design for this study, as all three relied on the use of a national dataset involving a substantial number of institutions. As critical thinking has never been assessed at the national level in Rwanda, there were no pre-existing datasets available for consultation.

Furthermore, the study objectives could not have been satisfactorily addressed through either a purely quantitative or a purely qualitative approach. As the study required the development of a new assessment tool that would be relevant for use in Rwanda (see Chapter 5), the implementation of a purely quantitative investigation, similar to the Arum & Roksa and Saavedra & Saavedra studies, would have relied entirely on the results from a previously untested assessment tool. There were too many validity concerns with this approach to consider a purely quantitative design. It would also have been very difficult to consider the impact of non-quantifiable factors, such as institutional culture, via purely quantitative analysis. However, it would have been impossible to assess general trends in critical thinking ability through a purely qualitative approach. Tsui's study demonstrated the feasibility of considering non-quantifiable factors through the use of case studies. On the other hand, she was only able to select theoretically

relevant cases because of the existence of pre-existing national data on critical thinking. Without any wider assessment of critical thinking ability, a qualitative exploration of institutional factors would have been of little use in the articulation of future higher education policy in Rwanda.

The study design was therefore based on a combination of the two approaches. A mixed methods design offered the prospect of both precision and in-depth understanding (Dowling & Brown, 2010). Furthermore, such an approach allowed for the consideration of questions requiring both quantitative and qualitative data (Creswell & Plano Clark, 2007; Mason, 1994). Johnson *et al* (2007) have defined mixed methods research as "the type of research in which a researcher…combines elements of qualitative and quantitative approaches…for the purposes of breadth and depth of understanding and corroboration" (p. 123). As the study aimed to contribute to higher education policy in Rwanda, the identification of macro-level trends in critical thinking ability within Rwanda's university population (breadth) was seen to be fundamental. At the same time, an exploration of how critical thinking is understood and valued in the Rwandan context (depth) was viewed as a crucial element in the elaboration of meaningful implications for policy.

The combination of quantitative and qualitative research can take myriad forms, depending on the emphasis of the study in question. In any mixed methods study, decisions must be made regarding the timing, weight and integration of the components (Creswell & Plano Clark, 2007). As this study aimed to influence national policy, the primary emphasis was on the quantitative analysis of critical thinking in the student population. Qualitative data were largely intended to serve an explanatory function in the analysis by offering potential explanations for the quantitative results (Bryman, 2006). Qualitative methods were also used to identify additional factors that were not included in the original conceptual framework, thereby complementing the data included in the primary data set (Greene, Caracelli & Graham, 1989).

Although integration for the purposes of complementarity can take a variety of forms, explanatory integration relies on a sequential research design, as the first phase clarifies which findings will require additional explanation during the second phase. Typically, collection of the data set of emphasis is completed first, so that the second data set can provide explanations for the results of primary interest. Studies emphasising quantitative results, therefore, often begin with a population survey, the results of which are then explored and clarified through qualitative techniques (Kemper, Stringfield & Teddlie, 2003; Yin, 2009). This study followed such a sequential format, beginning with a quantitative assessment of critical thinking ability and concluding with a qualitative investigation of the assessment results.

4.3 Research Methodology

The study was organised into three broad phases: Design, field-testing and piloting of the research instrument; Quantitative data collection and analysis; and Qualitative data collection and analysis.

4.3.1 Phase One: Design, field-testing and piloting of the research instrument

As all of the existing assessments of critical thinking were written for use in non-African contexts, the study began with the design of a contextually relevant assessment tool. This phase was therefore focused on the first of the study's research questions: How can critical thinking be appropriately assessed in the Rwandan context? During the assessment development phase, an appropriate assessment format was selected and adapted for use in the Rwandan context. The adapted tool was also pre-tested and piloted in order to verify its feasibility, validity and reliability as the central data collection instrument for the remainder of the study.

4.3.2 Phase Two: Quantitative data collection and analysis

Once an appropriate tool was identified, the second, quantitative stage began. During the quantitative phase, the adapted critical thinking assessment was administered to a random sample of first- and fourth-year students attending the public universities in Rwanda. Participants were also asked to complete a short demographic survey during this phase. The accompanying questionnaire collected information about the student-level input variables included in the conceptual framework, including participant gender, socio-economic status, educational background and family background. Participants were also asked to answer a number of basic questions about their university enrolment, including their field of study, year at university and enrolment status (full- or part-time). Assessment results were then analysed in order to address two of the study's research questions:

- Is there evidence that Rwandan students are improving in their critical thinking ability during their time at university?
- What individual and institutional factors appear to be associated with student critical thinking ability in Rwanda?

The structure of the second phase was based loosely on Arum and Roksa's study. However, unlike Arum and Roksa's design, this study did not follow a longitudinal structure. Most studies of university impact rely on either the analysis of longitudinal data

or the use of self-report measures of cognitive gains (Astin, 1991). Self-report measures are notoriously unreliable indicators of student growth (Pascarella, Seifert & Blaich, 2010), so such an approach was not pursued. The time frame for this study did not allow for the completion of a longitudinal study, so, instead, a cross-sectional approach was adopted. Cross-sectional designs can be problematic in contexts with high attrition rates (Saavedra & Saavedra, 2011). However, attrition is not a significant problem at Rwanda's public universities, given the high proportion of scholarship recipients. A cross-sectional design was therefore gauged to be a satisfactory method for investigating whether or not students in Rwanda appear to be improving in their critical thinking ability during university.

4.3.3 Phase Three: Qualitative data collection and analysis

The final stage of the study was inspired by Tsui's study design. Based on the results of the assessment in Phase Two, two of the participating institutions were selected for in-depth study during Phase Three. Analysis during the final stage aimed to move the study away from deficit model assumptions by focusing on stakeholder perspectives (Rogoff & Chavajay, 1995) and identifying potential differences between policy objectives and local realities (Marshall & Rossman, 2006). Within each institution, assessment results and elements of the conceptual framework were investigated through interviews with multiple stakeholders (including students, faculty members and senior administrators) and a review of institutional policies.

There are many qualitative methods that could have been used to probe the results of the quantitative stage. The case study approach was selected for a number of reasons. First, according to Yin (2009), case study inquiry "copes with the technically distinctive situation in which there will be many more variables of interest than data points, and as one result, relies on multiple sources of evidence, with data needing to converge in a triangulating fashion, and as another result benefits from the prior development of theoretical proposition to guide data collection and analysis" (p. 18). This definition resonated with this study, given that universities are complex, multivariate environments that require the consideration of multiple sources of evidence. Furthermore, as the study explicitly aimed to consider the applicability of a conceptual framework to a new empirical setting, the use of pre-existing theoretical propositions in case study analysis was seen to offer a distinct advantage. Case studies are also well suited for answering 'how' and 'why' questions, as they allow the researcher to consider the impact of various aspects of a setting on a given outcome (Yin, 2009). As the final stage focused on the fourth and final research question – How are Rwandan universities

helping (or hindering) the development of critical thinking skills in their students? – the ability of the case study method to address 'how' questions strongly recommended its use.

4.3.4 Integration of study results

Integration of the two data sets was completed prior to articulation of the study conclusions. Integration is often one of the most challenging aspects of mixed methods studies (Caracelli & Greene, 1993; Greene, 2007; Onwuegbuzie & Johnson, 2006). Integration in this study took inspiration from another investigation of the student learning experience in cross-cultural context. In their comparative study of students in Thailand and Australia, Aldridge *et al* (1999) used a large-scale assessment as a 'springboard' for further data collection. After administering a quantitative survey to students in the two contexts, interviews were conducted with participants from the assessment phase, as well as teachers from their classrooms. Qualitative analysis focused both on how participants understood the survey questions and on how students felt about their classroom environment. The final results, therefore, included both an elaborated version of the survey results and an in-depth understanding of the student experience in the two contexts. Aldridge *et al* used Denzin and Lincoln's concept of the researcher as "bricoleur" (1994, p. 2) to describe their integration technique, arguing that they were able to piece together different types of information, collected via different methods, in order to assemble a cohesive and consistent whole. This integration technique was selected for this study, as it allowed the use of data from one phase to both explain and complement the other, resulting in a more nuanced picture of how universities in Rwanda may be affecting the development of critical thinking skills in their students.

4.4 The Research Team

The strength of a mixed methods design is the depth of understanding that can potentially emerge from the data. However, the complex nature of the design also poses challenges. In particular, the design structure can be extremely challenging for one researcher to implement, given the difficulty of simultaneously collecting and analysing two entirely separate data sets. In order to address the logistical challenges posed by the study design, nine Rwandan university students were recruited to act as research assistants in the study.

In addition to helping with data collection and interpretation, the involvement of local research assistants introduced an insider perspective to the management of the study. Although familiar with the study context, the author's identity as a white, English-speaking Westerner positioned her as an outsider in relation to the study participants.

The question of the comparative advantage and disadvantage of outsiders and insiders in the research context has been discussed extensively in the literature (Hammersley & Atkinson, 1995; Liamputtong, 2010; Parker Webster & John, 2010; Robson, 2011). Hammersley and Atkinson (1995) have argued that insiders are often disadvantaged in ethnographic research, despite their understanding of local norms, given that it is sometimes harder for insiders to approach a topic with the objective stance necessary to view the diversity of factors affecting a given phenomenon. Outsider status also offers the advantage of relative neutrality and objectivity, and outsiders can be perceived as being more likely to protect the confidentiality of participants (Apentiik & Parpart, 2006). From a political standpoint, outsider status carries a potential advantage in Rwanda, as the closing of space for political opposition in Rwanda can prevent Rwandans from expressing negative opinions about current government policy (Hayman, 2009b). At the same time, the author's lack of fluency in Kinyarwanda, the Rwandan national language, was an obvious disadvantage, as it prevented the possibility of completing interviews or focus groups entirely in the first language of the study participants. The restriction of participation to those fluent in French or English would have significantly biased the results, as the link between linguistic proficiency and socio-economic status is quite pronounced in Rwanda (Samuelson & Freedman, 2010).

The use of local students as research assistants therefore allowed for the implementation of a complex study design, while also partially mitigating the potentially negative impact of the author's outsider status. As the research assistants were all recruited from the scholarship programme formerly directed by the author, the research team had a substantial pre-existing relationship prior to the implementation of the study. Trust is critical when operating in a cross-cultural context, particularly when the lead researcher is reliant on team members for interpretation and translation (Temple & Young, 2004), so familiarity within the research team was seen to be highly beneficial for the implementation of the study. Furthermore, all of the research assistants had participated in trainings offered by the scholarship programme, which had exposed them to pedagogy explicitly intended to develop their critical thinking skills. All of the research assistants were also highly proficient in English. The combination of their understanding of critical thinking and their linguistic proficiency made them uniquely qualified to assist with the study. Furthermore, all nine members of the research team were simultaneously enrolled as students at the public universities in Rwanda, making them full insiders in the study context. This offered a significant benefit, as they could easily facilitate contact with study participants and vouch for the sincerity of the lead researcher's intentions (Eide & Allen, 2005; Liamputtong, 2010).

4.5 Ethical Considerations

In addition to posing logistical challenges, the complex cross-cultural nature of the study design carried a number of ethical implications. First, the methods involved all required a significant investment of time from the study participants. In a resource-constrained environment such as a Rwanda, it can be problematic for a white, Western researcher to expect participation in a research study without acknowledgement of the time and effort provided by participants, especially given the history of social research as a tool of oppression under colonialism (Apentiik & Parpart, 2006; Burawoy, 1998; Liamputtong, 2010). At the same time, a study of this nature is best implemented through the willing participation of stakeholders. Incentives in the form of payment can have a biasing influence on the results, as participation can be driven by financial need, rather than motivation or interest (Liamputtong, 2010). This can be exacerbated by assumptions that researchers from the West always have access to substantial financial resources (Apentiik & Parpart, 2006). The research context provided some solutions to this ethical dilemma, as universities are traditionally supportive arenas for research. It was assumed that faculty members and administrators would be aware of the implications of participation in a research study, so incentives were judged to be unnecessary for these participants. However, the decision was made to provide refreshments and small non-monetary incentives in recognition of the time and effort provided by the student participants.

Another ethical concern was the potential for coercion when recruiting respondents. As student recruitment relied on the research assistants, it was vitally important for all members of the research team to thoroughly understand the ethical implications of participation in the study. It was therefore determined that a thorough training on the principles of informed consent would be a pre-requisite for officially joining the research team. Despite these efforts, the potential for involuntary coercion could not be entirely avoided (Berg, 2007). Foreigners, particularly Americans and English-speaking Europeans, hold positions of great privilege in Rwandan society. It is therefore possible that some selected students felt that they 'should' participate, whether or not they truly felt comfortable doing so. The perceived or actual interference of authority figures also posed a potential problem. As the central administration sanctioned the study at all participating institutions, faculty members may have felt that their participation was expected by the university authorities. Although administrators had no access to participant lists, some faculty members may have agreed to participate because they felt that they could not or should not refuse.

The issue of confidentiality is particularly acute in the Rwandan context. Although confidentiality is critical for all research, the political environment in Rwanda heightened the importance of protecting the identities of study participants. It is widely acknowledged that the Rwandan government actively discourages opposition (Hayman, 2009b). The closing of political space means that most Rwandans avoid disagreeing publicly with government policies. As it was possible that officials might interpret the phrase 'critical thinking' as implying the fostering of opposition (as discussed in Pinto & Portelli, 2009), confidentiality was a crucial consideration throughout the study. Although the names of the participating institutions could not be changed, given the small number of institutions in Rwanda, individual participant names were carefully protected. The research team was also thoroughly trained on the importance of participant confidentiality. This was particularly important given that some of the research assistants had pre-existing relationships with some of the respondents.

Dissemination was also an important component of the study design. Dissemination has ethical implications, as it officially acknowledges the contribution of respondents by granting them access to the work completed as a result of their participation (Binns, 2006; Liamputtong, 2010; Marshall & Rossman, 2006). However, it also plays another role in a study of this nature. As the study aimed to contribute to the higher education reform agenda in Rwanda, dissemination was seen to be a necessary prerequisite for any potential use of the study results.

Having thus outlined the rationale behind this study, the conceptual framework guiding it and the methodology proposed for its implementation, it is possible to proceed with the presentation of the study's central findings. The next chapter focuses on the selection and adaptation of an appropriate assessment for use in the Rwandan context. The remaining chapters are devoted to a discussion of the study results.

Chapter 5: Development of the Assessment Tool

The first phase of the study attempted to answer the study's first research question: How can critical thinking be appropriately assessed in the Rwandan context? This chapter outlines the process used to develop an assessment of critical thinking ability that could serve as the study's central data collection tool. The chapter begins with a review of the existing assessment formats available for use. A justification is then presented for the selection of the Collegiate Learning Assessment (CLA) as the most appropriate model for the study instrument. The process used to adapt the CLA for use in Rwanda is then described in detail. The chapter closes with an evaluation of the adapted assessment's feasibility, validity, and reliability.

5.1 Selecting an Appropriate Assessment Format

As discussed in Chapter 3, the use of un-adapted assessments across cultural contexts can be highly problematic, given the cultural nature of assessment tools. In order to avoid such cultural bias, it was necessary to begin the study with the identification of an assessment tool that could be validly used to assess critical thinking in the Rwandan context. This objective could have been accomplished through the development of an entirely new instrument. However, the development of a new assessment would have required substantially more complex design and testing procedures than the use of an existing format, as the tool would not have been previously evaluated for validity or reliability in other contexts. Given the time limits of this study, the creation of a new tool was not deemed to be a feasible option. Instead, the decision was made to select a pre-existing assessment format that could be adapted for use in Rwanda. This section presents the strengths and weaknesses of the available assessment options and outlines the rationale for selecting the CLA as the most appropriate model for use in the study context.

5.1.1 A review of existing assessment tools

The first step in the assessment selection process was to review the range of available tools. A number of recent studies of critical thinking assessment have included extensive discussion of existing assessment formats (e.g. Cook *et al*, 1996; Ennis, 1993;

Stein, Haynes & Unterstein, 2003; Williams, Wise & West, 2001). The review process for this study was largely informed by the analysis presented in these summaries[31].

5.1.1.1 Multiple-choice tests

The first assessments of critical thinking developed for use at the university level tended to be multiple-choice in nature. As these early instruments were developed by those equating critical thinking with formal logic (as discussed in Chapter 3), the tests in this category generally focus on the ability of students to evaluate arguments and identify logical fallacies.

Since the early days of critical thinking assessment, there has been a decided shift away from the use of purely multiple-choice tests for assessing critical thinking ability. The main critiques of multiple-choice critical thinking assessment highlight two significant limitations of the testing format. First, multiple-choice tests cannot test for dispositions or the process by which students arrive at a chosen answer (Johnson, 2009). As such, test scores may reflect the ability of students to predict the 'right' answers, rather than any ability to think critically about the reasons behind their selections. Second, multiple-choice tests do not test for creative thinking, as students are given no opportunity to propose their own alternatives (Hatcher, 2009). Multiple-choice tests are, however, easy to score, so they have often been used to assess large populations. It is also easier to test multiple-choice assessments for validity and reliability (Klein, Liu & Sconing, 2009). In recent years, theorists have considered the potential of a middle ground between strictly multiple-choice and strictly open-ended testing formats. Sobocan (2009), for instance, has suggested that current multiple-choice format tests could be improved by asking students to include a short explanation of a few of their selections, in order to demonstrate their thought process when answering questions.

The California Critical Thinking Skills Test (CCTST) is one of the most popular multiple-choice options. Built around the consensus definition of critical thinking that emerged out of the California Report in 1990, the CCTST includes 34 multiple-choice items that target the cognitive skills included in the report's definition (Facione, 1990). Although the CCTST is not intended to measure critical thinking dispositions, the developers of the CCTST created a parallel tool, known as the California Critical Thinking Disposition Inventory, which is typically administered alongside the CCTST (ibid.).

[31] All of the assessments referenced in this section were developed in the United States. Since the beginning of the Critical Thinking Movement, there has been a considerable market for critical thinking assessment in the U.S. Although new assessments have been created in other contexts, the tendency is for new assessments to follow one of the formats first proposed by American test developers. A focus on the range of test formats available in the U.S. was, therefore, seen as representative of wider trends in the field.

Despite the popularity of the CCTST, the test has a number of vocal critics. Groarke (2009), in particular, has argued that the CCTST is not an accurate measurement of any definition of critical thinking, given that one can reasonably defend multiple responses to many of the test questions. Furthermore, he claims that the test does not allow for alternative viewpoints and prevents any possibility of evaluating metacognitive ability, as respondents cannot demonstrate the thought process they use when making selections (ibid.). It has also been empirically demonstrated that the CCTST significantly disadvantages non-native speakers of English (Facione, 1990).

The Cornell Critical Thinking Test is another well-known critical thinking test that uses a multiple-choice format. It has not attracted the same criticisms as the CCTST, as the questions require respondents to rate the credibility of potential responses, rather than selecting one 'correct' answer to each question (Ennis, 2009). However, the multiple-choice format does prevent the possibility of using the test to measure critical thinking dispositions or the ability of students to self-regulate and/or propose alternative responses to questions.

5.1.1.2 Essay tests

The criticisms of multiple-choice critical thinking tests have inspired many institutions to adopt essay tests as a more valid format for assessing critical thinking ability. Many agree theoretically with Case, who has argued that "the key consideration [in assessment] is not whether teachers agree or disagree with the conclusions students reach but rather the quality of the thinking that supports their answers" (2009, p. 209). Essay tests allow evaluators to analyse this "quality of thinking", as respondents are asked to elaborate their reasoning process, rather than simply select an answer. Essay tests also allow students to demonstrate their creativity by proposing new solutions or arguments. However, inter-rater reliability can be problematic with essay tests, given potential subjectivity in assessing open-ended responses.

In the mid-1980s, Ennis and Weir attempted to build an essay test with high inter-rater reliability by developing a critical thinking essay test with a highly structured format. The Ennis-Weir Critical Thinking Essay Test (1985) was based on Ennis' early definition of critical thinking as "reasonable reflective thinking focused on deciding what to believe or do" (Ennis, 1985). The Ennis-Weir test assesses critical thinking "in the context of argumentation" by asking test takers to defend their judgments with reasons (Werner, 1991). Test takers are presented with an argumentative piece of prose. After reading the statement, test takers are asked to analyse each paragraph before responding to the text as a whole. Test evaluators use a detailed scoring rubric to assign a score to each

paragraph response. The individual scores lead to relatively standardised scoring and acceptable inter-rater reliability, despite the open-ended nature of the test format (Ennis & Weir, 1985). Generally, the reviews of the Ennis-Weir Test have been quite positive, given its relatively high reliability and its effective use as a diagnostic tool (Werner, 1991). After using both the CCTST and the Ennis-Weir Test with students at his institution, Hatcher argued that the Ennis-Weir Test "more closely resembles what we want our students as critical thinkers to do in real life" (Hatcher, 2009, p. 243). However, the test is not without its critics. Johnson (2009), for example, has argued that, although Ennis-Weir is a much better assessment than any of the multiple-choice options, it is not sufficient for truly gauging the ability of students to apply critical thinking skills to real-world situations.

5.1.1.3 Performance-task-based tests

The newest category of critical thinking assessment is the performance-task-based test, an assessment structure inspired by the "authentic assessment" movement (Wiggins, 1998). In performance-task-based assessments, students are presented with a real-world scenario and asked to use all available evidence to present the 'best' solution to a given problem. Praised for their ability to assess traditional critical thinking skills (such as evaluation and judgment), as well as creativity, self-regulation and disposition, performance-based tests have been lauded by some critics as the "most promising" method for assessing critical thinking (Pinto & Portelli, 2009). However, as with essay tests, the evaluation of responses can be subjective and, therefore, less reliable when used with multiple scorers (Lane & Stone, 2006). Performance tasks also take a long time to administer and evaluate, so they are not always a feasible option for large populations (Cook *et al*, 1996).

The Collegiate Learning Assessment (CLA), developed by the Council for Aid to Education (CAE), is the first critical thinking test to use a performance-task-based format. The CLA involves three components – two written essay sections, similar to the Ennis-Weir test, and one performance task[32]. The performance tasks included in the CLA are intended to simulate the "domain of real-world jobs suggested by activities found in education, work, policy, and everyday practice" (Klein *et al*, 2007, p. 419). Test takers are presented with some introductory material regarding a specific challenge in a workplace environment. They are also given a number of documents in various formats, some of which are relevant to the task and some of which are not. Test takers are then asked to

[32] As the essay components of the CLA resemble essay tests of critical thinking, this discussion focuses on the performance task element of the CLA.

respond by explaining how they would solve the challenge presented to them, based on the evidence provided. Much like the Ennis-Weir Test, scoring of the CLA is based on a detailed scoring rubric (2011).

The CLA represents a new methodology in the field of critical thinking assessment and, as a result, has attracted a significant amount of attention and support. A number of recent large-scale studies of critical thinking at American universities have opted to use the CLA to assess student critical thinking ability (e.g. Arum & Roksa, 2011; Council of Independent Colleges, 2008). The CLA has also generated interest outside the U.S. The OECD recently selected the CLA as the central tool in a feasibility study for developing an international Assessment of Higher Education Learning Outcomes (Arum & Roksa, 2011). This reflects a growing consensus around the validity of the performance task structure for assessing critical thinking. However, the CLA also has a number of vocal critics. Banta (2008) and Shermis (2008), in particular, have raised concerns with the low inter-rater reliability of the instrument.

5.1.2 Selecting an appropriate format for Rwanda

In his 2009 essay on critical thinking assessment, Ennis argues that three steps must be completed when evaluating the potential for using a particular assessment tool in a particular context. First, Ennis argues, one must make sure that the test is based on a "defensible conception of critical thinking and that the test does a reasonable job of covering that conception" (p. 75). Second, one must determine if the assessment is appropriate for the situation (ibid.). Third and finally, one must assess the validity of the scoring for the situation (ibid.). Ennis' steps were used as a guide when determining which of the available assessment formats offered the greatest potential for use in the Rwandan context.

5.1.2.1 Linking format to definition

According to Ennis' model, the first step in selecting an assessment tool is to determine the underlying definition of critical thinking that one is seeking to assess. As discussed in Chapter 3, this study relies on Chun's definition of critical thinking, as it was judged to be the best reflection of the study's underlying conceptualisation of critical thinking. Chun defines critical thinking as:

> "…considering an issue from multiple perspectives, critically examining evidence (and attending to information that may run counter to or disconfirm initial ideas), valuing claims that are backed by appropriate and adequate evidence, reasoning objectively and dispassionately, and arriving at informed judgments and decisions" (2010, p. 23)

The next step in Ennis' model is to identify an assessment that "does a reasonable job of covering" the chosen definition. Of all of the available assessments, the performance task element of the CLA offered the greatest potential for assessing the various components of Chun's definition. The document analysis involved in a CLA task requires test-takers to "critically examine evidence" and "attend to information that may run counter to or disconfirm initial ideas". As performance tasks are typically based around issues of relevance to everyday life, it is assumed that respondents will have some pre-conceived ideas about the subject area. Evaluation of respondent ability to attend to the presented information, rather than relying on anecdotal evidence or prior experience, is therefore a critical component of the performance task format. Respondents are also asked to identify which claims in the task documents are "backed by appropriate and adequate evidence". The assessment questions then demand that respondents "consider [the central issue of the task] from multiple perspectives" before writing an open-ended response that demonstrates their ability to "reason objectively and dispassionately" and to "arrive at [an] informed judgment [or] decision". The 'authentic' nature of the tasks allows for an assessment of how a given respondent applies general critical thinking skills to a particular scenario. Although disposition cannot be disaggregated from ability in the scoring of a performance task, the format also allows raters to assess a respondent's use of critical thinking skills outside of a traditional classroom setting. The testing format, therefore, simulates transfer by asking respondents to use their general competencies to address ill-defined problems in the 'real world'. Although essay tests, such as the Ennis-Weir Test, also cover a significant proportion of the study's underlying definition, the essay format does not offer the same authenticity as performance tasks. This undermines their effectiveness for assessing a respondent's ability and disposition to critically examine evidence when making a decision outside of a classroom setting. Essay tests also require students to make an argument based on the written claims of others. It seemed likely that Rwandan students would not have extensive experience with analysis of written arguments. The performance task format was, therefore, judged to be a more appropriate format for use in the Rwandan context.

5.1.2.2 Validity and reliability of the CLA

Although not an explicit component of Ennis' guide to selecting an assessment tool, the reliability and validity of a given assessment's performance in other contexts are critical indicators of its ability to appropriately assess critical thinking. It was therefore instructive to consider results from the administration of the CLA in other contexts. The evidence indicated that the CLA could be considered a valid testing format with sufficient

reliability to allow for its use as a model for the study assessment.

A. Validity

Shavelson (2010), one of the original developers of the CLA, discusses validity at length in his exposition on the merits of performance-task-based assessment. His arguments for the *content validity* of the CLA in the American context are convincing. When the CLA was initially under development, university faculty members were asked to evaluate whether or not the performance tasks were representative of what they felt university students should be able to do. Faculty members participating in the study were overwhelmingly positive in their evaluation of the test structure, noting that the tasks reflected important learning outcomes and were predictive of situations that students were likely to face after graduation (Hardison & Vilamovska, 2008). When new performance tasks are written for the CLA, they are independently evaluated through a number of "think aloud" sessions, during which university students verbally explain the thought processes that they would use when completing the task (Shavelson, 2010). Think aloud techniques are often advocated as an effective tool for evaluating tests of cognitive processes, as participants are able to 'demonstrate' their cognitive skills, thereby allowing developers to compare the skills used by participants with the skills that the assessment is intended to test (Lane & Stone, 2006). Shavelson (2010) argues that the CLA's think aloud sessions consistently demonstrate a clear link between the test specifications and the actual thought processes of the participants. Respondents in the think aloud sessions also tend to agree that performance tasks are a valid measure of their own ability to use critical thinking skills. Although assessments of content validity by test developers can sometimes be subjective (Kane, 2006), the arguments in favour of the CLA's content validity are persuasive.

The developers of the CLA have argued that the instrument also demonstrates *criterion validity*. In support of this claim, they point to the positive correlation (.55) between CLA scores and SAT scores (Shavelson, 2010, p. 61) and between the CLA, the CAAP and the MAPP, two other tests of critical thinking ability (Klein, Liu & Sconing, 2009). However, critics, such as Banta (2008) and Shermis (2008), have argued that the high correlation between CLA and SAT scores should not be used as proof of criterion validity. Rather, they argue that such results suggest that the CLA is predominantly a measure of prior learning, raising concerns about its potential for use in gain studies. Although this may be true, the more significant issue seems to be the likelihood that the definitions of critical thinking underlying the CLA and the SAT are substantially different, a probable scenario that raises concerns about the use of such a comparison to test criterion

validity. It is, in fact, unlikely that any existing assessment of critical thinking will demonstrate a high correlation with other critical thinking assessments unless based on the same underlying definition of critical thinking. Until multiple instruments are developed based on the same understanding of critical thinking, criterion validity would not appear to be a relevant measure of a critical thinking assessment's validity. Concerns about the CLA's criterion validity were, therefore, not considered to be a particularly convincing reason to avoid using the format as a model for the study assessment.

There is, to date, no published evaluation of the CLA's *construct validity* (Klein *et al*, 2007). However, it was determined that the performance task format of the CLA did reflect the conceptualisation of critical thinking underlying this study. As discussed earlier in the chapter, the 'authentic' nature of performance-task-based assessment assumes that a respondent will apply his or her general "meta-knowing competencies" to a particular scenario. The format, therefore, allows for the assessment of both an individual respondent's critical thinking ability and his or her disposition to apply such general abilities to ill-defined problems in the 'real world'.

B. Reliability

Tests of CLA results have also indicated a sufficient level of reliability. Test-retest, equivalent forms, internal consistency and inter-rater reliability have all been assessed and published by the developers of the CLA (Klein *et al*, 2007; Klein *et al*, 2005; Klein, Liu & Sconing, 2009). Mean and median internal consistency scores generally fall in the range of .79 to .88 for CLA performance tasks, while inter-rater scores tend to fall in the range of .67 to .84 (Shavelson, 2010, p. 55). Although lower than the reliability of multiple-choice critical thinking tests, these scores are considered to be acceptable for open-ended performance tasks (ibid.).

C. Differential Item Functioning

The developers of the CLA have also considered the potential for bias in the instrument by analysing the performance of different groups of test-takers. Penfield and Lam (2000) have argued that the potential for construct-irrelevant variance in performance assessment can contribute to significant differential item functioning between groups. However, during initial testing, no systematic differences were identified between the CLA scores of males and females or between the scores of different ethnic groups[33].

[33] Personal correspondence with CAE representative, May 29, 2012.

5.1.2.3 The use of the CLA in Rwanda

Having determined that the CLA format could be used to assess the study's underlying definition of critical thinking, the next question to consider was its potential to assess critical thinking *in Rwanda*. Ennis argues that an assessment of critical thinking must be valid "for students at the level of your students, in a situation like theirs" (Ennis, 2009, p. 75). Given the particulars of the study context, it was immediately apparent that the CLA would need to be significantly adapted for use in Rwanda. Most importantly, the content of CLA performance tasks is clearly culturally biased towards an American audience. Furthermore, the format of the CLA assumes both advanced competence in English (as test takers are asked to demonstrate their thinking via written responses) and advanced understanding of technology (as the test is administered via a computer) (Klein *et al*, 2007). Both assumptions are problematic in the Rwandan context. It was therefore not possible to administer the CLA in its original form. Both the content and the administrative parameters needed to be adapted for use in Rwanda.

The scoring methodology of the original CLA was also not appropriate for the study structure. The CLA was developed as an institutional-level assessment for use in 'value-added' studies of institutional impact (Klein *et al*, 2007). As aggregate scores are the central unit of analysis in such studies, the CLA relies on a 'matrix sampling approach', meaning that individual examinees only complete a portion of the overall test (Council for Aid to Education, 2008). Portions are allocated randomly, allowing the test administrators to shorten the length of the test, while continuing to sample widely from the domain (ibid.). One of the objectives of this study, however, was to consider the impact of individual-level variables on critical thinking ability. Individual-level analysis requires individual scores (Schmeiser & Welch, 2006). The cross-sectional design of the study also required a scoring structure that would allow for individual comparisons. Developers of the CLA and representatives from campuses that have used the assessment had indicated that scores on the CLA could feasibly be used to explore links between critical thinking and various individual and institutional-level variables (Benjamin *et al*, 2009; Council of Independent Colleges, 2008). It was therefore determined that the format of the CLA could work within the constraints of the study design. However, it was clear from the outset that a new scoring methodology would need to be created in order to allow for the individual-level comparisons required to fulfil the study's central objectives. It was therefore not possible to complete Ennis' final step during the selection process (assessing the validity of the scoring for the situation). Instead, consideration of the validity of the assessment's scoring methodology was delayed until a new scoring methodology had been devised.

5.2 Developing the Adapted Assessment

5.2.1 Adaptation strategy

A test adaptation strategy was designed to guide the adaptation process, based predominantly on the test development strategy first advocated by Schmeiser and Welch (2006). The Schmeiser and Welch procedure includes the following steps:

1) Affirm the test's underlying philosophy and central purpose
2) Define the target population
3) Outline any administrative constraints on the assessment
4) Specify the content-related evidence of validity for the knowledge and skills to be measured by the test
5) Draft the test specifications, including the type and order of test questions, the test length, the proposed scoring method and any particularities of test administration
6) Organise an expert review of the test specifications and make any necessary modifications
7) Develop the specific test items/questions
8) Review test items for content accuracy and editorial style
9) Complete a fairness review of the draft assessment
10) Field-test the individual test items and revise as necessary
11) Pilot the full assessment, noting any potential concerns with test administration
12) Using the pilot responses, complete extensive item analysis and overall test evaluation, including:
    a. Exploration of pilot descriptive statistics
    b. Analysis of item difficulty and differentiation
    c. Analysis of alternative responses
    d. Tests for reliability
    e. Final confirmation of test validity
13) Modify and, as necessary, re-pilot until satisfactory results are obtained

The developers of the CLA have published their own guidelines for creating performance tasks (Benjamin *et al*, 2009; Chun, 2008; Council for Aid to Education, 2011). These procedures involve the following steps:

1) Identify a central issue for the task that has a quantitatively measurable key outcome variable and two possible – and equally reasonable – options for achieving the outcome
2) Determine a task scenario, including an authentic situation and a role for the test taker to assume
3) Create documents that will serve as evidence for the test taker, incorporating multiple document formats, multiple types of evidence and some conflicting data
4) Write specific questions for the test taker to answer and draft an initial scoring rubric for assessing answers
5) Revise and modify the scenario, questions and documents based on initial feedback
6) Field-test the task through the use of think aloud sessions with the target population
7) Make edits to the task and documents as necessary
8) Complete a full pilot of the written task
9) Use the pilot responses to edit and finalise the scoring rubric
10) Once the rubric is finalised, train scorers and complete reliability checks

As these task development procedures were seen to be equivalent to the "item development" and "item analysis" components of the Schmeiser and Welch strategy, they were used in place of Schmeiser and Welch's steps 7, 8, 10, 11, 12b and 12c. The final test adaptation strategy also referenced guidelines proposed by the International Test Commission for adapting tests for different cultural contexts (Hambleton, 1994). The complete strategy is outlined in Appendix A.

5.2.2 Steps 1-5: Defining the assessment specifications

The process began with the elaboration of the adapted assessment's specifications. The intended purpose of the assessment was to diagnose or signal current trends in the Rwandan student population (Koretz, 2008); it was not intended for use as a high-stakes summative evaluation of individual student ability or institutional capacity. The target examinee population was defined as first- and fourth-year undergraduate students in Rwanda.

The study context presented a number of administrative constraints. As the target population would have had significant difficulties with an online delivery platform, given scarcity of computers, problems with power outages and a lack of familiarity with computer-based tests[34], the adapted assessment was conceptualised as a paper-and-pencil assessment that would be administered in a group setting with members of the research team available to answer questions and clarify instructions.

As for the test specifications, it was determined that the adapted assessment would be modelled on the performance task component of the CLA. CLA performance tasks are based around an overarching scenario in which respondents are presented with background context, a role to assume and a set of decisions to make. Respondents are also given a list of specific questions to answer and are provided with seven documents that present a range of information about the scenario. The question prompts are ordered so that the respondent begins by reading and attending to the information in one document before proceeding to evaluate and draw conclusions using information from multiple documents. The final product is an authentic document, such as a memo, that responds to all of the questions in the task (Chun, 2008).

The CLA is intended to assess four categories of higher-order thinking skills: Analytic Reasoning and Evaluation, Problem Solving, Writing Effectiveness and Writing Mechanics (Chun, 2008). As this study is focused on analytical and problem solving skills,

[34] Administrative constraints were identified by members of the research team during initial discussions of possible test formats in March 2011.

not writing ability, the intention was to assess student ability using only the first two categories. The developers of the CLA define these categories as follows:

- Analytic Reasoning and Evaluation: "Interpreting, analysing and evaluating the quality of information; Identifying information that is relevant to a problem, highlighting connected and conflicting information, detecting flaws in logic and questionable assumptions; Explaining why information is credible, unreliable, or limited."

- Problem Solving: "Considering and weighing information from discrete sources to make decisions (draw a conclusion and/or propose a course of action) that logically follow from valid arguments, evidence and examples; Considering the implications of decisions and suggesting additional research when appropriate." (Chun, 2008, p. 42).

The assessment of these individual skills relies on respondent use of the various documents included in each CLA performance task. Each document type plays a specific role within a given task. In the CLA performance task manual, each document type is accompanied by a description of the skills that respondents are expected to demonstrate as a result of analysing the document. Different documents may be expected to elicit any number of the following skills:

- Determine that it is not appropriate to generalise from one anecdotal case
- Consider the source of information
- Recognise the need to use appropriate scales to compare data
- Recognise irrelevant data
- Determine that results from one study might not apply to a different setting
- Recognise methodological flaws or concerns in research study design
- Recognise possible sources of bias
- Recognise the difference between correlation and causation (Chun, 2008)

As the underlying format and structure of the CLA tasks were assumed to be appropriate and feasible for use in the study context[35], it was determined that the adapted assessment would replicate the format and structure of the CLA tasks. The adaptation process was therefore focused on creating task scenarios and documents that would reflect situations that would be familiar to a Rwandan audience.

One important consideration during the task design phase was how to handle the potential for bias inherent in the particular content of a given task. As performance tasks are based on one overarching scenario, there is potential for differential performance depending on a given respondent's familiarity with the task content (Messick, 1994). Test formats that include different questions focused on various content areas do not have the

[35] Although the research team indicated that a performance task would not be a familiar test structure for Rwandan students, they felt that a task-based format would be feasible and comprehensible for the study population (group discussion, March 2011).

same potential for bias, as there is an equal likelihood that every respondent will be familiar with at least a portion of the test content. A performance task, however, is based entirely in one content area. The original CLA asks respondents to complete a performance task alongside two writing tasks focused on different subjects. In contrast, the adapted assessment was intended to take the form of one performance task as its central – and only – content. It therefore seemed likely that respondent familiarity with the content of the task could affect responses, thereby biasing scores. For instance, a task about a health scenario could systematically bias test results towards those respondents studying health.

As one performance task can take 90 minutes to complete, the incorporation of multiple tasks into one assessment structure was deemed to be unfeasible method for addressing this concern. Instead, the decision was made to create three tasks with identical structures, formats and scoring rubrics, each based around a different scenario in a different subject area. It would have been possible to give respondents a choice of task. However, it has been demonstrated that giving respondents a choice on open-ended assessments presents a high likelihood of construct-irrelevant variance, as a test score in such a circumstance could be seen to include both a measure of student proficiency and a measure of the respondent's ability to pick the easiest scenario (Wainer & Thissen, 1994). Choice has also been found to threaten the comparability of scores across assessments (ibid.). As a result of these concerns, the task options were not conceptualised as choices. Instead, the intention was to randomly distribute parallel forms of the assessment, thereby distributing the potential for bias within the sample. Although addressing the concerns associated with allowing students to choose their task, two additional issues remained with the task-based format. First, constructing the assessment around one task can somewhat limit the generalisability of the resulting score (Lane & Stone, 2006). Second, Klein *et al* (2005) observed modest correlations among individual tasks on the CLA, suggesting that it may be problematic to assume that individual tasks are truly parallel forms. Haertel and Linn (1996) have also argued that scores from different performance tasks are often not comparable, given the potential for different ancillary requirements, such as reading comprehension or content knowledge, on different tasks. As a result of this second concern, analysis of the comparability of the test versions was determined to be a crucial objective of the pilot phase.

Although the CLA scoring rubrics are not publicly available, the general methodology is discussed at length in CLA publications (e.g. Benjamin *et al*, 2009). The CLA scoring method relies on the assumption that complex cognitive tasks, such as performance tasks, require an integration of cognitive abilities that cannot be validly

scored as individual skills (Klein *et al*, 2007; Shavelson, 2010). In line with this general principle, the CLA does not assign individual scores for specific sub-skills (Benjamin, Chun & Shavelson, 2007). Instead, participants receive holistic scores for each of the four categories assessed. In order to ensure some reliability between scorers, CLA performance tasks are scored using a combination of analytic and holistic scoring (Benjamin *et al*, 2009). The analytic score is based on a list of all the possible information solicited by a given task. Students are given one point for including each piece of important information in their response (ibid.). After assessing a given student's analytic score, a holistic score is assigned, indicating the participant's overall ability in each of the general test content categories (ibid.). Within each content category, students are scored on a 4-7 point scale[36]. Unrelated or blank responses are assigned a 0 and flagged for removal prior to analysis (ibid.). The final score is recalibrated to a 1600-point scale that mirrors the SAT. This re-scaling allows for the use of the SAT as a control for incoming ability during analysis (Council for Aid to Education, 2008).

The use of such a scoring methodology would have undermined the construct validity of the assessment tool, as this study relies on a fundamentally different conceptualisation of critical thinking. This study assumes that critical thinking is an umbrella term used to describe a number of individual cognitive skills. There was therefore a need for a scoring methodology that would allow for differential scoring of individual skills. As the study also required individual-level scores (as discussed in Section 5.1.2), it was determined that an entirely new scoring methodology would need to be created for use with the adapted assessment.

The first step in constructing the new scoring rubric was to articulate a specific list of individual skills to be assessed via the adapted tool. The decision was made to focus on the following nine critical thinking skills:

1. The ability to recognise potential sources of personal bias
2. The ability to determine whether or not information is relevant to the situation
3. The ability to recognise when the source of information is not credible or reliable
4. The ability to identify statistical or methodological errors in presented information
5. The ability to determine whether or not information can be generalised and/or applied to other situations
6. The ability to recognise when there is a lack of information
7. The ability to evaluate whether or not information is connected and, if so, whether the data is conflicting or complementary
8. The ability to evaluate whether or not information supports or contradicts an argument
9. The ability to draw on valid evidence when formulating a decision

[36] The total number of points depends on the individual task (Benjamin *et al*, 2009).

These skills were selected as crucial components of Chun's definition of critical thinking. All of the skills were also specifically referenced in the CLA performance task manual, indicating that respondents should be able to demonstrate the skills through the completion of a performance task. Throughout this chapter and subsequent chapters, the nine skills will be referenced using the following shorthand:

- Skill A: Bias (The ability to recognise potential sources of personal bias)
- Skill B: Relevance (The ability to determine whether or not information is relevant to the situation)
- Skill C: Credibility (The ability to recognise when the source of information is not credible or reliable)
- Skill D: Errors (The ability to identify statistical or methodological errors in presented information)
- Skill E: Generalisability (The ability to determine whether or not information can be generalised and/or applied to other situations)
- Skill F: Missing Information (The ability to recognise when there is a lack of information)
- Skill G: Evaluation of Connections (The ability to evaluate whether or not information is connected and, if so, whether the data is conflicting or complementary)
- Skill H: Evaluation of Support (The ability to evaluate whether or not information supports or contradicts an argument)
- Skill I: Use of Evidence (The ability to draw on valid evidence when formulating a decision)

As this is the first study of critical thinking in Rwanda, norm-referenced scoring was not considered to be a feasible option. Instead, a criterion-referenced scoring methodology was proposed, with each response being assessed against a pre-determined scale (Astin, 1991). The plan was therefore to develop a scoring rubric for each of the nine individual critical thinking skills. After all nine scores were determined for a given assessment response, the intention was to total the scores and provide each response with one overall competence score that would be used in subsequent analysis.

5.2.3 Validity of the adapted assessment

Kane (2006) has asserted that an argument must be made for an assessment's validity before the assessment is formally developed. Although articulated in reference to the development of a new testing instrument, Kane's advice is also relevant for the adaptation of a pre-existing assessment. As discussed in Chapter 3, the validity of an assessment in one context does not automatically indicate that the assessment will be valid in another. The adaptation process can also alter an assessment's content validity. It was therefore important to consider the potential validity of using the adapted instrument in the study context prior to the development of specific performance tasks.

Loevinger (1957) partitions construct validity into three components: a substantive component, a structural component and an external component. The substantive component considers the connection between theories supporting the construct and the content of the assessment. There were no validity concerns anticipated in reference to the substantive component of the adapted assessment. The nine skills were based directly on Chun's definition, which in turn had been selected as a relevant definition within the Rwandan context. The selected skills were therefore assumed to be valid components of an assessment of critical thinking in Rwanda.

The next question was whether such skills could actually be assessed via a performance task. This is Loevinger's structural component, as the validity of the assessment structure is considered in relation to the underlying construct. As all nine skills have been assessed previously via CLA performance tasks, it seemed highly likely that participants would be able to demonstrate the skills through the completion of a performance task. Performance tasks require participants to analyse documents and use evidence when making a decision. It was assumed, therefore, that a respondent should be able to exercise the first six skills when analysing the documents and the last three when making his or her final argument or decision. The construct validity of the assessment structure is also strengthened by the clear connection between the requirements of a performance task and the components of Kuhn's (1999) theory of critical thinking. Kuhn has completed numerous experiments investigating how students can improve their "meta-knowing competencies". Many of these have resembled the structure of a performance task. In one experiment, for example, she asked students to "explain how data in the graphs supported the claim(s) made by the author of the text" (Kuhn, 2005, p. 134). Kuhn views such "authentic" tasks as valid assessments of individual ability to demonstrate general competencies, because respondents "cannot perform poorly on [the] test because they lack a sufficient knowledge base regarding the topic…[they need] only to assemble the argument, drawing on the material made available" (ibid., p. 133). Given the study's reliance on Kuhn's theory of critical thinking, the similarities between her experimental methods and the structure of performance tasks lent further support to the validity of using a performance task format in this study.

The final aspect to consider was the validity of the assessment's "external component" (Loevinger, 1957). This refers to the validity of the inferences or conclusions that can be drawn from assessment scores within the study context (Koretz, 2008). This component could not be feasibly evaluated until later in the assessment adaptation process. As an entirely new scoring methodology needed to be developed for the adapted

assessment, evaluation of the proposed scoring methodology could not be completed until sample assessment responses were obtained.

5.2.4 Step 7: Adaptation of the tasks

Once the assessment specifications were delineated, three adapted versions of the CLA's "Crime Reduction" Performance Task were created (Chun, 2008). Two versions were exact replications of the Crime Reduction task. In both tasks, the respondent was asked to assume the role of an intern in a government ministry. Participants were asked to review the relevant evidence included in a number of documents and to write a response, outlining the strengths and limitations of two policy options supposedly circulating in Parliament. One of the tasks was about road accidents in Rwanda, and the other was about malaria. The content of the third task was quite different, as the scenario was not presented as a national policy debate. Rather, respondents were asked to solve a disagreement between two business partners regarding the best strategy for their start-up company. However, the *format* of the third task was identical, as participants were asked to review the relevant evidence in the task documents and to write a response, outlining the strengths and limitations of the two business strategies.

All three tasks were developed following the CAE procedures and specifications. Each task relied on seven documents and included three assessment questions. The first question asked respondents to evaluate the accuracy of a claim based on the evidence included in *one* document. The second question asked them to evaluate the strengths and weaknesses of one of the policy options, based on relevant information included in *all* of the documents. The final question asked respondents if they could make a decision between the two policies, based on all of the evidence in the documents. If they could make a decision, they were asked to explain which option they supported and why. If they could not, they were asked if they could propose another solution, based on the evidence in the documents.

Draft scoring rubrics were created to accompany each task. The three assessment questions were considered one response for the purposes of scoring. For each of the nine individual skills, a five-point scale was developed, ranging from 1 (poor) to 5 (exemplary). It was anticipated that illegible or incomprehensible responses – or responses including no reference to any task document – would be automatically scored with a 0 and flagged for removal prior to analysis.

5.2.5 Steps 6, 8 and 9: Reviewing the adapted tasks

Once the tasks were developed, an initial round of feedback was solicited from the author's colleagues in Rwanda. Each task was sent to two volunteer reviewers, along with

a list of questions about the connection between the task and the assessment specifications, the clarity of the task itself (instructions, guiding questions and related documents), and the connection between the task and the scoring rubrics. The reviewers were also asked to judge the authenticity of the situation and related documents for use in Rwanda. All of the reviewers felt that there was a strong connection between the tasks and the assessment specifications and between the tasks and the scoring rubrics. The tasks were judged to be clear, although a few suggestions were made to clarify the wording of the instructions. All three tasks were given high marks for authenticity.

Following the initial review, minor modifications were made to the wording of the tasks and the visual organisation of the scoring rubrics. The tasks were then sent to two Rwandan colleagues for a "fairness review". The reviewers were asked to assess the tasks for potentially offensive or alienating content. Neither of the reviewers had any concerns with the fairness of the tasks.

Finally, the assessment specifications and the individual tasks were submitted to an "expert review". The expert review panel comprised two faculty members at American institutions: one who had used the CLA methodology in his own classroom, and one who had written extensively on the assessment of higher-order thinking skills in classroom settings. The expert reviewers were asked similar questions to the initial reviewers, although they were given all three task versions to review. Both reviewers were largely positive in their evaluation of the assessment's validity, clarity and authenticity. They agreed that the tasks were unambiguous and directly related to the test content, suggesting a high level of content validity. They also agreed that the versions appeared to be parallel in structure and scoring.

At the end of the review phase, the three adapted tasks were revised in order to incorporate any edits proposed by the reviewers.

5.2.6 Translation of the assessment

As discussed in Chapter 3, assessment in a foreign language tends to underestimate critical thinking ability, as respondents are required to devote a substantial proportion of their working memory to the comprehension of assessment questions. It would have been ideal to avoid this situation by administering the assessment in Kinyarwanda. However, the lead researcher's lack of fluency in Kinyarwanda prevented this possibility, as it would have been impossible for her to ensure the accuracy of assessment content written in Kinyarwanda. It was instead determined that the adapted assessment would be written in both English and French. As discussed in Chapter 2, English has only recently become the official language of instruction in Rwanda. As a result, some students are much more comfortable reading in French than English.

Although not the native language of any of the respondents, there was reason to believe that the availability of assessment materials in French would limit the cognitive load for many study participants. As the lead researcher is fluent in both French and English, it was feasible to produce the adapted assessment in the two languages.

The determination was made, however, to allow participants to *respond* to the assessment questions in any language of their choosing, including Kinyarwanda. This decision was made in order to minimise the potential for construct-irrelevant variance, as it was assumed that flexibility in language of response would decrease the likelihood of respondents receiving low scores as a result of poor language ability. Although the inclusion of Kinyarwanda as a language of response did increase the potential for researcher error during the translation process, the advantages in terms of improved construct validity seemed to outweigh the potential risks.

Once the final text of the adapted tasks was confirmed, a native French speaker translated the instructions and the individual tasks into French. English and French text were presented alongside one another in the task documents, so that participants could either select their language of preference or use both languages simultaneously to aid in comprehension.

5.3 Field-testing and Piloting the Adapted Assessment

5.3.1 Steps 10 and 11: Field-testing the performance tasks

"Think aloud" techniques are widely used for field-testing performance assessments (Chi, Glaser & Rees, 1982). Some psychologists have raised concerns with assumptions that spoken explanations are reliable descriptions of underlying cognitive strategies (as discussed in Ericsson & Simon, 1993). Nonetheless, think aloud techniques are generally considered to be the best option for evaluating assessment validity, as they allow a test developer to gain insights into the cognitive processes employed by respondents when answering assessment prompts (Ericsson & Simon, 1993; Koretz, 2008; Pellegrino *et al*, 2001). Typically, think aloud sessions are conducted individually, with one respondent sharing responses with a researcher (Forsyth & Lessler, 1991). However, CLA tasks are generally field-tested through group think aloud sessions with university students (Shavelson, 2010). A similar group think aloud technique was used to field-test the adapted tasks in Rwanda.

Each task was independently field-tested by a group of 3-5 Rwandan students. Field-testing sessions each lasted 90 minutes. At the start of each session, the overall objectives of the research were outlined, and the think aloud procedure was explained.

Informed consent forms were circulated, discussed and signed by all participants. The task under review was then distributed, and participants were given half an hour to read the instructions and the content of the task before the group discussion.

During the discussion, participants were first asked to paraphrase the overall task scenario (as advocated by Jobe & Mingay, 1989). Participants were also asked to re-articulate the assessment questions in their own words and to discuss, as a group, how they would have responded to each question. Think aloud sessions were video recorded for transcription purposes. At the end of each session, the group was asked to give feedback on the clarity and authenticity of the task[37].

All three groups easily explained the task scenario. Although all of the participants confirmed that the testing format was new to them, they universally found the instructions and format to be clear. The concept of evaluating the strengths and weaknesses of two options also appeared to be a familiar concept to all of the participants, suggesting that the performance task format would be feasible in the Rwandan context. Respondent explanations of the task included the following:

> My assignment here is somehow clear. I need to analyse the documents and ... make facts more clear for the speech that is going to be given in Parliament, so that even people who are going to be vote will go knowing like, "Why am I doing this? Why am I supporting this? Why am I voting for that?" So it is really very important that I show them, like, I make sure I understand the situation, I make sure that I know … what is the advantage of this and what is the disadvantage of that for … both sides …because I'm not the one who's going to vote, but I need to make sure they do understand. I think that would be my assignment.

> So, the problem is about fighting against malaria. So, it's prevention of malaria, which is a big problem especially in Africa. And … I think the task is we are supposed the help the Minister of Health, who is going to present in the Parliament, and there were two policies. There were two Parliament members, if I'm not mistaken, who had to suggest – you know, who had two ideas … that could be adopted. So, we are trying to look at, you know, both ideas and advise the Minister of Health who is going to present something to Parliament to come up with a policy and recommendations, showing the weaknesses of one policy…. Another. *(Making a balancing motion with his hands)* And then, of course, the Minister of Health is supposed to bring up something that will help and will be convincing to the Parliament. So, that's what I understood in my task.

Participants also demonstrated conceptual understanding of the specific questions, as evidenced by the following explanations:

> So, in the first question, we are commenting on the point of view of [one M.P.] who is against spending the government funds on trainings but you know spending money on buying mosquito nets. So, we are asked to show – what are the weaknesses of this point of view of [the M.P] and, of course, also the strengths. And I think also, if I'm not mistaken, we're also going to support our argument for our answer by using, of course, the reference that [M.P.] referred to… Document E where there is a chart.

[37] The complete field-testing and piloting protocol is included as Appendix B.

> The way I understand [the second] question is … [the M.P.'s] idea is using the money and buying mosquito nets, rather than using the money on funding trainings. So, we are going to comment on the limitations, of course, and the strengths of that point of view.
>
> [Question 3 is asking] which one is the best option, basically. They're trying to understand why.

Overall, participant responses suggested that the students had no difficulty understanding the context of the task or the specific questions that they were asked to answer.

Furthermore, the think aloud sessions indicated that the tasks did require the use of the intended cognitive skills. Participant explanations suggested that they understood that they needed to analyse and evaluate the evidence included in the documents before making a decision *based on the evidence*, rather than their own pre-understanding of the issue in the task:

> The question doesn't really only, like, doesn't only belong at me. This is a question that concerns the whole country, like, the nation, because it's about… It's not a matter of how I understand the situation. I need, like, in this document, they make it clear that there are documents provided … to show the truth, like the truth for the population, more than the truth in my head. Maybe I, I had an accident, but maybe that accident was not caused by the alcohol – and for me, the way it's touched me, I would always feel maybe, I think, driving with high speed would be a problem always. However, if I consider … the whole document, maybe I will find out that in the population … the population that took alcohol increased, also the accidents increased. And … what I should do, my best thing to do would be to analyse the documents and understand the reality of the situation. I think I wouldn't do a judgment according to my, like, directly according to myself. What I should do would be to analyse the documents and first understand the situation.

Participants demonstrated their use of analytical skills when considering the reliability, credibility and generalisability of the evidence included in the documents:

> High speed is one reason. But, it's not the only reason. You have places where you have high speed and low accidents. That means there are other hidden causes of accidents that we're not looking at.
>
> I have a doubt, I guess. … I'm more focused on how accurate this is. So, [this chart is from] Ethiopia and [the M.P] is using this chart as an example, but maybe there could be another study in Rwanda and then use that for the case of Rwanda. Of course, Ethiopia is in Africa, in East Africa, and you know, we can generalise, you know, to say this chart applies to Rwanda too, but what if Ethiopia is … has a different level of malaria issues, you know? That's my concern.

Participants also expressed how they would use evidence in the documents both to evaluate the claims under consideration and to make a decision as to which argument they should support:

> Yeah, based on these documents, they are strength for supporting the message of breathalyser tests. Like, from the Document A, it showed that the accident was caused by the drunk driver. And the Document B, they are showing that 60% … of road accidents were [because] they was drunk. And in the Document F also, they show that the people who drank frequently, they caused more accidents, like where, on Document F, they are showing that, like for instance, you can see the people who drunk 15, the number of road accidents are 500, and the people who drunk 8, the number of road accidents are 100… Yeah, maybe there is a problem of, a limitation of this argument, according to the Document G where they are saying that it's not, I can say, yeah, there are some studies which are showing that you can underestimate the blood alcohol content because of gender, body mass and physical exercise…. But, the overall, all documents are really showing the strengths of this method.
>
> There is this case of, you know, in South Africa, where the mosquito net system was very efficient, but then you come closer to Uganda, where they say that it's not very efficient, so Uganda and Ethiopia, you know, we're in the same region. So, the question is, how efficient would mosquito net system be in Rwanda? It was a success in South Africa – maybe also in Ethiopia – and not very successful in Uganda, which is closer to Rwanda, so that could be an objection.
>
> All the evidence that is in the documents are from other countries (South Africa, Ethiopia and Uganda). There is no evidence from here. So, what I was, like, my suggestion is maybe if we took one province as a sample and we have a system of like, we apply trainings or maybe we take two samples, two provinces as sample, and in each province, we apply one of the methods. In one, the mosquito nets, and another, trainings to see what the outcome, what it would give, because Uganda is a different country. The population is different from Rwanda. Ethiopia isn't the same, South Africa isn't the same – even the climate. Everything is different, so I think we should base on evidences found here, not from other countries.

The field-testing sessions confirmed that it would be feasible to use the adapted assessment in Rwanda. Participants seemed to have no difficulty understanding the assessment format or content, indicating that the language used in the directions and test questions were comprehensible to the target population. The assessment content also appeared to be appropriate for the population, as the task scenarios and related documents were familiar to all of the participants in the sessions. Participant responses also supported the content validity of the performance tasks, as participants demonstrated the use of all nine individual critical thinking skills when explaining how they would respond to the assessment questions.

Participants in the focus groups were also asked to share their general feedback on the task format (as advocated by Bowden *et al*, 2002). All three groups confirmed that the performance task format would be novel for all Rwandan students, suggesting that familiarity with the testing format should not be a confounding factor in the assessment scores. The participants also verbally confirmed the authenticity of the tasks and related

documents. They reported that the scenarios were realistic and easy to understand and that the document formats were familiar.

Participants did make a number of suggestions about the administration of the tasks. First, they raised concerns with the order of materials presented in the test booklet. In the original CLA, respondents are first presented with the task scenario. They are then given a list of questions that they are expected to read before they read and analyse the documents. Participants suggested that respondents in Rwanda would likely start to answer the questions as soon as they saw them, without realising that they should read the documents first. They proposed moving the questions page to the end, so that participants would read the scenario *and the documents* before reaching the questions. Participants also suggested separating the English and French text, in order to reduce the length of the task document. In addition to administrative suggestions, participants confirmed that it was best to allow respondents to answer each of the three questions individually. The original CLA format asks respondents to produce an 'authentic document', which is one overarching response addressing all of the individual assessment questions. The author had anticipated that this would be an unfamiliar and likely disorienting process for respondents in Rwanda, so the field-testing sessions focused on each question individually. When asked if they would have felt comfortable producing an 'authentic document' with an overarching response, all of the participants confirmed that they would not have understood what was required of them. Allowing respondents to answer each of the three questions individually was, therefore, confirmed as the more appropriate and feasible format.

Following the field-testing sessions, the tasks were reorganised to respond to the participants' suggestions. Separation of the English and French text resulted in six versions for use in the pilot (a French and an English version of each of the three tasks).

5.3.2 Step 12: Pilot testing

A full pilot test was then organised with 14 volunteer students. The pilot replicated the intended logistical parameters of the final assessment. As students arrived, they were given the assessment instructions, an information sheet explaining the overall study and an Informed Consent Form, which they were asked to read while others arrived. The session began with a review of the contents of the Information Sheet and the Informed Consent Form, which the participants then signed. Instructions were then presented orally, and participants were given an opportunity to ask questions. It took roughly 15 minutes to complete all of the administrative paperwork.

The three versions of the tasks were then distributed randomly to the group. Those participants who wished to use the French version were given a copy in French,

although the majority selected the English version. As the CLA gives students 90 minutes for the performance task, it was anticipated that the adapted version would take a similar length of time. On average, it took participants about 65 minutes to complete their assigned task. Only one participant used the full 90 minutes allowed.

Following the pilot, a few volunteers were asked to stay for a brief discussion of the assessment tool. Eight participants stayed for the voluntary session (seven male and one female). The group generally felt that the assessment was clear and interesting to complete. Those who had also participated in the field-testing confirmed that the revisions had improved the tasks.

Following the pilot, members of the local research team were asked to complete a version of the assessment, so that they could become familiar with the testing format prior to recruitment of study participants. As eight of the research assistants had not participated in the pilot session, this generated an additional eight pilot responses. Two research assistants also volunteered to take the alternate version of the assessment, bringing the total number of pilot responses to 24.

5.3.3 Steps 13 and 14: Scoring the pilot responses

An initial reading of the pilot responses indicated that one of the three parallel versions had not worked as well as the other two. All of the participants who completed the third task (the business scenario) had answered the questions based entirely on their prior knowledge, rather than using any of the task documents. It was therefore nearly impossible to score the responses against the rubrics. As it was assumed that similar results would occur in the main study, the third task option was eliminated entirely from the study[38]. This meant that the five pilot responses based on the third task could not be used for any testing of the pilot responses, leaving only 19 pilot responses for subsequent analysis.

The process of scoring pilot responses also highlighted a problem with one element of the scoring methodology. As discussed in Section 5.2.2, the initial methodology proposed assigning individual scores for each of the nine skills before totalling the scores to produce one overall score for each assessment. The assumption was that the overall score would be a competence measure of each participant's overall critical thinking ability. However, during the pilot scoring, it became clear that respondents with drastically different individual skill abilities (and scores) were receiving the same overall score. In the aggregate, this suggested that such respondents had an identical overall competence level. The total scores had been calculated as a simple sum, as it was assumed that the

[38] Final versions of the remaining two tasks can be found in Appendix C.

individual skills should be weighted equally. The similarity in overall scores suggested that this weighting might not be appropriate in the Rwandan context. During the development of the CLA, factor analysis was conducted on analytic scores, in order to determine whether the individual components of the task were related to one underlying construct. In their analysis, all of the analytic scores loaded onto one underlying factor. These results were taken as evidence that 'critical thinking' could be viewed as one holistic ability in the American context[39]. However, the pilot results suggested that it was premature to assume that the individual sub-skills would relate to one underlying construct in the same fashion in the Rwandan context. Rather than assuming that all nine sub-skills reflected one underlying latent variable, it became clear that it would be more appropriate to consider each of the sub-skills individually before reviewing evidence of any underlying constructs. Overall scores were therefore eliminated at this stage, leaving each pilot response with nine individual scores.

The pilot responses were also used to evaluate the validity of the scoring criteria (as advocated by Wiggins, 1998). Scoring of the pilot responses demonstrated that it was possible to score all of the responses using the rubrics, and the resulting scores were determined to be appropriate reflections of the ability demonstrated in each response. The scales also appeared to be comparably calibrated (i.e. the level of ability required for a 3 on one skill was similar to the level of ability required for a 3 on another). The only limitation in the testing process was the lack of pilot responses at the high end of some of the scales. Without responses at each level of each scale, it was not possible to test the appropriateness of all of the criteria. Following the recommendation of Baxter and Glaser (1998), a volunteer, purposively selected as likely to be an 'expert', was asked to complete the assessment. Her responses were then scored against the rubrics in order to test the top range of each scale.

Minor modifications were made to the scoring rubrics based on the pilot responses. A supplementary "Scoring Flowchart" was then created for each task to assist with the process of scoring. The flowcharts were designed to guide scorers through a series of decision points, thereby assisting with consistency in scoring between individual scorers[40].

Two volunteer scorers were then asked to provide feedback on the final rubrics and related flowcharts. The volunteers assessed both the clarity of the scoring materials and the perceived validity of the scoring criteria. Both reviewers agreed that the materials

[39] Personal correspondence with CAE representative, May 22, 2012.
[40] The scoring rubrics and flowcharts can be found in Appendix D.

were easy to use and that the descriptions of the individual scores reflected the range of abilities demonstrated within the pilot responses they examined.

Once the rubrics and flowcharts were finalised, the pilot responses were scored again, and the resulting scores were used during formal evaluation of the assessment tool.

5.3.4 Translation of the final assessment

No further modifications to the content of the assessment tool were suggested following the pilot. However, as the French and English text had been separated from one another following the field-testing, the French translations were subjected to verification via the back-translation method first advocated by Brislin (1970). A bilingual volunteer[41] translated all of the French text back into English. The lead researcher then compared the translated text to the original, in order to identify any discrepancies, before confirming the final French and English versions of both tasks. There were not enough participants in the pilot to allow for a statistical comparison of scores on the French and English versions of each task (as advocated by Hambleton & Kanjee, 1995). However, a few pilot participants opted to use the French version of their assigned task, so it was possible to review the resulting responses and verify that there were no obvious differences in responses between French and the English versions of the same task.

5.4 Assessment Evaluation (Step 15)

The final phase in the test adaptation process was a formal evaluation of the pilot responses. The evaluation focused on the following four questions:

1. Is the adapted assessment valid for the study context?
2. Is there sufficient variation in the pilot scores to allow for subsequent analysis of variance?
3. Can the two versions of the assessment be considered parallel forms?
4. Is the scoring methodology reliable?

Typically, assessment evaluation would also include tests of differential item functioning (DIF), in order to determine if certain components of the test are biased towards certain demographic groups (Schmeiser & Welch, 2006). DIF requires a large sample size in order to detect any differences between groups. It was therefore not possible to conduct DIF analysis on the adapted assessment. However, as no DIF concerns were identified in the

[41] Although a native English speaker, the volunteer is fluent in French and lived in Rwanda for more than two years. As Hambleton (1994) has argued that translators must have language skills *and* knowledge of the cultural context in order to adequately translate an assessment tool, the translators used for all of the translations were both fluent in French and familiar with the Rwandan context.

original CLA, it was assumed that the adapted tool would also not present any DIF concerns.

5.4.1 Final validity testing

Results of the think aloud sessions indicated that the adapted assessment tool did capture respondent ability to demonstrate the nine critical thinking skills outlined in the scoring rubrics, as participants exhibited use of the skills when verbally articulating their responses to the assessment questions. A review of the written responses completed during the pilot further supported the validity of the instrument, as all of the responses included evidence of at least some of the skills that the assessment was intended to evaluate. A clear connection between the definition of critical thinking underlying the study and the abilities and processes demonstrated by those taking the assessment suggested that the construct validity of the instrument was strong. Reactions to the assessment by pilot participants also indicated that the tool had high content validity in the Rwandan context. During the subsequent case study phase (described in more detail in Chapter 7), Rwandan faculty members were asked to give their opinions on the assessment format. All of them agreed that it would be likely that graduates would face similar situations to the task scenarios during their careers, demonstrating further support for the assessment's content validity.

Analysis of the pilot responses also indicated that scores from the assessment could reasonably be interpreted as reflecting the ability of Rwandan undergraduates to demonstrate particular critical thinking skills, indicating validity of the "external component" (Loevinger, 1957). However, the pilot process did underscore the potential for construct-irrelevant variance in the assessment scores in relation to differences in reading speed between test takers (Haertel & Linn, 1996; Messick, 1989). As the scoring of the tasks relied on the use of evidence, it was necessary for respondents to read a significant amount of material before answering the assessment questions. It therefore seemed likely that respondents with slow reading speeds might not be able to finish reading all of the supporting evidence in the task documents within the allotted time. This could have translated into a low score, regardless of respondent critical thinking ability. Aside from this concern, no other serious threats of construct irrelevance were identified during the pilot phase.

5.4.2 Variation in scores

If scores on an assessment are too homogeneous, it is difficult to draw useful conclusions about the relationship between the construct and any other relevant variables (Koretz, 2008). It was therefore important to examine the range of scores resulting from

the pilot in order to determine the feasibility of analysing the assessment scores during the quantitative phase. Table 5.1 outlines the mean score, range, standard deviation and variance for each skill assessed via the pilot.

Table 5.1: Range and Variance of Assessment Scores, Pilot (n=19) [42]

Skill	Mean	Minimum	Maximum	Standard	Variance
A: Bias	2.68	1	5	1.250	1.561
B: Relevance	3.38	2	4	.619	.383
C: Credibility	2.31	2	3	.479	.229
D: Errors	2.05	1	5	1.177	1.386
E: Generalisability	2.37	1	4	1.342	1.801
F: Missing Information	2.63	2	5	.895	.801
G: Evaluation of	3.16	1	5	1.642	2.696
H: Evaluation of Support	3.68	1	5	1.493	2.228
I: Use of Evidence	2.58	1	5	1.575	2.480

The overall variation in scores within the pilot sample was determined to be sufficiently heterogeneous to allow for the use of the assessment scores in analysis. However, the pilot results did indicate that variance in scores across the individual skills was not uniform. As the variance in scores for Skill B: Relevance and Skill C: Credibility was found to be relatively limited in the pilot sample, these skills were flagged as requiring further monitoring during the main phase of the study.

5.4.3 Testing for parallel forms

As discussed in Section 5.2.2, there was a potential danger that the two versions of the assessment might not be comparable, given the potential for construct-irrelevant variance in the scoring of performance tasks. Determining the comparability of the task versions was therefore an important objective of the pilot stage analysis.

The low number of pilot responses did not allow for a strong test of task comparability. A series of independent *t*-tests was conducted in order to compare the Task 1 and Task 2 scores for each sub-skill. The results of the tests are outlined in Table 5.2.

[42] Any individual score of 0 was considered missing data for the purposes of analysis. Three pilot responses received 0s on Skills B and C, leaving a valid n of 16 for these two skills. All other skills had 19 valid scores.

Table 5.2: Results of Task Comparison, Pilot (n=19)

Skill	Task	Mean	Standard Deviation	Standard Error of Mean	Equal variances assumed? (based on Levene's test for Equality of Variances)	Results of *t*-test	Significance (*p*-value)
A: Bias	1 (n=10)	2.50	.850	.269	No	$t(11.83) = -.646$	$p=.531$
	2 (n=9)	2.89	1.616	.539			
B: Relevance	1 (n=7)	3.71	.488	.184	Yes	$t(14) = 2.155$	$p=.05$**
	2 (n=9)	3.11	.601	.200			
C: Credibility	1 (n=7)	2.29	.488	.184	Yes	$t(14) = -1.91$	$p=.851$
	2 (n=9)	2.33	.500	.167			
D: Errors	1 (n=10)	2.40	1.35	.427	Yes	$t(17) = 1.390$	$p=.182$
	2 (n=9)	1.67	.866	.289			
E: Generalisability	1 (n=10)	2.60	1.265	.400	Yes	$t(17) = .784$	$p=.444$
	2 (n=9)	2.11	1.453	.484			
F: Missing Information	1 (n=10)	2.80	1.033	.327	Yes	$t(17) = .858$	$p=.403$
	2 (n=9)	2.44	.726	.242			
G: Evaluation of Connections	1 (n=10)	2.30	1.494	.473	Yes	$t(17) = -2.830$	$p=.012$*
	2 (n=9)	4.11	1.269	.423			
H: Evaluation of Support	1 (n=10)	3.20	1.619	.512	Yes	$t(17) = -1.547$	$p=.140$
	2 (n=9)	4.22	1.202	.401			
I: Use of Evidence	1 (n=10)	2.60	1.776	.562	Yes	$t(17) = .060$	$p=.953$
	2 (n=9)	2.56	1.424	.475			

** Significant at 5% level of significance; **Marginal at 5% level of significance*

As no significant differences in means were observed for eight of the nine skills assessed, the analysis suggested that the tasks were largely comparable. However, the results of the *t*-tests were not necessarily reliable, as normality could not be assumed in the underlying distribution. The sample distributions for most of the scores were quite skewed, and it was not possible to rely on the central limit theorem to assume normality in the population, as there were fewer than 30 cases in the sample (Field, 2009).

Despite these limitations, it seemed feasible to move forward with the initial intention to randomly distribute the two tasks in the main sample. However, concerns with the underlying assumptions indicated that it would be necessary to conduct additional task comparability tests on the full sample before completing any further analysis.

5.4.4 Reliability of the scoring methodology

The pilot responses were also used to test the reliability of the scoring methodology. Three volunteer scorers were trained on the methodology and then asked to score a subset of three pilot responses each. In total, six pilot responses (three from each task version) were scored by three independent scorers.

Inter-rater reliability of the scoring was tested through the calculation of intraclass correlation coefficients. Following the recommendation of Shrout and Fleiss (1979), a two-way random effects model was used, in which both the judges and the target ratings were treated as random effects. Analysis was conducted on the consistency of the ratings, rather than the absolute agreement (Yaffee, 1998), so that the reliability of a given judge's rating could be considered for each skill. The resulting coefficients are outlined in Table 5.3.

Table 5.3: Results of Reliability Analysis, Pilot (n=6)

Skill	Intraclass Correlation Coefficient for Single Measures	95% Confidence Interval	Significance
A: Bias	.899	.649, .984	$F(5,10) = 27.727$ ($p<.001$)*
B: Relevance	.714	.250, .948	$F(5,10) = 8.480$ ($p=.002$)*
C: Credibility	.558	.041, .911	$F(5,10) = 4.784$ ($p=.017$)*
D: Errors	.854	.529, .976	$F(5,10) = 18.483$ ($p<.001$)*
E: Generalisability	.799	.406, .966	$F(5,10) = 12.932$ ($p<.001$)*
F: Missing Information	.828	.468, .971	$F(5,10) = 15.400$ ($p<.001$)*
G: Evaluation of Connections	.789	.386, .964	$F(5,10) = 12.229$ ($p=.001$)*
H: Evaluation of Support	.700	-.090, .991	$F(2,4) = 8.000$ ($p=.04$)*
I: Use of Evidence	.701	.230, .946	$F(5,10) = 8.024$ ($p=.003$)*

** Significant at 5% level of significance*

The general rule of thumb is that tests with reliability coefficients of more than .70 can be used for individual-level analysis (Haertel, 2006). Although the reliability coefficient for Skill C: Credibility was lower than this threshold, the overall results indicated adequate reliability for the use of the scoring methodology in the main sample. However, as with the task comparability results, the small number of cases in the pilot presented a challenge for interpretation, as the assumption of normality may have been violated during analysis. Additional reliability testing in the main sample was therefore determined to be necessary in order to confirm reliability of the scores before analysis.

5.5 Summary of Phase One Results

Overall, the results of the study pilot suggested that the adapted assessment could feasibly be administered to study participants and that the resulting scores could be considered a valid indication of their critical thinking ability. Furthermore, initial testing suggested that the assessment was sufficiently reliable to yield trustworthy results. However, given the small number of participants in the pilot stage, some questions remained as to the reliability and comparability of the task versions. Additional assessment evaluation was therefore planned as a component of the main analysis.

Having thus determined that the adapted assessment tool could be used to validly assess critical thinking ability in the Rwandan context, it was possible to proceed with the central phases of the study. The results of these phases are outlined in the remaining chapters.

Chapter 6: Assessing Critical Thinking in Rwanda

As outlined in Chapter 4, the second phase of the study was designed to address two of the study's central research questions:

1) Is there evidence that Rwandan students are improving in their critical thinking ability during their time at university?
2) What individual and institutional factors appear to be associated with student critical thinking ability in Rwanda?

These questions were investigated by administering the adapted critical thinking assessment to students enrolled at three of Rwanda's public universities. This chapter outlines the methodology used during this phase of the study and highlights the major findings arising from the assessment results.

6.1 Detailed Methodology

6.1.1 Additional data collection tools

Participants were asked to complete a short demographic questionnaire alongside the critical thinking assessment. The questionnaire was developed explicitly for this study and included questions about all of the student-level inputs included in the conceptual framework. Questions regarding participant gender, family background, secondary school and university enrolment were drafted by the lead researcher. The questionnaire also included a number of questions from the most recent Demographic Health Survey (DHS) completed in Rwanda (Ministry of Health (MOH) [Rwanda], National Institute of Statistics of Rwanda (NISR) & ICF Macro, 2009). The DHS questions were used as a means of collecting information about participant socio-economic status.

There is considerable debate in the literature around the best way to capture socio-economic status in developing countries. The use of an asset-based index, in which an individual's economic status is determined based on the assets owned by his or her household, is generally considered to be the most consistent measure of long-term economic status. Other methods, such as indices based on family income or expenditure, are likely to be impacted by short-term economic shocks, such as cyclical income patterns or temporary expenses (Vyas & Kumaranayake, 2006). There is also a methodological justification for using asset-based indices, as participants are more likely to accurately recall the assets in their households than household income or expenditure (McKenzie, 2003; Rutstein & Johnson, 2004). Asset-based indices have been criticised as being biased towards urban residents, given that there is often greater access to infrastructure in urban

areas (Filmer & Pritchett, 2001). Despite this limitation, the construction of an asset-based index was gauged to be the best available method for determining the economic background of participants in this study[43]. The selection of which assets to include in an asset-based index is a complex endeavour requiring extensive piloting, and there was insufficient time to include such a process in this study. As the DHS uses an asset-based index to capture economic background, the decision was made to incorporate the previously tested questions from the DHS Rwanda survey into the study questionnaire[44].

All of the questions on the questionnaire were tested with participants during the pilot phase[45]. As with the critical thinking assessment, the final version of the questionnaire was translated into French and then back translated into English by a volunteer translator. The English translations were then compared to ensure accuracy of the final French version. Members of the research team also worked together to translate some of the possible responses to the DHS questions into Kinyarwanda to assist with participant comprehension.

6.1.2 Sampling

The critical thinking assessment and accompanying questionnaire were then administered to a random sample of first- and fourth-year students attending three of the public universities in Rwanda. As the study aimed to detect trends in the student population, probability sampling was judged to be the optimal method for selecting participants (Pellegrino *et al*, 2001). It was also a feasible method, as representatives from the participating institutions were willing to share university enrolment lists, which could be used as sampling frames.

In order to identify trends in the overall public university population, it would have been best to sample students attending all of the public institutions in Rwanda. However, this was an unfeasible strategy, given the time and financial resources available. Instead, a modified two-stage cluster sample strategy was used (Scheaffer, Mendenhall & Ott, 2006), with each public university being defined as a cluster. As only six clusters fit the study's eligibility criteria[46], it was illogical to draw a probability sample at the first stage of the cluster sampling process. Instead, three clusters were purposively sampled from the six in the sampling frame. Purposive sampling is typically used when a researcher wants to target certain groups for analysis, based on theory generated through

---

[43] Wealth indices actually seek to capture *economic background*, rather than socio-economic status (Gwatkin *et al*, 2007). 'Economic background' will therefore be used instead of 'socio-economic status' throughout the remainder of the discussion.

[44] The full English version of the questionnaire is included as Appendix E.

[45] Further details are included in the field-testing and piloting protocol in Appendix B.

[46] Public universities with more than 100 students

exploratory research or a review of the literature (Robson, 2011). As discussed in Chapter 2, Rwanda's public higher education sector includes one traditional comprehensive university and a number of technical colleges that focus on particular academic subjects. As 'niche' institutions and multi-faculty universities are very different models of higher education, it was decided that the study should focus on the National University (NUR), along with two of the technical institutes: the Kigali Institute of Science & Technology (KIST) and the School of Finance & Banking (SFB).

In addition to representing a diversity of institutional structures, the three institutions offered the advantage of a study population that was likely to be relatively homogeneous in terms of background academic ability. As they are widely considered to be the most prestigious public institutions in the country, the three institutions are largely populated with government scholarship recipients. As a result, the student populations are likely to exhibit similar levels of incoming academic ability while also representing a diversity of socio-economic and family backgrounds. When analysing variance within a sample, it is useful to limit the variance of one factor, as this reduces the overall variance in the sample, making it easier to detect the effect of other variables. As discussed in Chapter 3, incoming ability has been found to be one of the most significant factors influencing the development of critical thinking skills. It was not possible to assess the incoming critical thinking ability of students in the sample, given the cross-sectional nature of the study design. However, the purposive selection of a study population that was likely to be relatively homogeneous in this respect increased the likelihood of identifying the influence of other relevant background variables.

After the institutions were selected, power analysis conventions (advocated by Cohen, 1977) were used to determine the total number of participants to sample within each institution. As discussed in Chapter 5, testing of the original CLA had indicated that individual analytical scores loaded on one underlying latent variable. It was assumed that scores from the adapted assessment might also load onto one underlying factor. This would have allowed for the summation of individual scores, resulting in one total score for each participant. Multiple regression would have been the most efficient method for analysing the impact of various factors on one overarching score. A sample size was therefore calculated that would have been appropriate for multiple regression with five independent variables[47]. According to Cohen's statistical power formula, multiple regression with five independent variables requires a minimum of 96 cases in order to reach a desired power of .8 (Cohen, 1977, p. 384). Ninety-six participants would therefore

[47] The independent variables selected for analysis were the student-level input variables included in the conceptual framework: Gender, Socio-economic Status, Secondary School Type, Parental Education Level, and Academic Field.

have been required at each institution in order to consider the potential for differential effects within the three university populations. Based on the assumption that approximately 30% of those contacted would opt not to participate (a convention advocated by Israel, 1992), it was determined that 150 students would be recruited from each institution. As the central research question relied on a comparison between years, recruitment lists were divided evenly between classes. Seventy-five first-year students and seventy-five fourth-year students from each of the three institutions were therefore randomly selected for participation[48].

6.1.3 Recruitment

Recruitment of the study population was coordinated by the local research team. Each selected student was identified and given a personalised invitation letter, along with an information sheet about the project (written in both English and Kinyarwanda). The letter asked selected students to attend an informational meeting on the university campus. At the recruitment meeting, the lead researcher reviewed the contents of the Information Sheet, outlining the objectives of the study and the implications of participation. All of the content was simultaneously translated into Kinyarwanda by one of the research assistants. Attendees were encouraged to ask questions, most of which were fielded by the assistant in attendance. During the meeting, students were assured that they would be allowed to participate in the language of their preference and that fluency in English was not a requirement for participation. Interested students were asked to sign up to participate on a particular date. Prior to each assessment date, willing participants were contacted via text message with a reminder of the time and location of the assessment[49]. Details of the final sample are outlined in Section 6.2.

6.1.4 Data collection

Four or five data collection sessions were organised at each institution. All sessions were held in classrooms on the university campuses, in order to make access as easy as possible for participants. Each session was coordinated and supervised by the lead researcher, in the company of one or two research assistants. As study participants arrived, they were given a pen and a copy of the Informed Consent Form, which they were asked to read prior to the start of the session. Once a critical mass of students had arrived, the content of the Informed Consent Form was orally translated into French and

[48] As there was limited access to software at the study sites, random sampling was conducted via the use of a random numbers table (de Vaus, 2002).
[49] Recruitment materials are included in Appendix F.

Kinyarwanda. Students were then asked to sign the form, if they felt comfortable with participation[50].

After the Informed Consent Forms were collected, participants were provided with one version of the critical thinking assessment. The two versions of the assessment were distributed in an alternating fashion, so that the versions would be equally distributed in a random manner. Students were given the option of receiving their materials in French or in English. Any participant requesting both versions was allowed to use both, although participants were asked to write their responses on only one test booklet to avoid confusion during data entry. The assessment was not timed, as it was determined that a time limit could put undue stress on the testing situation, while contributing nothing to the research objectives (Floyd, 2011). It was also anticipated that the elimination of the time limit might mitigate the potential for construct-irrelevant variance arising from differential reading speeds (as discussed in Chapter 5). However, start and end times were recorded for each participant, so that the time taken could be calculated for subsequent analysis. Participants were informed that they could respond to the assessment in French, English, Kinyarwanda or some combination of the three[51].

Once participants finished the assessment, they were asked to complete the accompanying demographic questionnaire (numbered to match their assessment). As with the assessment itself, the questionnaire was available in either English or French. Students could consult both versions if they preferred, as long as they wrote their responses on the version corresponding to their assessment. At the end of the first data collection session, a few accompanying questionnaires were found to be incomplete. As a result, at all subsequent sessions, questionnaires were scanned as they were handed in, and participants were asked to complete any incomplete items before departure. On average, the assessment and questionnaire took approximately two hours to complete. In addition to keeping the pens they received at the start of the assessment, participants were given refreshments. No other incentives were provided.

6.1.5 Data reduction

6.1.5.1 Translation and scoring

Following data collection, the assessment responses were translated and scored. The particular challenge posed by translation in research has been extensively discussed

[50] A copy of the Informed Consent Form is included as Appendix G.

[51] Given the recent changes in language policy in Rwanda, it was highly likely that some participants had only "partial knowledge" of both French and English. In order to avoid inadvertently testing linguistic ability, participants were allowed to respond in a combination of any of the three languages (as suggested by Hambleton, 1994).

in the literature (Bradby, 2002; Edwards, 1998; Phillips, 1960; Temple & Young, 2004). Specifically, the interpretive act implicit in translation has raised significant concerns about the validity of using translated text as original, rather than "processed" data (Wengraf, 2001). As the assessment responses were analysed in the aggregate, it was determined that the interpretive element of translation should not pose a substantial problem for the integrity of the data in the study. However, the research team aimed for "conceptual equivalence" (Liamputtong, 2010) between original and translated versions of the responses in order to ensure validity of the resulting assessment scores. As was the case during the first phase of the study, the back-translation method advocated by Brislin (1970) was used to achieve such equivalence. Each Kinyarwanda assessment was translated from Kinyarwanda into English by one research assistant. Another assistant would then translate the translation back into Kinyarwanda, and a third would compare the two Kinyarwanda versions in order to detect any discrepancies. Differences were discussed by all three assistants, who then agreed on a final translation. Assessments completed in French were also back translated. The lead researcher completed the initial translations into English, which were then translated back into French by one of the research assistants. The back-translation and the original text were compared in order to verify the accuracy of the English translation.

Once translated, all of the completed assessments were scored by the lead researcher, based on the Scoring Flow Chart and Rubric for each test version. Given the number of details involved in each performance task, it was deemed prudent to focus on one task at a time, so all of the Task 1 responses were scored before the Task 2 responses. Scores were double-checked prior to data entry. Additional reliability checks were also completed at this stage (see page 150 for the results of the reliability checks).

Scoring of the full sample revealed one disadvantage of the scoring methodology that had not been obvious at the pilot stage. As the underlying definition of critical thinking used in the study was focused largely on the use and interpretation of evidence, the scoring required participants to reference at least some of the evidence included in the task documents. This method was judged to have high construct validity. However, in practice, this meant that participants referencing very few documents or very little specific evidence could receive a score of 3 on three of the skills (A: Bias, C: Credibility and E: Generalisability), not because they had exhibited stronger ability than those scoring a 1 or a 2 but because their avoidance of discussing any specific evidence meant that they had not demonstrated any evidence of *weak* ability. A score of 3 on these three skills, therefore, was not necessarily a reflection of average critical thinking ability. A score of 3 could indicate that the scorer could not effectively classify the participant's ability, given

their limited use of relevant information. In contrast, high scores (i.e. scores of 4 and 5) and low scores (i.e. scores of 1 and 2) always reflected a participant's strong or weak critical thinking ability on an individual skill. Although the scoring for Skill B: Relevance necessitated giving participants the freedom to choose which information to use in their response, scoring of the full sample revealed that it might have been possible to gain more precision in the middle range of the ability spectrum by insisting that participants refer to all documents in their responses. This drawback did not undermine the overall validity of the assessment scores. However, it did result in a lack of precision in the mid-range ability scoring of some of the individual skills.

6.1.5.2 Data entry

Once scoring was complete, all of the data from the demographic questionnaires, along with the final assessment scores, were manually entered into SPSS by the lead researcher[52]. Separate codes were created to differentiate between 'Don't Know', 'Not Applicable' and 'No Response', with illegible/unknown answers being coded as 'No Response'. Each entry was triple-checked to ensure data integrity, twice manually and once by verifying frequency tables in SPSS.

Codes for most variables had been pre-recorded on the questionnaire to aid with data entry. However, some variables needed to be constructed from a number of questionnaire responses. These variables were: Secondary School Type, Parental Education Level, Economic Background and Academic Field. Detailed discussion of how these variables were constructed can be found later in the chapter.

Once all of the data were entered and checked, analysis was conducted using SPSS.

6.2 Overview of the Sample

The final sample consisted of **220 participants**. Of these 220, 103 were from the National University of Rwanda (NUR), 82 from the Kigali Institute of Science and Technology (KIST) and 35 from the School of Finance and Banking (SFB).

Table 6.1: Participants by Institution (n=220)

Institution	Frequency	%
NUR	103	46.82
KIST	82	37.27
SFB	35	15.91

[52] Appendix H includes a list of decisions made during data entry, as a result of unanticipated participant responses.

During recruitment, 37 of the selected students could not be located (31 from NUR and six from KIST). All 150 selected students from SFB were identified. In total, therefore, 413 out of the 450 selected students were contacted regarding participation in the study. As 220 agreed to participate, this reflects a 47% non-response rate. This was significantly higher than the anticipated 30% and was likely due to the timing of the data collection. Last minute changes to the university calendar in Rwanda resulted in data collection coinciding with final examinations at the three institutions.

The non-response rate was particularly high at SFB, where a full 77% of the selected students opted not to participate. The most likely explanation for such high non-participation is that fourth-year SFB students were completing internships off-campus during the study period. This was not communicated prior to data collection, so the high non-response rate at SFB was not anticipated. Furthermore, as the research assistants were all students at KIST and NUR, they had the advantage of access to significant networks at those two institutions. Nobody on the team studied at SFB, so similar contacts were not available on the SFB campus. It is likely that this also contributed to the lower response rate at SFB. Non-response rates were much lower at the other two institutions (43% at KIST and 13% at NUR).

6.2.1 Assessment characteristics

Within the overall sample, 109 participants took the first version of the test (Task 1) and 111 took the second version (Task 2). Individual university samples also demonstrated an equal split between the two task versions, as outlined in Table 6.2.

Table 6.2: Distribution of Assessment Version, by Institution (n=220)

Institution	Frequency Task 1	% Task 1	Frequency Task 2	% Task 2
NUR	50	48.5	53	51.5
KIST	42	51.2	40	48.8
SFB	17	48.6	18	51.4
Total	109	49.6	111	50.4

As discussed previously, participants were allowed to answer the assessment in French, English, Kinyarwanda or some combination. In the overall sample, 95 participants responded in English, 91 in French, 17 in Kinyarwanda and 17 in some combination (12 in Kinyarwanda/French, 2 in Kinyarwanda/English and 3 in all three languages). Language of response was fairly similar at the three institutions. English was the language of response for 44% of NUR participants, 44% of KIST participants and 40% of SFB participants, while French was the language of response for 38% of NUR participants, 43% of KIST participants and 49% of SFB participants. Kinyarwanda (or some combination)

was the language of response for 18% of NUR participants, 13% of KIST participants and 11% of SFB participants.

Table 6.3: Participant Language of Response, by Institution (n=220)

Institution	Frequency English	% English	Frequency French	% French	Frequency Kinyarwanda or Combo	% Kinyarwanda or Combo
NUR	45	43.7	39	37.9	19	18.4
KIST	36	43.9	35	42.7	11	13.4
SFB	14	40.0	17	48.6	4	11.4
Total	95	43.2	91	41.4	34	15.4

Most participants took between one and two hours to finish the test (177 out of 220). Thirty-one participants took less than an hour (with the shortest time being 30 minutes), and nine took longer than two hours (with the longest time being three and half hours). The time was not recorded for three participants.

6.2.2 University enrolment characteristics

6.2.2.1 Year at university

The sample was divided relatively evenly between first- and fourth-year students at KIST and NUR. Most SFB participants, however, were in their first year.

Table 6.4: Participant Year at University, by Institution (n=220)

Institution	Frequency First-Year Participants	% First-Year Participants	Frequency Fourth-Year Participants	% Fourth-Year Participants
NUR	56	54.4	47	45.6
KIST	46	56.1	36	43.9
SFB	24	68.6	11	31.4
Total	126	57.3	94	42.7

Language of response did vary systematically with year at university. Of the first-year students in the sample, 30% opted to respond in English, 50% in French and 20% in Kinyarwanda (or some combination). Of the fourth-year students, 62% responded in English, 30% in French and only 8% in Kinyarwanda (or some combination). A chi-square test was conducted that identified a significant correlation between language of response and year at university [χ (2) = 23.475, p<.001, contingency coefficient = .311]. This suggests that fourth-year students may be more comfortable in English than their first-year counterparts, probably as a result of their use of English during university.

6.2.2.2 Enrolment status

The vast majority of participants were full-time students, as outlined in Table 6.5.

Table 6.5: Participant Enrolment Status, by Institution (n=220)

Institution	Frequency Full-time	% Full-Time	Frequency Part-time	% Part-time
NUR	98	95.1 % Full-time in NUR Student Body: 87.8	5	4.9 % Part-time in NUR Student Body: 12.2
KIST	81	98.8 % Full-time in KIST Student Body: 96.1	1	1.2 % Part-time in KIST Student Body: 3.9
SFB	31	88.6 % Full-time in SFB Student Body: 66.8	4	11.4 % Part-time in SFB Student Body: 33.2
Total	210	95.5	10	4.5

Source: NUR Department of Planning & Development (2012); KIST Office of Planning & Development (2012); SFB Directorate of Academic Services (2012)

In fact, there appears to have been a slight bias in the sample towards full-time participants. This is likely due to the fact that part-time students often work while attending school and may not have been able to find time to participate in the study.

6.2.2.3 Academic field

Participants were asked to complete a series of open-ended questions about their academic Faculty and department. The diversity of possible responses resulted in cells with very few (i.e. one or two) cases. Responses were therefore aggregated into two overall categories: Sciences & Engineering[53], and Social Sciences & Humanities[54]. Participant responses were then coded accordingly. The sample distribution is outlined in Table 6.6.

[53] Comprising participants studying Engineering, Applied Sciences, Sciences, Architecture & Design, Medicine, Pharmacy, Psychology and Agriculture

[54] Comprising participants studying Business (Economics, Management or Finance), Arts, Media, Social Sciences and Law

Table 6.6: Academic Field of Study, by Institution (n=220)

Institution	Frequency Sciences & Engineering	% Sciences & Engineering	Frequency Social Sciences & Humanities	% Social Sciences & Humanities
NUR	47	45.6 % Sciences & Engineering in NUR Student Body: 45.3	56	54.4 % Social Sciences & Humanities in NUR Student Body: 54.7
KIST	82	100.0 % Sciences & Engineering in KIST Student Body: 100	0	0.0 % Social Sciences & Humanities in KIST Student Body: 0
SFB	0	0.0 % Sciences & Engineering in SFB Student Body: 0	35	100.0 % Social Sciences & Humanities in SFB Student Body: 100
Total	129	58.6	91	41.4

Source: NUR Department of Planning & Development (2012); KIST Office of Planning & Development (2012); SFB Directorate of Academic Services (2012)

6.2.3 Background characteristics

6.2.3.1 Demographics

A. Age

The majority of participants were between the ages of 21 and 26 (170 out of 220). Eleven participants were younger than 21, with the youngest being 18. Twenty-four participants were between 27 and 39, and four were 40 or older, with the oldest being 55. Eleven participants did not share their age.

B. Gender

Most of the study participants were male (78.6% of the overall sample). Although the student populations of the three institutions are also predominantly male, the gender distribution of the study sample, as outlined in Table 6.7, reflects an under-sampling of female participants, particularly at SFB.

Table 6.7: Participant Gender, by Institution (n=220)

Institution	Frequency Male Participants	% Male Participants	Frequency Female Participants	% Female Participants
NUR	76	73.8 % Males in NUR Student Body: 68.6	27	26.2 % Females in NUR Student Body: 31.4
KIST	68	82.9 % Males in KIST Student Body: 75	14	17.1 % Females in KIST Student Body: 25
SFB	29	82.9 % Males in SFB Student Body: 58.2	6	17.1 % Females in SFB Student Body: 41.8
Total	173	78.6	47	21.4

Source: NUR Department of Planning & Development (2012); KIST Office of Planning & Development (2012); SFB Directorate of Academic Services (2012)

6.2.3.2 Secondary school background

Three pieces of information were collected regarding the secondary school backgrounds of the study participants: Secondary School Type (i.e. public or private, religious or secular), Secondary School Location (i.e. urban or rural), and National Examination Results.

In the overall sample, 112 (50.9%) had attended a public school, 92 (41.8%) a private religious school, and 16 (7.3%) a private non-religious school. One hundred and nineteen (54.1%) described their secondary school as being in a rural area, while 101 (45.9%) described the location as urban.

A typology was constructed by combining Secondary School Type and Secondary School Location, creating four possible categories: Urban Public School, Urban Private School (religious or non-religious), Rural Public School, and Rural Private School (religious or non-religious). Each participant was then recorded as belonging to one of the resulting groups. The sample was distributed quite evenly between the four categories[55].

Table 6.8: Participant Secondary School Type, by Institution (n=220)

Institution	Frequency and % Urban Public	Frequency and % Urban Private	Frequency and % Rural Public	Frequency and % Rural Private
NUR	24 (23.3%)	24 (23.3%)	32 (31.1%)	23 (22.3%)
KIST	17 (20.7%)	20 (24.4%)	18 (22%)	27 (32.9%)
SFB	10 (28.6%)	6 (17.1%)	11 (31.4%)	8 (22.9%)
Total	51 (23.2%)	50 (22.7%)	61 (27.7%)	58 (26.4%)

[55] There were no missing data for this field.

Participant National Examination data confirmed expectations that the overall sample would be relatively homogeneous in terms of incoming academic ability. The questionnaire asked participants about their National Examination subjects and scores. Participants were also asked if their National Examination score had made them eligible for a government bursary. The majority of the sample had sufficiently high exam scores to qualify for a scholarship in their year (183 participants, or 83.2% of the sample), confirming that the majority of participants had obtained the highest possible marks in their subject on the National Examination. This finding is likely to be correlated with the bias towards full-time students in the sample, as all government scholarship recipients are required to study full-time.

6.2.3.3 Family background

Of the 220 participants, 72 participants (32.7%) reported that they had either not lived with adults during their childhood or had lived with adults who had never attended any school. One participant did not answer the question. The remaining 147 participants were asked to indicate the highest level of education that any adult in their household had completed. The distribution of their responses is outlined in Table 6.9.

Table 6.9: Highest Level of Parental Education[56] (n=219[57])

Highest Level of Education	Frequency	%[58]
Primary School	47	21.5
Vocational Training	9	4.1
Secondary School to O Level	8	3.7
Secondary School to A Level	42	19.2
Technical College	5	2.3
University	28	12.8
Higher than University	3	1.4
Don't Know	5	2.3
No Adults in Household or No Adults with Education	72	32.9

Given the small numbers in some of these categories, a new variable was created to represent the highest level of parental education. Responses were grouped into four overall categories: Some Primary Education, Some Secondary Education, Some Tertiary Education, and No Education/No Adults in the Household. The distribution of these categories is outlined in Table 6.10.

[56] Participants were asked about adults in their households, rather than their parents, as many Rwandan students lost their parents during the genocide. However, the term 'parental education' will be used throughout the chapter, as this is the term most commonly used in the literature.

[57] One participant did not answer the question

[58] Percentages calculated out of 219

Table 6.10: Highest Level of Parental Education, by Institution (n=214[59])

Institution	Frequency and % Some Primary Education	Frequency and % Some Secondary Education	Frequency and % Some Tertiary Education	Frequency and % No Adults or No Adults with Education
NUR	30 (29.1%)	25 (24.3%)	11 (10.7%)	35 (34%)
KIST	21 (25.6%)	16 (19.5%)	21 (25.6%)	21 (25.6%)
SFB	5 (14.3%)	9 (25.7%)	4 (11.4%)	16 (45.7%)
Total	56 (26.2%)	50 (23.4%)	36 (16.8%)	72 (33.6%)

These data suggest that participants from KIST were the most likely to have come from families in which some of the adults had completed tertiary education. This may be related to the fact that more affluent families in Rwanda are more likely to have the means to send their children to the private, fee-paying secondary schools that specialise in science subjects (a pre-requisite for admission to KIST).

6.2.3.4 Economic background

The DHS portion of the questionnaire asked participants about assets in their regular family residence. The first question was whether participants had a regular family residence where they stayed during university holidays. Of the 220 in the sample, 203 participants (92.3%) said that they had a regular family residence. Sixteen said they did not, and one did not respond to the question. Those who did have regular family residences were then asked a series of questions about assets at their family home.

First, participants were asked whether any of eight listed assets could be found in their family home. The distribution of responses is displayed in Table 6.11.

[59] Five participants did not respond to the question, and one responded that he or she did not know the highest level of education. These six responses were classified as missing data for this field.

Table 6.11: Distribution of Household Assets (n=203[60])

Asset	Frequency Households with Asset	% Household with Asset[61]	Frequency Households without Asset	% Households without Asset	Number Missing Values[62]
Electricity	83	40.9	114	56.2	6
Radio	186	91.6	17	8.4	0
TV	53	26.1	147	72.4	3
Refrigerator	21	10.3	174	85.7	8
Bicycle	82	40.4	117	57.6	4
Motorcycle	27	13.3	173	85.2	3
Car	18	8.9	182	89.7	3
Phone	187	92.1	16	7.9	0

Participants also responded to four questions about their household characteristics. The distribution of these responses is outlined in Table 6.12.

[60] The 16 participants without a family residence were not required to answer questions about their household assets. One additional participant did not respond to any of the questions in the family residence section of the questionnaire. Table 6.11 outlines the distribution of assets amongst the 203 participants with family residences.

[61] Percentages calculated out of 203

[62] 'Missing Values' refers to any missing values for the 203 cases included in Table 6.11.

Table 6.12: Distribution of Household Characteristics (n=203[63])

Household Characteristic	Potential Responses	Frequency	%[64]
Primary Source of Energy for Cooking	Firewood/Straw	136	67
	Charcoal	60	29.6
	Kerosene	2	1
	Biogas	1	0.5
	Natural Gas	1	0.5
	Electricity	3	1.5
Primary Source of Drinking Water[65]	Rainwater	1	0.5
	Fresh Water Source	26	12.8
	Open Public Well	37	18.2
	Covered Public Well	10	5
	Covered Well in Plot	1	0.5
	Covered Well in House	1	0.5
	Public Tap	71	35
	Piped Water in Plot	37	18.2
	Piped Water in House	13	6.4
	Bottled Water	3	1.5
Toilet Facilities[66]	Pit Toilet	168	82.8
	Improved Pit Toilet	10	5
	Flush Toilet	23	11.3
Primary Flooring Material[67]	Mud & Dung	23	11.3
	Mud/Sand/Earth	93	45.8
	Palm Rushes	2	1
	Wood Planks	1	0.5
	Polished Wood	1	0.5
	Cement	73	36
	Ceramic Tiles	2	1
	Carpet	2	1

The calculation of an asset-based index relies on the use of principal components analysis (PCA) (as advocated by Filmer & Pritchett, 2001). Although PCA typically assumes continuous data, there is an established tradition within development economics to use PCA on dummy variables. Filmer and Pritchett's methodology converts responses to asset questions into a series of dummy variables, which are then subjected to PCA. Factor loadings are used to assign weights to the variables included in the index, and weighted totals are calculated for each participant (Gwatkin *et al*, 2007). The distribution of all weighted totals in the sample is then divided into quintiles, and individual participants are assigned to the appropriate wealth quintile. The advantage of the

[63] See Footnote 61

[64] Percentages calculated out of 203

[65] Three of the 203 cases were missing values for this field.

[66] Two of the 203 cases were missing values for this field.

[67] Six of the 203 cases were missing values for this field.

methodology is that quintiles are generated based on the distribution of assets within the sample, rather than being calculated against some absolute poverty line (Rutstein & Johnson, 2004).

As missing values can create significant problems for principal components analysis (Tabachnick & Fidell, 2007), it was imperative to assess the extent of missing values prior to the creation of the wealth quintiles. Overall, there were 24 cases with at least one missing value related to economic status, representing 10.9% of the overall sample. A dummy variable was created to identify these 24 cases (as advocated by Schafer & Graham, 2002). Independent samples *t*-tests were then conducted to compare the means of cases with missing data to the means of cases without missing data for each of the nine individual assessment scores[68]. As none of the differences between the groups were found to be significant at the 99% level, it seemed plausible that these data were missing at random in the sample.

A two-stage strategy was then used to address the problem of missing values in the data set. Of the 24 cases with missing data, 15 were missing only one value out of the twelve questions regarding household assets. For these 15 participants, the mean response to the question (from the overall sample) was substituted for the missing value (as advocated by Tolmie, Muijs & McAteer, 2011). The remaining nine cases were missing multiple values. As these nine cases represented less than 10% of the overall sample, they were removed from the distribution prior to analysis (as suggested by Tabachnick & Fidell, 2007). This approach allowed for the retention of the maximum amount of available data.

Responses to the asset questions were then converted into a series of binary variables (as suggested by Rutstein & Johnson, 2004). Given the number of potential responses to some of the questions and the total sample size of 195[69], there were a number of cells with very few (i.e. one or two) cases. This resulted in a non-positive definite matrix, which would have made PCA results highly unreliable. In order to rectify this situation, categories were combined until a correlation matrix was obtained that yielded only appreciable correlations (i.e. >.3). This process involved two major changes to the data set. First, responses to the four household characteristics questions (i.e. Primary Source of Energy for Cooking, Primary Source of Drinking Water, Toilet Facilities and Primary Flooring Material) were combined to create two possible categories for each question. The new categories were:

---

[68] See Appendix I for results of the *t*-tests.

[69] Twenty-five cases were not included in the principal components analysis. Sixteen participants did not have a family residence, so no values were available for the asset questions, and an additional nine had been removed because of multiple missing values.

a. Primary Source of Energy for Cooking: "Cooking with Firewood" (136 respondents) or "Cooking with Non-Wood Sources" – made up of those selecting Charcoal, Kerosene, Gas or Electricity (67 respondents)
b. Primary Source of Drinking Water: "Running Water in the Household" – made up of those selecting Piped Water in Plot, Piped Water in House or Bottled Water (53 respondents) – or "No Running Water in the Household" – made up of those selecting Rainwater, Fresh Water Source, Open Public Well, Covered Public Well or Public Tap (147 respondents)
c. Toilet Facilities: "Pit Toilet" – made up of those selecting Pit Toilet or Improved Pit Toilet (178 respondents) – or "Flush Toilet" (23 respondents)
d. Primary Flooring Material: "Manufactured Flooring" – made up of those selecting Wood Planks, Polished Wood, Cement, Ceramic Tiles or Carpet (79 respondents) – or "Non-manufactured Flooring" – made up of those selecting Mud & Dung, Mud/Sand/Earth or Palm Rushes (118 respondents)

Second, responses to the questions about ownership of a Radio, a Bicycle, a Motorcycle, a Car or a Phone were removed from the analysis. There were two reasons for this decision. First, these questions yielded a highly skewed (i.e. approximately 90% versus 10%) distribution across the sample. Rummel (1970) has suggested that any variable with a 90%-10% split should be removed from PCA, as any case falling in the 10% should be considered an outlier. Second, none of the five variables demonstrated appreciable correlations (i.e. >.3) with any of the other variables in the index. At the end of this two-step process, there were seven variables remaining which could be included in the PCA:

1. Access to Electricity
2. Ownership of a TV
3. Ownership of a Refrigerator
4. Source of Cooking Energy (Wood or Non-Wood)
5. Source of Drinking Water (Running Water in Household or No Running Water in Household)
6. Toilet Facilities (Pit or Flush)
7. Primary Flooring Material (Manufactured or Non-manufactured)

These decisions had also increased the sample size to 198. Three of the cases with multiple missing values were only missing values on the transportation (i.e. Bicycle, Motorcycle and Car) questions. As these questions were ultimately not included in the analysis, the cases could be reintegrated into the sample prior to analysis.

Principal components analysis was then conducted on the seven variables[70]. A one-factor solution was found to yield the least complexity, explaining 57.5% of the overall variance. A sharp 'elbow' in the scree plot confirmed a one-factor solution as the best solution for these data. A one-factor solution is also consistent with Filmer and Pritchett's (2001) theory that the first component can be considered a proxy for household long-run wealth, as it explains the maximum variance and co-variance in the asset variables. Table

[70] An evaluation of assumptions and detailed results of the PCA are included in Appendix I.

6.13 shows the resulting factor loadings. The alpha coefficient of the resulting scale was found to be α=.871, suggesting a highly reliable index[71].

Table 6.13: Factor Loadings for Wealth Index

Variable	Factor Loading
Dummy for Access to Electricity	.773
Dummy for Ownership of TV	.877
Dummy for Ownership of Refrigerator	.713
Dummy for Non-Wood Cooking Method	.758
Dummy for Any Running Water	.769
Dummy for Flush Toilet	.704
Dummy for Manufactured Flooring	.700

Wealth quintiles were then created using procedures outlined by Gwatkin *et al* (2007). For each participant, the response for each variable was multiplied by the variable's factor loading. The seven weighted scores were then summed to create the participant's overall score. To maintain integrity in the final analysis, any mean substitutions for missing data were removed before calculating individual scores. Although 24 cases were missing at least one value, only 18 were not assigned to a wealth quintile, as participants with missing data in the other categories could still be assigned a score. The 16 respondents with no family residence were automatically assigned to the lowest wealth quintile.

If the distribution had been somewhat normal, the sample would then have been divided into quintiles. However, the particular nature of the study sample prevented the use of a standard quintile arrangement, as the sample was highly skewed. In addition to the 16 participants with no family residence, 82 participants received identical weighted scores at the bottom end of the distribution. It would have been nonsensical to divide this group into different quintiles. Instead, all of these 98 participants were assigned to the lowest quintile. The remainder of the sample was divided into three groups based on the intended quintile cut-off scores. With a distribution of 202 cases[72], each quintile should have had 41 cases. With 98 cases classified in the lowest category, only 22 remained in the middle group. Forty-one cases could then be assigned to each of the highest wealth categories. The final distribution is outlined in Table 6.14.

[71] PCA was also run without the substituted missing values. As there was no difference in the solution reached, it was determined that the missing responses did not have an impact on the results of the PCA.

[72] As discussed earlier in the section, 18 cases could not be assigned to a wealth quintile, as they were missing values for at least one of the seven variables included in the calculation of the asset score.

Table 6.14: Wealth Quintiles, by Institution (n=220)

Institution	Frequency and % Highest Wealth Quintile	Frequency and % Second-to-Highest Wealth Quintile	Frequency and % Middle Wealth Quintile	Frequency and % Lowest Wealth Quintile	Frequency and % Missing Quintile
NUR	21 (20.4%)	16 (15.5%)	10 (9.7%)	49 (47.6%)	7 (6.8%)
KIST	16 (19.5%)	17 (20.7%)	6 (7.3%)	36 (43.9%)	7 (8.5%)
SFB	4 (11.4%)	8 (22.9%)	6 (17.1%)	13 (37.1%)	4 (11.4%)
Total	41 (18.6%)	41 (18.6%)	22 (10.0%)	98 (44.5%)	18 (8.2%)

6.3 Determination of the Final Sample Size for Analysis

As discussed in Chapter 5, each participant was assigned nine scores on his or her assessment, corresponding to the nine individual critical thinking skills included in the Scoring Rubric. The highest possible score for any sub-skill was a 5, while the lowest was a 1. Participants could also receive a 0 if their response could not be scored for any given skill. This would typically be because a response referenced no task documents, but it could also be due to incomprehensibility of response. As a score of 0 did not represent lower ability within the construct domain of the assessment, any scores of 0 were flagged as 'missing data'. Twenty-two participants received a 0 on at least one sub-skill, reflecting 10% of the overall sample. Missing Values Analysis was conducted on the twenty-two cases to determine the distribution of the missing data. As Little's MCAR test yielded a non-significant result (χ (9) = 6.765; p=.662), it was determined that the data were missing at random in the sample (Tabachnick & Fidell, 2007). The 21 cases missing values for all sub-skills could therefore be removed from the distribution prior to analysis. As one case received a 0 on only one of the nine skills, it was retained in the sample. The final sample for analysis therefore consisted of **199 cases**.

6.4 Further Assessment Evaluation[73]

As discussed in Chapter 5, the small number of pilot responses prevented the confirmation of either the reliability of the scoring methodology or the comparability of the two versions of the assessment. As both issues had implications for subsequent analysis, investigation of the assessment scores began with a return to these two test evaluation questions. The impact of other potentially confounding factors on the assessment results was also investigated at this stage.

[73] An evaluation of the assumptions behind all of the analytical techniques used in this chapter is included in Appendix I.

6.4.1 Reliability of the scoring methodology

In order to further verify the scoring methodology, a volunteer scorer was asked to do a "read behind" scoring of a random 10% of the overall sample (Koretz, 2008). The scorer was trained on the scoring methodology before scoring 22 test responses (11 of each version) [74].

In contrast to the reliability check at the pilot stage, the reliability check on the main sample only compared scores from two scorers (the lead researcher and the second scorer). As a result, reliability was evaluated using Cohen's kappa, rather than intraclass correlations (Tolmie, Muijs & McAteer, 2011). The resulting coefficients are outlined in Table 6.15.

Table 6.15: Results of Confirmatory Reliability Analysis (n=22)

Skill[75]	% Agreement	Kappa Value	Significance
A: Bias	81%	.747	p<.001*
B: Relevance	68%	.572	p<.001*
C: Credibility	90%	.872	p<.001*
D: Errors	77%	.693	p<.001*
E: Generalisability	73%	.556	p< .001*
F: Missing Information	77%	.565	p<.001*
G: Evaluation of Connections	73%	.654	p<.001*
H: Evaluation of Support	59%	.424	p<.001*
I: Use of Evidence	50%	.346	p=.001*

***** *Significant at 5% level of significance*

As agreement between the two scorers was found to be statistically significant for all nine skills, the results of the confirmatory reliability check analysis generally supported the pilot results. However, the percentage agreement was quite low for two of the nine skills (Skills H: Evaluation of Support and I: Use of Evidence). A review of the discrepancies indicated that the two raters had mostly assigned scores within one value of one another (i.e. one assigned a 1 while the other assigned a 2). This suggests that differences were due to marginal differences in rater ability to discriminate between categories, rather than extensive discrepancies in scoring. As a result, the assessment scoring methodology was determined to be sufficiently reliable to support subsequent analysis, despite the low percentage agreement on these two skills.

---

[74] Although the sample had been reduced from 220 to 199 cases, the reliability analysis was conducted on 10% of the full sample. This was because it was considered important to verify all of the scoring criteria, including the designation of responses receiving a score of 0.

[75] Individual skills will continue to be referenced by the shorthand descriptions outlined in Chapter 5.

6.4.2 Testing for parallel forms

The question of task comparability was also examined in the main sample. This question was investigated through the use of independent samples *t*-tests, as during the pilot stage. Table 6.16 displays the results of the task comparison in the main sample.[76]

[76] One assessment response could only be scored for Skill F: Missing Information, so there is an extra case included for Skill F. There is also one case missing a score for Skill I, so there is one fewer case for Skill I.

Table 6.16: Results of Task Comparison, Main Sample (n=199)

Skill	Task	Mean	Standard Deviation	Standard Error of Mean	Equal variances assumed? (based on Levene's test for Equality of Variances)	Results of *t*-test	Significance (*p*-value)
A: Bias	1 (n=101) 2 (n=98)	2.53 3.02	.855 .574	.085 .058	No	$t(175.43) = 4.718$	$p<.001$*
B: Relevance	1 (n=101) 2 (n=98)	3.26 2.87	1.197 1.249	.119 .126	Yes	$t(197) = 2.250$	$p=.026$*
C: Credibility	1 (n=101) 2 (n=98)	2.47 2.49	.609 .630	.061 .064	Yes	$t(197) = -.278$	$p=.781$
D: Errors	1 (n=101) 2 (n=98)	2.30 1.97	.701 .818	.070 .083	Yes	$t(197) = 3.038$	$p=.003$*
E: Generalisability	1 (n=101) 2 (n=98)	1.59 1.58	.851 .759	.085 .077	Yes	$t(197) = .109$	$p=.914$
F: Missing Information	1 (n=101) 2 (n=99)	2.20 2.14	.490 .429	.049 .043	Yes	$t(198) = .868$	$p=.386$
G: Evaluation of Connections	1 (n=101) 2 (n=98)	2.78 2.33	1.487 1.322	.148 .134	No	$t(195.52) = 2.282$	$p=.023$*
H: Evaluation of Support	1 (n=101) 2 (n=98)	2.42 2.34	1.608 1.268	.160 .128	No	$t(189.14) = .385$	$p=.700$
I: Use of Evidence	1 (n=101) 2 (n=97)	2.46 2.33	1.261 1.134	.125 .115	Yes	$t(196) = .736$	$p=.463$

***** *Significant at 5% level of significance*

The *t*-test results were not entirely conclusive, as they indicated a significant difference between the two tasks on four of the nine skills. An additional test was therefore performed in order to confirm the results. Multivariate analysis of variance (MANOVA) tests whether mean differences between groups on a combination of dependent variables are likely to have occurred by chance (Tabachnick & Fidell, 2007). A one-way MANOVA was conducted, using Task as the grouping variable. The analysis yielded significant results, using Pillai's trace $F(9, 188)=4.784$, $p<.001$. However, when Bonferroni's correction was applied to adjust for the effect of running multiple tests (as suggested by Tolmie, Muijs & McAteer, 2011), a significant difference was only observed for two skills (Skill A: Bias and Skill D: Errors). The effect size for Skill A (assessed using partial eta squared) was only modest (.100), while the effect size for Skill D was quite weak (.045).

These results indicate that, although the tasks were not strictly parallel, they were sufficiently comparable to allow for an assumption of parallel forms. This meant that scores from the two assessment versions could be combined into one distribution for analysis.

6.4.3 Additional assessment evaluation

Tests were also conducted to evaluate whether additional test-related variables may have had a confounding effect on the assessment scores. The two variables of interest at this stage in the analysis were the time taken to complete the assessment and the language of response selected by the participants.

6.4.3.1 The effect of time taken

Time taken was recorded as a continuous variable. However, to simplify analysis, possible values were grouped into four overarching categories: Fewer than 60 minutes; 60-89 minutes; 90-119 minutes; and 120 minutes or more.

MANOVA was performed on the nine dependent variables[77], using Time Taken as the grouping variable. The overall effect of time taken on the pattern of scores was found to be non-significant at the 95% confidence level, using either Pillai's trace ($F(27,555)=1.498$, $p=.052$, partial eta squared=.068) or Wilks' lambda ($F(27,535.097)=1.495$, $p=.053$, partial eta squared=.068). However, as the *p*-values were very close to the 5% level of significance, these results could not be taken as firm evidence of a lack of effect. In fact, although all of the individual effect sizes were weak (<.07), univariate follow-ups indicated significant results for two of the nine skills: Skill B:

[77] The nine assessment scores serve as the dependent variables in this analysis.

Relevance ($F(3)=4.577$, $p=.004$, partial eta squared=.067) and Skill G: Evaluation of Connections ($F(3)=2.863$, $p=.038$, partial eta squared=.043). For both skills, additional time taken was found to have a positive effect on the scores.

These results indicate that the time taken to complete the assessment may have been a confounding factor in the assessment results. The results of the follow-ups suggest that time taken had a particularly significant effect on the scores for Skill B: Relevance. This is likely because Skill B was the only skill in which a correlation could be drawn between a high score and a high number of documents referenced. Participants received higher scores on Skill B if they referenced all of the relevant documents, while those participants referencing only one or two documents scored lower on the scale (even if the referenced documents were relevant). It seems probable that those taking more time to complete the assessment would be more likely to have read all of the documents in the task and would therefore be more likely to have received higher scores on Skill B. The evidence does not suggest that this was a general issue aside from Skill B, as the effect sizes suggest that time taken did not have a strong influence on the scores for any other skills in the assessment.

Tests were also conducted to ascertain any potential correlation between time taken and year at university. A significant effect could, in fact, be observed, as first-year students appear to have taken longer to complete their assessments than their fourth-year counterparts. Of the first-year students in the sample, 7% finished the assessment in less than an hour, 45% spent 60-89 minutes, 39% spent 90-119 minutes, and 9% took more than two hours, while 23% of the fourth-year students finished in less than an hour, 53% spent 60-89 minutes, 19% spent 90-119 minutes, and 4% took more than two hours. A chi-square test was conducted that identified a significant correlation between time taken and year at university [χ (3) = 19.058, $p<.001$, contingency coefficient = .284]. This could imply that first-year students were more motivated to perform well on the assessment or more willing to spend the time necessary to read all of the test materials. However, it could also indicate that first-year students took longer to read and comprehend the materials than their fourth-year counterparts.

6.4.3.2 The effect of language

MANOVA was also performed on the nine dependent variables, using Language of Response as the grouping variable[78]. The overall effect of language on the pattern of

[78] Language of Response was grouped into three overarching categories: English, French or Kinyarwanda/Combination (comprising those who responded entirely in Kinyarwanda and those using a combination of Kinyarwanda and at least one other language).

scores was found to be non-significant, using either Pillai's trace (F(18,376)=.697, p=.814, partial eta squared=.032) or Wilks' lambda (F(18,374)=.695, p=.817, partial eta squared=.032). These results may not be reliable, as Box's M was found to be significant at the .01 level (p=.007), indicating that the assumption of homogeneity of variance was violated in the analysis (Tolmie, Muijs & McAteer, 2011). However, univariate follow-ups also indicated no significant results, and all of the individual effect sizes were found to be weak (<.03).

Results therefore indicate no systematic difference in the overall pattern of assessment scores between groups of students electing to use different languages of response. This suggests that language preference did not have a confounding effect on the assessment results.

6.5 Evaluation of Latent Variables

Once the test evaluation process was complete, it was possible to move forward with analysis of the assessment results. As discussed in Chapter 5, it was necessary to begin with an investigation of whether the nine individual skills reflected any underlying constructs. Principal components analysis (PCA) was selected as the most appropriate strategy to use for this investigation, as this study was exploratory, rather than confirmatory, in nature. It was therefore necessary to consider all of the variance in the observed variables, rather than focusing solely on the shared variance between variables (Tolmie, Muijs & McAteer, 2011).

The default option in SPSS is to test for the optimal solution of a PCA using Kaiser's criterion, a method which retains only those factors with eigenvalues greater than one (Tolmie, Muijs & McAteer, 2011). However, Cattell (1966) has suggested that this method extracts too many factors if there are fewer than 20 variables, as was the case in this analysis. Instead, the optimal solution was determined using average variable complexity, a method that considers the average number of factors on which variables have an appreciable loading (i.e. >.3). Using this method, the solution with average variable complexity closest to one is assumed to be the optimal solution (Tolmie, Muijs & McAteer, 2011). The data were also subjected to varimax rotation, as it tends to minimise variable complexity (ibid.). After rotation, all nine potential solutions were compared, and the option producing the average variable complexity closest to one was identified.

Of the nine possible solutions, a seven-factor solution with varimax rotation was found to yield the least complexity, explaining 91% of the variance[79]. Three of the sub-

[79] More detailed results, including an evaluation of assumptions, can be found in Appendix I.

skills (Skills G, H and I) were found to load on one latent factor, with the remaining six sub-skills loading independently. Table 6.17 shows the factor loadings after rotation.

Table 6.17: Factor Loadings after Rotation, Analysis of Latent Variables

	Component						
Skill	**1**	**2**	**3**	**4**	**5**	**6**	**7**
A: Bias							.973
B: Relevance		.903					
C: Credibility					.976		
D: Errors				.960			
E: Generalisability			.960				
F: Missing Information						.993	
G: Evaluation of Connections	.716	.463					
H: Evaluation of Support	.846						
I: Use of Evidence	.764						

This solution was determined to have theoretical validity, as the three clustering skills (Skill G: Evaluation of Connections; Skill H: Evaluation of Support; and Skill I: Use of Evidence) could all be described as "evaluation", as opposed to "inquiry", skills (Kuhn, 2005). It, therefore, seems plausible that a student proficient in any one of the three skills would be proficient in all three. Grouping of the three items generated an alpha coefficient of α=.739, indicating reliability of the scale (Tolmie, Muijs & McAteer, 2011).

The results of the PCA suggested that the nine skills assessed in the study did not reflect one underlying construct in the Rwandan student population. Although there did appear to be a systematic correlation among the three evaluation skills, no similar correlation was identified among the remaining six skills. It therefore seemed inappropriate to calculate an overall score for each assessment. Instead, the decision was made to retain the individual skill scores during analysis. It was, however, possible to combine the scores on the three skills that loaded onto the same latent factor in order to reduce some analytical complexity. As the three skills demonstrated near-equivalent factor loadings (between .72 and .85), weighting of the individual scores did not appear to be necessary. Instead, individual scores on each of the three skills were averaged as one composite "evaluation" score for each participant.

Following PCA, each study participant therefore remained with seven individual scores on the critical thinking assessment. The seven scores reflected participant ability to demonstrate the following seven skills:

- Skill A: Bias (The ability to recognise potential sources of personal bias)
- Skill B: Relevance (The ability to determine whether or not information is relevant to the situation)
- Skill C: Credibility (The ability to recognise when the source of information is not credible or reliable)
- Skill D: Errors (The ability to identify statistical or methodological errors in presented information)
- Skill E: Generalisability (The ability to determine whether or not information can be generalised and/or applied to other situations)
- Skill F: Missing Information (The ability to recognise when there is a lack of information)
- Composite Skill: Evaluation (The ability to evaluate how information connects – to other information and to a central argument – and to use information when making a decision)

6.6 Overview of Assessment Results

The final distribution of the assessment results is outlined in Table 6.18.

Table 6.18: Distribution of Assessment Scores (n=199)

Skill	Mean	Standard Error of Mean	Median	Mode	Standard Deviation	Range
A: Bias	2.77	.054	3	3	.768	Low 1; High 5
B: Relevance	3.07	.088	3	4	1.235	Low 1; High 5
C: Credibility	2.48	.044	3	3	.618	Low 1; High 4
D: Errors	2.14	.055	2	2	.776	Low 1; High 4
E: Generalisability	1.59	.057	1	1	.805	Low 1; High 4
F: Missing Information	2.17	.033	2	2	.461	Low 1; High 4
Composite: Evaluation	2.45	.078	2.33	2	1.104	Low 1; High 5

Although the standard deviation differs between individual skills, results indicate relatively uniform variation in the range of scores obtained on the seven individual skills. This suggests that the diversity in scores across the range of skills is likely to be a valid representation of the range of ability within the sample. It also suggests that the discrepancies identified in the pilot sample (discussed in Chapter 5) were likely to be random effects, rather than a reflection of systematic differences in the underlying scales for individual skills.

First- and fourth-year results were also disaggregated, so that the range of ability of entering and graduating students in the sample could be considered in more detail. These results are presented in Table 6.19.

Table 6.19: Distribution of Assessment Scores, by Year (n=199)

Skill	Year	Mean	Standard Deviation
A: Bias	First Years (n=112)	2.79	.725
	Fourth Years (n=87)	2.75	.824
B: Relevance	First Years (n=112)	3.01	1.234
	Fourth Years (n=87)	3.14	1.241
C: Credibility	First Years (n=112)	2.48	.600
	Fourth Years (n=87)	2.47	.644
D: Errors	First Years (n=112)	2.12	.780
	Fourth Years (n=87)	2.16	.776
E: Generalisability	First Years (n=112)	1.64	.804
	Fourth Years (n=87)	1.52	.805
F: Missing Information	First Years (n=113)	2.11	.470
	Fourth Years (n=87)	2.25	.437
Composite: Evaluation	First Years (n=112)	2.30	1.090
	Fourth Years (n=86)	2.64	1.104

The assessment results suggest that critical thinking ability within the study population is quite weak, as the average participant received between 1.5 and 3 out of a possible score of 5 for each skill. To use the normative language of the Scoring Rubric, most participants demonstrated average or below average ability on each of the seven skills assessed. Although it was anticipated that first-year students would demonstrate low critical thinking ability, the poor performance of fourth-year participants implies that students may be leaving the universities in the sample with weak critical thinking skills.

The range of scores also indicates that students in the sample are more proficient at demonstrating some skills than others[80]. Participants seem to be most proficient at demonstrating the ability to determine whether or not information is relevant to a situation (assessed via Skill B) and least proficient at demonstrating the ability to

[80] It is important to note that this interpretation relies on an assumption that the scales for the individual skills were appropriately calibrated. As discussed in Chapter 5, this was a consideration during the assessment development process. The nine scales were compared during development and following the pilot, and a qualitative assessment was made that the cut-off points for each score on the individual scales were similar. The comparability of scores on the two tasks also indicates some stability in the scales.

determine whether or not information can be generalised and/or applied to other situations (assessed via Skill E).

6.7 Results of Analysis of Variance and Discussion of Research Questions

In addition to providing an overview of critical thinking ability in the sample, assessment scores were used to investigate the central research questions underlying the study. In this section, each research question will be considered independently, with statistical results preceding discussion of each topic.

6.7.1 What institutional and individual factors appear to be associated with critical thinking ability in Rwanda? [81]

6.7.1.1 The influence of institutional structure

As discussed in Chapter 3, there is little evidence from research in other contexts that institutional type or structure has any impact on the development of critical thinking ability. It was assumed that a similar lack of effect would be identified in the Rwandan context. This was tested by using MANOVA to compare the overall pattern of assessment scores within each of the participating institutions.

MANOVA was performed on the seven dependent variables[82], using Institution as the grouping variable. The overall effect of institution on the pattern of scores was found to be non-significant, using both Wilks' lambda ($F(14, 378)=1.619$, $p=.071$) and Pillai's trace ($F(14,380)=1.621$, $p=.071$). Univariate follow-ups also indicate no significant results, except in the case of Skill B: Relevance ($F(2)=6.172$, $p=.003$, partial eta squared=.060). The significant result for Skill B appears to be due to the particularly low average score for Skill B within the SFB population. (The mean score for Skill B was 2.34 at SFB, as opposed to 3.2 at NUR and 3.18 at KIST). Given the small sample of SFB students, this is likely to be a random fluctuation, rather than an actual effect related to institutional structure, particularly as all of the individual effect sizes were found to be minimal (<.06 for Skill B and <.03 for all other skills).

Results therefore indicate no systematic difference in the overall pattern of assessment scores between the three institutions in the sample. This suggests that

---

[81] For reasons explained later in the chapter, the third research question was investigated prior to the second. The questions in this section are, therefore, presented out of the original order indicated in Chapter 4.

[82] Results of the remaining MANOVAs only consider the seven assessment scores retained following PCA.

institutional type or structure is not systematically related to demonstrated critical thinking ability in the Rwandan context. These results support the findings from other contexts discussed in Chapter 3.

The lack of significant between-group differences identified in the sample also confirmed that it would be possible to combine the scores from the three institutions into one overall distribution for analysis. This allowed for a consideration of the rest of the research questions in the aggregate, as well as within the individual university populations.

6.7.1.2 The influence of individual student characteristics

As discussed in Chapter 3, research in other contexts has indicated that incoming critical thinking ability has a strong influence on the improvement of critical thinking skills during university. The cross-sectional nature of this study did not allow for an investigation of how incoming ability may affect gains in critical thinking ability during university. However, assessment results do suggest that students in Rwanda are entering university with generally low critical thinking skills. Although some individual participants could exhibit high ability on certain skills, the overall results (outlined on page 157) indicate a low range of incoming critical thinking ability within the sample.

Research in other contexts also suggests that the individual characteristics outlined in the conceptual framework (gender, economic background, parental education level and secondary school type) have an impact on both incoming critical thinking ability and growth in critical thinking skills during university. The relationship between these individual-level variables and the demonstrated critical thinking ability of participants in the sample was explored through the use of analysis of variance techniques. MANOVA was first used to assess differences in the overall pattern of scores between groups. Follow-ups were then conducted using analysis of variance (ANOVA) or, in the case of binary variables, independent samples *t*-tests. Analysis was first performed on the full sample so that overall trends could be assessed. The same techniques were then applied individually to the first- and fourth-year populations in order to determine the impact of the inputs on incoming critical thinking and on growth. Each individual set of analyses is presented below. The section concludes with a summary discussion of the results.

A. Analysis of the influence of gender

MANOVA was first performed on the seven dependent variables using Gender as the grouping variable. The overall effect of Gender on the pattern of scores was found to be non-significant, using Pillai's trace ($F(7,190)=1.631$, $p=.129$)[83].

Levene's test for homogeneity of variance was found to be significant for two of the seven skills. As Gender is a binary variable, univariate follow-ups were conducted via independent samples *t*-tests, rather than via individual ANOVAs, because *t*-tests can be corrected for non-homogeneity of variance (Tolmie, Muijs & McAteer, 2011). Results from the *t*-tests on the overall sample are displayed in Table 6.20. T-tests were also performed on the first-year and fourth-year aggregate populations, and these results are presented in Tables 6.21 and 6.22.

[83] Wilks' lambda is not included for binary variables, as the results of the Wilks and Pillai tests are identical when used to assess binary categories (Tolmie, Muijs & McAteer, 2011).

Table 6.20: Analysis of Differences in Critical Thinking Ability by Gender (n=199)

Skill	Gender	Mean	Standard Deviation	Standard Error of Mean	Equal variances assumed? (based on Levene's test for Equality of Variances)	Results of *t*-test	Significance (*p*-value)
A: Bias	F (n=42) M (n=157)	2.93 2.73	.513 .819	.079 .065	No	$t(103.44) = 1.91$	$p=.059$
B: Relevance	F (n=42) M (n=157)	2.71 3.16	1.215 1.227	.188 .098	Yes	$t(197) = -2.091$	$p=.038$*
C: Credibility	F (n=42) M (n=157)	2.60 2.45	.627 .614	.097 .049	Yes	$t(197) = 1.394$	$p=.165$
D: Errors	F (n=42) M (n=157)	2.14 2.13	.683 .801	.105 .064	Yes	$t(197) = .067$	$p=.946$
E: Generalisability	F (n=42) M (n=157)	1.62 1.58	.697 .833	.108 .066	Yes	$t(197) = .281$	$p=.779$
F: Missing Information	F (n=42) M (n=158)	2.05 2.20	.439 .462	.068 .037	No	$t(67.17) = -2.01$	$p=.048$*
Composite: Evaluation	F (n=41) M (n=157)	2.07 2.54	.982 1.116	.156 .089	Yes	$t(196) = -2.461$	$p=.015$*

** Significant at 5% level of significance*

Table 6.21: Analysis of Differences in Critical Thinking Ability by Gender, by Year – First-Year Students (n=112)

Skill	Gender	Mean	Standard Deviation	Standard Error of Mean	Equal variances assumed? (based on Levene's test for Equality of Variances)	Results of *t*-test	Significance (*p*-value)
A: Bias	First F (n=26)	2.92	.628	.123	Yes	$t(110) = 1.032$	$p=.304$
	First M (n=86)	2.76	.750	.081			
B: Relevance	First F (n=26)	2.58	1.065	.209	Yes	$t(110)=-2.067$	$p=.041$*
	First M (n=86)	3.14	1.257	.136			
C: Credibility	First F (n=26)	2.65	.562	.110	Yes	$t(110) = 1.679$	$p=.096$
	First M (n=86)	2.43	.605	.065			
D: Errors	First F (n=26)	2.15	.675	.132	Yes	$t(110) = .281$	$p=.779$
	First M (n=86)	2.10	.812	.088			
E: Generalisability	First F (n=26)	1.81	.749	.147	Yes	$t(110) = 1.196$	$p=.234$
	First M (n=86)	1.59	.817	.088			
F: Missing Information	First F (n=26)	1.96	.445	.087	Yes	$t(111) = -1.807$	$p=.073$
	First M (n=87)	2.15	.471	.050			
Composite: Evaluation	First F (n=26)	1.897	.837	.164	Yes	$t(110) = -2.182$	$p=.031$*
	First M (n=86)	2.419	1.126	.121			

** Significant at 5% level of significance*

Table 6.22: Analysis of Differences in Critical Thinking Ability by Gender, by Year – Fourth-Year Students (n=87)

Skill	Gender	Mean	Standard Deviation	Standard Error of Mean	Equal variances assumed? (based on Levene's test for Equality of Variances)	Results of *t*-test	Significance (*p*-value)
A: Bias	Fourth F (n=16) Fourth M (n=71)	2.94 2.7	.250 .901	.062 .107	No	$t(81.56) = 1.884$	$p=.063$
B: Relevance	Fourth F (n=16) Fourth M (n=71)	2.94 3.18	1.436 1.199	.359 .142	Yes	$t(85) = -.713$	$p=.478$
C: Credibility	Fourth F (n=16) Fourth M (n=71)	2.50 2.46	.730 .629	.183 .075	Yes	$t(85) = .196$	$p=.845$
D: Errors	Fourth F (n=16) Fourth M (n=71)	2.13 2.17	.719 .793	.180 .094	Yes	$t(85) = -.204$	$p=.839$
E: Generalisability	Fourth F (n=16) Fourth M (n=71)	1.31 1.56	.479 .857	.120 .102	Yes	$t(85) = -1.128$	$p=.262$
F: Missing Information	Fourth F (n=16) Fourth M (n=71)	2.19 2.27	.403 .446	.101 .053	Yes	$t(85) = -.660$	$p=.511$
Composite: Evaluation	Fourth F (n=15) Fourth M (n=71)	2.378 2.695	1.16 1.09	.299 .129	Yes	$t(84) = -1.011$	$p=.315$

* *Significant at 5% level of significance*

Although broadly reflective of the MANOVA results, the results of the follow-ups do indicate a few isolated gender effects. These are discussed in more detail in Section E below.

B. Analysis of the influence of economic background

MANOVA was then performed on the seven dependent variables, using Wealth Quintile as the grouping variable. The overall effect of economic background on the pattern of scores was found to be non-significant, using either Pillai's trace (F(21,522)=1.337, p=.145, partial eta squared=.051) or Wilks' lambda (F(21, 494.441)=1.344, p=.141, partial eta squared=.052). However, as Box's M was found to be significant at the .01 level (p<.001), these results may not be reliable. The assumption of homogeneity of variance is likely to have been violated because of the small number of cases in some of the cells. As there was no obvious way to collapse the categories further, there was little that could be done to strengthen these results.

However, univariate follow-ups could be conducted using individual ANOVAs. The results of the ANOVAs from the overall sample are displayed in Table 6.23. ANOVAs were also performed on the first-year and fourth-year populations. These results are presented in Table 6.24.

Table 6.23: Analysis of Differences in Critical Thinking Ability by Economic Background (n=183[84])

Skill	**Results of *F* test**	**Significance** (p-value)	**Effect Size** (partial eta squared)
A: Bias	2.609	p=.053*	.042
B: Relevance	.753	p=.522	.012
C: Credibility	.887	p=.449	.015
D: Errors	1.212	p=.307	.020
E: Generalisability	.538	p=.657	.009
F: Missing Information	.409	p=.747	.007
Composite: Evaluation	2.059	p=.107	.034

* *Marginal at 5% level of significance*

[84] Of the 199 cases with valid assessment scores, only 183 were assigned to a wealth quintile. The remaining 16 were among the 18 missing asset data.

Table 6.24: Analysis of Differences in Critical Thinking Ability by Economic Background, by Year (n=183)

Skill	Year	Results of *F* test	Significance (*p*-value)	Effect Size (partial eta squared)
A: Bias	First	1.919	*p*=.132	.056
	Fourth	1.623	*p*=.191	.059
B: Relevance	First	2.858	*p*=.041	.081
	Fourth	.832	*p*=.480	.031
C: Credibility	First	1.301	*p*=.279	.039
	Fourth	1.611	*p*=.194	.058
D: Errors	First	1.217	*p*=.308	.036
	Fourth	.327	*p*=.806	.012
E: Generalisability	First	.467	*p*=.706	.014
	Fourth	1.005	*p*=.395	.037
F: Missing Information	First	.368	*p*=.777	.011
	Fourth	.088	*p*=.967	.003
Composite: Evaluation	First	2.899	*p*=.039*	.082
	Fourth	1.813	*p*=.152	.066

* *Significant at 5% level of significance*

As in the gender analysis, the results of the follow-ups largely reflect the MANOVA results, as only one significant effect could be identified. This result is discussed in more detail in Section E.

C. Analysis of the influence of parental education level

MANOVA was then performed on the seven dependent variables, using Parental Education Level as the grouping variable. The overall effect of parental education level on the pattern of scores was found to be non-significant, using either Pillai's trace ($F(21,552)=1.139$, $p=.302$, partial eta squared=.042) or Wilks' lambda ($F(21, 523.156)=1.155$, $p=.287$, partial eta squared=.042).

Univariate follow-ups were conducted via individual ANOVAs. Table 6.25 presents the results of the ANOVAs from the overall sample, while Table 6.26 presents the results by year.

Table 6.25: Analysis of Differences in Critical Thinking Ability by Parental Education Level (n=193)

Skill	Results of *F* test	Significance (*p*-value)	Effect Size (partial eta squared)
A: Bias	.460	*p*=.710	.007
B: Relevance	.623	*p*=.601	.010
C: Credibility	1.748	*p*=.159	.027
D: Errors	.530	*p*=.662	.008
E: Generalisability	.435	*p*=.728	.007
F: Missing Information	.135	*p*=.939	.002
Composite: Evaluation	3.655	*p*=.014*	.055

** Significant at 5% level of significance*

Table 6.26: Analysis of Differences in Critical Thinking Ability by Parental Education Level, by Year (n=193)

Skill	Year	Results of *F* test	Significance (*p*-value)	Effect Size (partial eta squared)
A: Bias	First	.585	*p*=.626	.017
	Fourth	.156	*p*=.926	.006
B: Relevance	First	1.781	*p*=.155	.049
	Fourth	1.093	*p*=.357	.039
C: Credibility	First	1.194	*p*=.316	.033
	Fourth	1.335	*p*=.269	.047
D: Errors	First	.101	*p*=.959	.003
	Fourth	1.536	*p*=.211	.054
E: Generalisability	First	.060	*p*=.981	.002
	Fourth	.749	*p*=.526	.027
F: Missing Information	First	.315	*p*=.815	.009
	Fourth	.887	*p*=.451	.032
Composite: Evaluation	First	1.729	*p*=.166	.048
	Fourth	1.326	*p*=.272	.047

* *Significant at 5% level of significance*

Follow-up analysis of the influence of family background confirms the MANOVA results, as no significant effects could be identified, either in the aggregate or within the first- or fourth-year populations[85].

D. Analysis of the influence of secondary school background

MANOVA was finally performed on the seven dependent variables, using Secondary School Type as the grouping variable. The overall effect of secondary school type on the pattern of scores was found to be non-significant, using either Pillai's trace (*F*(21,570)=.422, *p*=.990, partial eta squared=.015) or Wilks' lambda (*F*(21, 540.385)=.419, *p*=.990, partial eta squared=.015).

[85] Although the results of the follow-up ANOVA for the Composite Skill demonstrate a significant difference between groups, Levene's test was significant for this skill, so the results may not be reliable.

Univariate follow-ups were conducted via individual ANOVAs. The results of the ANOVAs from the overall sample are displayed in Table 6.27. Results by year are presented in Table 6.28.

Table 6.27: Analysis of Differences in Critical Thinking Ability by Secondary School Type (n=199)

Skill	Results of *F* test	Significance (*p*-value)	Effect Size (partial eta squared)
A: Bias	.702	*p*=.552	.011
B: Relevance	.053	*p*=.984	.001
C: Credibility	.428	*p*=.733	.007
D: Errors	.575	*p*=.632	.009
E: Generalisability	.653	*p*=.582	.010
F: Missing Information	.044	*p*=.988	.001
Composite: Evaluation	.033	*p*=.992	.001

** Significant at 5% level of significance*

Table 6.28: Analysis of Differences in Critical Thinking Ability by Secondary School Type, by Year (n=199)

Skill	Year	Results of *F* test	Significance (*p*-value)	Effect Size (partial eta squared)
A: Bias	First	1.193	*p*=.316	.032
	Fourth	.623	*p*=.602	.022
B: Relevance	First	1.291	*p*=.281	.035
	Fourth	1.561	*p*=.205	.053
C: Credibility	First	.307	*p*=.820	.008
	Fourth	.546	*p*=.652	.019
D: Errors	First	1.976	*p*=.122	.052
	Fourth	2.243	*p*=.089	.075
E: Generalisability	First	.875	*p*=.456	.024
	Fourth	.352	*p*=.788	.013
F: Missing Information	First	1.542	*p*=.208	.041
	Fourth	1.883	*p*=.139	.064
Composite: Evaluation	First	.474	*p*=.701	.013
	Fourth	.552	*p*=.648	.020

* *Significant at 5% level of significance*

As in the analysis of family background, follow-up analysis of the influence of secondary school background confirms the MANOVA results, as no significant effects could be identified.

E. The influence of individual characteristics on demonstrated critical thinking ability

Results indicate no systematic difference in the overall pattern of assessment scores between men and women in the sample. Univariate follow-ups do suggest marginal differences in the ability of male and female participants to demonstrate Skill B: Relevance, Skill F: Missing Information and the Composite Skill: Evaluation. In all three instances, males in the sample performed slightly better than females. However, although the same findings were replicated within the first-year population, similar results could

not be identified within the fourth-year population. This suggests that any gender effects that do exist may not persist throughout the university experience.

No systematic difference could be observed in the overall pattern of assessment scores or the individual skill abilities of participants from different economic backgrounds[86]. Results also indicate little systematic difference between the individual skill levels of participants from different economic backgrounds within either the first- or the fourth-year population. A significant effect was identified within the first-year population in relation to the Composite Skill: Evaluation. However, this effect appears to have been the result of an unusually high average for those first-year students placed in the second-to-highest wealth quintile (average score = 2.89, versus 2.11 for the first quintile, 2.2 for the middle quintile and 2.15 for the highest quintile). As there is no theoretical explanation for why those in the second-to-highest wealth quintile would be able to demonstrate higher ability than others on this skill, it can be assumed that this was a random effect within the sample.

Results also indicate no systematic difference in the overall pattern of assessment scores between participants with different parental education levels. There also appears to be no difference between these participant groups in terms of their ability to demonstrate individual skills, either in the aggregate or within the first- or fourth-year populations.

Finally, there appears to be no systematic difference in the overall pattern of assessment scores or the individual skill abilities of participants from different secondary school backgrounds. Differences between participants from varying school backgrounds could also not be identified within either the first- or the fourth-year populations.

In the aggregate, therefore, results indicate that the background characteristics of participants seem to have no systematic influence on demonstrated critical thinking ability within the sample. The only variable found to impact incoming critical thinking ability was gender, and this effect does not appear to persist throughout university. No other background characteristics were found to systematically affect critical thinking ability within either the incoming or the graduating populations.

6.7.1.3 The influence of institutional experiences

The institutional experiences found to influence critical thinking ability in other contexts were generally assumed to be too complex to assess during the quantitative

[86] Although the results of the follow-up ANOVA for Skill A demonstrated a marginally significant difference between groups, Levene's test was significant for Skill A, so the results may not be reliable.

component of this study. However, information was collected about participant Academic Field, so it was possible to investigate the potential influence of field of study on demonstrated critical thinking ability within the sample. As with the individual-level variables outlined in the previous section, effects were considered at both the aggregate and the individual level. Overall patterns were analysed using MANOVA, while individual skill scores were compared using independent samples *t*-tests. Analysis was performed on the overall sample and within the first- and fourth-year populations.

The overall effect of Academic Field on the pattern of scores was found to be non-significant, using Pillai's trace ($F(7,190)=1.879$, $p=.075$). However, the significance level of this effect ($p=.075$) is close to the conventional significance level of .05, suggesting a possible marginal effect.

The results of the follow-up *t*-tests on the overall sample can be found in Table 6.29. Results by year are displayed in Tables 6.30 and 6.31.

Table 6.29: Analysis of Differences in Critical Thinking Ability by Academic Field (n=199)

Skill	Academic Field	Mean	Standard Deviation	Standard Error of Mean	Equal variances assumed? (based on Levene's test for Equality of Variances)	Results of *t*-test	Significance (*p*-value)
A: Bias	Sciences (n=120)	2.75	.822	.075	Yes	$t(197) = -.539$	$p=.590$
	Soc. Sci. (n=79)	2.81	.681	.077			
B: Relevance	Sciences (n=120)	3.27	1.207	.110	Yes	$t(197) = 2.886$	$p=.004$*
	Soc. Sci. (n=79)	2.76	1.222	.137			
C: Credibility	Sciences (n=120)	2.44	.562	.051	No	$t(142.34) = -.962$	$p=.338$
	Soc. Sci. (n=79)	2.53	.695	.078			
D: Errors	Sciences (n=120)	2.08	.811	.074	Yes	$t(197) = -1.362$	$p=.175$
	Soc. Sci. (n=79)	2.23	.715	.080			
E: Generalisability	Sciences (n=120)	1.56	.828	.076	Yes	$t(197) = -.639$	$p=.524$
	Soc. Sci. (n=79)	1.63	.771	.087			
F: Missing Information	Sciences (n=121)	2.15	.477	.043	Yes	$t(198) = -.806$	$p=.421$
	Soc. Sci. (n=79)	2.20	.435	.049			
Composite: Evaluation	Sciences (n=120)	2.61	1.089	.099	Yes	$t(196) = 2.602$	$p=.010$*
	Soc. Sci. (n=78)	2.197	1.087	.123			

** Significant at 5% level of significance*

Table 6.30: Analysis of Differences in Critical Thinking Ability by Academic Field, by Year – First-Year Students (n=112)

Skill	Academic Field	Mean	Standard Deviation	Standard Error of Mean	Equal variances assumed? (based on Levene's test for Equality of Variances)	Results of t-test	Significance (*p*-value)
A: Bias	First Sci (n= 68)	2.79	.744	.090	Yes	$t(110) = -.009$	$p=.992$
	First Soc (n=44)	2.80	.701	.106			
B: Relevance	First Sci (n= 68)	3.32	1.190	.144	Yes	$t(110)=3.523$	$p=.001$*
	First Soc (n=44)	2.52	1.151	.174			
C: Credibility	First Sci (n= 68)	2.47	.532	.064	No	$t(74.53) = -.238$	$p=.812$
	First Soc (n=44)	2.50	.699	.105			
D: Errors	First Sci (n= 68)	1.97	.810	.098	Yes	$t(110) = -2.513$	$p=.013$*
	First Soc (n=44)	2.34	.680	.103			
E: Generalisability	First Sci (n= 68)	1.50	.782	.095	Yes	$t(110) = -2.387$	$p=.019$*
	First Soc (n=44)	1.86	.795	.120			
F: Missing Information	First Sci (n= 69)	2.07	.495	.060	Yes	$t(111) = -.955$	$p=.342$
	First Soc (n=44)	2.16	.428	.065			
Composite: Evaluation	First Sci (n= 68)	2.544	1.115	.135	Yes	$t(110) = 3.102$	$p=.002$*
	First Soc (n=44)	1.917	.926	.140			

** Significant at 5% level of significance*

Table 6.31: Analysis of Differences in Critical Thinking Ability by Academic Field, by Year – Fourth-Year Students (n=87)

Skill	Academic Field	Mean	Standard Deviation	Standard Error of Mean	Equal variances assumed? (based on Levene's test for Equality of Variances)	Results of *t*-test	Significance (*p*-value)
A: Bias	Fourth Sci (n=52) Fourth Soc (n=35)	2.69 2.83	.919 .664	.127 .112	No	$t(84.53) = -.803$	$p=.424$
B: Relevance	Fourth Sci (n=52) Fourth Soc (n=35)	3.19 3.06	1.237 1.259	.172 .213	Yes	$t(85) = .496$	$p=.621$
C: Credibility	Fourth Sci (n=52) Fourth Soc (n=35)	2.40 2.57	.603 .698	.084 .118	Yes	$t(85) = -1.193$	$p=.236$
D: Errors	Fourth Sci (n=52) Fourth Soc (n=35)	2.21 2.09	.800 .742	.111 .126	Yes	$t(85) = .740$	$p=.461$
E: Generalisability	Fourth Sci (n=52) Fourth Soc (n=35)	1.63 1.34	.886 .639	.123 .108	No	$t(84.55) = 1.783$	$p=.078$
F: Missing Information	Fourth Sci (n=52) Fourth Soc (n=35)	2.25 2.26	.437 .443	.061 .075	Yes	$t(85) = -.074$	$p=.941$
Composite: Evaluation	Fourth Sci (n=52) Fourth Soc (n=34)	2.692 2.559	1.058 1.183	.147 .203	Yes	$t(84) = .546$	$p=.587$

** Significant at 5% level of significance*

The aggregate results suggest no systematic difference in the overall pattern of assessment scores between participants studying different academic disciplines. However, univariate follow-ups do indicate a significant difference between the scores of participants studying Sciences and those studying Social Sciences in their ability to demonstrate Skill B: Relevance and the Composite Skill: Evaluation. In both instances, those studying Sciences performed better than those studying Social Sciences.

The same findings were replicated within the first-year population. A systematic difference was also identified between first-year participants from different disciplines in their ability to demonstrate Skill D: Errors and Skill E: Generalisability. For these skills, those studying Social Sciences performed better than those studying Sciences. However, none of these differences could be identified within the fourth-year population. This suggests that, although students entering university in science disciplines seem to demonstrate a different range of individual critical thinking skills than those entering in social science disciplines, incoming differences may 'equal out' over the course of four years at university.

6.7.2 Is there evidence that Rwandan students are improving in their critical thinking ability during their time at university?

As discussed in Chapter 4, this research question was investigated through cross-sectional analysis, in which assessment results of first-year students were compared to those of fourth-year students as a proxy for change over time. Typically, cross-sectional studies of critical thinking ability include controls for individual-level characteristics, so that the confounding effects of individual-level characteristics can be eliminated (e.g. Saavedra & Saavedra, 2011). However, as discussed in the previous section, results indicate no evidence of any systematic relationship between individual-level characteristics and overall critical thinking ability in the sample. The cross-sectional comparison was therefore completed without any controls for incoming characteristics. As in the analysis of individual effects, differences in the overall pattern of assessment scores were analysed using MANOVA, while individual skill scores were compared using independent samples *t*-tests.

The overall effect of Year on the pattern of scores was found to be non-significant, using Pillai's trace ($F(7,190)=1.488$, $p=.174$). The results of the follow-up *t*-tests are presented in Table 6.32.

Table 6.32: Analysis of Differences in Critical Thinking Ability by Year at University (n=199)

Skill	Year	Mean	Standard Deviation	Standard Error of Mean	Equal variances assumed? (based on Levene's test for Equality of Variances)	Results of *t*-test	Significance (*p*-value)
A: Bias	Year 1 (n=112) Year 4 (n=87)	2.79 2.75	.725 .824	.068 .088	Yes	$t(197) = .432$	$p=.666$
B: Relevance	Year 1 (n=112) Year 4 (n=87)	3.01 3.14	1.234 1.241	.117 .133	Yes	$t(197) = -.730$	$p=.466$
C: Credibility	Year 1 (n=112) Year 4 (n=87)	2.48 2.47	.600 .644	.057 .069	Yes	$t(197) = .123$	$p=.902$
D: Errors	Year 1 (n=112) Year 4 (n=87)	2.12 2.16	.780 .776	.074 .083	Yes	$t(197) = -.403$	$p=.687$
E: Generalisability	Year 1 (n=112) Year 4 (n=87)	1.64 1.52	.804 .805	.076 .086	Yes	$t(197) = 1.093$	$p=.276$
F: Missing Information	Year 1 (n=113) Year 4 (n=87)	2.11 2.25	.470 .437	.044 .047	No	$t(190.99) = -2.277$	$p=.024$*
Composite: Evaluation	Year 1 (n=112) Year 4 (n=86)	2.298 2.639	1.085 1.104	.102 .119	Yes	$t(196) = -2.181$	$p=.030$*

** Significant at 5% level of significance*

Overall results indicate no systematic difference in the pattern of assessment scores obtained by first- and fourth-year students in the sample. Univariate follow-ups do suggest a significant difference between entering and exiting students in their ability to demonstrate the Composite Skill: Evaluation, as fourth-year students performed slightly better than first-year students on this skill. However, differences could not be observed between the demonstrated ability of first- and fourth-year students on any other skill[87].

Cross-sectional comparisons are not considered to be valid if any additional systematic differences are identifiable between the participant populations (Saavedra & Saavedra, 2011). As discussed earlier in the chapter, first-year students in the sample did take longer to complete their assessments than their fourth-year counterparts. If this was the result of higher levels of motivation within the first-year population in comparison to the fourth-year population, the lack of observable difference between the first- and fourth-year scores may reflect a difference in motivation level between the two groups, rather than a *bona fide* lack of improvement in critical thinking ability.

However, there is little evidence that the difference in time taken to complete the assessment actually represents a systematic difference in motivation levels between first- and fourth-year students. Fourth-year students in the sample would have been more familiar with an open-ended assessment format than first-year students. It is therefore possible that first-year students took longer to complete their assessments because of a lack of familiarity with the testing format. It is also possible that fourth-year students completed their assessments faster than their first-year counterparts because of higher proficiency in reading English and/or French. The systematic relationship between language of response and year would support such an interpretation. Neither of these factors would have negatively affected the comparability of the assessment scores.

In fact, the range of scores within the fourth-year population suggests that the lack of significant difference between the first- and fourth-year participant scores does reflect an actual lack of improvement in critical thinking ability. The scores of fourth-year students in the sample indicate a generally low level of critical thinking ability. Although fourth-year students appear to be more adept than first-year students at evaluating and using information in decision-making, graduating students seem to be no more proficient than incoming students at demonstrating any of the other individual critical thinking skills assessed in the study. It therefore seems reasonable to conclude that undergraduates at

---

[87] Although a significant difference was also detected between first- and fourth-year students in terms of their ability to recognize a lack of information (Skill F), Levene's test was significant for this skill, suggesting that the results of the *t*-test may be unreliable.

the institutions in the sample are not substantially improving in their overall critical thinking ability during their university careers.

6.8 Concluding Thoughts

Prior to concluding this chapter, it is important to acknowledge that no strong or moderate effects were observed in any of the analyses conducted during this phase of the study[88]. Although this could be the result of genuinely small differences in the means of the various comparison groups, it could also indicate that substantial within-group variance interfered in the results by making it difficult to detect any strong effects. As the assessment results were used to inform the protocol for the final stage of the study, it was necessary to investigate this concern before proceeding. The issue was explored by recalculating the difference between the means and the standard deviation around each mean as a percentage of the overall scale. This allowed for a comparison of between-group differences and within-group variation. To take one example, the overall difference in means between male and female students on Skill A: Bias was .196. This represents a difference of 3.92% of the five-point scale. The standard deviation of the female mean on this skill was .513, while the standard deviation of the male mean was .819. These values represent 10.26% and 16.38% of the overall scale respectively. This indicates that the within-group variance for this skill was larger than the difference in means between male and female students. The results of the rest of the recalculations comprising this analysis can be found in Section E of Appendix I.

Overall, the within-group variance was found to be significantly larger than any difference between group means. This indicates that other sources of variance are affecting the results outlined in this chapter. From the outset, the study did presume that other sources of variance would have a more significant impact on demonstrated critical thinking ability than many of the variables assessed during the quantitative phase of the study. As discussed in Chapter 3, the literature indicates that institutional experiences have a substantial impact on the development of critical thinking skills during university. As none of these experiences were assessed during the quantitative phase, it seemed likely that sources of additional variance might be identified during the final phase of the study.

However, large within-group variance can also be an indication of problems with the reliability of an assessment tool, so it was important to reconsider the reliability of the study instrument in light of such a possibility. Results from the main analysis suggest little evidence of such an interpretation. The comparison of individual skill scores between

[88] Partial eta squared was never greater than .1 for any test (Tolmie, Muijs & McAteer, 2011).

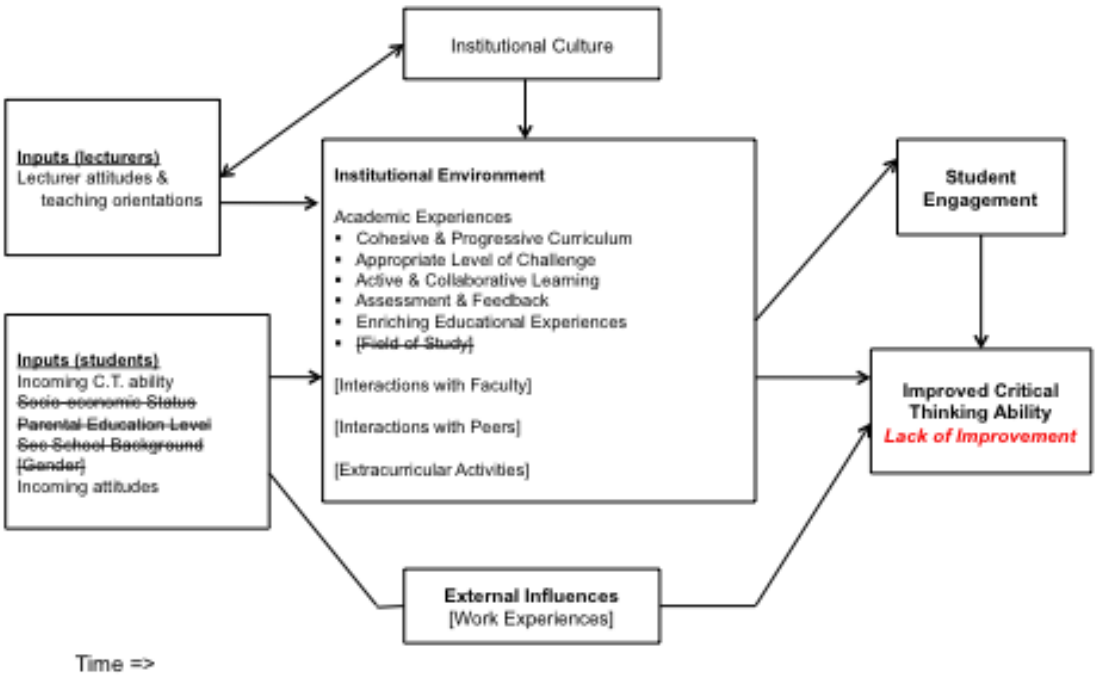
participants in different fields of study indicates differences in ability consistent with what one would expect to find, given results from past research. Similarly, the observed difference in the ability of first- and fourth-year students to demonstrate Evaluation skills is consistent with results from other contexts. Given that the assessment tool was able to capture these between-group differences, it can be concluded that the large within-group variance identified in the analysis is the result of the interference of other variables not considered in the quantitative phase, rather than the effect of an unreliable study instrument.

6.9 Summary of Phase Two Results

Results from the quantitative phase of the study suggest that students in Rwanda may not be improving in their critical thinking ability during university, as both first- and fourth-year participants demonstrated relatively weak critical thinking skills on the study assessment. Participants did, however, appear to be more proficient at some skills than others, suggesting some diversity in student ability to demonstrate individual critical thinking skills. Furthermore, results indicate that individual background characteristics seem to have little influence on either incoming critical thinking skills or improvements in critical thinking in the Rwandan university context.

Some of these results differ from the findings of studies conducted elsewhere, which suggests that some aspects of the conceptual framework may be inapplicable to Rwanda. Figure 6.1 depicts a modified version of the conceptual framework, incorporating the results of the quantitative phase.

Figure 6.1: Modified Conceptual Framework



As discussed in Chapter 4, the final phase of the study had both an explanatory and a complementary function. Although the complementary objectives were outlined from the inception of the study, the explanatory aspect of the final phase was dictated by the assessment results. Two unanticipated findings arose from the quantitative phase: 1) the apparent differential ability of Rwandan students to demonstrate individual critical thinking skills; and, 2) the apparent lack of impact of background characteristics on demonstrated critical thinking ability. One of the objectives of the case study phase was therefore to further investigate these unanticipated results. These findings, along with other findings from the final phase of the study, are outlined in the next chapter.

Chapter 7: Investigating the Institutional Context

The third and final stage of the study was constructed as a series of in-depth case studies. As discussed in Chapter 4, the case study phase played two crucial roles in the study design. First, the case studies had an "explanatory" function (Bryman, 2006), as they offered the possibility of deeper exploration of some of the unanticipated results arising during the assessment phase. Second, the case studies were intended to complement the quantitative results by considering additional inputs and experiences that might play a role in the cultivation of critical thinking skills at Rwanda's universities. The case study phase also considered the final research question guiding the study: How are Rwandan universities helping (or hindering) the development of critical thinking skills in their students? This chapter focuses on this final phase, outlining the specific methodology and presenting the main findings resulting from the case study analysis.

7.1 Detailed Methodology

7.1.1 Case selection

Prior to commencing data collection, it was necessary to determine which institutions to involve in the final phase of the study. Yin (2009) has argued that the study of multiple cases is akin to the use of multiple experiments in a study, as the use of identical methodologies when analysing multiple cases can be considered a means of replication. As multiple case studies allow researchers to verify results across cases, they offer the benefit of an additional method for validating study findings. Although the results of case study analysis are never strictly generalisable beyond the boundaries of the study, it is more likely that theoretical explanations arising from multiple cases will be applicable to other cases and contexts.

The decision was therefore made to include more than one institution in the case study phase. However, it only seemed reasonable to incorporate those institutions contributing a substantial number of cases to the quantitative analysis. As only 35 SFB students participated in the assessment phase, there were not enough participants to allow for an effective investigation of the full range of research questions, and the institution had to be removed from the sample. Focusing on KIST and NUR carried the additional benefit of logistical feasibility, as all of the research assistants were current or former students at one of the two institutions.

Unfortunately, KIST and NUR are not comparable institutions, given dramatic differences in enrolment numbers and the diversity of academic subjects offered on the two campuses. In order to reduce the impact of such heterogeneity on the study results, a multiple embedded case design was pursued (Yin, 2009). In a multiple embedded design, the unit of analysis is not the institution itself. Rather, the design focus is on sub-units within institutions. Students in Rwanda enrol in a particular Faculty and have no substantive interaction with the other Faculties at their institution. As a result, the near entirety of their institutional experience is limited to their Faculty. The experiences of faculty members tend to be similarly restricted. The decision was therefore made to focus on individual Faculties as the unit of analysis. In addition to mitigating concerns related to differences in the structure of the two institutions, such a design allowed for insights into a diversity of participant experiences, despite the limited number of institutions involved (Astin, 1991).

KIST has only three Faculties on its campus, while NUR has seven. Rather than overemphasising the NUR experience by including all of the NUR Faculties in the sample, three NUR Faculties were purposively selected for inclusion in the study. Two of the Faculties – the Faculty of Science and the Faculty of Applied Sciences – were selected because of their comparability to the Faculties of Engineering and Applied Sciences at KIST. The Faculty of Economics and Management was chosen as the third case because it is the largest Faculty on the NUR campus. It also seemed that the study would benefit from the inclusion of a social science discipline in the sample. The case study phase therefore focused on six Faculty cases: the Faculties of Science, Applied Sciences, and Economics & Management at NUR; and the Faculties of Engineering, Applied Sciences, and Architecture & Environmental Design at KIST.

7.1.2 Data collection

Once the Faculty cases were specified, a Case Study Protocol was developed, in order to guide data collection and maintain consistency between cases (as recommended by Yin, 2009). The Case Study Protocol outlined the objectives of the case study phase and clarified the data collection procedures to be used for all cases[89].

7.1.2.1 Objectives of the case study phase

As discussed previously, the case study phase was guided by two overarching objectives. First, it was anticipated that findings from the case studies would clarify some of the results obtained during the assessment phase. Specifically, investigation focused on

[89] The complete Case Study Protocol is included as Appendix K.

the two unanticipated results outlined at the end of Chapter 6: why students in Rwanda appear to demonstrate differential ability on individual critical thinking skills, and why there appears to be little connection between student background and demonstrated critical thinking ability in the Rwandan context.

The second objective was to complement the assessment results by considering the potential role of additional inputs and institutional experiences not considered during the quantitative component of the study. Data collection during the final phase also focused on the final research question governing the study: How are Rwandan universities helping (or hindering) the development of critical thinking skills in their students?

7.1.2.2 Data collection procedures

Triangulation between results obtained from different sources and via different methods is an important validation procedure in case study analysis (Stake, 1995; Yin, 2009). This study incorporated three different data collection methods: document review, group interviews and individual interviews. Observational methods were not included, given the high likelihood that the presence of the lead researcher would alter the dynamics of the classroom under observation. Although Faculties were the central unit of analysis, data collection was conducted simultaneously within all Faculties at a given institution. In practice, this meant that the research team completed all data collection at one institution (KIST) before beginning data collection at the second (NUR).

Interview guides for both the group and the individual interviews were drafted and piloted prior to data collection[90]. A mock focus group was first organised with five volunteer participants[91]. The research team ran the mock focus group as a "dress rehearsal" (Yin, 2009), so that the flow and timing of the questions and participatory activities could be tested. At the end of the session, participants were asked to give feedback on the process and identify any questions or activities that were unclear. A mock student interview was also conducted with one volunteer student. As with the mock focus group, the mock interview took the form of a complete interview, so that timing could be tested. The participant was also asked to share his feedback at the end of the session. The research assistants participated in both the mock focus group and the mock interview as part of their training. In addition, the lead researcher piloted the faculty interview guide with the assistance of a volunteer faculty member from SFB. Few changes were made to the interview guides following the pilot sessions. However, the number and length of questions included in the focus group guide was substantially reduced following the pilot,

[90] The final Focus Group and Interview Guides are included as Appendices K and L.
[91] All volunteers for the pilot were Rwandan undergraduate students.

as the mock session took significantly longer than had been anticipated. Some of the wording of particular questions was also clarified in response to participant feedback.

Once the pilot phase was complete and the interview guides were confirmed, data collection began at the first institution. Data collection included four components: informal document review, individual student interviews, student focus groups, and interviews with faculty members and administrators. NVivo software was used throughout the case study phase to assist with data management and analysis.

A. Document review

Data collection at each institution began with an informal document review. As students at the selected institutions, the members of the research team were personally familiar with the institutional contexts. However, the team did not include representatives of all six Faculties under investigation, nor was it assumed that the members of the research team would be familiar with all of the policies and practices governing the academic experience within the individual Faculties. A document review was therefore deemed necessary in order to identify the relevant policies on learning, teaching and assessment within each Faculty. At both institutions, the review process captured documents produced by the institution's Academic Quality Office, as well as strategic plans disseminated by the Office of Development & Planning. Institutional websites were also examined, both for central administrative policies and for information pertaining to academic programmes within the individual Faculties (as advocated by Kuh & the Documenting Effective Educational Practice Project, 2005). In addition to generating contextual information, the document review served an important purpose as the one unobtrusive method utilised in the study (Marshall & Rossman, 2006).

The document review was not exhaustive, nor did it aim to specifically address any of the research questions. Rather, relevant documents were rapidly reviewed prior to data collection in order to contextualise the resulting data and inform the use of particular probes during interviews. Information about learning and teaching policies and practices was not always available online, and it occasionally proved difficult to find up-to-date printed versions on campus. This limitation was partially mitigated by asking interview participants about relevant policies and related documentation. However, it is likely that a few documents may have been missed. The sample of documents included in the study can, therefore, only be considered a purposive sample of the available literature at each institution (Dowling & Brown, 2010).

Once identified, documents were reviewed for any reference to critical thinking. Any such references were compiled and organised by Faculty. Document content was also

examined for contextual information about institutional teaching and assessment policies. Document Summary Sheets were then completed for each document (as suggested by Miles & Huberman, 1994). If a particular document generated questions for investigation during interviews, a list of probes was written to accompany the interview guides. Otherwise, the Summary Sheets and direct text from relevant documents were entered as data sources in NVivo to be used during analysis.

B. Individual student interviews

Following the document review, interviews were conducted with individual students who had performed particularly well on the critical thinking assessment. The student interviews were intended to support the explanatory function of the case study phase by exploring individual participant backgrounds in more detail. 'Positive outliers' were recruited for participation in this phase, as it seemed that it might be possible to identify certain experiences or characteristics of their individual backgrounds that might help to explain their ability to demonstrate stronger critical thinking skills than other participants in the sample. Findings could therefore provide insight into the apparent lack of connection between student background characteristics and critical thinking ability observed during the quantitative phase. Negative outliers were not recruited, as it was assumed that low critical thinking ability was more likely to be the result of a *lack* of particular experiences, which would have been very difficult to ascertain through an interview.

Positive outliers could be identified within both the first-year and the fourth-year populations. As a result, the student interviews involved participants who had been first-year students at the time of the assessment and participants who had been fourth-year students[92]. Although interviews with both sets of participants explored background experiences, there were some differences in the interview questions depending on the participant's year in university. For those who had been in their first year at the time of the assessment, the interviews focused primarily on the participant's secondary school experiences and family backgrounds. Interviews with graduating students included more questions about the participant's experiences at university[93].

---

[92] Due to changes in Rwanda's academic calendar in 2012, the case study phase was unexpectedly conducted at the beginning, rather than the end, of an academic year. As a result, those students who had been in the first year at the time of the assessment had moved up to the second year by the time of the interviews, while those who had been in the fourth year had mostly graduated. FAED has five-year degree programmes, so the fourth-year FAED students in the sample were all in their fifth year at the time of the interviews.

[93] Complete interview guides are available in Appendix M.

During the assessment phase, participants had been asked to share their contact details with the lead researcher if they were willing to participate in a follow-up interview. The sampling frame for the student interviews was therefore restricted to those students at KIST or NUR who had taken the assessment and agreed to be contacted regarding an interview. From this list of willing participants, students were selected for participation in an individual interview if they had received an outlying score on at least one of the seven skills on the critical thinking assessment[94]. Selected students were contacted by the lead researcher regarding participation in the interview stage. Recruitment generally involved both a text message and an emailed invitation, including a detailed participant information sheet[95]. Invitations were written in both French and English.

All students meeting the eligibility criteria were asked to participate in an interview. At KIST, six students met the criteria. Of these six, three were second-year students, two were fifth-year students, and one was a recent graduate. Five of the six responded and agreed to be interviewed. One fifth-year student did not respond. All six positive outliers from KIST were male, so all of the interviewees from KIST were male. All three of the KIST Faculties were represented in the interview sample. Three participants were from the Faculty of Engineering (FOE), one was from the Faculty of Applied Sciences (FAS), and one was from the Faculty of Architecture & Environmental Design (FAED). At NUR, 10 students were contacted (six second-year students and four recent graduates). Of these 10, five responded and agreed to be interviewed. The final group from NUR included three second-year students and two recent graduates. The NUR participants included three males and two females and represented two of the three Faculties included in the analysis: the Faculty of Economics & Management (FEM) and the Faculty of Applied Sciences (FAS). None of the positive outliers from NUR were students in the Faculty of Science (FOS).

Student interviews were all held at public cafés, located within walking distance of the university campus[96]. The research team felt that a café environment would be less threatening than a location on campus, as students might feel more able to speak freely about their experiences away from the physical university environment. The location was also convenient for those still enrolled at university. The use of the same café for the majority of the interviews at a given institution also minimised the potential for bias resulting from differing interview environments (discussed in Dowling & Brown, 2010).

---

[94] An outlying score was defined as any score higher than two standard deviations above the mean for a given skill.

[95] Recruitment materials are available in Appendix F.

[96] The one exception was one second-year student at NUR who also had a job on the NUR campus. Her interview was conducted in her office.

Recruitment of recent graduates was restricted to those participants still living near their university, as there was no funding available to support transport costs for those living further afield. Fortunately, this limitation does not appear to have substantially biased the interview results, as all but two of the recent graduates agreed to participate.

The lead researcher conducted all of the interviews with the assistance of one member of the research team[97]. Interviews were held in either English or French, depending on participant preference. However, the presence of the research assistant allowed participants to respond to questions in Kinyarwanda or ask for further clarification at any time during the interview. Interviews were semi-structured in nature (Robson, 2011). As a result, although all of the interviews followed the interview guide, additional discussion around areas of particular interest to individual participants was allowed. Although all participants had already signed an Informed Consent Form at the time of the assessment, participants were asked to sign an additional Informed Consent Form before beginning the interview. With participant consent, interviews were audio-recorded to assist with analysis[98]. The lead researcher also took minimal notes during the interviews, largely to protect against the potential loss of data due to equipment failure (as suggested by Willis, 2006). Most interviews took about an hour to complete, although a few lasted only 45-50 minutes. Refreshments were provided to thank respondents for their participation.

At the end of each interview, the lead researcher and the research assistant briefly discussed any translation difficulties (as suggested by Bujra, 2006; Liamputtong, 2010). The lead researcher also wrote detailed field notes and summarised her initial reactions in a one-page Contact Summary Sheet at the end of each interview (as suggested by Miles & Huberman, 1994).

C. Focus groups

Concurrent with the student interviews, focus groups were organised with other students from the same institution. The focus groups were intended to launch the complementary component of the case study phase by exploring additional factors that may be helping or hindering the development of critical thinking skills at the two institutions. The group interviews focused particularly on student attitudes toward university and academic experiences within the various Faculties under investigation. Group interviews were selected as an appropriate methodology for this objective, as they

[97] In order to maintain consistency, the same research assistant participated in all 10 student interviews.
[98] Nine of the 10 students agreed to the recording.

allowed for the consideration of a wide diversity of student experiences, while also providing the possibility of immediate follow-up and clarification of unexpected information (Marshall & Rossman, 2006). It was also anticipated that group interviews might be less intimidating for some participants, particularly those less comfortable in English or French.

Group interviews were not limited to positive outliers, as the objective of the group interviews was to gain an increased understanding of the full range of academic experiences within the relevant Faculties. However, participation in the focus groups was restricted to those who had been fourth-year students at the time of the assessment, as one important area of inquiry was participant perceptions of the knowledge and skills that they had acquired during university.

Although not a strictly generalisable method, Krueger (1993) has argued that focus groups may generate results that are representative of the population from which the participants are drawn, provided the participants are sampled randomly and the results are obtained through a number of independent groups. As random sampling was used for the assessment phase, it seemed reasonable to assume that the results of the focus groups might reflect trends in the wider population if multiple groups were organised on each campus. A number of groups were therefore assembled that represented a wide range of disciplinary backgrounds. Although it would have been possible to organise separate groups consisting of students from each Faculty, it was anticipated that diversity of academic background within individual groups would allow for a more lively discussion. It also seemed likely that the use of diverse groups would reduce the potential for more reserved participants to allow others from their Faculty to speak on their behalf, as participants were likely to be one of the only representatives of their Faculty within a given session (Lloyd-Evans, 2006).

As with the individual student interviews, the sampling frame for the focus groups was the list of participants from the assessment phase who had volunteered to be contacted for a follow-up interview. All recent graduates (or fifth-year students)[99] from a given institution were contacted regarding participation in a focus group[100]. Potential participants received a text message and an emailed invitation with an attached information sheet. Any interested individual was then contacted by telephone by a

[99] As with the individual interviews, focus group participants had all been fourth-year students at the time of the critical thinking assessment. With the exception of the FAED students – all of whom were fifth-year students – participants in the focus groups were recent graduates of their universities.

[100] The only students not asked to participate in a focus group were those outlier students who had already been contacted regarding an individual interview.

member of the research team, in order to ascertain their availability to participate in a particular session.

At KIST, 35 students were contacted, and 17 agreed to participate[101]. The 17 participants were grouped into three sessions of 4-7 students each. Although the first two focus groups had sufficient numbers (six and seven, respectively), only two participants attended the final session, so the third group was instead structured as a two-person 'key informant' interview. Of the 15 participants[102], 13 were male and two were female[103]. All three Faculties were represented: five participants were from FOE, seven from FAS, and three from FAED. The focus groups at KIST were held in the Student Government Office. Two took place in the evening, and one was organised on a Saturday morning, so that the sessions would not conflict with the operating hours of the office.

In order to maintain consistency between data collection at the two institutions, the research team aimed to organise the same number of focus groups with a similar number of participants at NUR. In total, 38 potential respondents from NUR were contacted, and 22 agreed to participate. However, geography proved to be a significant barrier to participation for the NUR focus groups. As all undergraduate programmes at NUR are four years in length, all of the potential participants had recently graduated. Groups were initially organised in both Butare (the city where NUR is located) and Kigali (the capital of Rwanda), as it was assumed that most participants would have settled in one of the two locations after graduation. However, once recruitment began, it became apparent that the vast majority of the participants were living in rural areas, far from either city. NUR is the largest public university in Rwanda, and a large number of students come from rural backgrounds. Nonetheless, it was not anticipated that so many participants would have returned to their homes following graduation, instead of relocating to Kigali, where most graduates ultimately find employment. Although not originally budgeted, it became clear that focus groups with NUR students would only be possible if transportation costs could be provided. Ultimately, three focus group sessions were organised in Kigali, and all participants outside of Kigali were offered a transportation allowance to enable their participation. Despite the promise of reimbursement, only 11 participants managed to attend one of the three sessions, so the groups were smaller than would have been desirable (three, five and three participants,

[101] Of these 17, three were fifth-year students in FAED, and 14 were recent graduates of the other two Faculties at KIST.

[102] Two of the 17 participants did not ultimately attend any of the sessions.

[103] This proportion is representative of the gender distribution during the assessment phase, as 82.9% of KIST participants in the overall sample were male.

respectively). Of the 11 participants, 10 were male and one was female[104]. Two of the three Faculties under investigation were represented, with five participants coming from FEM and two from FOS[105]. Unfortunately, no FAS students attended the focus groups. Analysis of student perspectives on academic experiences within FAS was, therefore, restricted to the responses of FAS students during the individual interviews. The NUR focus groups were all held in a training room on NUR's Kigali campus[106]. As at KIST, two sessions were held on weekday evenings, and one was organised on a Saturday morning.

All of the focus groups were organised into two parts. The first portion of the session was structured as a guided data collection exercise, while the second half resembled a more traditional focus group discussion. As discussed in Chapter 3, past research has identified a number of educational practices that have been found to both stimulate student engagement and foster cognitive development in university students. The National Survey of Student Engagement (NSSE) was designed to capture the implementation of such practices at universities in the U.S. Administration of the NSSE in its entirety would have been nonsensical for this study, as some of the questions would have had little relevance in the Rwandan context. However, it was determined that administration of a revised version of the survey might suggest relevant trends within the Faculties under investigation, thereby illuminating potential areas for follow-up during subsequent data collection. After a preliminary 'ice breaker' activity, an abbreviated version of the NSSE was therefore administered to the focus group participants[107]. The NSSE questions asked participants about their coursework and assignments, the time they spent doing various academic and non-academic activities, and the skills that they felt they had learned during university[108]. Each participant was given an individual questionnaire

---

[104] This proportion is not representative of the gender distribution during the assessment phase, as 26.2% of NUR participants in the overall sample were female.

[105] Four of the eleven participants came from other Faculties at NUR (three from Agriculture and one from Arts, Media and Social Sciences). As their experiences did not pertain to any of the Faculty cases under investigation, their responses were not included in the cross-case analysis.

[106] NUR maintains a branch campus in Kigali.

[107] The NSSE questions used in the focus groups can be found in the Focus Group Guide in Appendix L. Some of the questions from the original NSSE were removed from the outset, given their perceived inapplicability to the study context. One additional question was removed following the focus group pilot (see Footnote 108). A complete version of the original NSSE questionnaire has been included in Appendix L for comparison.

[108] Initially, the focus group guide also included a question from the NSSE about the mental processes emphasised by university coursework in Rwanda. Response options included memorising, analysing, synthesising, making judgments and applying theories to concrete problems/new situations. This question was removed following the pilot session, as it took over fifteen minutes to explain the differences between the options to the pilot participants. Although this cannot be considered study 'data' *per se*, the lack of comprehension of this question within the pilot group suggests that students in Rwanda may be unfamiliar with the range of mental processes assumed by the NSSE to be included in university coursework.

booklet in which to record his or her answers. This allowed for anonymity of response, which was assumed to have a positive influence on the validity of the resulting data. Each survey question was read aloud in English before being orally translated into Kinyarwanda by one of the research assistants[109]. Participants were also encouraged to ask clarifying questions when necessary. Once the modified survey was complete, participants were given a short break and provided with some refreshments, while the lead researcher tallied the responses[110] and noted areas for follow-up during the subsequent discussion. During the second half of the session, the lead researcher facilitated a discussion about academic experiences within the individual Faculties. The facilitated discussion generally followed the structure of the pre-determined focus group guide, which included more specific follow-up questions about particular academic experiences referenced in the NSSE questions. However, as it was not possible to cover all of the questions in the guide in any detail, the tallied survey responses informed the selection of which topics to prioritise with each individual group.

All of the focus groups were facilitated by the lead researcher, in the company of at least one research assistant. Participants were encouraged to use any language of their preference, so as not to limit participation to those more comfortable in English. Typically, groups used a combination of English and Kinyarwanda. When participants opted to use Kinyarwanda, the research assistants simultaneously translated participant responses, in order to allow the lead researcher to ask follow-up questions (as suggested by Esposito, 2001). Although this method generally worked well, discussions occasionally took place entirely in Kinyarwanda, leaving the research assistant with no choice but to wait for a break in the conversation before translating the content. Although such instances left the lead researcher with little control of the direction of the discussion (as discussed in Marshall & Rossman, 2006), the team opted to allow such discussions to develop naturally in order to generate richer data. All of the sessions were audio-recorded, so there was little danger of losing data during these exchanges. The audio recording also helped to guard against inadvertent "filtering out" of information seen as incorrect or unimportant by the translators (Bujra, 2006). Whenever possible, a second research assistant took detailed notes during the session. The note-taker was asked to record any relevant group dynamics or reactions that could not be captured by the recorder (as suggested by Robson, 2011; Tilley, 2003).

---

[109] The research team determined that there was no need to translate the questions into French, as all participants would understand either English or Kinyarwanda. French was only used during the assessment phase because the assessment tool could not be written in Kinyarwanda, given the lead researcher's limited comprehension of the language.

[110] The tally sheet is included in Appendix L.

One danger inherent in the use of a group interview format is the potential that one or two participant perspectives will dominate the group (Lloyd-Evans, 2006; Marshall & Rossman, 2006; Mayoux, 2006). In order to avoid such an effect, questions were generally asked in such a manner as to encourage at least one response from each participant. Typically, this was accomplished by directly asking each participant if he or she had had similar experiences within his or her Faculty to those outlined by other participants in the session. As most groups had only one student representing each Faculty, this method was generally a successful way to include all participants in the discussion. Each session also began with a review of some 'Focus Group Guidelines' which were posted on the wall during the session and included suggestions such as not interrupting other participants and respecting diversity of experience (as suggested by Lloyd-Evans, 2006).

As with the individual interviews, focus group sessions began by asking participants to review and sign a second Informed Consent Form. Once the informed consent procedures were completed, the focus groups generally lasted two hours. The survey component took about 45 minutes to complete. Participants were then given a 15-minute break. The facilitated group discussion typically lasted one hour. No incentives were provided, aside from refreshments distributed during the break. At the end of each session, the members of the research team briefly discussed any translation difficulties, and the lead researcher summarised her initial reactions to the interview content in a Contact Summary Sheet for use during analysis.

D. Interviews with faculty members & administrators

Data collection at each institution concluded with a number of individual semi-structured interviews with faculty members and administrators. Faculty and administrator interviews focused largely on pedagogical practices and attitudes towards institutional learning and teaching policies. The faculty interviews therefore included questions about the lecturer's responsibilities, teaching philosophy and pedagogical techniques. Participants were also asked to discuss any challenges they faced in terms of their teaching responsibilities. Each interview concluded with a discussion of critical thinking. Faculty members were asked to give their own definition of critical thinking and to share their perspectives on the relevance of the individual critical thinking skills captured in the assessment. Administrator interviews were broader in scope, focusing more on institutional policies and challenges to academic quality at the institution.

Administrators were also asked to discuss their understanding of critical thinking and their perception of its importance for undergraduate students[111].

The intention was to use random purposive sampling for the faculty interviews by randomly selecting participants from Faculty directories (as suggested by Kemper, Stringfield & Teddlie, 2003). This technique would have allowed for the selection of participants representing certain relevant characteristics (e.g. specific departments or academic ranks), while also preventing against sampling bias (ibid.). However, as directories were unavailable in most Faculties, it was necessary to rely on more opportunistic sampling methodologies.

At KIST, a senior administrator offered to contact members of the institution's Quality Assurance Committee regarding participation in the study. The list of individuals included male and female faculty members of varying ranks, representing a diverse range of academic departments. Although the email was sent by the administrator, interested participants were encouraged to contact the lead researcher directly so as to protect their confidentiality. Seven faculty members agreed to participate as a result of the administrator's invitation. Throughout data collection, the lead researcher maintained a Data Accounting audit as a means of actively reflecting on the characteristics of study participants (as suggested by Miles & Huberman, 1994; Willis, 2006). The audit was used to guide additional recruitment by identifying perspectives that were underrepresented in the sample. Early data accounting at KIST identified a lack of participation by junior members of the faculty. In order to redress this imbalance, the lead researcher visited individual academic departments to obtain lists of faculty email addresses. Potential participants were selected at random from the email lists and contacted regarding participation in the study. An additional two participants were recruited in this manner. The total number of faculty participants at KIST was dictated by the analytical process, as additional participants were recruited until confirmatory evidence from two or more sources was obtained for most topics (as suggested by Yin, 2009). At KIST, nine faculty interviews were judged to be a sufficient number for addressing the major themes of the study. One member of the senior administration was also interviewed. The final sample at KIST included three participants from each of the Faculties, plus the senior administrator. Participants represented a variety of academic ranks (two Deans, four Heads of Department, two junior faculty members and one tutorial assistant). Four were native Rwandans, and six were expatriate staff. Of the 10 participants, three were female.

In order to maintain consistency across cases, the lead researcher aimed to recruit the same number of participants from the Faculties at NUR as had been recruited from the

[111] Complete interview guides are available in Appendix M.

Faculties at KIST. Unlike at KIST, the research team was not provided with an initial list of potential participants at NUR. Two of the Faculties under investigation (FEM and FOS) had faculty directories available. Random purposive sampling was therefore a feasible strategy to use for recruitment within these Faculties. Potential participants were selected at random from the Faculty directories and emailed an invitation with information about the study. Six participants were recruited using this method. The third Faculty (FAS) did not have a directory available, so it was necessary to rely on more opportunistic sampling methods. Recruitment within FAS was managed by a research assistant who was a current student in the Faculty. He suggested a number of candidates for participation who were then contacted by the lead researcher. Three participants were recruited in this manner. As at KIST, a senior administrator was also recruited to participate in the study. The final sample at NUR therefore included three participants from each of the Faculties, plus the senior administrator. As at KIST, participants represented a variety of academic ranks (one Dean, two Heads of Department, two senior faculty members and four junior faculty members). Nine were native Rwandan, and one was an expatriate. None of the participants at NUR were female.

Every effort was made to ensure consistency between interview settings, while also respecting participant preferences. Interviews at both institutions were held either in faculty offices or in a public café near campus. As junior faculty tend not to have individual offices in Rwanda, four interviews were held in shared office space, raising some concerns about confidentiality. In all four instances, the participant was given the option of relocating to a more private space.

Unlike the student interviews, the lead researcher conducted all of the interviews independently. There were two reasons for this decision. First, it was assumed that all faculty members would be proficient in either English or French, so there was no anticipated need for translation by a research assistant. Second, it seemed likely that faculty members would be more comfortable discussing teaching practices without a current student in attendance. Interviews were held in either English or French, depending on participant preference. As with the student interviews, faculty interviews were semi-structured in nature. Each interview began with a review of the study's Information Sheet for Participants. Each participant then signed an Informed Consent Form before beginning the interview. With participant consent, interviews were audio-recorded to assist with analysis. In contrast to the student interviews, a substantial number of faculty participants opted not to be recorded[112]. When the recorder was not used, the researcher took detailed interview notes. Interviews were generally one hour in

[112] Of the 20 interviews, 12 were recorded (seven at KIST and five at NUR).

length, and no incentives were provided. As with the student interviews, the lead researcher summarised her initial reactions in a one-page Contact Summary Sheet directly following each interview.

7.1.3 Data reduction

As the majority of interviews were audio-recorded, transcription played a prominent role in the data reduction process. The lead researcher opted to transcribe all of the English and French content of the interviews herself, as a means of becoming more intimately familiar with the data (as suggested by Robson, 2011). However, any content in Kinyarwanda had to be transcribed by other members of the research team. Twinn (1998) has argued that interview content should always be transcribed in the original language prior to translation, in order to maintain the integrity of participant responses. Following this recommendation, any Kinyarwanda content was transcribed by a research assistant in its original form. Each completed transcript was then forwarded to another member of the team to check for accuracy prior to translation.

As during the assessment phase, the team used multiple translators when translating transcribed content (as recommended by Esposito, 2001). However, it was determined that back-translation was too time-intensive a methodology for use during this phase. Instead, one research assistant completed an initial translation, which was then checked by another member of the research team. This method was deemed sufficient for ensuring conceptual equivalence, as all Kinyarwanda content was reviewed by four individual members of the team (two at the time of transcription and two during translation).

Once finalised, transcripts were emailed to the participants for their review. In a few instances, participants requested that certain segments of their transcript be removed. All such requests were granted prior to analysis.

7.1.4 Analytical strategy

Analysis of the case study data was completed in two stages: institutional analysis and cross-case analysis. Within each institution, data collection and analysis occurred simultaneously. The analytical process within institutions was iterative, as the lead researcher cycled between the collection of new data and ongoing analysis of existing content. At the end of data collection at each institution, tentative results were drafted and shared with members of the research team. Data from the two institutions were then combined, so that data from the six Faculties could be subjected to cross-case analysis.

Analytical strategies suggested by Miles and Huberman (1994) were used throughout both stages. Prior to data collection, an initial code list was generated, based

on the conceptual framework guiding the study. Similar to the strategy used by Belenky *et al* (1986) in their study of epistemological orientations, the analytical strategy for this study was both deductive and inductive in nature. Analysis was deductive in that data were examined for connections with the orienting constructs included in the conceptual framework. At the same time, new themes were considered as they emerged from the data itself (as suggested by Dowling & Brown, 2010; Miles & Huberman, 1994; Strauss & Corbin, 1998). This introduced an inductive element to the analysis. Although any deductive methodology carries the potential danger of overly prioritising pre-existing themes, the hope was that a combination of deductive and inductive methods would allow for the identification of explicit connections between the study data and the pre-existing literature, while also privileging the emergence of context-specific concepts.

7.1.4.1 Institutional analysis

Analysis within each institution began with an examination of the focus group data, as data from the questionnaire component of the focus groups required an additional data reduction stage. Tallied responses to the NSSE questions were first clustered by Faculty. Questionnaire responses were then described in written summary sheets, so that they could be uploaded into NVivo as data sources for analysis.

The questionnaire summary sheets, in addition to transcripts and field notes from the group and individual interviews, were then coded for descriptive themes. Data were first coded using the initial code list. However, as new themes emerged from the data, new codes were created. The addition of new codes required the author to frequently revisit previously coded material and recode when necessary.

As suggested by Miles and Huberman (1994), an interim summary was drafted at approximately the mid-point of data collection for each institution, which included a description of the institutional setting and some exploratory responses to the main research questions. The summary was used to develop a number of pattern codes, which were used to further elaborate the existing data. This process moved the analysis from the descriptive to the explanatory or "meaning making" phase (Miles & Huberman, 1994). During explanatory analysis, matrices were used to display the emerging themes and monitor whether sufficient data had been collected from each Faculty about each topic[113]. Causal network diagrams were then constructed to describe the major processes that appeared to be occurring within the selected Faculties. Brief narratives were written to accompany the diagrams and begin the process of building potential explanations (Yin,

[113] It was this process that dictated the final number of faculty participants at KIST.

2009). The diagrams also helped to identify potential intervening variables that may not have been considered during initial coding (Miles & Huberman, 1994).

At the end of data collection at each institution, an overarching causal network diagram, representing all of the major explanatory themes, was constructed and shared with those research assistants who had attended the institution under investigation. The verification meeting was used to both solicit feedback on emerging explanations and generate alternative perspectives on the study data. The team member feedback helped to inform any further analysis. A final summary was then written, which outlined the major findings from the institution.

7.1.4.2 Cross-case analysis

Once data collection was complete at both institutions, it was possible to begin cross-case analysis. Much like the institutional analysis, cross-case analysis included both a descriptive and an explanatory phase. During the descriptive phase, matrices were used to compare and contrast data from the six Faculties in relation to the topics suggested by the conceptual framework and the questions generated during the assessment phase of the study. During the explanatory phase, additional matrices were created which regrouped the data according to emerging explanations for any observed differences and similarities. A series of causal models was then designed which aimed to represent the different causal processes that appeared to be occurring in the different Faculties. As in the institutional stage, narratives were written to accompany the models and outline initial explanations.

At the end of the analytical process, a final validation meeting was held with the full research team. As during institutional analysis, the research assistants were asked to provide feedback on the causal models presented and to suggest any alternative explanations that the lead researcher might have missed[114]. The validation meeting was recorded, so that the lead researcher could incorporate all team member feedback into the final analysis.

7.1.4.3 Verification of findings

Throughout the case study process, findings were continuously examined for their credibility and trustworthiness, largely through the use of verification techniques advocated by Miles and Huberman (1994) and Yin (2009). Each aspect of data collection and analysis was extensively documented, and data quality was continuously assessed in terms of sampling adequacy, sampling appropriateness and the potential interference of

[114] The final causal diagrams are included as Figures 7.1 and 7.2 at the end of this chapter.

researcher effects (Robson, 2011). Disconfirming evidence and alternate explanations were actively considered through investigation of extreme cases and the search for negative examples (Miles & Huberman, 1994). "Member checking" (Yin, 2009) during the team meetings also served a valuable verification function.

7.2 Setting the stage

This section gives a brief overview of KIST and NUR in order to contextualise the results outlined in the remainder of this chapter. Important characteristics of each institution are highlighted, and the assessment results from each population are briefly reviewed. A discussion of similarities and differences between the two institutions has also been included[115].

7.2.1 The Kigali Institute of Science & Technology (KIST)

KIST was established in 1997 as an "Institute of superior Science, Technology and Management education" intended to respond to the country's acute shortage of qualified and experienced technicians, engineers and managers in the years directly following the genocide (KIST Directorate of Planning & Development, 2007). Funded initially by the United Nations Development Programme and the governments of Japan and the Netherlands, KIST today functions through the financial support of the Government of Rwanda and monies from student fees and institutional income generation (ibid.). KIST is located in Kigali, the capital of Rwanda.

7.2.1.1 Institutional structure

KIST has three academic faculties: the Faculty of Applied Sciences, the Faculty of Engineering, and the Faculty of Architecture & Environmental Design. The institution employs approximately 240 academic members of staff.

The Faculty of Engineering (FOE) is the largest faculty, housing four central departments: Civil Engineering & Environmental Technology, Computer Engineering & Information Technology, Mechanical Engineering, and Electrical & Electronics Engineering. In 2012, 1319 full-time and 183 part-time undergraduate students were enrolled in the Faculty's four degree programmes. The Faculty also administers five Masters programmes, which graduates approximately 30 candidates each year. In addition, the Faculty manages a number of diploma, certificate and training programmes.

[115] Unless otherwise indicated, all of the information in this section was found on institutional websites or in public institutional documents.

The Faculty of Applied Sciences (FAS) houses five departments: Food Science & Technology, Applied Chemistry, Applied Biology, Applied Physics, and Applied Mathematics. In 2012, 741 undergraduate students were enrolled in the Faculty (all full-time).

The Faculty of Architecture & Environmental Design (FAED) is the newest faculty, established in 2009. The Faculty houses four departments: Architecture, Creative Design, Construction Management, and Estate Management & Valuation. In 2012, 386 students were enrolled in the Faculty (all full-time).

KIST has strict quality assurance regulations. Biennial module reviews and annual programme reviews are required for each degree programme. The review process is overseen by an active Directorate of Quality Assurance. Student evaluations are a substantial component of the programme review process.

7.2.1.2 The KIST student body

The vast majority of KIST students are funded by government scholarships. In order to gain admission to KIST, students must have studied sciences in secondary school.

The KIST student body is disproportionately male (75% of the undergraduate population). During the assessment phase, it was observed that participants from KIST were more likely to have parents with higher academic qualifications than participants from NUR or SFB. Within the KIST sample, 26% of participants had parents with a tertiary level of education, in contrast to only 11% at NUR. Similarly, KIST participants were more likely to come from more affluent economic backgrounds, as 40.2% of the KIST sample fell in the top two wealth quintiles, in contrast to only 35.9% at NUR. Although data on these background variables is not available at the institutional level, the use of random sampling during the assessment phase suggests that these trends are likely to be representative of the overall KIST student body.

7.2.1.3 Assessment results at KIST[116]

Results of the critical thinking assessment at KIST were very similar to results from the total study population. As in the overall sample, the assessment evidence suggests that students at KIST have relatively weak critical thinking ability, with students being most proficient at demonstrating Skill B: Relevance and least proficient at demonstrating Skill E: Generalisability. The evidence suggests that students at KIST are not improving in their critical thinking ability during their time at university.

[116] Detailed within-institution assessment results from KIST and NUR are included in Appendix J.

7.2.2 The National University of Rwanda (NUR)

NUR was the first institution of higher learning in Rwanda. Established in 1963 just after independence, NUR is Rwanda's only comprehensive public institution. The university is supported largely by the Government of Rwanda, with additional income coming from student tuition fees, donor funding and a few small income generation initiatives, such as the university fish ponds (NUR Department of Planning & Development, 2008). The National University is the largest public university in Rwanda, enrolling over 11,000 undergraduate and postgraduate students. It is also the most internationally connected university in Rwanda, boasting a large number of donor projects, inter-university partnerships and memberships in various international organisations. NUR is located in the city of Butare, which is approximately 150 km south of Kigali in Rwanda's Southern Province[117]. Although the site of the former colonial capital, Butare is a relatively small city, comprising one central commercial area and a number of residential neighbourhoods.

7.2.2.1 Institutional structure

NUR has seven Faculties: Economics & Management; Arts, Media & Social Sciences; Science; Applied Sciences; Agriculture; Medicine; and Law. There is also a School for Foundation Language Studies and a School of Public Health (which enrols only post-graduate students). In addition to the central Faculties, there are a number of research centres, including the Centre for Conflict Management, the Centre for Geographic Information & Remote Sensing, and the University Centre for Arts and Drama.

NUR also has branch campuses in Kigali and Cyangugu (a small town on the border with the Democratic Republic of Congo). The Cyangugu campus mostly offers evening courses, while the Kigali campus is home to the Department of Journalism & Communication and a number of evening programmes. Medical students also relocate to Kigali in their fourth year in order to complete rotations in the central hospitals.

The Faculty of Economics & Management is by far the largest faculty, enrolling about 3,000 undergraduates. Most of the other Faculties enrol between 1,000 and 1,500 undergraduate students, while the Faculty of Law hosts approximately 500. Most Faculties also have a number of post-graduate programmes. Overall, there are approximately 2,000 postgraduate students at NUR. There are approximately 500 academic members of staff.

---

[117] Butare was officially renamed Huye in 2006. However, Rwandans generally continue to refer to the city as Butare.

The three Faculties included in the case study phase were the Faculty of Economics & Management (FEM), the Faculty of Science (FOS), and the Faculty of Applied Sciences (FAS). FEM includes three departments: Economics, Management, and Applied Statistics. FOS has five departments: Biology, Chemistry, Physics, Geography, and Applied Mathematics, while FAS has three departments: Civil Engineering, Electrical & Electronics Engineering, and Computer Science. FAS and FEM both offer evening programmes in addition to the central day programme.

There are a number of official quality assurance regulations in place at NUR. However, NUR is very decentralised in its administration. As a result, most quality assurance activities are conducted within the individual Faculties, rather than at the central level. As at KIST, Faculties are expected to review their teaching and assessment methods each year and to review their programmes every four years. Review paperwork is submitted to the Directorate of Quality Assurance. However, unlike at KIST, participants indicated that the Directorate at NUR does not demand strict compliance with these regulations. Programme evaluation practices therefore differ significantly between Faculties. Within the three Faculties in the case study phase, FEM appears to have the strictest review policies in place.

7.2.2.2 The NUR student body

The majority of government scholarship recipients in Rwanda attend NUR, as it is considered to be the most prestigious university in the country. As a result, the NUR student body is quite diverse in terms of socio-economic background. Diversity is also increased due to the evening programmes, which attract a number of older students, many of whom work at the university.

NUR accepts students who have studied the full range of secondary school subjects. However, placement into Faculties is determined both by secondary school subject and by National Examination score. Students that studied sciences are generally the most desired applicants, with the highest scoring students entering Medicine and Applied Sciences. 47% of NUR students are enrolled in degree programmes requiring a science background.

The NUR student body is disproportionately male (68.6% of the undergraduate population). In contrast to KIST, the NUR student body appears to be less affluent, with fewer students coming from families with a substantial amount of education.

7.2.2.3 Assessment results at NUR

Results of the critical thinking assessment at NUR were very similar to the results in the overall study population. As at KIST, the evidence suggests that students at NUR are not improving in their ability to demonstrate any of the critical thinking skills included in the assessment.

7.2.3 Relevant governing policies

As both institutions are public universities, they are governed by a number of centralised academic policies. The central administration of both universities includes a Rector (who is appointed by Parliament), a Vice-Rector in charge of Academics, and a Vice-Rector in charge of Administration & Finance. Individual Faculties are managed by a Dean and a Vice-Dean, both of whom are also active members of the faculty. All academic staff members, including Deans and Heads of Department, are required to teach a minimum of one module per term.

In 2011, Rwanda's higher education system adopted a semester system, in order to align with other systems in the EAC. As a result, the academic year now begins in September and continues until June. Each semester consists of twelve weeks of teaching and an examination period.

Most undergraduate degree programmes are four years in duration, although some specialised programmes can last five or six years. Academic programmes are overseen by the Vice Rector Academic, with the support of the Directorate of Quality Assurance.

Final examinations are moderated, both internally and through the use of external examiners from outside Rwanda. An Examinations Board approves the final publication of results at the end of each semester. A minimum score of 50% is required for a student to pass any course. Those scoring between 30% and 50% on a final examination are allowed to sit for a supplementary examination, provided they have acceptable continuous assessment marks for the course. Those who fail the course must repeat in the following year. Students who fail more than two courses must repeat the whole year. Scholarship recipients are not funded for repeat courses. Regulations at both institutions specify that the final examination be worth 60% of the final grade in any course. The additional 40% is earned through continuous assessment during the semester.

7.2.4 Differences in curricular structure

Despite similarities in the overarching policies, there are significant differences in the curricular structure of the two universities.

In 2008, NUR adopted a modular system, in compliance with the norms of the Bologna Process in Europe. The system was designed to encourage student-centred pedagogy by shifting the emphasis of the curriculum from teacher-led instruction to independent student learning (Gahutu, 2010). Modules at NUR are divided into three components: contact hours, independent work and examinations. A 20-credit module at NUR consists of 200 hours, which are divided into these three components. Of the total 200 hours, only 72-84 hours are spent in the classroom with the instructor. Three hours are dedicated to the final examination. The remainder is devoted to independent learning and the completion of assignments outside of class. Some modules consist of a number of individual sections, each taught by a different faculty member. Others are taught entirely by one lecturer or are co-taught throughout the semester by a team of instructors. Each module has one final examination. Continuous assessment expectations vary depending on the module and the instructor(s).

The curricular structure at KIST is also 'modular.' However, modules at KIST were largely implemented as an efficiency mechanism, a practice advocated as a component of many higher education reforms in the region (Ashcroft & Rayner, 2011), rather than as a pedagogical tool. The KIST modular system requires that courses common to multiple departments be available to students across the institution, rather than restricting enrolment to students in a particular programme. This improves efficiency by eliminating the need to teach the same courses in multiple departments within a given semester. The modular system at KIST does not, however, imply a specific distribution of course hours or emphasise independent learning outside of the classroom.

7.3 Contextualising the Assessment Results

As discussed at the beginning of the chapter, data from the case studies were analysed with an eye to two overarching objectives, one explanatory and one complementary to the assessment phase. The remainder of the chapter presents the findings from the case studies in terms of these two objectives.

This section discusses the first objective, which was to identify contextual factors that might explain some of the more surprising results of the assessment phase of the study. Analysis aimed specifically at answering the following questions:

1) Why do Rwandan students appear to differ in their ability to demonstrate the individual critical thinking skills included in the assessment?
2) Why does there appear to be no systematic relationship between student background characteristics and demonstrated critical thinking ability in Rwanda?

As the data from the six Faculties did not differ significantly regarding these two questions, the results presented here reflect the perspectives of participants from all six Faculties.

7.3.1 Differences between individual critical thinking skills

Both when viewed in the aggregate and when considered within individual institutions, the range of assessment scores suggests that students in Rwanda are more proficient at demonstrating certain critical thinking skills than others. As discussed in Chapter 6, Rwandan students appear to be the most proficient at determining whether or not information is relevant to a situation (assessed via Skill B) and least proficient at determining whether or not information can be generalised and/or applied to other situations (assessed via Skill E). Improvement in critical thinking ability also appears to differ by skill. Although fourth-year students are still quite weak at evaluating and using information to make decisions (assessed via the Composite Skill), the difference between the average ability of first-year students and fourth-year students to demonstrate this skill was found to be statistically significant in the overall sample[118]. A similar difference was not observed between the abilities of entering and graduating students on any other individual skill.

Feasible explanations for some of these differences can be found in the theoretical literature. Cognitive skills that require individuals to connect multiple pieces of information are inherently more cognitively complex than skills involving only one piece of information at a time. Determining the relevance of a piece of information is therefore less cognitively taxing than evaluating connections between multiple pieces of evidence. Determining the generalisability of information is a particularly difficult skill to master, as it requires both an understanding of the context in which a given piece of information was produced and a sense of how that context might differ from other contexts. As discussed in Chapter 3, cognitive load theory suggests that cognitive performance can be impaired if a given process is particularly complex or if an individual has particularly inefficient processing methods (Daneman & Carpenter, 1983; Daneman & Tardif, 1987). Conversely, individuals are more likely to perform better on tasks requiring less complex processing and/or less detailed information. Cognitive load theory therefore offers a reasonable

[118] The difference between first- and fourth-year scores on the Composite Skill was not statistically significant within either of the university populations. However, the overall result was still investigated during this phase, given that KIST and NUR participants made up the majority of the overall sample (170 out of 199 cases). It seemed reasonable to assume that a similar effect was occurring within the individual student populations, even if a statistically significant effect could not be observed at the 95% confidence level.

explanation for why students in Rwanda may be more proficient at determining relevance than other critical thinking skills, given the skill's relatively low cognitive complexity.

One could also argue that the determination of relevance may be easier for students because it is a skill used regularly in conversation (as discussed in Clark & Wilkes-Gibbs, 1986; Grice, 1975). In contrast, most of the skills included in the assessment are unlikely to be fostered automatically in daily life. As discussed in Chapter 3, the development of critical thinking requires exposure to explicit modelling and extensive opportunities to practice the use of such skills in a range of contexts (Kuhn, 1999). Without the benefit of such experiences, individuals are unlikely to demonstrate critical thinking skills when confronted with ill-structured problems, such as those presented in the study assessment. Without explicit instruction, therefore, it is unlikely that students in Rwanda would be able to demonstrate most of the skills included in the study.

The document review described in Section 7.1.2 revealed some differences in the stated importance of critical thinking as a learning outcome at the two participating universities. At KIST, there is a central Learning, Teaching and Assessment Strategy in place that emphasises the importance of critical thinking skills. The definition of critical thinking in the policy references a number of individual skills, noting that students at KIST should gain "an understanding of the limits of knowledge" and "an ability to evaluate knowledge" while also being able to "draw on a range of sources in making judgments" (KIST Directorate of Quality Assurance, 2012). Critical thinking is also mentioned as a specific learning outcome of almost every academic programme at KIST. In contrast, NUR has no central Learning, Teaching and Assessment Strategy. Individual Faculties at NUR are responsible for determining all of the learning outcomes for students enrolled in their programmes. A review of the websites of the three participating Faculties at NUR revealed only one reference to critical thinking as a learning outcome of one undergraduate programme within FEM.

Regardless of differences in the official policies, faculty members at both institutions indicated that critical thinking is considered to be an important component of a university education in Rwanda. It was anticipated that some instructors might express doubts as to the importance of critical thinking as a learning outcome, given perceived limits on individual agency in Rwanda's authoritarian political climate. However, only one participant mentioned societal pressure to conform to government policies as a potential limitation on the disposition to use critical thinking skills outside of the classroom setting. In contrast, most of the faculty participants seemed to believe that graduates would be more likely to aid in Rwanda's economic development if they were able to use critical

thinking skills to make informed decisions or propose new solutions to problems within their chosen industry.

However, when asked to define critical thinking, most lecturers described only the elements of the Composite Skill: Evaluation. Out of the 20 faculty members interviewed, only six mentioned other skills, such as questioning the credibility of a source. This result could be a reflection of the diversity in definitions of critical thinking discussed in Chapter 3. However, it is also possible that most faculty members in Rwanda equate critical thinking with the use of information when making a decision or constructing an argument. If inquiry skills, such as questioning evidence, are not a component of the general definition of critical thinking in Rwanda, it is unlikely that such skills would receive explicit attention in the university curriculum.

Evidence from the faculty interviews also generated some important contextual information regarding the apparent difference in student ability to demonstrate Skill A: Bias on the two versions of the critical thinking assessment. When presented with the list of individual skills considered in the study, two junior faculty members at NUR expressed confusion around the meaning of the term "personal bias". The term was clarified by describing the examples of bias included in the two versions of the critical thinking assessment. Although both participants understood the value of recognising when an individual may have a financial incentive to support a given decision (the type of bias described in Task Version 2), they were less convinced of the value of recognising when an individual may be supporting a given decision because of his or her personal or family circumstances (the type of bias described in Task Version 1). Another participant from the KIST FAS agreed, arguing that it is unimportant for students to learn how to detect personal bias. As discussed in Section 6.4.2, a highly significant difference was identified between the Skill A scores of participants completing different versions of the critical thinking assessment. Those completing Task 1 had an average score of 2.53 on Skill A, while those completing Task 2 had an average score of 3.02. The Government of Rwanda maintains a strict anti-corruption policy. The assessment scores and related interview responses suggest that the societal pressure to avoid corruption may translate into a general understanding of financial bias. In contrast, recognition of other kinds of personal bias, as required in Task 1, does not appear to be widely understood as a relevant critical thinking skill in the Rwandan context.

When considered together with the theoretical explanations outlined earlier in the section, the perspectives of the faculty participants provide insights into the apparent differences in student ability to demonstrate the individual critical thinking skills included in the study. As all of the other skills in the assessment are more cognitively complex than

Skill B: Relevance, the expectation would be that students would score lower on all of the remaining skills, unless they had the benefit of experience using the skills in a variety of contexts. This appears to have been the case with Skill A: Bias, as those completing Task Version 2 received a similar average score on Skill A as they received on Skill B, likely due to their familiarity with the concept of financial bias and the likelihood of their having experienced recognising it in other circumstances. Average scores were lower for all of the other skills, which suggests a general lack of exposure to the use of the other individual skills included in the assessment.

In addition to offering insight into the reasons behind the differential skill ability observed in the assessment results, interview responses provide a reasonable explanation for why students appear to be improving in their ability to evaluate information and use evidence when making decisions (as assessed via the Composite Skill), while not appearing to improve on any other individual skill. As faculty members consistently defined critical thinking as the ability to make an independent decision or to use evidence when making an argument, it is more likely that these skills feature in undergraduate curricula in Rwanda. If students in Rwanda are, indeed, gaining some experience with these skills during university, this could explain the observed difference in ability to demonstrate the Composite Skill between first-year and fourth-year students in the overall sample.

7.3.2 Student backgrounds and critical thinking

As discussed in Chapter 6, assessment results indicate that student background characteristics – including secondary school type, economic background and parental education level – have no systematic association with critical thinking ability in the Rwandan context. These results were unanticipated, given the results from other contexts described in Chapter 3. An investigation of the reasons behind these results was therefore included as an objective of the individual student interviews. During the interviews, participants were asked to discuss their secondary school experiences and share some information about their individual family backgrounds. Graduating students were also asked to discuss their experiences at university.

As discussed in Section 7.1, five students were interviewed at each institution, yielding a total sample of 10 student interviews. The five participants from KIST were broadly similar to one another, as all were male students between the ages of 18 and 25 who had enrolled in university directly following secondary school. In contrast, only three of the participants from NUR matched this profile. The remaining two NUR participants were mature female students who had elected to complete their university studies on a

part-time basis while continuing employment at the university. Both women had completed their secondary school education many years before enrolling at NUR, one having completed her secondary studies outside of Rwanda.

The 10 participants represented a range of different secondary school backgrounds. Of the eight more 'typical' students, five had attended Catholic schools (four in rural areas and one in Kigali), two had attended state-run secondary schools (one in Kigali and one in Butare), and one had attended a state-run technical school in a large town in the north of the country. One of the older participants had also attended a Catholic school in Kigali, although she completed her studies just after the genocide in the late 1990s. The final participant went to secondary school in the Democratic Republic of Congo in the 1970s. The 10 participants also represented a range of different secondary school subjects.

Despite such a diversity of institutional structure and academic subject, participant responses suggested striking similarities in the pedagogical practices of their secondary schools. All 10 participants reported being assessed almost entirely via examination during secondary school. Likely as a result of this assessment-based structure, all of the participants described their courses as following a traditional lecture format, in which students took notes following instructor explanations. Studying was universally described as a process of memorising the content of notes and class handouts. All of the younger students in the sample had attended schools in which instructors incorporated a discussion element into their classes. However, all eight participants indicated that such discussions were used only for clarification of points covered during lecture. Other more interactive learning experiences, such as laboratory experiments or student projects, were not common at any of the institutions. Although this was clearly exacerbated by a lack of materials at some of the less-resourced schools, even the wealthier secondary schools appear to prioritise lecture-based classes over interactive pedagogies. Although the content of the National Examination varies depending on the academic subject, success in all subjects seems to rely on the memorisation of facts, rather than the demonstration of any independent thought. As secondary schools are judged based on their National Examination pass rate, it is logical that schools would therefore prioritise the memorisation of information over the cultivation of critical thinking skills. Although some of the participants were very impressed with their secondary schools and felt they had benefited from committed, qualified teachers, many observed that their instructors struggled with high student enrolment and a large quantity of material that had to be covered each year. More than half mentioned having to learn aspects of the curriculum on

their own, because their teachers had not been able to cover everything during class time. Most reported having little relationship with their teachers outside of the classroom.

Evidence from the student interviews, therefore, suggests that the lack of observed relationship between student background characteristics and critical thinking ability is likely to be the result of a lack of emphasis on critical thinking skills in Rwandan secondary schools. Studies identifying links between student background characteristics and critical thinking ability have largely been conducted in Western contexts. In most Western countries, socio-economic status, parental education level and secondary school type tend to be highly correlated (Coleman, 1966; Lauder *et al*, 2006). Students of higher socio-economic status are often those with highly educated parents. In addition to being more likely to experience dialogue and access intellectual resources, such as books, within their homes (Sylva *et al*, 2012), students matching this profile are more likely to be enrolled in 'good' secondary schools that explicitly cultivate critical thinking ability from a young age. Graduates of such secondary schools then tend to enter university with higher levels of critical thinking ability than students from less-resourced secondary school backgrounds. As research has indicated that students with higher levels of incoming critical thinking ability also tend to demonstrate the largest gains in critical thinking during university (see Chapter 3), it is unsurprising that student background characteristics are systematically related to the critical thinking ability of university students in Western contexts. Data from this study suggests that the same pattern does not hold in Rwanda. Although secondary schools clearly vary in terms of quality, there appears to be little difference in pedagogical practices between schools. As the 'best' schools are considered to be those that achieve the highest pass rates on the National Examination, even the most resourced institutions appear to emphasise rote learning in their curricula, rather than implementing instructional techniques found to encourage cognitive development. These findings replicate the results of other studies of secondary school education in Rwanda (Freedman *et al*, 2011; McLean Hilker, 2011; Rutayisire, Kabano & Rubagiza, 2004; Walker-Keleher, 2006).

7.4 Critical Thinking and the Academic Experience at KIST and NUR

The second objective of the case study phase was to complement the assessment results by investigating other factors that may influence critical thinking at universities in Rwanda. Although the first phase of the study allowed for an analysis of the relationship between critical thinking ability and a number of input variables suggested by the conceptual framework, it was not possible to use the assessment results to examine the

effect of student attitudes or explore additional inputs that may influence critical thinking in the Rwandan context. Furthermore, assessment results could not be used to consider the impact of institutional experiences on critical thinking ability. The case study phase was intended to address these gaps by exploring the aspects of the academic experience that may be helping or hindering the development of critical thinking skills at KIST and NUR. This section outlines the findings related to this second case study objective.

7.4.1 The role of student attitudes and motivations

As discussed in Chapter 3, studies have indicated that student attitudes and motivations are an important aspect of student approaches to learning and engagement with the university experience (e.g. Entwistle, 1997). As student engagement is a necessary component of cognitive development, student motivations play a vital, if indirect, role in the development of critical thinking skills. Data from the focus groups and interviews were therefore analysed for insights into how student attitudes and motivations may affect the development of critical thinking skills at KIST and NUR.

All of the focus groups and student interviews began with a discussion of why students in Rwanda choose to go to university. The data suggests that Rwandan students view university education in largely utilitarian terms. Participants overwhelmingly mentioned two primary reasons for pursuing higher education:

1) To get jobs with good salaries (and, therefore, improve or maintain their own – and their family's – living standards)
2) To contribute to Rwanda's development

Many also mentioned the motivation inherent in receiving a government scholarship. Although a few mentioned learning new skills, the concept of learning for learning's sake was not referenced as a reason to attend university. Participants indicated that university education is viewed as valuable because it increases the likelihood of a higher standard of living through the acquisition of "good" jobs (defined as jobs providing a "big" salary). Faculty participants confirmed that students at KIST and NUR tend to view university education in such terms. One faculty participant explained the reasons behind these incoming motivations:

> For somebody coming here…I think what they foresee is mostly uplifting their life. Being … an African country – education is more like a means to transfer resources – financial, that is – and then when they come here, first and foremost, they see it as an opportunity to become that which will enable them to have better lives – them and their families – and less of an academic pursuit…I don't know whether

> it's the same everywhere, but usually it's more as a means to that end which they consider desirable and feel that they're going to achieve through this learning. (Junior Faculty Member, Department of Architecture, FAED, KIST)

Rwandan students are certainly not unique in equating university education with increased employment prospects. Similar motivations have been identified at universities throughout the world (see Arum & Roksa, 2011; Gouveia, 2010; Marton, Hounsell & Entwistle, 1997). However, the job market in Rwanda appears to exacerbate such utilitarian attitudes towards university education. Participants indicated that, for the majority of jobs in Rwanda, any applicant meeting the basic requirements of a position is invited to complete a written examination. Those receiving the highest scores on the examination are given the job. As a result of such employment practices, the most valuable product of a university education is, in many instances, the credential itself. Although many jobs are restricted to those with a university degree, individual *performance* at university is not widely valued. Participants indicated that entrance examinations are unlikely to feature questions related to their undergraduate programmes, so those excelling at university rarely have much of an advantage in the examination process. As one participant in a KIST focus group explained, "No one cares if you learned anything. Just you have to have the paper." As a result of these recruitment practices, there is little incentive for students to excel in their classes or to engage deeply with the content of their coursework, as those finishing first in their class receive the same degree certificate as those passing with the minimum marks.

Furthermore, many students at KIST and NUR are enrolled in academic subjects that they did not choose to study. Although incoming students at both institutions can select their preferred Faculty at the time of application, there is no guarantee of enrolment in the Faculty of one's choice. Applicants with high scores on the National Examination, particularly in the sciences, are likely to have little difficulty entering their chosen discipline. However, most departments are limited in the number of incoming students that they can accept, resulting in reallocation of many students to their second or third choice subjects. In keeping with their motivations for attending university, most students select academic subjects based on perceived employment prospects. Departments such as Civil Engineering, Computer Science and Medicine are regularly inundated with applicants, as students see engineering and health as growth industries in Rwanda, while other subjects, such as the pure sciences, receive very few applicants. Participants at both institutions indicated that enrolment practices can de-motivate incoming students, as many find themselves enrolled in disciplines that they did not select and which they do not see as offering sufficient prospects for the future.

Marton and Saljo (1976) have argued that attitudes such as those articulated by many of the participants often result in a surface approach to learning, as the primary motivator is the completion of each stage in the educational process, rather than any deeper engagement with the subject material. As discussed in Chapter 3, students taking a surface approach tend to exert the minimum effort necessary in order to complete each stage, as there is little incentive to perform at a higher level than the minimum required. The data confirms that students at KIST and NUR do appear to take such a surface approach to learning. For example, participants explained that students tend to calculate the minimum amount of studying required to obtain a sufficient number of marks to pass a given course:

> Because, see, at our institution, there are two exams – [a continuous assessment test (CAT)] and a main exam. So, for example, when, after doing CATs, you see … I have some marks in each course. You see the courses you succeed, and you choose the one [where] you have less marks, so that you make some effort to study those courses in order to succeed … in the main exam. (Graduate of FAS, KIST)

> I think that a student at KIST is more interested in hearing about their scores. We want to know if we succeeded or failed, to know how much effort we need to use and to celebrate if we succeeded. That's why students ask more about their scores more than anything...Yeah. So, appealing[119], it's like not very common. Yeah, it comes when you have like very less marks. Like you have [to take] a supplementary [exam]. That's when people argue… [If you got] less marks, like 51, you've survived [without a] supplementary. You don't care about that. (Graduate of FOE, KIST)

When asked to describe their studying strategies, many mentioned memorising notes and referencing examinations from past years, a surface approach to studying identified in other university contexts (Entwistle & Entwistle, 1997):

> I concentrated on doing exams done in the previous years of that module – questions previously asked – and compare all the lecturers in that module and know the one who is the most complicated. Then I studied his part first – the questions he asked in the previous years – and focus on that and study the rest after. (Graduate of FEM, NUR; initially expressed in Kinyarwanda)[120]

> Also, other ways I used is to pass to some questions asked during the CATs or exams of previous years, so that they can help me to see how lecturers ask questions and guide me to study, to revise. (Graduate of FAS, KIST)

In one particularly extreme example, one focus group participant confessed to not attending any classes throughout her university career. She instead chose to revise the course handouts provided by her instructor, coming to campus only to take the final

[119] "Appealing" is the practice of officially questioning a mark received on an assessment.

[120] All quotes are written in English to maintain consistency within the chapter, but any translated text is explicitly identified.

examinations at the end of each term. These findings support the results of a recent study of assessment at two public universities in Rwanda, in which Rwandan undergraduates were predominantly found to demonstrate surface approaches to learning (Mugisha, 2010).

The likelihood of students taking such a surface approach is almost certainly exacerbated by the financial difficulties experienced by many students as a result of the recent reductions in government scholarship benefits (described in Chapter 2). Participants highlighted the impact of reduced scholarships on the student experience, explaining that many students are only able to remain enrolled by taking extreme measures, such as living far away from campus or limiting themselves to one meal a day. The daily struggles experienced by students in such circumstances almost certainly exacerbate the tendency to exert the minimum effort required to complete a degree programme.

It is widely acknowledged that surface approaches to learning do not stimulate any deep engagement with the academic experience. Given that student engagement has been identified as a vital factor in the development of critical thinking skills, it is likely that student attitudes and motivations are negatively impacting the cultivation of critical thinking at KIST and NUR. Student attitudes may therefore help to explain some of the lack of improvement in critical thinking observed at the participating institutions.

7.4.2 The role of language

Data from the case study phase also suggests an additional input variable that is likely to affect the development of critical thinking skills in Rwanda. As discussed in Chapter 2, English was adopted as Rwanda's official language of instruction in 2010. Given the speed of the policy change, many students were expected to transition from a secondary school education conducted entirely in French to a university education conducted entirely in English. Although English had long been a required academic subject in secondary schools, many students entered university having never completed any academic coursework in English. The linguistic environment in Rwanda is also likely to prevent students from improving their English ability outside of the classroom, as daily communication occurs almost exclusively in Kinyarwanda, the native language common to all Rwandans. As one faculty participant explained, "this is, you know, a general problem. Once you have one language, like Kinyarwanda, which everyone understands, so people don't focus on these other difficult languages." Evidence from the interviews suggests that the low English language ability of students may be hindering the development of critical thinking ability at the participating institutions.

Although most participants claimed that students have generally coped well with the transition to English, it is clear that many continue to struggle with their language skills. Low levels of English ability were evident throughout the implementation of the study. As discussed in Section 7.1.2, student participants were encouraged to use the language of their preference during the individual and group interviews. All six of the focus groups included frequent use of Kinyarwanda. Six of the 10 student interviews were conducted predominantly in French, while one student interview was completed entirely in Kinyarwanda[121]. A majority of the faculty participants also mentioned student language ability as one of their biggest teaching challenges.

An extensive body of literature suggests that low levels of language ability can significantly hinder the development of higher-level cognitive skills, such as critical thinking (e.g. Vygotsky, Cole & Luria, 1978). Language ability was not assessed in any systematic way in this study, but the evident linguistic difficulty demonstrated by many student participants suggests that some students may simply not have a sufficient level of English to allow them to engage with the activities required for the development of higher-level cognitive skills. Low language ability is also likely to exacerbate low levels of motivation, as students who need to devote the majority of their working memory to the comprehension of course content are unlikely to have the confidence to engage with the more challenging aspects of the learning process. These diverse effects are all likely to hinder the development of critical thinking skills. Student language ability therefore appears to be an important input variable affecting critical thinking in the Rwandan context.

7.4.3 Exploring the implementation of effective educational practice

As discussed in Chapter 3, academic experiences appear to play a fundamental role in the development of critical thinking skills. The academic experiences provided by participating Faculties at KIST and NUR were therefore explored in detail during the case study phase. At each institution, information obtained during the informal document review, along with responses to the NSSE-inspired questionnaires administered during focus groups, provided an overview of current educational practice within the participating Faculties. These data were analysed alongside data from the student and faculty interviews in order to determine whether or not students in the participating

[121] It is interesting to note that most students did not choose to respond to the critical thinking assessment in Kinyarwanda. This is likely due to participant familiarity with taking assessments in either English or French, given that Kinyarwanda is not an official language of instruction at the secondary or tertiary level.

Faculties are regularly exposed to any of the effective educational practices found to encourage the development of critical thinking skills in other contexts.

7.4.3.1 Missing elements of the conceptual framework

Given that students at KIST and NUR do not appear to be improving in their critical thinking ability during university, it was assumed that the effective educational practices outlined in the conceptual framework would not be evident in either institutional context. In some respects, this assumption was confirmed by the data, as most of the Faculties do not appear to have implemented a number of the most important elements described in the literature.

For example, although the curricula of individual academic programmes are cohesive in their content, there is little evidence of interaction between courses or intentional structuring of course sequences in five of the six Faculties included in the study. This is not surprising at NUR, as the institution has very few centralised academic policies. However, the KIST administration has dedicated a significant amount of time to the development of programme curricula. KIST faculty members are expected to clearly outline the learning objectives and assignments of each course, while Heads of Department are required to coordinate annual reviews of each undergraduate programme under their purview. Despite these efforts, there is little evidence to suggest that programme curricula are designed to intentionally foster connections between courses or to progressively build individual student skills.

Furthermore, most of the Faculties do not appear to have found an appropriate balance between academic challenge and support for their students. Rwandan students do seem to dedicate a substantial amount of time to their coursework. The majority of participants indicated spending between 40 and 50 hours a week on academic work outside of class. However, there appears to be little exposure to the types of assignments found to stimulate higher-order cognitive skills. For instance, most of the programmes require almost no reading or writing. Although courses tend to include a minimum of three graded assignments, evaluation typically consists of one or two tests, some quizzes and a group project. Individual writing assignments are extremely rare. Although reading is required, students typically read summaries of lecture notes provided by the instructor and/or chapters from the course textbook reviewed during class. Additional assigned reading is unusual. Faculty participants indicated that examinations and group projects are largely preferred because of the high student-to-faculty ratio in most Faculties[122]. The

[122] Participants at NUR reported teaching courses of up to 200 students without any tutorial assistant support.

lack of required reading, meanwhile, was largely attributed to low levels of student language ability and the lack of a "culture of reading" within the student population. Faculty participants bemoaned the lack of student interest in reading, arguing that the Rwandan secondary school system does not encourage students to read or learn independently. As faculty members generally believe that students will not complete any required reading, most seem to avoid reading assignments altogether[123]. It does appear that students in Rwanda enter university with little understanding of how to read for class. However, the lack of required reading throughout the university curriculum almost certainly exacerbates such limitations. Similarly, the avoidance of individual writing assignments is likely to prevent students from improving in their writing ability. Tsui (2002) identified reading and writing as two of the most valuable ways to improve the critical thinking skills of undergraduates. It is probable that the avoidance of either type of assignment within most participating Faculties has a negative influence on the acquisition of critical thinking skills at KIST and NUR.

Although the modular system at NUR was established with the explicit intention of encouraging student-centred learning, the evidence suggests that the modular structure in its current form may actually be exacerbating this situation for NUR students. Although most of the faculty participants at NUR appear to support the idea of the modular system, all agreed that the system is not yet being implemented as it was intended. Faculty participants expressed frustration that students tend to complete only the contact hours of a given module. One faculty participant explained that students generally view the time scheduled for independent work as their "free time":

> We try to tell them. Give them a lot of things to do. But the question is…lies on whether they really do it or not. That's what we don't know … I think… students need the attention in that. The university needs to put in more effort on sensitising, because sometimes when you give them, and you tell them, you have 15 hours as part of the module to do exercises. For them, they take it to be free! They don't concentrate on it. And then you ask something related to what you are teaching, which they were supposed to look at as assignment, sometimes they don't know. I'm not going to generalise, but the majority would not consider that time as part of their [coursework]. They consider it as free time. (Junior Faculty Member, Department of Economics, FEM, NUR)

Despite their frustration, faculty members do not appear to penalise students who do not complete assigned work outside of class. In fact, most faculty members seem to expect many of their students to avoid their independent assignments. Some implied that

[123] Some faculty participants also mentioned that the university libraries have very few resources. However, both institutions have benefited from expanded digital libraries in recent years, so access to resources no longer appears to be the main limiting factor.

students were unlikely to complete assigned work given their secondary school backgrounds:

> This kind of system is ideal – is very good – is ideal when you look on the side of the students. The students can take a very large benefit – but if they are serious! If they are serious! They have to be serious. How will they be serious? That's another problem. Because the seriousness is coming from their past. From their secondary school, ok? If a student is coming from a secondary school where he's never learned to read, you see? So, the problem becomes… The problem starts with, you know, the secondary school, primary school, where they don't have enough materials to learn – which may help them to develop a culture of reading, ok? So, if a student at university level now doesn't have that culture of reading, you give him or her a free time to go and just do self-reading … It becomes a problem which is… Who is responsible for that? (Senior Administrator, NUR)

Others suggested that students could not be expected to spend a significant amount of time doing independent work, given the financial difficulties that many of them face. These participants expressed empathy with students who choose to spend their time in paid employment, instead of completing their coursework.

Apparent confusion around the intended objectives of the modular system also appears to be limiting the level of academic challenge of the classes themselves. Many faculty participants reported that the introduction of the modular system had made it harder for them to cover the required course content. Prior to the implementation of the modular system, each course at NUR comprised 120 contact hours. Most faculty participants indicated that they had responded to the new curricular requirements by simply condensing the content previously covered in 120 hours into the new 72-84 hour format. Although the new system was intended to introduce an entirely new way of teaching at NUR, very few instructors appear to have fundamentally revised their courses. In fact, instructors may now be *less* likely to experiment with new pedagogical techniques, as they are now attempting to cover the same course content in a shorter period of time.

Students also appear to find the modular system disorienting and frustrating. There seems to be little scaffolding in place to help students adapt to the unstructured nature of the curriculum, a format which is likely to stand in marked contrast to their secondary school experiences. Although NUR requires students to take a course on Study Skills in their first year, the course does not appear to support students by modelling research skills or helping them determine how to evaluate the relevance or utility of different sources of information. A number of participants described modules in which the final examination included content neither covered in lecture nor assigned as independent work. These participants expressed frustration with what they perceived as a need to guess what the lecturer had expected them to learn outside of class.

As discussed in Chapter 3, research has indicated that students learn best when they are assigned challenging work and provided with adequate support to allow them to succeed. In many modules at NUR, the balance appears to be reversed. Students are required to do very little substantive independent work during the term but are expected to pass examinations that may include a significant amount of unassigned content. Although they are informed that they should use their time outside of class to prepare for such examinations, students are provided with very little guidance or support regarding how to do so. Such an inappropriate balance between challenge and support is likely to be hindering the development of critical thinking ability in the NUR student population.

7.4.3.2 Existing elements of the conceptual framework

Although the majority of participating Faculties are missing the important elements of the conceptual framework described in the previous section, the data largely contradicts the assumption that students at KIST and NUR have not been exposed to the kinds of pedagogical techniques found to encourage the development of critical thinking in other contexts. In fact, the evidence suggests that most of the effective educational practices outlined in the conceptual framework are required elements of the academic experience at both KIST and NUR. At both institutions, for example, students reported participating regularly in group assignments and engaging frequently in class discussions. Many courses at KIST incorporate "enriching experiences" (Kuh & the Documenting Effective Educational Practice Project, 2005), such as site visits, guest lecturers and laboratory work. All KIST and NUR students complete an internship and a final culminating project, and examinations at both institutions are now required to include at least one open-ended, application-based question, in which students must synthesise their learning and propose a solution to a complex problem.

Despite the prevalence of such practices, the assessment results suggest that they are not having the positive effect on critical thinking ability suggested by research in other contexts. When the implementation of the various practices was analysed in detail, an explanation for this lack of connection between strategies and outcomes became apparent: in five of the six Faculties under investigation, the educational practices are not being implemented as intended and advocated by the literature. As a result, there is a substantial difference between the practices outlined in the literature and the experiences of students in Rwandan classrooms.

For example, although group work is required by all undergraduate programmes, the group assignments described by the study participants cannot truly be considered 'collaborative learning'. Participants indicated that group assignments are typically quite

small in scope, meaning that they can feasibly be completed by one or two people. Generally, group projects seem to include a short written assignment and a class presentation. Although some faculty members require all members of a group to participate in the presentation, many allow groups to select their own presenters. Faculty and student participants at both institutions acknowledged that the majority of students do not participate in group assignments. Given the surface approach to learning taken by most Rwandan students, it is perhaps unsurprising that many students view group assignments as opportunities to get "free marks", rather than truly collaborative learning opportunities. Participants indicated that one or two students generally complete the assignment for a group, while other group members contribute money for photocopying fees and add their names to the final report. Collaborative learning has been found to influence critical thinking *when the collaborative element introduces students to alternating viewpoints and perspectives*. The group projects currently implemented at KIST and NUR are therefore unlikely to have any effect on cognitive development, as there is very little actual collaboration involved. Furthermore, those student participants that did report participating in group assignments appeared to view collaborative work as beneficial because of the increased likelihood of 'getting it right' as a group. Rather than viewing group work as an opportunity to learn about differing perspectives, students in Rwanda seem to reach out to their peers as a way to verify their understanding of course content.

Similarly, class discussions, although commonplace, do not appear to offer the benefits outlined in the literature. Participants generally described class discussions as being an opportunity for students to ask questions and clarify concepts, rather than a forum for exploring new ideas or differing perspectives. One graduate of the KIST FOE, for example, defined class discussions as a time when "each [student] brings his ideas, then we choose those ones we all agree with", while another explained that discussions allow faculty members to clarify difficult concepts:

> It depends on the strategy of some lecturers. A lecturer can first… He can ask a question… And try to capture the points of views of many people. And many people think differently. And, we have a lecturer – he used to do that. He asked question, and he select some people to answer. And thereafter, because we are many, we give different responses. And the teacher decides which one is true. Which are false. So, like that. (Graduate of FOE, KIST)

NUR participants described classroom discussions in similar terms. For example, one faculty participant defined class discussions as follows:

> When I'm teaching maybe three hours, I teach for like two hours and 15 minutes. Then we have 45 minutes of different questions on the overall coverage of what we did. If there are queries there. If there are questions. If there are things that they did not understand well. Then we clear them up before we close on the chapter. (Junior Faculty Member, Department of Economics, FEM, NUR)

Although referred to as "discussions", there is little evidence that such practices involve any debate or dialogue between students. As discussions focused on getting the 'correct answer' do not encourage students to question their own understanding of a concept (Dewey, 1933 (renewed 1960); Kuhn, 2005), such practices are unlikely to have a positive impact on critical thinking skills. The benefits of discussion are likely to be further reduced by low levels of English language ability. Student participants indicated that student language ability often prevents active participation in class discussions:

> Normally [we don't have discussions]… For students who studied in the Francophone system, it's very difficult to speak in English. So, [discussions] could waste a lot of time. (Graduate of FEM, NUR; initially expressed in French)

Furthermore, some participants, particularly at NUR, reported that lecturers occasionally avoid class discussions altogether when feeling pressured to complete course content. As one participant explained:

> Sometimes [the lecturer] take the time of discuss about the course. Not always... As you know, the time for learning in class is not too much … Not always the lecturer come and take the discussion. There is some times they come and we discuss, but there are a lot of time, they come and teach without discussion. (Graduate of FOS, NUR)

The implementation of 'active' teaching methods, such as laboratory work, also seems to differ from the practices advocated in the literature. At KIST, students are generally asked to follow a prescribed laboratory protocol, rather than generating their own experiments. At the end of a given experiment, students write a summary laboratory report, either individually or in groups. A tutorial assistant from the FAS explained that plagiarism is a common issue with laboratory reports, as students copy one another's results and conclusions. Students at both institutions indicated that students are often prevented from using expensive laboratory equipment. If the use of equipment is a necessary component of an experiment, lab technicians are asked to conduct the experiment while the students observe. Furthermore, a number of faculty participants indicated that the timetable does not allow for lengthy experiments. Some choose to address this issue by organising one or two laboratories during the semester, instead of regularly including laboratory work throughout the term. A lack of basic materials, such as reagents, and high student-to-faculty ratios also limit the inclusion of laboratory work in the curriculum. Although students undoubtedly benefit from occasionally observing experiments, it is unlikely that such methods stimulate much student engagement or foster skills of inquiry. The limited amount of 'hands-on' experience also seems to have a negative impact on student motivation. Many focus group participants expressed frustration over the lack of "practical experience" that they gained during their four years

at university. Students could see little connection between the theoretical content of their courses and job opportunities in their chosen field. Research has indicated that students are much more likely to feel engaged with their academic coursework if they can see the relevance of the content (Entwistle, 1997), so the perceived lack of connection between the undergraduate curriculum and the job market is likely to have a negative impact on student engagement.

The "enriching experiences" required by all programmes also seem to offer little benefit as currently implemented. Students are generally positive about the internship requirement, as many view internships as practical and potentially useful for gaining future employment. However, participants complained that many students do not manage to get internships in their field of study. Furthermore, there appears to be little connection between the internship experience and the rest of the university curriculum. Although students are required to complete a final internship report, assessment of the report seems to be based primarily on adherence to the required reporting format, rather than a student's ability to link the contents of their coursework to their internship experiences. A faculty participant responsible for assessing the internship reports for his department explained the marking process as follows:

> The first thing, of course, we look for the stamp. And we have to make sure that they are following the format. (Junior faculty member, Department of Mechanical Engineering, FOE, KIST)

The participant mentioned no further marking criteria. In at least this department, the emphasis is clearly on the completion of the internship requirement, rather than any deeper connection between the internship experience and the theoretical content of the programme.

Studies have indicated that a culminating final project can help students to synthesise and integrate the information that they have learned during their time at university (Bok, 2006; Kuh & the Documenting Effective Educational Practice Project, 2005). The final projects completed at KIST and NUR, however, bear little resemblance to such educational practice. As discussed earlier in the chapter, most students have very few opportunities to practice academic writing during their time at university. Aside from one course in Research Methodology, it appears that little additional guidance is provided to students to help them prepare for their final projects. However, most programmes require a final written project of at least 20,000 words. The lack of preparation for such an undertaking, combined with low levels of English, often results in blatant plagiarism, a fact acknowledged by student and faculty participants alike. Participants indicated that the recent decision to eliminate scholarship funding for final projects has resulted in an

increase in the number of students copying data from other sources, instead of collecting it themselves. Projects requiring transportation or the provision of incentives to study participants are not possible given the financial implications. Neither are projects based on laboratory experiments, as they typically require the use of expensive private laboratories due to restrictions on student use of university laboratory facilities. The high student-to-faculty ratio in most departments also limits the possibility for meaningful supervisory relationships, as faculty members can be responsible for supervising up to 50 student projects each year. In its current format, it is highly unlikely that the final project would have a significant influence on student critical thinking skills.

The new requirement that every final examination must include at least one open-ended synthesis question is also unlikely to have much of an impact on the development of critical thinking skills in most of the participating Faculties. The norm at both institutions is for examinations to include only one of these types of questions. As a result, it is possible for students who have received good marks throughout the semester to avoid the open-ended question altogether and still pass the examination and the course. Given their demonstrated approaches to learning, it is highly likely that many students at KIST and NUR would make such a choice. Furthermore, students do not seem to have any opportunity to practice answering similar questions during the semester, given the lack of writing assignments and the closed nature of most class discussions. Participants also indicated that they rarely receive feedback on their examination responses. As discussed in Chapter 3, synthesis-type examination questions can be helpful for developing critical thinking skills, provided students are given sufficient feedback and opportunities to practice during the semester. Given the apparent lack of such support within most of the participating Faculties, it is highly unlikely that the introduction of synthesis-type examination questions has had any positive effect on student critical thinking skills.

7.4.3.3 The role of academic experiences

All of these individual components combine to create a learning environment that is highly unlikely to stimulate the development of critical thinking skills. As discussed in a recent study of student engagement at American universities, the simple adoption of educational practices is never sufficient for affecting systemic change. Practices must be effectively implemented in order to positively affect student engagement and influence learning outcomes (Kuh & the Documenting Effective Educational Practice Project, 2005). The evidence from the case study analysis indicates that none of the educational practices included in the conceptual framework have been effectively implemented in the majority of participating Faculties. Although many of the practices are specifically included in

institutional policies on pedagogy and assessment, they have been fundamentally altered during implementation. The resulting adaptations are unlikely to have a positive effect on either student engagement or cognitive development. The evidence therefore suggests that the apparent lack of improvement in critical thinking at KIST and NUR is likely to be attributable to ineffective implementation of the educational practices found to support the development of critical thinking in other contexts.

However, there appears to be an exception to this rule. Although observed in five of the six participating Faculties, a similar problem of implementation was not observed in the Faculty of Architecture & Environmental Design (FAED) at KIST. In fact, in contrast to other Faculties in the sample, FAED appears to have effectively implemented all of the educational practices outlined in the conceptual framework.

7.5 Investigating the Outlying Case

Data from the student and faculty interviews indicates that there is a fundamental difference between the learning environment within FAED and the learning environment within the other five participating Faculties. In this section, the academic experiences provided within FAED are described in some detail. The section concludes with an analysis of how these experiences appear to be influencing the learning orientation and critical thinking ability of FAED students.

7.5.1 Academic experiences in FAED

The first characteristic that differentiates FAED from other Faculties in the sample is the progressive nature of the programme curricula[124]. Architecture students, for example, follow a curricular sequence that was explicitly designed to foster incremental improvement in a number of overarching learning objectives. Improvement in writing, for instance, is a learning objective of nearly every course in the Architecture curriculum. First-year courses require short writing assignments of only a paragraph or two, but, by the third and fourth year, students are completing a substantial amount of individual writing in their classes. Similarly, courses require more in-class participation as students progress through the programme. The sequence of courses is also intended to help students learn how to defend their design choices. This ability is fostered through weekly oral critiques, in which students are asked to explain and defend their work to their peers and instructors. Like writing assignments, oral critiques increase in length and difficulty as students progress through the programme. Critical thinking is also included as an

[124] Given time constraints, data collection focused exclusively on the departments of Architecture and Creative Design.

explicit learning objective of all programmes in the Faculty. Nearly every course is designed to encourage the development of critical thinking skills through assignments and/or classroom practices. Throughout their time at KIST, students in the Faculty are expected to incrementally assume increased responsibility for their own learning. In practice, this means that students spend more time with instructors in the first few years of the programme than they do later in their university careers. By their final year, students spend the majority of their time working independently on their final projects.

There are a number of benefits to the implementation of such a progressive curriculum. First, the curricular structure is based on a realistic assessment of incoming student ability. Rather than assuming that students will be able to demonstrate higher-order cognitive skills from the beginning of their undergraduate career, the curricular structure builds upon a supposed low level of linguistic and cognitive ability. Such an explicit connection with prior learning has been found to have a positive influence on cognitive development in other contexts (as discussed in Chapter 3). Second, the progressive structure provides a scaffold for student learning by gradually exposing students to challenging content and expectations, while simultaneously providing sufficient opportunities for them to practice and improve. Finally, the progressive sequence means that students have the benefit of exposure to effective educational practice from the beginning of their time at university. As discussed in Chapter 3, it can be detrimental to wait until the final years of university before expecting students to complete activities requiring the use of higher-order cognitive skills, such as critical thinking.

In contrast to other Faculties in the sample, FAED courses seem to require an appropriate level of academic challenge. Participants indicated that students in FAED are expected to do a significant amount of independent work. For instance, theory classes often include weekly readings which form the basis of class discussions. Rather than assuming that low levels of English ability and lack of preparation for a reading-based curriculum will prevent students from completing assigned readings, faculty participants from FAED described using a number of pedagogical techniques to help students overcome such barriers. One faculty participant, for example, mentioned using a "memo" system, in which students are required to respond to a series of written questions about their reading in advance of class. In addition to ensuring completion of the assigned work, the memos help students to clarify their understanding of the readings, thereby increasing the likelihood that they will be able to engage meaningfully in class discussion. The memos also help the instructor to plan for class, as he is able to ascertain the aspects of the reading that students struggled to comprehend. Students in FAED are also required to

complete frequent independent writing assignments and numerous design projects. As discussed earlier in the section, design projects are subject to weekly oral critiques, in which students must defend their design choices and respond to questions from their peers and instructors. Students are likely to be challenged by their writing assignments, as there is a strict anti-plagiarism policy in place in the Faculty, which is enforced by all instructors. If a student is found to have plagiarised an assignment, he or she is given the opportunity to correct it. However, a second offense can result in expulsion from the Faculty. As one graduating student explained,

> In the beginning, nobody was taken out from the campus because of [plagiarism], but they call you, they tell you, you have to do this. They want to know if you do it aiming to do it – or you did it because you didn't know it. But, they tell you, things are done like this. Never do it again. (Fifth-year student, Department of Architecture, FAED, KIST)

While demanding a lot from the student population, FAED also seems to provide a significant amount of support and guidance to help students succeed. For instance, faculty participants mentioned giving frequent and detailed feedback to students on their designs and written work. As one faculty participant explained:

> At the end of the semester, we have a report for every student, like what areas you are weak in, you need to improve in this area, and this and that. So, students get that, and it's really been helpful. For the past three years, we've been doing it. Yeah, we've seen students improve, because they are told – while the work is still up there, the … examiner tells them why this work is not as good as this work, and they get to see. So, come the next semester, then they have that idea. (Senior Faculty Member, Department of Creative Design, FAED, KIST)

The progressive nature of the curriculum itself is also likely to act as a scaffold for learning, as it allows students to gradually gain confidence in their abilities.

Like courses in other participating Faculties, courses in FAED seem to involve the frequent use of active and collaborative learning techniques. However, unlike in other Faculties, such methodologies appear to be implemented as the literature suggests. Studio courses form the core of the FAED curriculum, providing students with substantial opportunities to practice their craft. During the weekly oral critiques, peers are encouraged to critique one another's work, a practice that is likely to help students form their own opinions about design while also learning to listen to – and incorporate – differing perspectives. The majority of assignments in FAED are individual. However, when group work is assigned, there are strict parameters that ensure participation by all students. Participants indicated that assignments are vast in scope, so it would not be feasible for one or two students to complete the work on behalf of others in the group. In fact, students are often assigned individual tasks on group projects, so that everybody in the group has a particular function to perform. Participants also stressed that all group

members are involved in presenting and defending group work, so there is distributed responsibility and incentive for everybody to participate actively throughout the project. In some courses, group assignments lead directly into individual assignments, meaning that students can only succeed individually if they have been engaged throughout the group component. An explicit emphasis on diverse perspectives also appears to guide the implementation of class discussions within the Faculty. Discussions seem to be used to generate debate about controversial topics in architecture and design, rather than to clarify lecture content. In contrast to the descriptions of class discussions held in other Faculties, FAED student participants described class discussions as opportunities for forming opinions about significant issues:

> Everybody has time to raise up his ideas and views, and then afterward, whether a student or an instructor … come up with the conclusion and you can agree or disagree about some, the conclusions. In discussion, they aren't… There isn't a concrete understanding for everybody. That is why I think discussions are important. Everybody brings his opinion, and you try to hear from everybody, and then on your own, you can make up some decisions. (Fifth-year Student, Department of Architecture, FAED, KIST)

Participants also indicated that discussions are frequently based on assigned readings, allowing for dialogue about differing perspectives.

The final projects completed by FAED students are likely to have a more positive impact on critical thinking ability than the final projects completed by students in other Faculties, as FAED students have the benefit of a substantial amount of guidance and support. Students in Creative Design, for example, are required to select the topic for their final project in their third year, so that they can use the required Research Methodology class to prepare their project design. This practice allows students to benefit from significant feedback prior to the actual implementation of their project. Having practiced academic writing throughout the curriculum, students in FAED are likely to be prepared for the work involved in writing their final projects. The strict anti-plagiarism policy in the Faculty also prevents students from replicating results or copying conclusions from other studies. Students in FAED struggle with the same scholarship reductions as the rest of the KIST student population. However, participants indicated that faculty members work with students to identify local projects that do not require significant expenditure, thereby removing one of the most substantial barriers to effective implementation, while still retaining the purpose of the assignment.

There also appears to be minimal use of written examinations within FAED. In studio courses, students are evaluated on their design projects and oral critiques. Theory courses do tend to include a final examination, but participants indicated that final student grades are based primarily on individual assignments and participation during the

semester. In line with KIST's assessment policy, all final examinations in FAED must include at least one open-ended synthesis question. However, unlike in other Faculties, there is evidence that FAED instructors work with students during the semester to help them improve in their ability to answer such complex questions. As one faculty participant explained,

> … at the same time, we avoid a situation where they are getting these questions for the first time in the main exams. So, they get used to these kind of responses…the questions that have anything to do with analysis … We felt that maybe one way that we could improve that was if they had similar settings, similar questions, similar vocabulary, as often as possible, to build up their vocabulary and also to get them familiar with those terms. (Junior Faculty Member, Department of Architecture, FAED, KIST)

Such integration between classroom practice and assessment design is likely to increase student engagement and substantially improve student ability to synthesise and apply their understanding of course content.

7.5.2 Evidence of impact

According to the conceptual framework, the effective implementation of the educational practices described in the previous section should result in increased student engagement and improved critical thinking ability within the student population. This suggests that students graduating from FAED should differ substantially from other graduating students in the sample, both in terms of their learning orientations and their critical thinking ability. This assumption was tested in two ways during analysis. First, interview responses between fourth-year FAED and non-FAED students were compared in order to identify any potential differences in learning orientations between students graduating from FAED and students graduating from other participating Faculties. Second, follow-up analysis was conducted on the critical thinking assessment responses completed by FAED and non-FAED students in order to ascertain any qualitative difference in demonstrated critical thinking ability between the two groups.

7.5.2.1 Differences in approaches to learning

There is no evidence to suggest that students entering FAED differ substantially from students entering any other participating Faculty. Prior to their enrolment at KIST, most FAED students completed secondary school in Rwanda, so they are likely to have entered university with similar levels of critical thinking and language ability to students entering other Faculties. Faculty participants also describe incoming FAED students as

having the same utilitarian attitudes toward university education as other incoming students at KIST.

However, analysis of interview data indicates a significant difference between the learning orientations of fifth-year FAED students and the learning orientations of students graduating from other Faculties in the sample. Five fourth-year FAED students took the critical assessment during the first phase of the study[125]. Of these five, four agreed to participate in either a focus group or an interview during the case study phase. When compared to comments made by other students in the sample, the responses of these four students clearly indicate a deeper approach to learning than the approach demonstrated by their counterparts. For example, one fifth-year student explained how his attitude towards reading assignments changed during his time in the Faculty:

> Before, I didn't like reading, because I really – I was this person who liked just watching movies! But, because I had to do it, I did it. At the beginning, I didn't like it at all. But, finally, because I was forced to do it, I find it's something interesting and teaching also. I was gaining other things outside of the classes, I think. It was good. So, I'm reading more than I was used to do in the past. (Fifth-year Student, Department of Architecture, FAED, KIST)

The same participant articulated a clear understanding of how individual assignments contribute to student learning:

> All of these assignments are like another part of the course. Because, I can say the course is a half and the assignment is also another half, so both together, they form the whole. So, if you miss the assignment, it means you miss half of the course, so… It's really something important, and I think everybody understands why this needs to be done like that.

Another student participant described the FAED student culture as being one in which students encourage each other to give maximum effort on their assignments:

> Everybody [is] committed, so … when it is in group works, maybe there are some people who don't work as hard as others do, but we try to talk as a group and say, "you know if you're tasked this one, you have to come up with this information!" And among ourselves in the group, we are able to organise ourselves, so that…we get our best. (Fifth-year Student, Department of Architecture, FAED, KIST)

Such attitudes contrast sharply with the perspectives of graduating students from other Faculties described in Section 7.4.1. Faculty participants also identified a palpable change in the motivations of incoming and finalist students within FAED:

> But you could see from the class that indeed … money, as an end in its self – or resource, I mean, getting livelihood – has now stopped being the main focus. You could see them – some are really interested in specific areas – which are not exactly what we thought they were looking for initially … Ultimately, when they move along the process, they now get more interests besides that architectural bit – which are multi-discipline … Some get really interested in research. Some get

[125] All fifth-year FAED students in the sample were students in the Department of Architecture.

> interested in community – because there are faculty who are interested more in the non-governmental field – so some get more into community participation. Issues to do with working with the resource deficient places. So, I think they get more than they even imagine when they're coming in… And we see that with the way that they grow up and what they become. (Junior Faculty Member, Department of Architecture, FAED, KIST)

These data suggest that fifth-year FAED students have a deeper approach to learning than students graduating from other Faculties. It seems likely that the educational practices described in the previous section have contributed to the apparent change in the learning orientations of FAED students.

Faculty participants also indicated that these same educational practices seem to have a positive effect on student language ability. Although linguistic improvement was not assessed formally in this study, it was evident that the fifth-year FAED students involved in the case study phase could demonstrate higher levels of English proficiency than many of their counterparts. Both faculty and student participants from FAED mentioned that students graduating from FAED had improved dramatically in their English skills during their time at KIST. Lecturers and administrators from other KIST Faculties also acknowledged that FAED students seem to improve in their language proficiency more than other students at KIST.

7.5.2.2 Differences in critical thinking ability

In addition to increased levels of student engagement, it would seem likely that FAED students would demonstrate greater gains in critical thinking ability than students from other participating Faculties, given their exposure to the educational practices described in the previous section. One would therefore expect FAED students to have outperformed other students on the critical thinking assessment.

Unfortunately, it was not possible to formally test this assumption, as there were an insufficient number of fourth-year FAED students in the sample to allow for a statistical comparison of assessment scores between fourth-year FAED students and fourth-year students from other Faculties. However, using the principle of case control analysis, it was possible to investigate whether FAED students had demonstrated a qualitatively different level of critical thinking ability in their assessment responses than other students in the sample. In case control analysis, a technique typically used in quantitative studies, cases are matched with other cases that have identical values on potentially confounding variables (Tolmie, Muijs & McAteer, 2011). In the follow-up analysis, the assessment responses written by the five fourth-year FAED students in the sample were compared to the responses of five 'matched' fourth-year students from other Faculties at KIST. As

analysis of the assessment results had indicated that the version of the assessment and the gender of the participant were potentially confounding factors (see Chapter 6), FAED students were matched with participants of the same gender who had completed the same assessment task. Matching was also conducted along linguistic lines, as comparisons were deemed to be more valid when made between two assessment responses written in the same language. Other variables were judged to be less important, as the main study results had indicated little systematic difference between students from different backgrounds. However, in order to select an appropriate match, if there were multiple potential matches on gender, language and task number, the most similar match to the original case was selected in terms of family background, economic background and secondary school background. Most cases matched on all but one variable. The final matching profile is outlined in Table 7.1.

Table 7.1: Characteristics of Matched Pairs for Case-Control Follow-up Analysis

Pair	FAED Student	Non-FAED Student	Matched Characteristics	Non-Matched Characteristics
1	Male Fourth Year; French Response; Task #1	Male Fourth Year; French Response; Task #1	Same Secondary School Background; Both missing SES Quintile	FAED student = Parents with tertiary education; Non-FAED student = Parents with secondary education
2	Male Fourth Year; English Response; Task #1	Male Fourth Year; English Response; Task #1	Same Secondary School Background; Same SES	FAED student = Parents with secondary education; Non-FAED student = Parents with primary education
3	Male Fourth Year; English Response; Task #2	Male Fourth Year; English Response; Task #2	Same Secondary School Background; Same Family Background	FAED student = Highest SES Quintile; Non-FAED student = Second-to-Highest SES Quintile
4	Male Fourth Year; English Response; Task #2	Male Fourth Year; English Response; Task #2	Same Secondary School Background; Same Family Background	FAED student = Highest SES Quintile; Non-FAED student = Second-to-Highest SES Quintile
5	Female Fourth Year; French Response; Task #2*	Female Fourth Year; French Response; Task #2	Same Family Background	FAED student = Urban Public Sec. School + Highest SES Quintile; Non-FAED student = Urban Private Sec. School + Second-to-Highest SES Quintile

** Only one possible match was a female fourth year with a French response to Task #2*

Once paired, basic content analysis was performed on the 10 matched responses. Responses were analysed in terms of the following six categories:

1. Demonstration of the Basic Skills of Argument
2. Use of Evidence
3. Use of Logic
4. Consistency of Response
5. Certainty of Response
6. Epistemological Orientation

The selected categories were based largely on Kuhn's study of the skills of argument (1991), in which Kuhn assessed the ability of adults to make an argument or assess the strength of other peoples' arguments about ill-structured problems. Given the parallels between Kuhn's study and the structure of the adapted critical thinking assessment, categories from her study were seen to offer a valid alternative method for evaluating assessment responses. In Kuhn's study, participant responses were classified in terms of the type of evidence used to support an argument (i.e. actual evidence, "pseudo-evidence" or no evidence), the consistency of an argument, and the certainty with which a

respondent stated his or her argument. Kuhn's study also referenced the importance of epistemological orientation, so this was constructed as a separate category, using the stages from King and Kitchener's theory of reflective judgment (1994). The final two categories emerged inductively from the data, as a review of the content exposed differences in student ability to demonstrate basic skills of argument (i.e. to be able to make and evaluate an argument) and to use logic when assessing a presented argument.

The results of the case control analysis are not entirely conclusive, as FAED students did not outperform the non-FAED students on every single category. However, in the aggregate, the FAED students did demonstrate stronger evidence of critical thinking ability than students from other Faculties at KIST. The results of the case control analysis are as follows:

1. Basic Skills of Argument: Most of the FAED students could demonstrate the ability to both evaluate another person's argument and make their own argument using evidence. In two pairings, the FAED student demonstrated stronger ability in this category than their paired response. In two pairings, the responses were equal. In one, the FAED student performed poorly in comparison with the non-FAED response.

2, Use of Evidence: FAED responses were stronger in terms of the use of evidence to support claims, as they were all less likely than their paired respondent to use anecdotal evidence or make unsupported claims. Four out of five of the pairs reflected this trend. Only one FAED response demonstrated some unsubstantiated claims and some use of anecdotal evidence, while all five of the non-FAED responses demonstrated at least some unsubstantiated claims and/or claims supported by anecdotal and/or external evidence.

3. Use of Logic: The first question on the assessment asked respondents to evaluate the validity of an illogical claim. Four out of five FAED students questioned the validity of the claim, with three out of five offering potential explanations. The one FAED student who did not explicitly question the validity of the claim did go on to offer an explanation for the supporting evidence. Of the five non-FAED students, only one explicitly questioned the claim, with two offering explanations of the evidence used to support it.

4. Consistency of Argument: There is no conclusive result in this category. Of the 10 responses, only five were consistent in their argument throughout their response. Three of these five were FAED students. However, this result does not indicate that FAED students are necessarily stronger at putting together a consistent response than non-FAED

students, as two non-FAED students were consistent in their argument, while two FAED students were not.

5. Certainty of Argument: The responses written by FAED students were more likely to acknowledge other viewpoints and/or reference conflicting evidence in their responses. In fact, all five FAED students did this, while three out of the five non-FAED students ignored conflicting evidence in their responses. The other two non-FAED students acknowledged conflicting evidence but adapted the evidence in order to incorporate it into their own pre-existing arguments, rather than tolerating uncertainty within their responses. None of the FAED students ignored conflicting information in the documents. FAED students were also more likely to recognise the need for more information. Three out of five explicitly stated the need for more information, while none of the non-FAED students did so.

6. Epistemological Orientation: All ten of the responses suggested a "pre-reflective" epistemological orientation, according to King and Kitchener's classification. As "pre-reflective" thinkers assume that answers are certain and obvious, they do not believe it is necessary to provide justifications for their beliefs. Individuals demonstrating such an epistemological orientation tend to seek answers from authorities, and, in the event that authorities disagree, to classify one expert as incorrect or misguided, rather than acknowledging the validity of differing perspectives (King & Kitchener, 1994). All 10 assessment responses indicated a belief that knowledge is certain and that information obtained either through direct observation or from an authority should be believed to be true. However, three out of five FAED responses indicated some comfort with the idea of temporary uncertainty, acknowledging that not everything is known at the present moment. Although not suggesting that knowledge itself might be uncertain, the responses reflected an understanding that the search for more information helps to improve knowledge about a given topic. In contrast, all five non-FAED students discussed the information included in the documents – as well as any knowledge obtained through their own direct observation – as being certain, absolute and unchanging.

To supplement these results, initial assessment scores were compared for each of the pairings. Although the matched pairs had received similar scores on many categories, scores on the Composite Skill: Evaluation were dramatically different between the two sets of scores. Four out of five FAED students had received a 4 or a 5 on Skill G: Evaluation of Connections; three out of five FAED students had received a 5 on Skill H: Evaluation of

Support; and three out of five FAED students had received a 4 or a 5 on Skill I: Use of Evidence. In comparison, only one non-FAED student received a 4 or a 5 on any of the three Composite Skill sub-skills. As most faculty participants in the study equated the Composite Skill with critical thinking, it is likely that the explicit critical thinking instruction described as a component of the FAED curriculum would emphasise these sub-skills. Given that other Faculties do not appear to explicitly incorporate critical thinking instruction into their curricula, it is not surprising to find that fifth-year FAED students could demonstrate stronger ability on the Composite Skill than their colleagues from other Faculties.

The evidence from the follow-up case control analysis appears to confirm that students graduating from FAED are able to demonstrate higher levels of critical thinking ability than their colleagues graduating from other Faculties. This suggests that, in addition to encouraging deeper levels of student engagement, the educational practices implemented within FAED may indeed have a positive influence on the development of critical thinking in the student population.

7.6 Educational Practice and Departmental Culture

The marked differences in the implementation of effective educational practices between FAED and the other participating Faculties presented an additional question for consideration during the final analysis: Why are lecturers in FAED able to effectively implement such practices, while others adapt them in such a way as to substantially limit their impact? In this final section, evidence from the interviews will be explored in relation to this question, and conclusions will be drawn regarding the role of departmental culture in the process of pedagogical change.

7.6.1 Adapting, not adopting

As discussed earlier in the chapter, effective educational practices – although actively promoted and, in some cases, required – have not been fully adopted by individual faculty members in five of the six participating Faculties. Instead, practices are being significantly adapted during implementation. Analysis of the evidence suggests three reasons for this adaptation process: low faculty language ability, a lack of understanding of pedagogical innovations, and a lack of faculty motivation.

7.6.1.1 Faculty language ability

First, the language ability of some faculty members appears to be preventing the effective implementation of some of the proposed innovations. Much like students in

Rwanda, most faculty members completed their studies in French. Many have also taught in French throughout their professional careers. Although the change in language of instruction did result in the retirement of some Francophone instructors, many opted to take advantage of government-sponsored English language instruction, hoping to improve their English ability and remain in their positions. Student and administrator participants did indicate that instructor levels of English are improving. Government bursaries are available for faculty members to pursue postgraduate studies outside of Rwanda, and those who are able often choose to complete their degrees in English-speaking countries. Others seek to improve their English within Rwanda itself, taking advantage of the increasing use of English as the medium of communication in public conferences, meetings and events. However, the fact remains that many instructors can only demonstrate what Blommaert has called a "locally good" level of English (quoted in Samuelson & Freedman, 2010). Although able to communicate with their students, instructors in this category cannot act as resources for improving student language levels, as their own English ability would be considered quite low according to international standards. Faculty and student participants alike acknowledge that there have been difficulties with the implementation of English as the language of instruction. Some faculty participants explained that they frequently choose to communicate with students in Kinyarwanda outside of the classroom, in order to aid with comprehension of class content. One participant from the NUR FOS even acknowledged that students in his department are allowed to respond to examination questions in a mix of English, French and Kinyarwanda, as instructors in the department feel they should be marking student understanding of content, rather than language ability. Such practices are unlikely to foster improvement in English for either instructors or students.

Low levels of faculty language ability are also likely to result in the avoidance of many of the educational practices found to encourage the development of critical thinking skills, as many of the most effective pedagogical techniques require active communication. Facilitation of a classroom discussion, for example, requires immediate comprehension and ease of oral communication on the part of the instructor. Instructors with low levels of confidence in their English ability may feel more comfortable preparing lecture slides and reading the content to their students, so as to ensure that all delivered subject matter is articulated 'correctly'. Similarly, instructors with low levels of English ability are unlikely to assign lengthy writing assignments or engage substantially with the content of a student's final project.

Although a significant barrier to implementation, language ability does not appear to be the most significant driver of *adaptation* of educational practice. Faculty members

with low levels of English ability seem to be more likely to avoid effective practices altogether, rather than adapting them to suit their pedagogical preferences.

7.6.1.2 Faculty comprehension of proposed innovations

The evidence suggests instead that adaptation is driven predominantly by a fundamental lack of understanding of the educational practices themselves. This appears to be the result of both a lack of exposure to such methods of pedagogy and a lack of understanding of the rationales behind the proposed innovations.

It is widely acknowledged that teachers tend to teach as they were taught. Simply requiring instructors to change their practice is never sufficient for effective implementation, as instructors must understand the purpose of a given innovation and the means through which they can effectively accomplish the desired outcomes (de Corte, 2000). Most faculty members in Rwanda were taught through traditional methods of instruction. Having never been personally exposed to more innovative techniques, traditional methods are likely to be synonymous with the concept of 'education' in the minds of most instructors. Data from the interviews support this assumption, as most faculty participants indicated that they view education as being a process of transmitting information and knowledge to students. Most participants emphasised student comprehension of course content as being their primary teaching objective. Assessments, for instance, were described as measures of student understanding that serve the single function of objectively judging student retention of the information presented in a given course. In other words, the interview data suggests that instructors in Rwanda tend to have a "knowledge transmission" orientation towards teaching (Kember & Gow, 1994). It is therefore unsurprising that pedagogical innovations are being adapted during implementation, as most of the effective educational practices outlined in the conceptual framework rely on a fundamentally different teaching orientation.

Baxter Magolda (1999) has argued that instructors are likely to adapt proposed innovations to their pre-existing understanding of education. This explanation supports the process that appears to be occurring in Rwandan classrooms. To take one example, the requirement that instructors incorporate classroom discussions into their teaching assumes that instructors can see the inherent value in such a pedagogical approach. Without exposure to effective classroom discussions, instructors with knowledge transmission orientations are likely to construct classroom discussions to support their pre-existing primary objective, which is the transmission of course content to students. Discussions are therefore likely to take the form of question-and-answer sessions in which students can clarify important concepts. This is exactly what is occurring in most of the

participating Faculties at KIST and NUR. In order to construct the type of classroom discussion that is likely to encourage critical thinking skills, instructors would need to modify their conceptualisation of what it means to be an instructor. Instead of the expert transmitting knowledge, the instructor would need to assume the role of facilitator, creating an environment supportive of student exploration of a range of different perspectives on a given topic or issue. Similarly, open-ended synthesis-type assessment questions are likely to be viewed as a means of evaluating student comprehension of course notes, unless faculty members view assessment as a formative tool in the learning environment (Entwistle & Entwistle, 1997).

The evidence suggests that faculty members will need to shift away from the traditional conception of teaching as knowledge transmission in order to effectively implement many of the educational practices found to encourage the development of critical thinking skills. However, such a fundamental shift requires an understanding of the rationale(s) behind proposed pedagogical innovations. This can only be accomplished through the provision of explicit instruction and support for instructors. It can be threatening for an individual to question the techniques used during his or her own education, as recognising the flaws in one's own academic experiences raises questions about one's own learning. It is therefore unrealistic to assume that instructors will be able to alter their pre-existing beliefs about education without a significant amount of guidance. Much as students need to be supported when facing disorienting concepts, instructors are likely to require support when challenging deeply rooted perceptions about their own profession (Baxter Magolda, 1999). Once instructors are able to question their pre-existing assumptions about teaching, however, there is increased potential for them to understand the theory and evidence underlying more innovative pedagogical techniques. Instructors are also likely to need guidance to help them execute new teaching methods. Many of the effective educational practices included in the conceptual framework require careful planning and facilitation. Group projects, for example, are most effective when the groups are carefully constructed (Arum & Roksa, 2011). Successful collaborative learning often requires thoughtful facilitation by the instructor, particularly when students are unfamiliar with collaborative techniques or when educational practice challenges traditional notions of authority (as discussed in Tikly & Great Britain Department for International Development, 2003). In addition to affecting teaching orientations, past academic experiences result in deeply engrained perceptions of education within students. In the Rwandan context, students generally expect to receive knowledge from their instructors (Mbabazi Bamwesiga, Dahlgren & Fejes, 2012). Given the pedagogical methods used in Rwandan secondary schools (discussed earlier in the

chapter), students entering university are unaccustomed to seeking out new perspectives or viewing their peers as reliable sources of information (Baxter Magolda, 1999). Pedagogical methods that challenge such engrained understandings of education can cause feelings of disorientation and resentment in students (Moon, 2008). Instructors need to be equipped with the tools to anticipate such negative reactions and to effectively support students through the more challenging aspects of the learning process.

Pedagogical training therefore has a critical role to play in the process of pedagogical change. However, faculty participants indicated that such training is unusual at Rwandan universities. Workshops and trainings are occasionally offered to staff members, but they tend to be stand-alone events, rather than a sustained source of support for instructors. Although the Kigali Institute of Education offers courses in pedagogy for university instructors, enrolment in such courses is entirely optional, and very few faculty members choose to take advantage of them. Given the evident lack of pedagogical training, it is unsurprising that instructors at KIST and NUR are adapting innovative educational practices to suit their pre-existing teaching orientations.

One final challenge in the Rwandan context is the likelihood that some faculty members may themselves have only limited critical thinking ability, given their educational backgrounds. Theorists have argued that instructors need to model critical thinking in their own behaviour, in order to encourage development of critical thinking skills in their students (Meyers, 1986). If the instructors themselves cannot model critical thinking skills, it is unlikely that they will be able to encourage such skills in others. Although this study did not attempt to assess the critical thinking ability of instructors, it was observed that some faculty participants struggled to define critical thinking or to link the individual skills included in the assessment with their teaching responsibilities. A few participants suggested that critical thinking skills could be taught to students during Induction Week or the one-credit Study Skills course required within each Faculty, indicating a lack of understanding of what critical thinking is or how it should be cultivated in the student population. However, such responses were not representative of all faculty participants in the sample. Many easily defined critical thinking and articulated why it is important for students to develop such skills during university. In general, the participants demonstrating the weakest understanding of critical thinking tended to be junior faculty members who had only ever studied in Rwanda. In view of the study results, it seems likely that instructors with similar backgrounds to their students would have similarly low levels of critical thinking ability. The majority of faculty participants in the sample, however, had spent at least some of their academic career outside of Rwanda.

There was little evidence to suggest that those instructors with more varied educational backgrounds would have difficulty demonstrating critical thinking skills.

Research from other contexts has demonstrated that educational practice cannot be effectively modified unless instructors comprehend both the rationale behind any proposed changes and the most effective methods for implementation (see Gouveia, 2010, re Portugal; Green & Sakamoto, 2001, re Singapore). The data from this study further confirms this finding, as it appears that a lack of understanding is limiting the effective implementation of innovative educational practices in the Rwandan context.

7.6.1.3 The issue of motivation

Lack of understanding does not appear to be the only barrier to implementation in Rwanda. Many faculty participants demonstrated a clear passion for teaching and a high level of motivation to ensure student success. For these instructors, the provision of pedagogical training would likely have a significant positive impact on their ability to adopt effective educational practices in their classrooms. However, for other participants in the sample, the adaptation of practices appears to be driven more by a lack of motivation than a lack of understanding. For this second group of instructors, it seems that there would need to be a significant change in their circumstances in order for them to contribute the effort required to effectively implement any labour-intensive pedagogical innovations.

All of the practices found to encourage the development of critical thinking skills require significantly more time and effort from instructors than traditional lecture-based courses (Moon, 2008; Tsui, 2002). Most require a substantial amount of preparation, as well as the active involvement of the instructor during implementation. The provision of feedback on written assignments and open-ended assessments also demands significantly more time than marking multiple-choice or short answer examinations. The effective implementation of such practices is therefore dependent on the motivation of the faculty. However, the evidence suggests that faculty morale is low at both KIST and NUR.

At both institutions, participants indicated that the primary source of faculty frustration is the low salaries provided by public institutions in Rwanda. Faculty participants, particularly in the field of engineering, frequently expressed resentment at the fact that their former students earn higher salaries in the private sector than they could ever hope to earn at a university. Although administrators are aware that salaries are a source of frustration, there appears to be little that can be done to improve the situation, as the government has indicated that it is likely to reduce, rather than increase, funding to public higher education in the future.

In addition to low salaries, incentive structures at the two institutions contribute to low levels of faculty morale. The incentive structure at KIST is similar to most international research universities, in that promotions and pay rises are generally determined by research productivity. However, there is almost no research culture on the KIST campus. Faculty members cannot access funding for research projects, and administrators have indicated that the quality of research generated by KIST faculty members tends to be quite low. As a result, very few KIST faculty members manage to get published in peer-reviewed journals. This, in turn, means that very few faculty members are eligible for promotion. Furthermore, KIST has banned its full-time faculty members from pursuing any employment outside of their responsibilities to the institution. Although there is evidence that some instructors have continued to seek supplementary work despite the policy, many have complied. The compliant instructors appear to feel penalised by the university's decision, as they view the policy as having taken away their individual agency and ability to improve their standard of living. In contrast, there is a more vibrant research culture at NUR. Funding is available for faculty research projects, and, as a result, the number of publications is rising. However, promotions and pay rises at NUR are determined either by research productivity or by qualification. As faculty members are eligible to apply for bursaries to support postgraduate studies, typically overseas, most lecturers choose to gain promotions through the acquisition of an additional qualification, rather than through research output. Instructors can see that increased qualifications carry long-term advantages, as they can be used in the search for alternate employment outside of the academy, while research output is less likely to affect future job prospects. Unlike at KIST, lecturers at NUR are allowed to seek external employment. Although this seems to have a positive impact on morale, the culture of pursuing supplementary income appears to have a detrimental effect on student learning. Many NUR lecturers have teaching or consultancy responsibilities at other institutions in Rwanda. As the vast majority of supplementary work is not located in Butare, instructors are required to spend a significant amount of time away from campus, further limiting their ability to actively engage with the student learning experience.

In addition to salary-based concerns, participants expressed frustrations with changing conditions on their campuses. Student enrolment is increasing dramatically every year, particularly at NUR. However, departments do not have the funding to hire more staff or to improve teaching facilities. As a result, participants described extremely challenging infrastructural constraints, such as having to teach over 200 students in a classroom with only 100 chairs. Against a backdrop of frustrating financial incentives, such challenges are likely to contribute to declining morale. Some participants also

explained that they feel overwhelmed and pressured by the constantly changing policy environment in Rwanda. Since 2008, the public universities have changed their language of instruction, their admissions practices, their academic calendar and their curricular structure. Faculty participants who have worked within the system for a long time feel that they have been asked to change every aspect of their practice without any compensation or support from the government or their institution. Such an environment is likely to further contribute to declining levels of morale.

Although many participants indicated a strong passion for teaching and a desire to assist their students, it is likely that declining levels of motivation are contributing to the adaptation of educational practice at the classroom level. When required to change their teaching practices, faculty members with low morale are likely to implement only the most basic components of a new innovation, while avoiding any element requiring significant amounts of time or effort. To take the example of collaborative learning, group projects like those described earlier in the chapter require minimal effort on the part of the instructor. By implementing such projects, an instructor can claim he is following university requirements, as he is technically "incorporating group work" into his course content, even though the adapted practice bears little resemblance to the stated policy and is unlikely to achieve the desired learning outcomes.

The adaptation of effective educational practices within the majority of Faculties in the study appears to be occurring as a result of the combined effects of a lack of faculty understanding of pedagogical innovations and a low level of faculty motivation to implement such practices. Low levels of English ability also seem to exacerbate this adaptation process. However, the evidence suggests that similar adaptation is not occurring within FAED. This chapter concludes with a discussion of why FAED does not seem to have struggled with similar barriers to implementation.

7.6.2 Adopting, not adapting

It is important to start by acknowledging that FAED has a number of inherent characteristics that sets it apart from other Faculties in the sample. First, the Faculty was only recently established[126]. As a new Faculty, FAED has been able to develop an entirely new curriculum, benefiting from the gradual year-by-year increase in students to pilot new courses and experiment with instructional techniques. It is possible that motivation levels have remained high – for both instructors and students – because of the novelty of being involved in the first-ever architecture and design programme in the country.

[126] FAED started accepting students in 2009.

Furthermore, as FAED subjects were not previously taught in Rwanda, the vast majority of faculty members were recruited from outside the country. The largely expatriate composition of the faculty carries two potential advantages. First, most faculty members are comfortable speaking in English, so there are limited linguistic barriers to implementing educational practices. As most faculty members do not speak Kinyarwanda, there is also no possibility of communication with students in any language aside from English, which is likely to contribute to improvements in student language ability. Second, the diversity of educational backgrounds within the Faculty is likely to contribute to more varied teaching orientations and, potentially, higher levels of faculty critical thinking ability. One faculty participant from FAED also postulated that most lecturers in the Faculty came to Rwanda with the explicit purpose of teaching at KIST, suggesting a high level of instructor motivation.

FAED is also one of the most highly resourced Faculties in the sample. As a new Faculty, FAED has benefited from a significant amount of start-up funding, which may or may not continue in the future. Studio courses cost more money to implement than lecture-based courses, and administrators acknowledged that it would be difficult to find sufficient funding to support all Faculties at a similar level. In addition, participants from other Faculties indicated that FAED has the advantage of a particularly low student-to-faculty ratio. Departments in FAED typically accept only 25-30 students per year, while other departments at KIST, such as Computer Engineering, accept up to 70 students per year. Although some departments are clearly burdened by large numbers of students, a detailed analysis of student-to-faculty ratios reveals that the ratio within FAED is not widely disproportionate for the KIST campus. Some of the science departments, for example, accept even fewer students than FAED each year (KIST Office of Planning & Development, 2012). The student-to-faculty ratio in itself, therefore, does not appear to be a sufficient explanation for the successful implementation of educational practice within FAED, although there is no question that many of the methods are easier to execute with smaller class sizes.

Although each of these characteristics may contribute to the success of the Faculty, the data point to three other aspects of the Faculty culture that are likely to be driving the effective implementation of educational practice within FAED: 1) the collegial nature of the working environment; 2) the Faculty's focus on the exploration, rather than the transmission, of knowledge; and 3) the existence of a culture of shared learning within the Faculty.

7.6.2.1 A shared understanding of the academic experience

Faculty participants from FAED described a highly collegial working environment within the Faculty. When the Faculty was first established, all of the programme curricula were devised collaboratively, incorporating contributions from all faculty members within a given department. By actively soliciting the opinions of other lecturers during the establishment of programme curricula, departmental leaders are likely to have cultivated a shared understanding of the objectives of each programme and a shared vision of how the individual elements of a given programme should work together towards the accomplishment of programme goals. Participants also described a culture of continuous improvement within the Faculty, in which departmental leaders frequently organise opportunities for lecturers to propose revisions, both to the overall programme curriculum and to individual course components. As one senior lecturer explained,

> Before we start teaching, every semester, we are required to submit the syllabus or the course outline one month before the commencement of the semester. … And then after … that, two weeks to the opening date, we have a staff meeting to go through what each one of us prepared. So, you share with the rest of the team on what you're going to teach in this module, why you think it's relevant, how it's filling in the main vision or the purpose or aims of the department. So, we evaluate that. We give you feedback. We discuss. And then we give you time to go and sort-of improve on that. And then, we believe then what you're going to teach is a little bit more comprehensive, as opposed to just one person's thoughts. (Senior Faculty Member, Department of Creative Design, FAED, KIST

Faculty members are also encouraged to share 'best practices', visiting one another's classrooms to observe different methods and provide constructive feedback. One example of the results of such exchange is the introduction of the memo technique referenced earlier in the chapter:

> Mostly, I think the faculty seems to try to share best practices and learn. If I must tell you, the memo is not even my issue. I never learned through memos. But, I picked it from, I think, another faculty, and I thought it was going to work. And I think, so far, it's working. (Junior Faculty Member, Department of Architecture, FAED, KIST)

Tsui (2001) has argued that departmental collegiality can have a substantial positive impact on the kinds of faculty attitudes found to support student critical thinking ability. In addition to learning from one another, instructors benefit from mutual support within a collaborative working environment. This can be a particularly important resource in the face of student resistance to challenging instructional techniques. A collegial working environment can also act in the place of explicit pedagogical training by collectively bolstering faculty understanding of non-traditional teaching methods. The working environment within FAED appears to serve both of these functions, thereby

helping to avoid some of the barriers to implementation evident in other Faculties in the sample.

Leadership has a vital role to play in the cultivation of a truly collegial working environment. FAED appears to benefit from such strong leadership. Faculty participants from FAED expressed a high level of respect for leaders within the Faculty, indicating that those in leadership positions are both outstanding in their professional practice and effective in their ability to generate consensus. Junior lecturers from other Faculties did not tend to mention their Faculty or departmental leaders during interviews. When leaders themselves were interviewed in other Faculties, many indicated a lack of trust in their colleagues, suggesting that the role of a leader is primarily to inform his or her subordinates of decisions and deadlines and to ensure compliance with regulations. In contrast, leaders within FAED tended to describe their colleagues as important resources with a vital role to play in improving academic experiences within the Faculty. Likely as a result of these apparent differences in the perception of colleague competence, none of the other departments appear to work collaboratively to revise their programmes or improve their courses. Instead, instructors in other Faculties seem to work independently from one another, focusing exclusively on their own course content and instructional practice.

7.6.2.2 Epistemological orientation of the Faculty

The programmes within FAED also appear to have been constructed with the explicit intention of encouraging students to explore knowledge within their field. Such a focus on the 'unknown' stands in stark opposition to the traditional knowledge transmission approach to education observed in other Faculties.

As discussed earlier in the chapter, instructors from other Faculties in the sample demonstrated the characteristics of a knowledge transmission approach, as they tended to emphasise the importance of covering course content and testing student understanding of core information within a given subject area. In contrast, instructors from FAED emphasised fostering independent thought and creativity by encouraging students to think in new ways and propose new "design solutions". Although design-based subjects are arguably more creative than some other academic disciplines, the process of inquiry, in which students question established knowledge and challenge their own understanding of central topics, is fundamental to all subject areas taught in the participating Faculties. The difference is therefore not one of academic subject but one of departmental orientation.

As discussed in Chapter 3, a focus on exploration of the unknown is likely to support the development of critical thinking by challenging students to test their

epistemological understanding and practice their metacognitive skills. At the same time, such an orientation is likely to further bolster instructor understanding of the rationale behind many of the pedagogical techniques found to increase student engagement and contribute positively to cognitive development.

7.6.2.3 Creating a culture of shared learning

The evidence also suggests that the FAED culture encourages instructors and students to work together as partners in the learning process. Research in other contexts has underscored the importance of the learning environment for increasing student engagement and contributing to cognitive development. The culture of shared learning evident within FAED is likely to offer similar benefits to both the student and the faculty population.

A culture of shared learning must first be based on a foundation of mutual respect between instructors and students. The required studio courses in FAED seem to cultivate such an environment, as instructors appear to consistently emphasise the importance of student beliefs and perspectives. During the oral critique process, both instructors and peers give opinions on presented designs, a process which validates student perspectives and positions the instructor as an expert resource, rather than an unapproachable authority figure. Participants also indicated that critiques are often given by more than one instructor, a practice which allows students to see that 'experts' do not always share the same opinion and reduces the tendency for students to look for one 'correct' answer or perspective. Within the Department of Creative Design, instructors run a weekly "Open Studio", in which students are invited to share their work with any instructor in the department. Such practice is likely to encourage students to interact with all members of the departmental community, while also allowing for engagement with multiple perspectives and design preferences. This stands in marked contrast to the lecture-based courses offered in most other Faculties in the sample. Although modules at NUR are often taught by teams of instructors, participants indicated that instructors tend to teach their course content independently of one another, rather than teaching together and potentially modelling differences in the perspectives of authorities in their field.

The FAED community also seems to extend beyond the traditional classroom boundary. FAED sponsors a number of annual student competitions, in which students are rewarded for exemplary work. Such structured incentives are likely to foster deeper approaches to learning, as students can see a tangible benefit to in-depth understanding of course objectives. FAED also holds special events for students and faculty members, such as the recent opening of a gallery space in which student work is exhibited for members of

the public. Such events are likely to make students feel like included and respected members of the community. Student competitions and events do not seem to feature in any of the other participating Faculties.

Faculty participants from FAED universally indicated a belief in the ability of their students to exceed expectations and demonstrate high levels of proficiency and independent thought. Although some participants from other Faculties indicated a similar belief in their students' abilities, many described their students as generally "lazy" or lacking in "seriousness". Such attitudes imply a lack of belief in student potential, which is likely to be felt by the students themselves. This, in turn, is likely to contribute to low student confidence and a lack of connection between students and instructors in other Faculties.

In addition to potentially increasing student engagement, it is likely that the collegial community environment within FAED has a positive impact on faculty motivation. The root causes of low morale identified in other participating Faculties also apply to FAED, as all instructors at KIST receive comparatively low salaries and work within the same incentive structure. The largely expatriate composition of FAED does have an effect on incentives, as expatriates are typically paid more than local staff in Rwanda. However, expatriates have very little job security, as expatriate contracts are usually short-term[127]. Although faculty participants from FAED tended not to discuss salary levels during their interviews, one did express dissatisfaction with the length of his contract and the lack of reward for exemplary teaching. Despite such frustrations, there was little evidence that the incentive structure is significantly diminishing faculty motivation to implement effective educational practices within FAED. This may be because many FAED lecturers do not rely on their university salaries as their sole source of income[128]. However, this cannot be the sole explanation, as the pursuit of supplementary contracts does not appear to result in increased dedication to teaching in other participating Faculties. FAED instructors appear to be highly dedicated to the work of the Faculty, as most seem to be willing to contribute a significant amount of extra time and effort to their Faculty responsibilities. It is likely that such dedication is reinforced by the community environment evident within the Faculty.

7.6.3 The role of departmental culture

The findings described in this section are presented visually in Figures 7.1 and 7.2.

---

[127] In contrast, local staff contracts tend to be open-ended.

[128] Participants indicated that many faculty members in FAED continue to work as professionals in their field, in addition to their responsibilities at KIST.

Figure 7.1: Cross-Case Causal Model (Most Faculties)

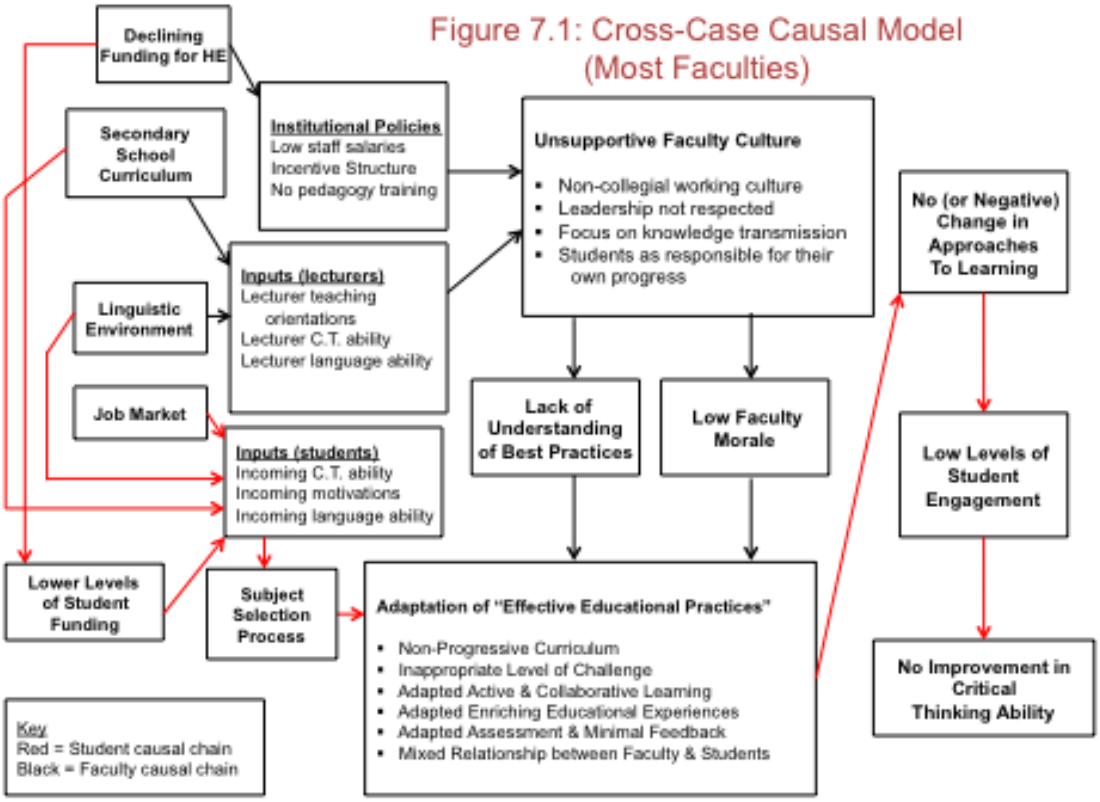
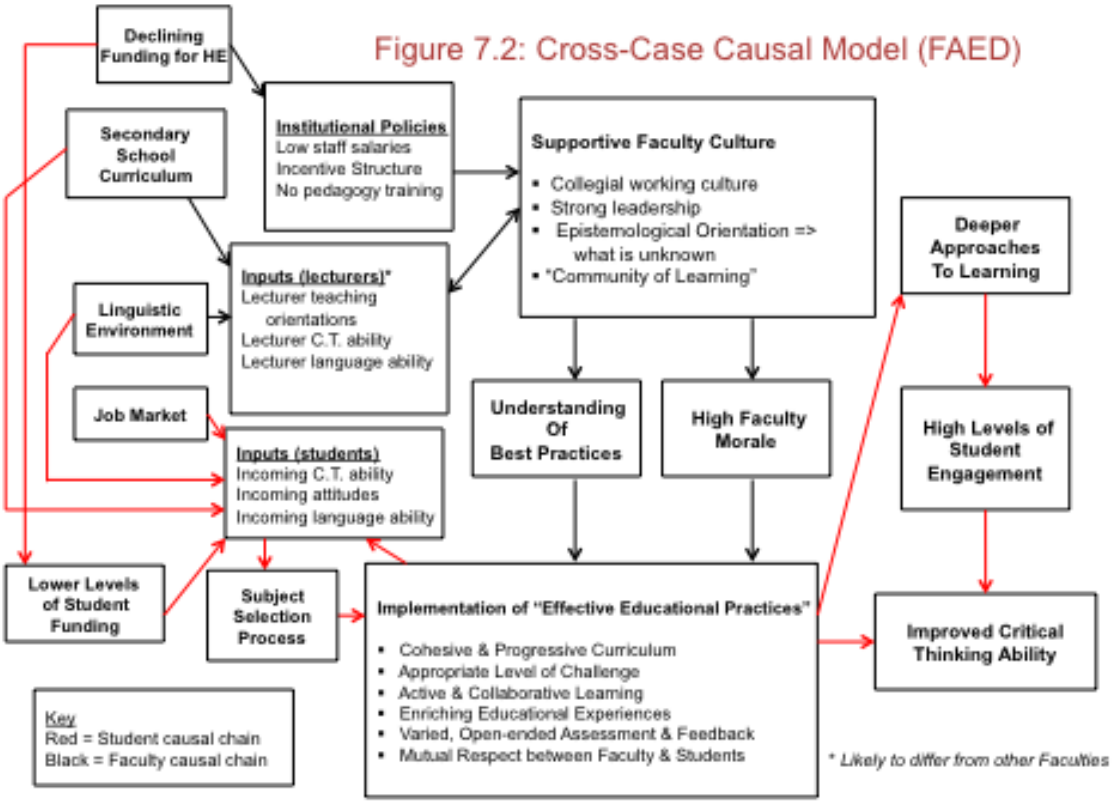


Figure 7.2: Cross-Case Causal Model (FAED)

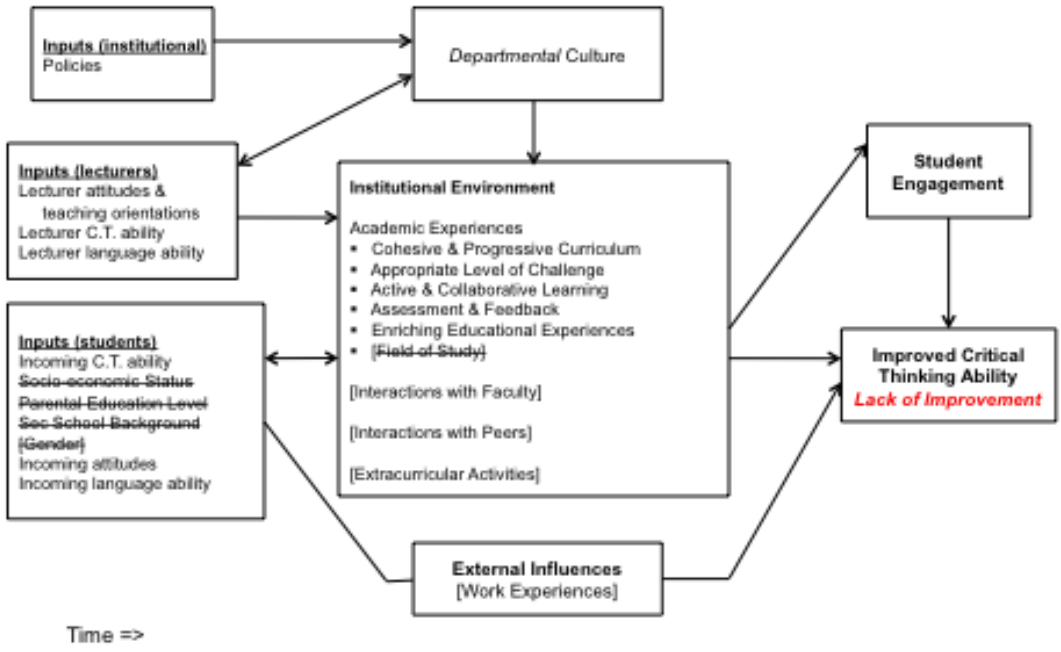


The diagrams highlight the vital role that departmental culture seems to play in the cultivation of critical thinking skills in Rwanda. Despite similarities in governing policies and incoming student characteristics, there appears to be a significant difference between FAED and other participating Faculties in terms of both student engagement and demonstrated student critical thinking ability. Tsui's (2000) study of American institutions yielded strikingly similar results, although her analysis concluded that *institutional* culture plays a significant role in the cultivation of critical thinking ability. Data from this study indicates that cultures within institutional sub-units can differ substantially, implying that it is *departmental*, rather than institutional, culture that seems to make the greatest difference in the Rwandan context.

7.7 Summary of Phase Three Results

The findings from the case study phase provide rich insights into the institutional context behind the assessment results. Evidence from this phase suggests that academic experiences are the most significant factor affecting the development of critical thinking skills at KIST and NUR. When implemented effectively, many of the innovative educational practices outlined in the conceptual framework seem to have the potential to yield significant gains in both student engagement and critical thinking ability. However, the results indicate that the implementation of such practices cannot be assumed, even if required by centralised academic policies. Without faculty understanding of such practices or motivation to adopt them, required innovations are more likely to be adapted at the classroom level, thereby reducing their potential impact on student learning. The results from FAED suggest that successful implementation of such practices is possible in the Rwandan context. However, it appears to be unlikely without a supportive departmental culture, which simultaneously supports faculty development and includes students in a mutually respectful community of learning. Building on the diagram presented at the end of Chapter 6, Figure 7.3 depicts a further modified version of the study's conceptual framework, incorporating the results of this phase.

Figure 7.3: Modified Conceptual Framework



Chapter 8: Conclusions and Implications

This study aimed to assess one aspect of the academic quality of Rwanda's public universities by evaluating the extent to which students seem to be improving in their critical thinking skills during their undergraduate careers. Investigating the individual and institutional factors that appear to affect the improvement of critical thinking skills in the Rwandan context, the study considered how universities in Rwanda are helping or hindering the development of such abilities in their student populations. In this final concluding chapter, the main findings of the study will be discussed in terms of the implications for Rwanda and for the regional reform agenda. Possible areas of future research will also be explored.

8.1 Study Conclusions

Findings from the study contribute to the existing body of knowledge in two important ways. First, the results provide insight into the current critical thinking ability of Rwandan undergraduates. The assessment results indicate that students attending NUR, KIST and SFB are generally quite poor at demonstrating critical thinking skills. Although there was some variation in individual skill proficiency, students in the sample generally demonstrated below average ability on all of the critical thinking skills assessed in the study. Assessment results also suggest that students at the three institutions may not be improving in their critical thinking ability during university. Although there was some evidence of improvement in the ability to evaluate and use information when making decisions, fourth-year students did not appear to be more proficient than their first-year counterparts at demonstrating any other critical thinking skills. Although not strictly generalisable to the overall student population in Rwanda, given that the sampling strategy relied on a purposive selection of participating institutions, the results are likely to be indicative of national trends, as the three universities included in the study are generally considered to be the most prestigious institutions in the country. If anything, one would assume that the students involved in the study would be *more* proficient at demonstrating critical thinking skills than their counterparts at other institutions. The results therefore imply that limited critical thinking ability may be a general issue affecting all of Rwanda's tertiary institutions. As this study was the first empirical investigation of critical thinking at the university level in Rwanda, the results are a substantial contribution to the evidence base informing Rwanda's higher education policy.

Second, findings from the study offer an explanation for the apparent lack of improvement in critical thinking at Rwanda's public universities. Unlike in other contexts, background student characteristics seem to have very little impact on the demonstration of critical thinking skills in Rwanda. The evidence suggests that this may be the result of an examination-focused secondary school curriculum, in which memorisation and rote learning are emphasised over the cultivation of critical thinking skills. The negative implication of this finding is that students in Rwanda seem to be enrolling at university with universally low levels of incoming critical thinking ability. However, the positive implication is that students from all backgrounds appear to have an equal likelihood of improving their critical thinking ability during university.

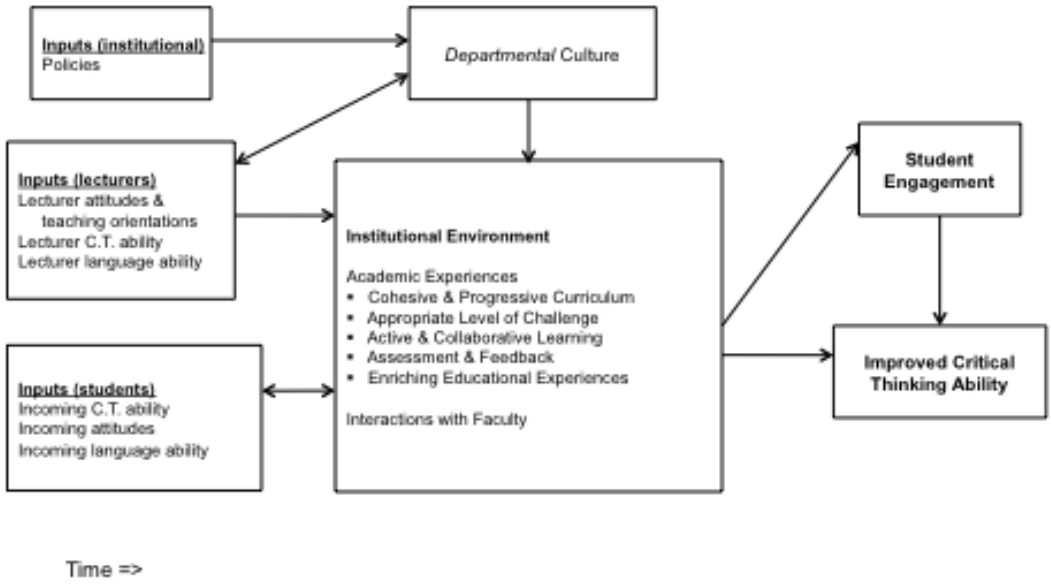
Research from other contexts has indicated a number of specific educational practices that seem to have a positive influence on critical thinking. Many of these practices are already enshrined in the teaching and learning policies governing Rwanda's public universities. However, the results of the study suggest it is likely that students are not improving in their critical thinking ability because these practices are not being implemented effectively within individual Faculties at KIST and NUR. Instead, it appears that pedagogical innovations are being substantially adapted at the classroom level, as a result of limited faculty understanding of such practices and low levels of faculty motivation. This adaptation process appears to be undermining the potential of such practices to impact student critical thinking ability. Results from the study suggest that such adaptation will continue to be a problem unless the underlying issues of faculty understanding and motivation are addressed.

However, the results also indicate that adaptation does not always occur and that, if implemented effectively, innovative educational practices can have a positive influence on critical thinking ability in the Rwandan context. By constructing a collegial and collaborative working environment for both instructors and students, the leadership within FAED seems to have increased both instructor understanding of the Faculty's educational philosophy and instructor motivation to implement challenging and time-consuming teaching methods. As a result, practices are being successfully adopted, rather than adapted, by FAED instructors, resulting in increased student engagement and apparent improvement in student critical thinking ability. The study results suggest, therefore, that departmental culture plays a vital role in the development of student critical thinking skills.

These findings expand existing understanding of how critical thinking can be improved at the university level by identifying the factors that seem to be helping – and hindering – the development of critical thinking ability in Rwanda. As this is the first

study of its kind in an African context, the results also make a substantial contribution to existing evidence of academic quality at tertiary institutions in the region. Figure 8.1 presents a visual representation of the study's central findings. A modification of the original conceptual framework[129], Figure 8.1 highlights the inputs and elements of the institutional environment that seem to have an influence on the development of critical thinking ability at Rwandan universities.

Figure 8.1: Improving Critical Thinking in Rwanda



8.2 Implications for Rwanda

The study findings have vital implications for higher education reform within Rwanda, as the apparent lack of critical thinking skills amongst university graduates raises significant concerns for Rwanda's national development strategy. As discussed in Chapter 2, the assumption underlying *Vision 2020* is that university graduates will fuel the creation of a service-based knowledge economy in Rwanda, thereby attracting foreign investment, increasing domestic capital and allowing the government to reduce its reliance on aid and

[129] In addition to the modifications documented in Chapters 6 and 7, Figure 8.1 is missing three of the variables included in the original conceptual framework. Two were removed because of likely irrelevance in the Rwandan context. As students generally enter university with low levels of critical thinking ability, it seemed unlikely that Interaction with Peers or Involvement in Extracurricular Activities would have much of an impact on critical thinking skills. Work experiences were also removed, as the potential influence of work experiences on critical thinking could not be addressed in the study, as there were an insufficient number of participants with professional experience to allow for any valid analysis.

foreign technical assistance. One crucial component of this vision is the assumption that university graduates will be able to propose new solutions to entrenched problems in Rwandan society. This vision relies fundamentally on the cultivation of critical thinking skills, as graduates need to be able to interpret and use existing evidence in order to make decisions or suggest solutions to problems. The lack of critical thinking ability observed within Rwanda's most prestigious, most resourced public universities therefore presents a serious challenge for the national strategy.

However, findings from the study also suggest a number of ways that Rwanda may be able to address this challenge. First, the findings support results from other studies that have underscored the importance of cultivating critical thinking skills at lower levels of education. Singapore, the model for Rwanda's development plan, has struggled with a similar lack of critical thinking ability in its university graduates. In the Singaporean context, this has largely been ascribed to the culture of "rote learning and close supervision" that permeates all levels of the country's education system (Green & Sakamoto, 2001, p. 220). The same culture appears to exist within the lower levels of education in Rwanda. In recent years, there has been an increased understanding of the implications of limited critical thinking ability for Singapore's economic development, as stakeholders have started to understand the importance of critical thinking for innovation and technological adaptation. In response, Singapore has proposed a number of reforms, such as the introduction of problem-based learning into its secondary school curriculum and the diversification of university admissions requirements (ibid.). Similar reforms may have a role to play in improving critical thinking ability within the Rwandan context.

Furthermore, as depicted in Figure 8.1, the study suggests that Rwanda's universities should be able to support the development of critical thinking ability in their student populations by implementing a number of pedagogical innovations, such as the establishment of progressive and cohesive curricula, the construction of an appropriate balance between academic challenge and student support, the use of varied and open-ended assessment formats, and the incorporation of meaningful opportunities for active and collaborative learning. However, the study findings also underscore the crucial importance of instructor understanding and motivation. Many of these practices are already included in the policies governing Rwanda's public institutions, but there is a substantial disconnect between the policies and actual classroom practice. Although there is often an assumption that policies will be adopted as they were intended, policies tend to be adapted by relevant stakeholders if there are significant barriers to their implementation (Brock-Utne, 2000; Leach, 1994). The results of this study demonstrate the consequences of this process of adaptation. Despite the existence of policies

promoting pedagogical innovation at Rwanda's public universities, results from this study indicate that most instructors continue to teach as they have always taught, thereby undermining the potential impact of such policies on student learning outcomes.

At the same time, the study offers a positive example of how a Faculty in Rwanda has managed to effectively implement many of the pedagogical innovations found to help students develop stronger critical thinking skills. FAED is clearly unique in Rwanda, given its novelty and its high proportion of expatriate staff. However, it appears that the FAED culture has also had a significant impact on educational practice within the Faculty. Many aspects of the FAED culture could be replicated elsewhere in Rwanda if departmental cultural change were to be viewed as a priority by those in positions of leadership, both within individual Faculties and at the central institutional level.

At the Faculty level, the FAED experience suggests that the creation of a collegial and collaborative environment can improve lecturer understanding of the rationale behind curricular and pedagogical choices, while also allowing instructors to benefit from the exchange of best practices. The creation of a culture of shared learning, which values student contributions as well as those made by faculty members, seems to have a positive impact on faculty motivation, even in the face of resource constraints. Members of the FAED community also appear to benefit from a broader conceptualisation of university education than can be found in many other university Faculties. Most students and instructors in Rwanda seem to assume that the purpose of a university education is 'train' students to perform the particular technical skills required in their chosen profession. This assumption reflects a traditional 'manpower planning' vision of university education, as it suggests that universities can anticipate what precise skills a graduate will need in a given career. However, rapid changes in the global economy make such a vision untenable. The objective of a university can no longer be the transmission of known information from instructors to students. The sheer amount of available information makes it impossible to achieve such a goal, while the rapidly changing nature of most industries diminishes the relevance or effectiveness of such an educational model. FAED appears to have changed this paradigm by shifting the emphasis of the curriculum from the 'known' to the 'unknown'. Other Faculties in Rwanda would likely benefit from a similar epistemological shift.

At the central level, institutions could assist in the process of cultural change by supporting pedagogy as a critical component of institutional quality. The provision of regular pedagogical training would seem to be a minimal first step, particularly given that many lecturers in Rwanda have not had any personal experience with the kinds of pedagogical innovations found to have a positive impact on critical thinking. Rather than

providing transportable qualifications for such training, it could be helpful to link pedagogical training to faculty incentives, in order to prevent the 'brain drain' of trained instructors to other, higher-paying industries. A change to the existing incentive structure could also make a substantial difference, given that excellence in pedagogy is not currently rewarded within the system. Although institutions often take student evaluations into consideration when considering staff incentives, such tools are not always an accurate reflection of teaching quality (as discussed in Everett & Zinser, 1998). The introduction of methods for more accurately assessing and rewarding excellence in pedagogy could motivate many instructors to improve the quality of their teaching.

Changing institutional culture is never easy, as stakeholders typically have a vested interest in certain aspects of the status quo (Leach, 1994). However, the study results indicate that cultural change may be necessary in order for Rwanda's institutions to foster the development of critical thinking skills in their students.

8.3 Implications for the Regional Reform Agenda

The study results also have wider implications beyond the Rwandan context. As discussed in the introductory chapter, academic quality features very little in the debate around how to revitalise Africa's universities, with current reforms tending to focus primarily on improvements in institutional efficiency and research capacity. The results of this study suggest that there is a profound cost to ignoring the importance of student learning outcomes within the reform agenda. Rwanda's public universities are some of the most resourced institutions in the region. They are populated almost entirely with high-performing scholarship recipients, and they have benefited from careful and extensive planning in the years since the genocide. Nonetheless, despite this encouraging environment, Rwandan students appear to be graduating from these institutions with limited ability to evaluate the quality of information or use evidence when making decisions. As most other institutions in the region have not benefited from similar levels of support, it is likely that they are experiencing similar trends in their student populations.

The implications are significant for the reform agenda. The literature linking higher education and development largely assumes that university graduates have the ability to think critically about problems. The study results raise doubts as to the viability of this assumption in the African context. The study findings therefore demonstrate the necessity of moving academic quality to the centre of the revitalisation debate. Although problems of academic quality at African universities have been acknowledged for decades,

there continues to be very little support for initiatives aimed at improving pedagogy at African institutions. Instead, reforms have focused on improving institutional and system efficiency, despite a growing body of evidence that some efficiency measures may actually be contributing to further declines in academic quality at institutions across the region (Cloete *et al*, 2011; Mamdani, 2007).

The emphasis on alternative modes of delivery is particularly problematic in light of the study results. Development organisations have increasingly argued that the problem of access at Africa's universities can be solved through the provision of online education. In a recent publication by the Partnership for Higher Education in Africa, improved internet access was lauded for its ability to "[empower] students to drive their own learning, giving them the freedom to follow their curiosity and gather new material from varied online sources of their own choosing, no longer constrained by the linearity of a textbook" (Lindow, 2011, p. 43). Implicit in this perspective is an assumption that African students have the capacity to "drive their own learning" and interact productively with "varied online sources". However, the results of the study suggest that such assumptions may not hold. In fact, if students in the region are entering university with similarly low levels of critical thinking ability to participants in this study, the expansion of online education as a solution to low levels of access may actually serve to further disadvantage African students by limiting their exposure to the kinds of academic experiences that could positively influence their cognitive development.

The recent focus on improving research capacity at universities across the region might help to address some underlying issues, as research experience could help instructors to develop their own critical thinking skills. However, as there is little proof that research productivity is positively correlated with teaching quality (Feldman, 1987; Marsh & Hattie, 2002), an exclusive focus on research outputs is unlikely to have much of an effect on student learning outcomes in the region. In fact, it is possible that the increased emphasis on research productivity may further reduce the amount of time that faculty members dedicate to teaching, particularly given the tendency for faculty members to hold supplementary teaching and consulting positions outside of their regular posts.

It is not surprising that the current reform agenda has largely avoided the subject of student learning outcomes. International development organisations are under increasing pressure to document the impact of their interventions. As a result, they have increasingly prioritised interventions that lead to easily measurable outcomes. Although there is a general acknowledgement of the importance of academic quality, many within the development community imply that the difficulty of measuring student learning outcomes makes it all but impossible to successfully demonstrate impact.

However, this argument reflects a lack of understanding of the potential benefits of collaboration between the international development community and the field of educational research. Educational researchers in other contexts have generated numerous methodologies that could be used to help African institutions analyse learning outcomes. This study demonstrates the feasibility of adapting an existing methodology for use in an African context, as it represents a first attempt to empirically investigate critical thinking ability at African universities using an adapted research instrument. The study results suggest that the methodology was both feasible and appropriate for use in Rwanda. This is likely because careful attention was paid to the cultural context throughout the study's implementation. The central research instrument was explicitly adapted for use in Rwanda, and the local members of the research team made significant contributions throughout the entirety of the study. This experience suggests that methodologies developed in other contexts can be feasibly be used to evaluate complex issues within African institutions, provided appropriate research instruments are developed and local perspectives are included. In the long term, sustainable reform will require the development of local instruments and methods. However, this study suggests that, in the meantime, the adaptation of existing methodologies could play an important role in improving academic quality across the region by providing contextually relevant empirical data on student learning.

The study results also suggest a role for the international development community in addressing the issue of academic quality at African universities. Interviews with administrators indicated that the leaders of public institutions often recognise the significance of academic quality but are frustrated in their ability to affect change, due to their reliance on unpredictable levels of government support. The provision of funding to support improvements in pedagogy would significantly increase the ability of such institutions to address academic quality concerns. For instance, funding could be used to support ongoing pedagogical training for faculty members or to establish new incentives to reward excellence in teaching. The provision of funding could also help to address some of the underlying barriers to the effective implementation of pedagogical innovations. Stand-alone trainings and workshops are highly unlikely to affect systemic pedagogical change. However, they are often the most feasible option for resource-constrained institutions. The provision of funding by the international community for ongoing initiatives focused on improving pedagogy could go a long way toward improving academic quality at institutions across the region.

8.4 Implications for Future Research

Although the results of the study make a substantial contribution to the existing literature, it must be acknowledged that there were limitations related to the study's design and implementation. This section provides an overview of these limitations and considers potential areas of future research in this domain.

8.4.1 Limitations of the study

There are limitations inherent in any research design. In this study, the most significant design-related limitation was the potential influence of confounding variables on the study results, given the use of a cross-sectional design. As discussed in Chapter 6, first-year students in the sample generally took longer to complete their assessments than fourth-year students. If this were the result of systematically higher levels of motivation within the first-year population, the results of the cross-sectional comparison would have been negatively affected. However, the data suggest that the potential of such an effect is unlikely to have resulted in any erroneous conclusions regarding student improvement in critical thinking. As the range of scores within the fourth-year population indicates that students in Rwanda are leaving university with generally weak critical thinking skills, it is reasonable to conclude that students are not substantially improving in their critical thinking ability during university.

There were also limitations related to the reliance on one primary researcher in the implementation of a multiple embedded case design. Although the research team assisted with data collection, reduction and verification of findings, analysis was completed entirely by the lead researcher. During the case study phase, pre-existing knowledge of the findings from the first institution may have subconsciously affected the prioritisation of themes at the second institution. In order to mitigate the potential effect of this limitation on the study results, verification methods were employed at each stage of the case study phase. The use of these methods appears to have successfully prevented any negative influence on the validity of the study results, as the research assistants involved in the validation meetings universally indicated that the findings resonated with their experiences at KIST and NUR.

A number of unanticipated events during implementation of the study must also be viewed as limitations. The original intention was to complete both Phase Two and Phase Three of the study within the same academic year. However, due to sudden changes to the academic calendar, the case study phase was implemented a few months after the end of the year. As a result, the majority of fourth-year students in the sample had graduated before qualitative data collection began. As discussed in Chapter 7, this meant

that it was more difficult to recruit a representative sample of participants for the focus groups, as the research team was restricted to identifying participants who were able to attend interviews in Kigali. It is possible that this biased the sample towards those from urban backgrounds who had remained close to campus following graduation. The change presented a particular problem for data collection at NUR, as it resulted in fewer focus group participants in comparison to KIST. The perceived validity of the case study findings suggests that this limitation did not have a substantial effect on the study results. However, it would clearly have benefited the analysis to incorporate the perspectives of additional NUR students. Unanticipated changes in the fourth-year internship schedule at SFB also impacted recruitment during the assessment phase. Although limited participation at SFB did not prohibit institutional-level analysis, low participant numbers did prevent the inclusion of SFB in the case study phase. The original intention was to use the case study phase to analyse academic experiences at the three institutions included in the sample. The impossibility of including SFB in the qualitative phase was one of the primary motivations for altering this initial intention by focusing on Faculties, instead of institutions, as the unit of analysis. The emphasis on Faculties may have increased the validity of the case study results, as it was possible to examine student experiences in more detail by concentrating on institutional sub-units. However, the mid-study change in the intended unit of analysis meant that insufficient numbers of students had been recruited within each Faculty to allow for any Faculty-level comparison of assessment results. If the final design had been anticipated from the outset, a different sampling methodology could have been used, which would have allowed for further integration between the quantitative and qualitative results of the study.

Finally, the assessment tool itself presented some limitations. As discussed in Chapter 6, the two versions of the assessment were not strictly parallel. In particular, follow-up analysis suggested that there were systematic differences in the scores on Skill A: Bias between the two versions. As discussed in Chapter 7, this is likely to have occurred because one task required participants to detect personal bias, while the other required the detection of financial bias. There is no evidence that this limitation had any substantial impact on the study results, but it would have improved the comparability of the tasks if such an effect had been avoided. It is also clear that the use of an English/French assessment may have limited the range of critical thinking ability demonstrated by students in the sample. As discussed in Chapter 5, the lead researcher's lack of fluency in Kinyarwanda prohibited the use of an assessment tool written in the participants' native language. The results of prior research on the L2 language deficit (discussed in Chapter 3) suggest that participants in the study may have been able to

demonstrate higher critical thinking ability if the assessment had been written in Kinyarwanda. Although the language of *response* does not seem to have affected scores on the assessment, it is not possible to verify if participants might have demonstrated higher ability on an assessment written in Kinyarwanda, given that all of the participants were restricted to reading assessment materials written in either English or French.

8.4.2 Suggestions for future research

Given the implications of the study results discussed earlier in the chapter, it is clear that further research into critical thinking at the university level would be beneficial for both the Rwandan higher education sector and the higher education reform agenda in the region. The results and limitations of the current study indicate that there is substantial scope for further research in this domain.

Within the Rwandan context, it would be possible to expand upon the results of the current study in a number of ways. First, it would be beneficial to replicate the study methodology, using a similar assessment tool written in Kinyarwanda. Results of such a study would provide important insights into the potential impact of language on demonstrated critical thinking ability in Rwanda. It would also be useful to develop alternative critical thinking assessment formats that could be used in the Rwandan context. The use of multiple measurement methods would help to replicate and potentially validate the study results.

It would also be useful to conduct a similar study with a larger research team, in order to expand the sample to other tertiary institutions in Rwanda. It would be particularly interesting to consider potential differences in critical thinking ability between students attending public and private institutions, given the rapid expansion of the private sector. With a larger research team, it would also be feasible to recruit a much larger sample within institutions, which would allow for analysis at the Faculty, or even departmental, level. Such an approach could provide additional evidence of the impact of particular academic experiences on critical thinking ability in the Rwandan context.

The implementation of additional studies relying on the use of different research designs would offer further insight into the question of how critical thinking can be improved at the university level in Rwanda. A longitudinal design would be particularly useful for clarifying the influence of various institutional factors on the improvement of critical thinking ability. Although the current study suggests that effective educational practices can have a positive influence on critical thinking skills, the current design did not allow for any investigation of *how* particular practices influence the development of critical thinking. It would therefore be useful to implement a series of evaluation studies

to investigate the impact of particular innovations on the demonstrated critical thinking ability of Rwandan students.

Expanding beyond the Rwandan context, the results suggest that the study methodology could be successfully implemented in other countries. The adaptation of the methodology for use in other African contexts could provide insight into the critical thinking ability of undergraduate students elsewhere in the region and expand understanding of how the region's tertiary institutions may be helping or hindering the development of critical thinking in their student populations.

The results of this study make a significant contribution to the available evidence base guiding higher education reform in sub-Saharan Africa by providing much-needed empirical data about a crucial student learning outcome – the acquisition of critical thinking skills. The findings also underscore the important role that educational research can play in the revitalisation of higher education across the region. By highlighting the educational practices that appear to help – and to hinder – the cultivation of critical thinking skills at the university level, the study offers important insight into the kinds of innovations that could affect lasting change at Africa's tertiary institutions. The implementation of such innovations may not be easy, but the potential benefits – both for individual university students and for wider development objectives across the region – suggest that it would be well worth the effort.

Appendices

A. Assessment Adaptation Strategy
B. Field-Testing and Piloting Protocol
C. Complete English-Language Versions of Assessments
   C1: Assessment Task 1
   C2: Assessment Task 2
D. Scoring Rubrics and Flowcharts
   D1: Basic Scoring Instructions
   D2: Scoring Cover Sheet
   D3: Scoring Flow Chart (Task 1)
   D4: Scoring Rubric (Task 1)
   D5: Scoring Flow Chart (Task 2)
   D6: Scoring Rubric (Task 2)
E. English-Language Version of Demographic Questionnaire
F. Recruitment Materials
   F1: Sample Invitation Letter
   F2: Participant Information Sheet (Quantitative Phase)
   F3: Participant Information Sheet (Qualitative Phase)
G. Informed Consent Form
H. Inventory of Decisions Made During Data Entry
I. Evaluation of Assumptions and Additional Quantitative Results
J. Assessment Results for KIST and NUR
K. Case Study Protocol
L. Focus Group Materials
   L1: Focus Group Guide
   L2: Focus Group Handouts
   L3: Focus Group Tally Sheet
   L4: Questions from National Survey of Student Engagement (2012)
M. Interview Guides
   M1: Student Interview Guide
   M2: Faculty Interview Guide
   M3: Administrator Interview Guide

Appendix A: Assessment Adaptation Strategy

Based on Schmeiser and Welch (2006) and Chun (2008)

Test Development

1) Affirm the test's underlying philosophy and central purpose
2) Define the target population
3) Outline any administrative constraints on the assessment
4) Specify the content-related evidence of validity for the knowledge and skills to be measured by the test
5) Draft the test specifications, including the type and order of test questions, the test length, the proposed scoring method and any particularities of test administration
6) Organise an expert review of the test specifications and make any necessary modifications
7) Create the performance task(s)
    a. Identify a central issue for each task that has a quantitatively measurable key outcome variable and two possible – and equally reasonable – options for achieving the outcome
    b. Determine a task scenario, including an authentic situation and a role for the test taker to assume
    c. Create documents that will serve as evidence for the test taker, incorporating multiple document formats, multiple types of evidence and some conflicting data
    d. Write specific questions for the test taker to answer and draft an initial scoring rubric for assessing answers
8) Revise and modify the scenario, questions and documents based on initial feedback
9) Complete a fairness review of the draft assessment

Field-Testing and Piloting

10) Field-test the draft assessment through the use of think aloud sessions with the target population, ensuring that "language use in the directions, rubrics, and items themselves…are appropriate for…the cultural…population for whom the instrument is intended" and that "item content and stimulus materials are familiar for [the] intended population" (Hambleton, 1994)
11) Make edits to the task and documents as necessary
12) Complete a full pilot of the draft assessment
13) Use the pilot responses to edit and finalise the scoring rubric
14) Once the rubric is finalised, train scorers and score pilot responses

Assessment Evaluation

15) Using the scored pilot responses, complete overall test evaluation, including:
    a. Exploration of pilot descriptive statistics
    b. Tests for reliability
    c. Final confirmation of test validity, including "evaluation of validity in … the target population" (Hambleton, 1994)
16) Modify and, as necessary, re-pilot until satisfactory results are obtained

Appendix B: Field-Testing and Piloting Protocol

Prompts for Field-Testing Sessions:

1) Explain the structure of the actual data collection sessions and give instructions as would give in real session
2) Explain think aloud technique
3) Have students read the full assessment
4) Start by having students paraphrase their understanding of each question, as well as the overall task
5) Have students take turns thinking aloud through how they would answer the assessment. Have other students comment, add and critique if they have different interpretations
6) Prompted questions for follow ups:
    a. Is this test format familiar to you? Do you think it would be familiar to anybody in the target population?
    b. Task 1: Do you anticipate that some respondents will have trouble answering the questions due to their religious preferences regarding alcohol?
    c. Is anything in the instructions unclear or ambiguous?
    d. Does the scenario feel authentic and familiar?
    e. Do the documents feel authentic and familiar?
    f. Would anything in the assessment make respondents feel uncomfortable?
    g. Is the topic potentially offensive? Is any of the language potentially offensive?
    h. Might the assessment unfairly advantage or disadvantage certain members of the target population?
    i. How was the translation? Did you find anything problematic?
    j. Is this type of task representative of the type of task you are expected to be able to complete after graduation?

=> Would anybody be willing to come back tomorrow to field-test the accompanying demographic questionnaire?

=> When analysing, test timing between question and response

Prompts for Questionnaire Field-Testing Sessions:

1) Use same think aloud technique to report thoughts throughout questionnaire
2) Prompted questions for follow ups:
    a. Is the questionnaire format familiar to you? Do you think it would be familiar to anybody in the target population?
    b. Is anything in the instructions unclear or ambiguous?
    c. Are any of the questions unclear or ambiguous?
    d. Is the format unclear or ambiguous?
    e. Would anything in the questionnaire make respondents feel uncomfortable? What questions raise sensitive issues? Is there another way that the question could be rephrased?
    f. Is any of the language potentially offensive?
    g. How was the translation? Did you find anything problematic?

Pilot Session:

1) Make sure that at least one person responds in Kinyarwanda and one person responds in French
2) Distribute three tasks randomly
3) Administer pilot exactly as would administer real session => give instructions verbally in two languages, allow for questions, give 90 minutes to complete, distribute demographic questionnaire at the end (same for test-retest pilots)
4) Ask if anybody is willing to stay after for focus group
5) Questions for focus group:
    a. How did you find the assessment?
    b. Was the test format familiar to you? Do you think it would be familiar to anybody in the target population?
    c. Task 1: Do you anticipate that some respondents will have trouble answering the questions due to their religious preferences regarding alcohol?
    d. Was anything in the instructions unclear or ambiguous?
    e. Did the scenario feel authentic and familiar?
    f. Did the documents feel authentic and familiar?
    g. Would anything in the assessment make respondents feel uncomfortable?
    h. Were any of the topics potentially offensive? Was any of the language potentially offensive?
    i. Might the assessment unfairly advantage or disadvantage certain members of the target population?
    j. Was the questionnaire format familiar to you? Do you think it would be familiar to anybody in the target population?
    k. Was anything in the instructions unclear or ambiguous?
    l. Were any of the questions unclear or ambiguous?
    m. Was the format unclear or ambiguous?
    n. How was the translation? Did you find anything problematic?
    o. Is this type of task representative of the type of task you are expected to be able to complete after graduation?

Appendix C1: Assessment Task 1

Instructions

This test was designed to assess your critical thinking, analytic reasoning and problem solving skills. The results of this assessment will be used in a research study, so your work will remain anonymous and will not be seen by anybody at your university. Please try your best! Our results will only be meaningful if you complete the test to the best of your ability.

Test Format

In this assessment, you will be asked to imagine that you are in a hypothetical but realistic situation. You will be given a series of seven documents that represent a range of different types of information, and you will need to use the information in the documents to support your written responses.

Although your personal values and experiences are important, you should base your responses *only on the evidence provided in these documents*. Responses that only reference personal experience cannot be assessed!

When writing your response, you may choose to use French, English or Kinyarwanda. *Si vous trouvez le français plus facile à lire, vous pouvez aussi demander une version française de l'examen.*

You will have 90 minutes for this assessment.

Test Logistics

You are welcome to make notes on the test booklet when planning your response. There is space on the question pages to write your final responses.

We can only provide one test booklet per student, but please remember that the presentation of your responses is not important for this assessment. We are interested in your thinking skills. Don't waste time recopying your responses!

Thank you very much for your assistance with this research study. Please do not hesitate to ask if you have any questions or concerns.

The Situation[130]

Imagine that you have started an internship at MININFRA (Ministry of Infrastructure) in Kigali. Your role at the Ministry is to assist the chief policy advisor to the Minister of Infrastructure.

One of the government's highest priorities is reducing the number of fatal road accidents in Rwanda. The president and the Parliament have announced that they plan to reduce the number of road fatalities by 30% by the beginning of next year. If they are successful, Rwanda would be the first country in Africa to achieve such a dramatic reduction in road fatalities.

However, the government is suffering from a budget problem. There is not enough money available to implement the entirety of the government's plan to reduce road fatalities. The Parliament is now debating what to do to meet the government's goals.

The Parliamentary Working Group on Road Safety, led by Mr. Ngabo Joseph, has argued that the money would be best used to hire more police officers to enforce the speed limit on the major roads across the country.

Another group within Parliament, led by Ms. Nkurunziza Marie Louise, disagrees. They believe that the money should be used on breathalyzer tests ("alco-tests") to monitor drunk driving. This group has argued that speed limits actually lead to more fatalities and claim that the use of breathalyzer tests is the best solution for reducing road fatalities in Rwanda.

There is a vote scheduled on this issue next week. In advance of the vote, Parliament has asked the Minister of Infrastructure to give a speech, evaluating the two proposed policies.

---

[130] This assessment is an adaptation of the Crime Reduction Performance Task published as part of the Collegiate Learning Assessment (CLA). The CLA is a registered trademark of the Council for Aid to Education. The situation and all documents in this assessment were written explicitly for this task, with the exception of Document D, which is an excerpt from an actual New Times article (accessed at http://www.irexrwanda.com/irexrwanda/?p=21).

Your Task

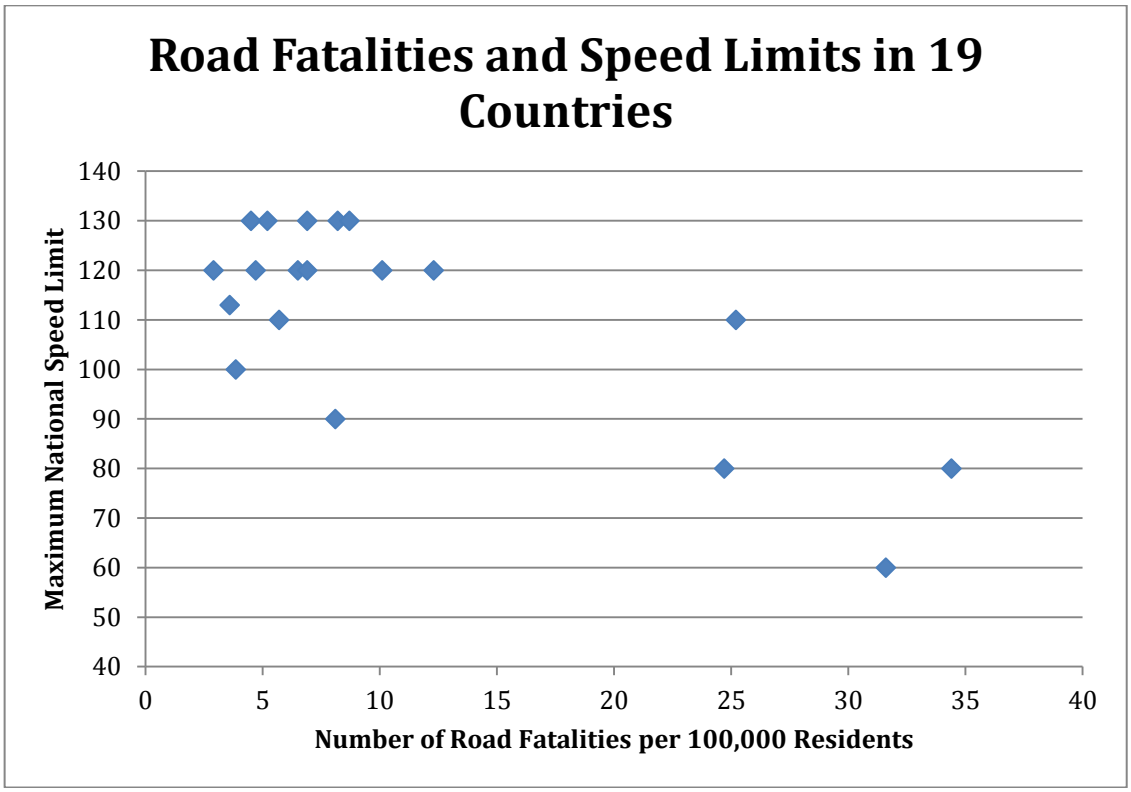
The Minister must address the strengths and limitations of the two policies proposed by the members of Parliament, so your supervisor has asked you to conduct some research on the two options. You identified seven documents, which are found on the following pages.

Your supervisor has asked you to **spend the next ninety minutes reviewing the documents and responding to a number of questions**. The Minister will use your responses as the basis of his speech to Parliament.

DOCUMENT A

Road Fatality Statistics by Country: 2002

The figure below shows the relationship between the maximum speed limit and the number of road fatalities in the 19 countries that participated in the International Road Safety Conference in May 2002.



DOCUMENT B

THE NEWTIMES
Rwanda's First Daily

Member of Parliament's Daughter Dies in Car Accident

A prominent Member of Parliament, Ms. Nkurunziza Marie Louise, has lost her only daughter to a fatal car accident. Ms. Nkurunziza's daughter was killed following a collision with a drunk driver on the road to Kibungo. Although doctors at the Central Hospital worked valiantly to save the young woman, both she and the car's driver died early on Sunday morning. A funeral will be held in her honor on Thursday in Nyamirambo.

DOCUMENT C

AMERICAN INSTITUTE FOR TRANSPORT SAFETY

Research Brief

Study shows breathalyzer tests help to reduce road fatalities

A recent study sponsored by the American Institute for Transport Safety has found that breathalyzer tests can help to dramatically reduce deaths due to road accidents.

The study was carried out in the city of Anytown, U.S.A. Last year, the city decided to increase its use of breathalyzer tests in an attempt to reduce the number of fatal road accidents in the city.

The number of road accidents has dropped dramatically since the initiative began. Last year, there were 525 road accidents (41 fatal) in Anytown. This year, there were 500 accidents in Anytown but only 17 were fatal. This represents a 60% reduction in road fatalities since the department increased the use of the breathalyzer tests.

Research has shown that road accidents are often caused by drunk driving. By keeping drunk drivers off the streets, Anytown has taken control of its road fatality problems.

DOCUMENT D

THE NEWTIMES
Rwanda's First Daily

Another One for the Road?

By Esther Kirabo, The New Times

In November, the traffic police introduced the breathalyzer –commonly known locally as 'alco-test', which is used to test for alcohol levels in drivers. The small, gun-like instrument was introduced due to the persistent increase of road accidents and fatalities that police attribute to the influence of alcohol. Between February and March last year alone, according to police records, some 349 accidents occurred in Rwanda. Statistics from the Kigali University Central Hospital show that more than 60 per cent of those involved in road accidents were found to have been under the influence of alcohol. Among them, 45 are reported to have died.

Sharif Gatete, a taxi manager in the main taxi park agreed that in urban centers, such as Kigali, Gakenke, Rubavu, some drivers drink too much and yet have to drive all day or night. "Night times are particularly dangerous because of the excess speed resulting generally from alcohol consumption," he says.

Some drivers say in rural areas, where roads are in a far worse condition than those in urban centers, speed levels are slower.

Ategekimana Theobald a doctor from the Kigali main referral hospital while consulted about the new technique said that the system was very good and that it will contribute positively in reducing the accidents in the country.

DOCUMENT E

RWANDA ROAD SAFETY COMMISSION

Traffic fatalities and alcohol abuse in Rwanda

The two tables below present data about traffic accidents in Rwanda's five provinces. The information found in Table 1 is presented in chart format in Document F.

TABLE 1: ROAD FATALITY STATISTICS BY PROVINCE

Province	Percentage of residents who frequently drink alcohol[131]	Number of road fatalities	Total population	Number of road fatalities per 100,000 residents
Kigali City	15%	527	745,000	70.7
Northern Province	5%	90	1,740,000	5.17
Eastern Province	12%	150	1,640,000	9.14
Southern Province	14%	226	2,070,000	10.92
Western Province	8%	92	1,940,000	4.74

TABLE 2: DEMOGRAPHIC CHARACTERISTICS BY PROVINCE

Province	Percentage of residents who frequently drink alcohol	Percentage of adult residents in the district who are employed
Kigali City	15%	72%
Northern Province	5%	85%
Eastern Province	12%	84%
Southern Province	14%	85%
Western Province	8%	90%

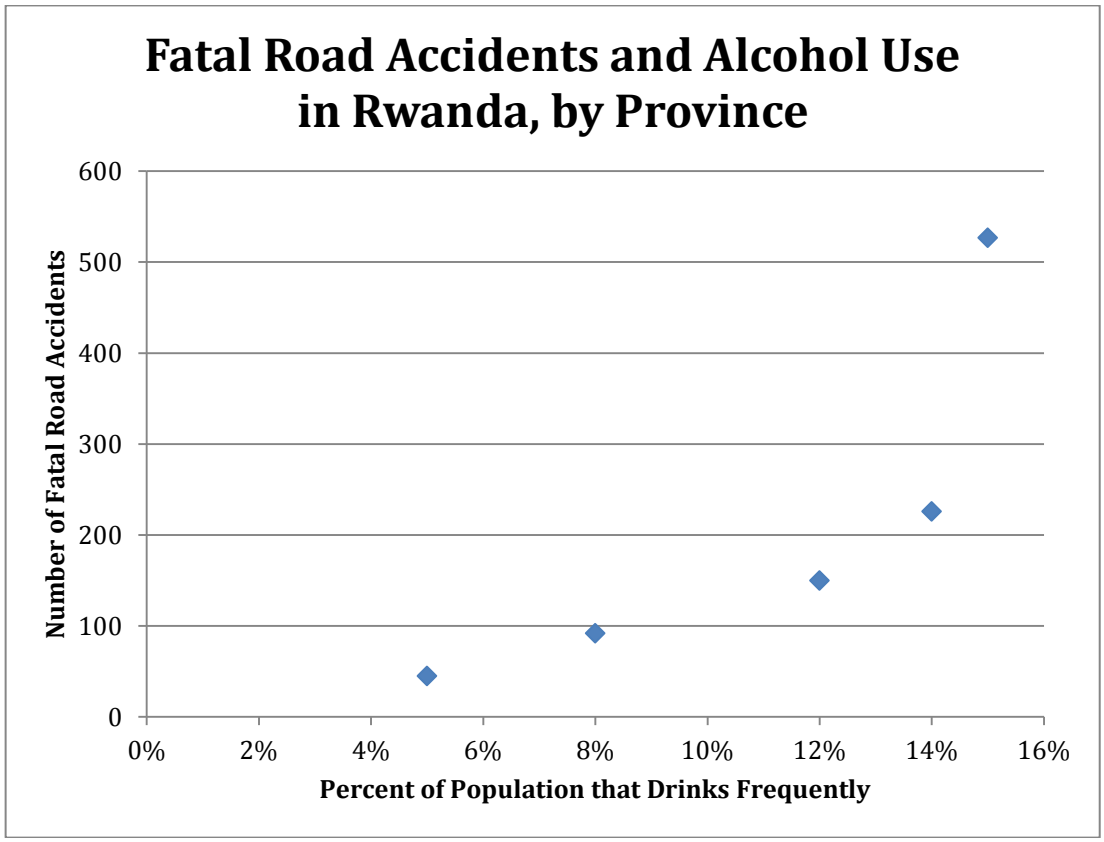
[131] The percentage of residents who drink alcohol was obtained from a survey. In the survey, "frequent use" was defined as more than five drinks per week.

DOCUMENT F

RWANDA ROAD SAFETY COMMISSION

Traffic fatalities and alcohol use in Rwanda: Chart

This chart is based on the data found in Document E.



DOCUMENT G

Results of Online Search for Research Studies

Google scholar | Breathalyzer and decreased road | Search

Potential error in breathalyzer readings
R Schendel – Drug and Alcohol Review, 2001 – JSTOR
…In this analysis, Schendel examines a number of studies of breathalyzer accuracy in determining blood alcohol content. Overall, the studies included in the analysis suggest that breathalyzer tests often underestimate the blood alcohol content of examinees. Factors such as gender, body mass and recent physical exertion were found to impact the results of breathalyzer tests, suggesting that the tests are not 100% accurate in determining blood alcohol content.…

Can breathalyzer tests make a difference? The case of Uganda
A Koleros – African Affairs, 2005 – Social Science Research Database
…This study aims to analyze the effectiveness of using breathalyzer tests to reduce road accidents in Uganda. Individuals arrested for speeding or reckless driving were given a breathalyzer test before completing an in-depth interview with the research team. Results of the study indicate that alcohol was the cause of only some of the dangerous driving. Many of the drivers reported other potential causes in their interviews, such as sleeplessness, anger, road conditions or vehicle maintenance problems…

The Minister's Questions

Now that you have reviewed the documents, you must respond to the Minister's questions (found below).

Your responses should include **all of the relevant evidence** necessary to support your positions.

While your personal values and experiences are important, you should explain and justify the reasons for your conclusions by referring only to the information included in the documents.

When you reference the documents in your response, please **be explicit about what document you are discussing** and please **explain why you selected to reference the evidence that you selected**.

Your response will be assessed on the accuracy of the information you provide and the strength of your argument.

Please write your answers in the space provided. You may also use the back of the paper if needed. Please write legibly! Remember, you may respond in French, English or Kinyarwanda. The three languages will be evaluated equally.

<u>Questions</u>

1. The group of politicians led by Ms. Nkunrunziza has argued that lower speed limits will not help the problem, because lower speed limits actually cause an *increase* in road fatalities. They have referred to Document A as proof of their argument.
    a. Based on the evidence in Document A, do you think it is accurate to claim that lower speed limits cause an increase in road fatalities?
    b. Why or why not?

2. Ms. Nkurunziza's group has argued that spending the available money on alco-tests will allow the government to reduce road fatalities in Rwanda by the desired 30%.
    a. Based on the evidence in all of the documents, what are the strengths and limitations of this argument?

3. Does the evidence in the documents convince you that one of the two options is the best solution for addressing the problem of road fatalities in Rwanda?
    a. *Please answer one of the following questions, depending on your response:*
        i. If yes, which option do you support and why?
        ii. If not, is there another solution that you can propose, based on the evidence?

Appendix C2: Assessment Task 2

Instructions

This test was designed to assess your critical thinking, analytic reasoning and problem solving skills. The results of this assessment will be used in a research study, so your work will remain anonymous and will not be seen by anybody at your university. Please try your best! Our results will only be meaningful if you complete the test to the best of your ability.

Test Format

In this assessment, you will be asked to imagine that you are in a hypothetical but realistic situation. You will be given a series of seven documents that represent a range of different types of information, and you will need to use the information in the documents to support your written responses.

Although your personal values and experiences are important, you should base your responses *only on the evidence provided in these documents*. Responses that only reference personal experience cannot be assessed!

When writing your response, you may choose to use French, English or Kinyarwanda. *Si vous trouvez le français plus facile à lire, vous pouvez aussi demander une version française de l'examen.*

You will have 90 minutes for this assessment.

Test Logistics

You are welcome to make notes on the test booklet when planning your response. There is space on the question pages to write your final responses.

We can only provide one test booklet per student, but please remember that the presentation of your responses is not important for this assessment. We are interested in your thinking skills. Don't waste time recopying your responses!

Thank you very much for your assistance with this research study. Please do not hesitate to ask if you have any questions or concerns.

The Situation[132]

Imagine that you have started an internship at the Ministry of Health (MOH) in Kigali. Your role at the Ministry is to assist the planning officer responsible for malaria programs.

One of the government's central priorities is reducing the number of deaths due to malaria in Rwanda. The president and the Parliament have pledged to reduce the number of deaths due to malaria by 50% in the next two years. If they are successful, Rwanda would be the first country in Africa to achieve such a dramatic reduction in malaria fatalities.

However, the government is suffering from a budget problem. There is not enough money available to implement the entirety of the government's plan to achieve the 50% reduction target. The Parliament is now debating what to do to meet the government's goals.

The Parliamentary Working Group on Malaria Prevention, led by Mr. Ngabo Joseph, has argued that the available money should be used to fund training sessions with pregnant women to teach them how to prevent their children from contracting malaria.

Another group within Parliament, led by Ms. Nkurunziza Marie Louise, disagrees. They believe that the money should be used to buy one mosquito net per household in Rwanda. This group has argued that trainings actually increase the number of deaths due to malaria and claim that providing bed nets is the best solution for reducing malaria fatalities in Rwanda.

There is a vote scheduled on this issue next week. In advance of the vote, Parliament has asked the Minister of Health to give a speech, evaluating the two proposed policies.

---

[132] This assessment is an adaptation of the Crime Reduction Performance Task published as part of the Collegiate Learning Assessment (CLA). The CLA is a registered trademark of the Council for Aid to Education. The situation and all documents in this assessment were written explicitly for this task, with the exception of Document D, which is an excerpt from an actual blog (accessed at: http://blog.compassion.com/malaria-in-africa/#ixzz1VrS7tc2M).

Your Task

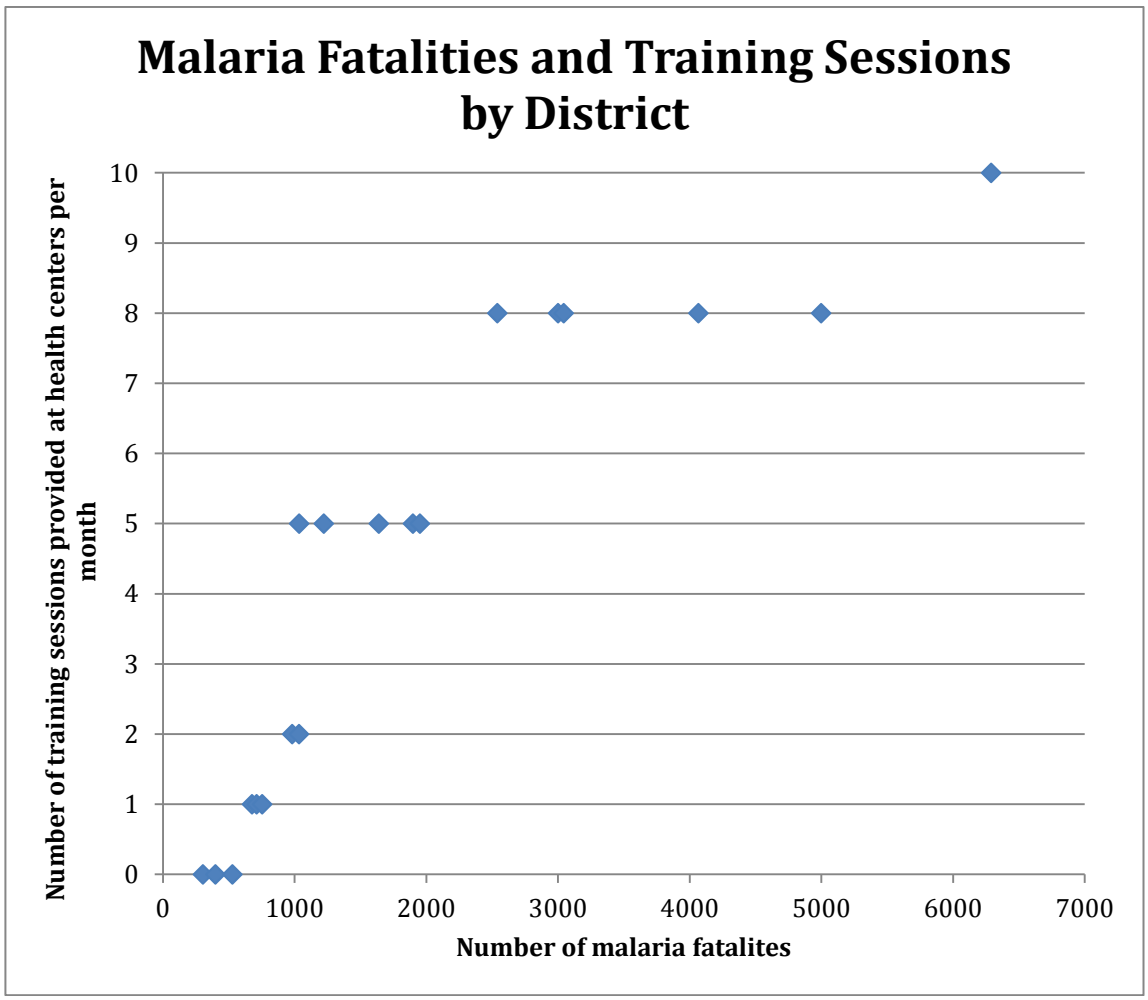
The Minister must address the strengths and limitations of the two policies proposed by the members of Parliament, so your supervisor has asked you to conduct some research on the two options. You identified seven documents, which are found on the following pages.

Your supervisor has asked you to **spend the next ninety minutes reviewing the documents and responding to a number of questions**. The Minister will use your responses as the basis of her speech to Parliament.

DOCUMENT A

Malaria Education Data from Ethiopia

The chart below was created by a non-profit organization based in Ethiopia. It shows the relationship between the average number of malaria prevention training sessions provided by district health centers each month and the number of malaria fatalities in the district. Each point on the chart represents one of the districts that participated in the Ethiopian Malaria Conference in Addis Ababa in 2002.



DOCUMENT B

THE NEWTIMES
Rwanda's First Daily

Rwandan company succeeds in obtaining WHO approval for mosquito bed nets

The World Health Organization (WHO) has announced it will add the Kigali-made *No More Mosquitoes* bed net brand to the WHO approval list of mosquito nets that can be distributed by foreign aid agencies. The acceptance of *No More Mosquitoes* represents the first time that an African company has succeeded in producing a line of bed nets to obtain WHO approval.

The No More Mosquitoes brand is manufactured by FABRWA, a Rwandan company that has been in business in Kicukiro since 2001. The company specializes in producing fabric, netting and other woven products. The company's president, Mr. Mugisha Jean de Dieu, is the husband of Ms. Nkurunziza Marie Louise, who has served as a Member of Parliament since 2007.

DOCUMENT C

SOUTH AFRICAN INSTITUTE FOR PUBLIC HEALTH POLICY
Research Brief

Study shows provision of treated nets can reduce malaria deaths

A recent study sponsored by the South African Institute for Public Health Policy has found that the provision of insecticide-treated mosquito nets can help to dramatically reduce the number of deaths due to malaria.

The study was carried out in the city of Anytown in the Western Cape. Last year, the city decided to triple the number of free bed nets it distributed to low-income families.

The number of deaths due to malaria has dropped dramatically since the initiative began. Last year, there were 6 fatalities in the city due to malaria. This year, there were only 2. This represents a 60% reduction in malaria deaths since the city increased its provision of bed nets.

Research has shown that malaria can be prevented through regular, accurate use of insecticide-treated mosquito nets. By providing nets free of charge to low-income families, Anytown has taken control of its malaria fatality problem.

DOCUMENT D

Releasing children from poverty Compassion in Jesus' name

BLOG

On that Thursday I visited, Nana had been at the center since the morning. After lunch, the team leader took Nana home. When the team leader and Nana reached home, the boy's father was sitting in the shadow of one of the two huts that compose the household.

Some months ago, Nana's family was going through hard times. Nana was sick from malaria. It was harvest time, and Nana's father was totally short of money and could not pay for any treatment. "The Compassion development center paid for all medical fees, fortunately. Otherwise, I would have needed to borrow some money from a friend of mine to treat my child," Nana's father says.

When in September 2008 the student center workers informed Nana's parents and all the registered children's parents that their children were going to receive mosquito nets, it was such a relief to at last get one of the most reliable prevention methods of malaria.

The mosquito net that Nana received was the very first net of the family and is the only one that they have to this day.

Before the mosquito net distribution, there were many children falling sick from malaria – so much so that the center was spending more money than was allocated to health.

"With mosquito nets and sensitization campaigns that we launched, we now have less and less registered children suffering from malaria," a development center team leader says.

Thanks to malaria intervention, Nana was treated and recovered from malaria. Now he sleeps under a mosquito net, happy to be out of the reach of mosquitoes.

DOCUMENT E

PREVENTION INTERNATIONAL

Malaria fatalities and use of bed nets in Rwanda

The two tables below present data about malaria fatalities in Rwanda's five provinces. The information found in Table 1 is presented in chart format in Document F.

TABLE 1: MALARIA FATALITIES BY PROVINCE

Province	Percentage of households lacking mosquito bed nets	Number of deaths due to malaria in 2009	Total population	Number of deaths due to malaria per 100,000 residents
Kigali City	5%	45	745,000	6.04
Eastern Province	15%	527	1,740,000	30.29
Northern Province	14%	226	1,640,000	13.8
Southern Province	8%	92	2,070,000	4.44
Western Province	12%	150	1,940,000	7.73

TABLE 2: DEMOGRAPHIC CHARACTERISTICS BY PROVINCE

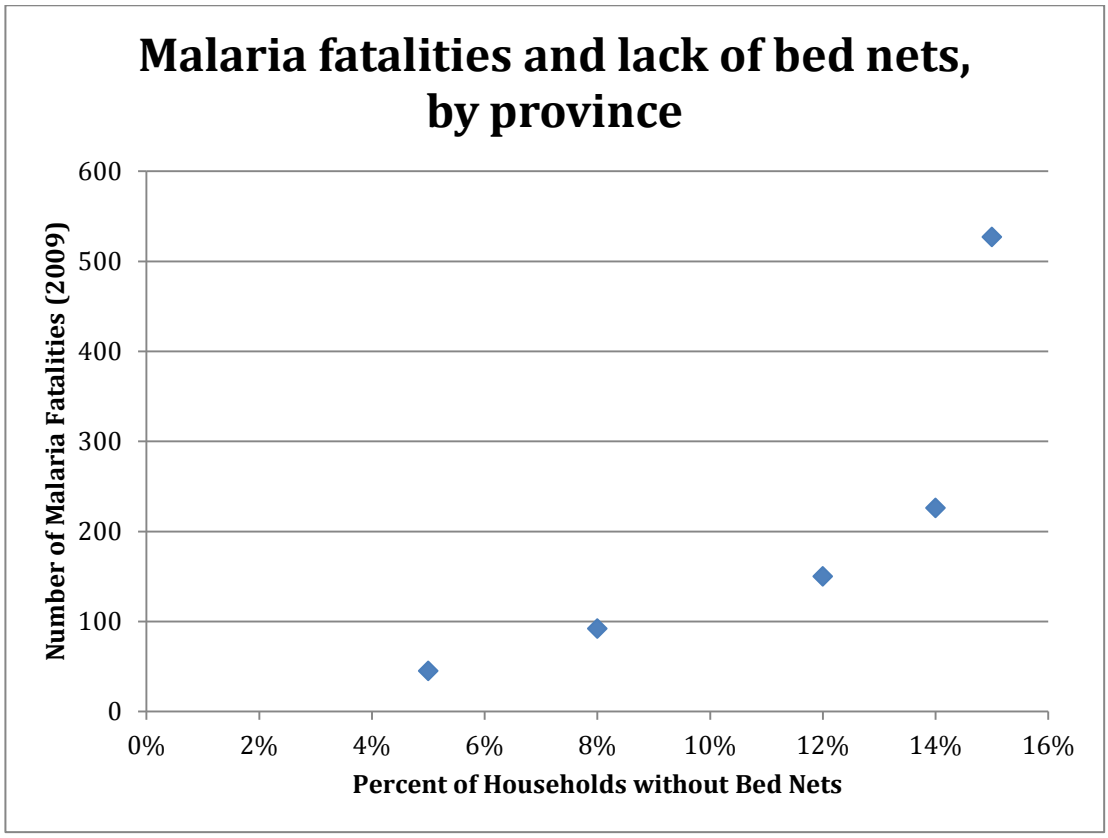
Province	Percentage of deaths due to malaria in 2009	Percentage of adult residents in the district who were employed in 2009
Kigali City	35%	72%
Eastern Province	42%	85%
Northern Province	45%	84%
Southern Province	38%	85%
Western Province	41%	90%

DOCUMENT F

PREVENTION
INTERNATIONAL

Malaria fatalities and use of bed nets in Rwanda: Chart

This chart is based on the data found in Document E.



DOCUMENT G

Results of Online Search for Research Studies

Google scholar | **Mosquito nets and reduced malaria deaths** | **Search**

Combating malaria: A systematic review of malaria prevention policies in Africa
A Koleros – The American Journal of Public Health, 2001 – JSTOR
…In this analysis, Koleros examines a number of evaluations of malaria prevention policies in Africa. Although the majority of the studies demonstrated that provision of free bed nets has played a critical role in reducing the number of malaria fatalities in the region, many also raised the possibility that bed nets are still not well understood in rural areas. Some studies indicated that donated nets are not always used as bed nets. Others identified that nets were used irregularly or inaccurately, which reduced their effectiveness in combating the disease…

Can nets make a difference? The case of southwestern Uganda
R Schendel – Uganda Health Report, 2008 – African Journals Online
…This study sought to analyze the potential for providing free mosquito nets as a prevention strategy for reducing malaria fatalities in southwestern Uganda. Researchers conducted focus groups and interviews with public health workers and medical personnel across the region. Participants were asked to discuss their perceptions of the primary causes of malaria, based on their discussions with infected patients. Although many participants did mention that those infected with malaria often lacked access to bed nets, the majority also highlighted other factors that contribute to the epidemic, including the geographic location of houses, lack of access to treatment and widespread reliance on traditional medicine to cure disease…

The Minister's Questions

Now that you have reviewed the documents, you must respond to the Minister's questions (found below).

Your responses should include **all of the relevant evidence** necessary to support your positions.

While your personal values and experiences are important, you should explain and justify the reasons for your conclusions by referring only to the information included in the documents.

When you reference the documents in your response, please **be explicit about what document you are discussing** and please **explain why you selected to reference the evidence that you selected**.

Your response will be assessed on the accuracy of the information you provide and the strength of your argument.

Please write your answers in the space provided. You may also use the back of the paper if needed. Please write legibly! Remember, you may respond in French, English or Kinyarwanda. The three languages will be evaluated equally.

<u>Questions</u>

1. The group of politicians led by Ms. Nkunrunziza has argued that trainings will not help the problem, because trainings actually cause an *increase* in malaria fatalities. They have referred to Document A as proof of their argument.
    a. Based on the evidence in Document A, do you think it is accurate to claim that trainings cause an increase in malaria fatalities?
    b. Why or why not?

2. Ms. Nkurunziza's group has argued that spending the available money on free bed nets will allow the government to reduce malaria fatalities in Rwanda by the desired 50%.
    a. Based on the evidence in all the documents, what are the strengths and/or limitations of this argument?

3. Does the evidence in the documents convince you that one of the two options is the best solution for addressing the problem of malaria fatalities in Rwanda?
    a. *Please answer one of the following questions, depending on your response:*
        i. If yes, which option do you support and why?
        ii. If not, is there another solution that you can propose, based on the evidence?

Appendix D1: Basic Instructions for Scoring Assessments

1. Read the task and documents thoroughly.
2. Read each response in full.
3. Make a note of any erroneous claims made by the respondent (i.e. incorrect use of evidence).
4. Then, go back through the response with the first page of Scoring Flow Chart and make a note of which documents the student used in his or her response.
    a. Remember, the respondent is not required to reference the document by number. However, it must be clear which document is being used.
5. Using the Flow Chart has a guide, score the response, noting the sub-skill scores on the Score Cover Sheet.
    a. I suggest also making a note of your answers to the questions in the Flow Chart, in case you need to go back for any reason.
    b. Remember that quality of writing is not relevant for this assessment!
6. Once you've finished, you should have nine sub-scores.
7. Check the scores against the descriptions on the Scoring Rubric to double-check that the score makes sense and seems to be valid.
8. When you're finished, send me the test number and the nine scores – and, of course, contact me if you have any questions!

Appendix D2: Scoring Cover Sheet

Score Sheet
Test #_____ (Task 1 / 2)

Cognitive Skill	SCORE
Skill A: Bias Respondent can recognise potential sources of personal bias	
Skill B: Relevance Respondent can determine whether or not information is relevant to the situation	
Skill C: Credibility Respondent can recognise when a source of information is less credible or reliable	
Skill D: Errors Respondent can identify statistical or methodological errors in presented information	
Skill E: Generalisability Respondent can determine whether or not information can be generalised and/or applied to other situations	
Skill F: Missing Information Respondent can recognise when there is a lack of information	
Skill G: Evaluation of Connections Respondent can evaluate whether or not information is connected and, if so, whether the data is conflicting or complementary	
Skill H: Evaluation of Support Respondent can evaluate whether or not information supports or contradicts an argument	
Skill I: Use of Evidence Respondent can draw on evidence when formulating a decision	

Appendix D3: Scoring Flow Chart (Task 1)

1a. Is the overall response illegible, incomprehensible or unrelated to the task question?

Yes => If yes, automatically score entire response as "0"	No => If no, continue scoring

1b. Does the response include at least one reference to a task document?

Yes => If yes, continue scoring	No => If no, automatically score entire response as "0"

2. Which documents are referenced in the response?

Used?	Doc	Relevance?	Credible?	Includes erroneous data?	Reliance on other national contexts?	Use of individual examples?	Missing info?
	A	Very relevant (Q1) Less relevant (Q2, Q3)	Yes	No	Yes	No	Yes (other variables that may contribute to trend)
	B	Less relevant (irrelevant for Q1)	No	N/A	No	Yes	N/A
	C	Very relevant (irrelevant for Q1)	Yes	No	Yes	Yes	Yes (other variables that may impact reduction)
	D	Very relevant (irrelevant for Q1)	No	N/A	No	No	N/A
	E	Table 1 – Very relevant (irrelevant for Q1) Table 2 – Irrelevant	Yes	Yes (potential bias in use of self-report data, if use # instead of %)	No	No	Yes (regional factors)
	F	Very relevant (irrelevant for Q1)	Yes	Yes (% versus #)	No	No	No
	G	Very relevant (irrelevant for Q1)	Yes	No	Yes	No	No

3. Using this chart as a guide, continue to the scoring questions on page 2.

Scoring Questions

Skill A: Recognition of Personal Bias

Does the response acknowledge Ms. Nkurunziza's potential personal bias towards the use of breathalyzer tests?

If yes =>
Does the response also acknowledge that her personal experience makes her argument more compelling?

If yes => Score "5"

If no (response rules out the use of Doc B altogether)
=> Score "4"

If no =>
Does the respondent use Document B as evidence?

If yes =>
Does the respondent use Ms. Nkurunziza's personal experience as proof that alcohol is the leading cause of road accidents?

If yes => Score "1"

If no (simply uses document to demonstrate that she has a compelling reason to support breathalyzer tests)
=> Score "2"

If no (Doc B not mentioned) => Score "3 "

Skill B: Determination of Relevance

Does the response reference any irrelevant documents? (See chart on page 1)

If yes =>
Score "1"

If no =>
Does the response reference Document A in Question 1?

If yes =>
Does the response reference any other relevant documents?

If yes => How many other relevant documents?

If one => Score "3"

If two or three* => Score "4"

If more than three => Score "5"

If no => Score "2"

If no =>
Does the response reference any other relevant documents?

If yes =>
Does the response reference only one relevant document?

If yes => Score "2"

If no (response references multiple relevant documents aside from Document A) => Score "3"

If no => Score "1"

* For this skill, Documents E & F should not be counted individually (i.e. referencing both only counts as one relevant set of evidence)

Skill C: Recognition of Source Credibility

Does the response reference any less credible documents? (Registered as "No" on the chart on p. 1)

If yes =>
Does the response also reference credible documents (aside from the use of Document A in Question 1)?

If yes =>
Does the respondent discuss the relative credibility of the documents?

If yes => Score "4"

If no => Score "2"

If no => Score "1"

If no =>
Does the respondent discuss the relative credibility of the documents?

If yes => Score "5"

If no (selection of documents could be by chance) => Score "3"

Skill D: Identification of Methodological Errors in Information

Does the response argue against the claim in Question 1?

If yes =>

Does the respondent discuss the difference between causation and correlation by acknowledging that there may be other variables that could affect the graph?

If yes =>

Does the response rely on any erroneous data as evidence?*

If yes => Score "3"

If no =>

Does the response explicitly question erroneous evidence presented in at least one other document (aside from Doc A)?

If yes => Score "5"

If no => Score "4"

If no =>

Does the response rely on any erroneous data as evidence?*

If yes => Score "2"

If no =>

Does the response explicitly question erroneous evidence presented in at least one other document (aside from Doc A)?

If yes => Score "4"

If no => Score "3"

If no =>

Does the response rely on any erroneous data as evidence?*

If yes => Score "1"

If no => Does the response explicitly question erroneous evidence presented in at least one other document (aside from Doc A)?

If yes => Score "3"

If no => Score "2"

* Data refers to specific data within a document, not to the entire document

Skill E: Determination of Generalisability

Does the response use any documents that refer to other national contexts and/or that rely on individual cases? (See chart on page 1)

If yes =>
Does the response explicitly question the generalisability of either type of document?

If yes =>
Does the response question both the use of other contexts *and* the use of individual cases?

If yes => Score "5"

If no (only questions one) => Score "4"

If no =>
Does the response uncritically use both documents referencing other contexts *and* those relying on individual cases as evidence?

If yes => Score "1"

If no (only uses one) => Score "2"

If no => Score "3"

Skill F: Recognition of Missing Information

Does the response explicitly state that there is sufficient information available to adequately answer each question, including proposing an overall solution?

If yes =>
Score "1"

If no =>
Does the response acknowledge the need for more information overall?

If yes =>
Does the response also acknowledge the need for more information within at least one task document?

If yes =>
Does the response indicate that *multiple* documents are missing critical information?

If yes => Score "5"

If no (only mentions one doc) => Score "4"

If no => Score "3"

If no =>
Does the response acknowledge the need for more information within at least one task document?

If yes => Score "3"

If no => Score "2"

Skill G: Evaluation of How Information Connects

Does the response use more than one piece of evidence to respond to Questions 2 and 3?

If yes =>
Does the response use more than 4 pieces of evidence in the overall response (i.e. Questions 1-3)?

If yes =>
Does the response acknowledge that some of the evidence in the documents conflicts and some is complementary?*

If yes => Score "5"

If no (only discusses how evidence is complementary *or* how the evidence contradicts) => Score "3"

If no =>
Does the response acknowledge that some of the evidence in the documents conflicts and some is complementary?*

If yes => Score "4"

If no (only discusses how evidence is complementary *or* how the evidence contradicts) => Score "2"

If no => Score "1"

* For this skill, the respondent does not need to have identified errors, issues with credibility, etc, as discussed in other skills. Rather, this skill takes each piece of evidence at face value (i.e. does a given piece of evidence, *if taken to be correct as presented*, support or contradict the claims that the respondent says it supports or contradicts)?

Skill H: Evaluation of How Information Supports/Contradicts an Argument

Is all of the evidence from the documents used correctly to support or contradict claims* (i.e. is evidence used in support of a claim actually supportive of the claim)?

If yes =>
Does the response include both evidence that supports *and* evidence that contradicts the claims presented in the questions?

If yes =>
Does the respondent discuss *why* the selected evidence supports or contradicts the claims?

If yes => Score "5"

If no (or sometimes no) => Score "4"

If no =>
Does the respondent discuss *why* the selected evidence supports or contradicts the claims?

If yes => Score "3"

If no (or sometimes no) => Score "2"

If no => Score "1"

* For this skill, the respondent does not need to have identified errors, issues with credibility, etc, as discussed in other skills. Rather, this skill takes each piece of evidence at face value (i.e. does a given piece of evidence, *if taken to be correct as presented*, support or contradict the claims that the respondent says it supports or contradicts)?

Skill I: Ability to Draw on Evidence when Formulating Decisions

Does the respondent propose a decision or an alternative solution?

If yes =>

Is the decision/solution based entirely on anecdotal evidence and/or personal experience?

If yes => Score "1"

If no =>

Does the response rely on more than one task document?

If yes =>

Does the evidence used to support the decision/solution actually support the decision/solution?

If yes =>

Does the response acknowledge differing perspectives and make a case for why the proposed decision/solution is the best, given the presented evidence?

If yes =>

Does the respondent make any use of anecdotal evidence or personal experience?

If yes => Score "4"

If no => Score "5"

If no => Score "3"

If no => Score "2"

If no => Score "2"

If no => Score

Appendix D4: Scoring Rubric (Task 1)

Cognitive Skill	Exemplary 5 points	Good 4 points	Average 3 points	Weak 2 points	Poor 1 point	Un-scorable 0 points
<u>Skill A</u> Respondent can recognise potential sources of personal bias	Response acknowledges Ms. Nkurunziza's potential personal bias Also acknowledges that her personal situation may make her argument more compelling (although not more empirically sound)	Response acknowledges Ms. Nkurunziza's potential personal bias Respondent rules Document B out as a possible source, given the potential bias	Response does not mention Document B	Response uses Document B but not in support of Ms Nkurunziza's claim	Response actively uses Document B as evidence in support of Ms. Nkurunziza's claim	Response is illegible, incomprehensible or unrelated to the task question *and/or* Response references no task documents
<u>Skill B</u> Respondent can determine whether or not information is relevant to the situation	Response references all of the most relevant documents, in addition to referencing Document A in Question 1	Response references two or three of the most relevant documents, in addition to referencing Document A in Question 1	Response references one relevant document, in addition to referencing Document A in Question 1 ---------- OR ---------- Response references more	Response references one relevant document (*either* Document A in Question 1 or one relevant document in Questions 2/3)	Response references no relevant documents ---------- OR ---------- Response references at least one *irrelevant*	Response is illegible, incomprehensible or unrelated to the task question *and/or* Response references no task documents

			than one relevant document but does not reference Doc A in Question 1		document	
<u>Skill C</u> Respondent can recognise when a source of information is less credible or reliable	Response uses only credible sources of evidence Respondent explicitly discusses the credibility of the sources he or she uses	Response includes a mixture of credible and less credible documents (aside from the use of Document A in Question 1) Respondent explicitly discusses the credibility of the sources he or she uses	All documents included in the response are credible Less credible evidence is not used Respondent does not discuss the credibility of the documents he or she selected to include, indicating that documents may have been selected by chance	Response includes a mixture of credible and less credible documents (aside from the use of Document A in Question 1) Respondent does not discuss the credibility of the documents he or she selected to include, indicating that documents may have been selected by chance	Only less credible documents are included in response (aside from the use of Document A in Question 1)	Response is illegible, incomprehensible or unrelated to the task question *and/or* Response references no task documents
<u>Skill D</u> Respondent can identify statistical or methodological errors in presented information	Respondent argues against the claim in Question 1 by discussing the difference between causation and correlation Response does not use erroneous data	Respondent argues against the claim in Question 1 by discussing the difference between causation and correlation Response does not use erroneous data	Respondent argues against the claim in Question 1 by discussing the difference between causation and correlation Response uses erroneous data as	Respondent argues against the claim in Question 1 but does not explain why the claim is problematic (or uses an untenable reason) Response uses	Respondent accepts the claim made in Question 1 Response uses erroneous data as evidence	Response is illegible, incomprehensible or unrelated to the task question *and/or* Response references no task

	as evidence Response explicitly questions erroneous evidence presented in at least one other document	as evidence Respondent does not explicitly question evidence in any other document --- OR --- Respondent argues against the claim in Question 1 but does not explain why the claim is problematic (or uses an untenable reason) Response does not use erroneous data as evidence Response explicitly questions erroneous evidence presented in at least one other document	evidence --- OR --- Respondent argues against the claim in Question 1 but does not explain why the claim is problematic (or uses an untenable reason) Response does not use erroneous data as evidence, but also does not explicitly question evidence in any other document --- OR --- Respondent accepts the claim made in Question 1 Response does not use erroneous data as evidence, but does explicitly question evidence	erroneous data as evidence --- OR --- Respondent accepts the claim made in Question 1 Response does not use erroneous data as evidence Respondent does not explicitly question evidence in any other document		documents

			in at least one other document			
<u>Skill E</u> Respondent can determine whether or not information can be generalised and/or applied to other situations	Response explicitly questions the generalisability of presented evidence from other contexts to Rwandan context *And* Response explicitly questions the use of individual examples as generalisable evidence	Response explicitly questions the generalisability of presented evidence from other contexts to Rwandan context *Or* Response explicitly questions the use of individual examples as generalisable evidence	Response does not use individual examples or other national examples as evidence Response includes no discussion of generalisability, indicating that the respondent may have avoided evidence by chance	Respondent uses evidence from other national contexts uncritically as evidence in the response *Or* Respondent uses evidence from individual examples uncritically as evidence in the response	Respondent uses evidence from other national contexts uncritically as evidence in the response *And* Respondent uses evidence from individual examples uncritically as evidence in the response	Response is illegible, incomprehensible or unrelated to the task question *and/or* Response references no task documents
<u>Skill F</u> Respondent can recognise when there is a lack of information	Response explicitly mentions the general need for more information in order to respond to the scenario *And* Response explicitly mentions that *multiple* documents	Response explicitly mentions the general need for more information in order to respond to the scenario *And* Response explicitly mentions that *one* document is	Response explicitly mentions the general need for more information in order to respond to the scenario *Or* Response explicitly mentions that at least one document	Response does not acknowledge the need for additional information (either overall or within individual documents)	Respondent explicitly states that there is sufficient information (overall and within each document) to respond to the task	Response is illegible, incomprehensible or unrelated to the task question *and/or* Response references no task documents

	are missing critical information	missing critical information	is missing critical information			
<u>Skill G</u> Respondent can evaluate whether or not information is connected and, if so, whether the data is conflicting or complementary	Response draws on multiple pieces of evidence from the documents (>4) *And* Response acknowledges that some of the evidence in the documents conflicts and some of the evidence is complementary	Response draws on a few different pieces of evidence from the documents (2-4) *And* Response acknowledges that some of the evidence in the documents conflicts and some of the evidence is complementary	Response draws on multiple pieces of evidence (>4) *And* Response only discusses how the evidence is complementary *or* how the evidence contradicts (response does not discuss both areas of agreement and disagreement)	Response draws on a few different pieces of evidence from the documents (2-4) *And* Response only discusses how the evidence is complementary *or* how the evidence contradicts	Response only uses one piece of evidence to respond to each question	Response is illegible, incomprehensible or unrelated to the task question *and/or* Response references no task documents
<u>Skill H</u> Respondent can evaluate whether or not information supports or contradicts an argument	Response includes evidence that supports *and* evidence that contradicts the claims presented in the questions Evidence from the documents is used correctly to support or contradict the claims (i.e.	Response includes evidence that supports *and* evidence that contradicts the claims presented in the questions Evidence from the documents is used correctly to support or contradict the claims	Response only discusses *either* how evidence supports *or* contradicts the claims presented in the questions Evidence from the documents is used correctly to support or contradict the claims	Response only discusses *either* how evidence supports *or* contradicts the claims presented in the questions Evidence from the documents is used correctly to support or contradict the claims	At least some of the evidence included in the response is used incorrectly (i.e. it does not support an argument it is intended to support), reflecting an incorrect evaluation of how the evidence relates to the argument	Response is illegible, incomprehensible or unrelated to the task question *and/or* Response references no task documents

	evidence used in support of a claim actually does support the claim) Response includes an explicit discussion of why the evidence presented supports or contradicts the claim	Response includes no (or occasionally no) discussion of why evidence supports or contradicts the claim	Response includes an explicit discussion of why the evidence presented supports or contradicts the claim	Response includes no (or occasionally no) discussion of why evidence supports or contradicts the claim		
<u>Skill I</u> Respondent can draw on evidence when formulating a decision	Response proposes a decision – or an alternative solution Decision/alternative is entirely supported with data from multiple documents (No reference to anecdotal evidence or personal experience) Response acknowledges differing perspectives and makes a case for why the proposed	Response proposes a decision – or an alternative solution Decision/alternative is mostly supported with data from multiple documents (Occasional reference to anecdotal evidence/ personal experience) Response acknowledges differing perspectives and makes a case for	Response proposes a decision – or an alternative solution Decision/alternative is mostly supported with data from multiple documents (Occasional reference to anecdotal evidence or personal experience) Response ignores any conflicting evidence in the documents	Response proposes a decision – or an alternative solution Evidence used to support the decision or solution does not actually support the decision or solution (i.e. response uses conflicting evidence to support the proposed decision) ---------- OR ---------- Decision or solution is	Response does not propose a decision or present an alternative solution ---------- OR ---------- Decision/alternative presented but either based entirely on anecdotal evidence/personal experience or formulated without reference to task documents	Response is illegible, incomprehensible or unrelated to the task question *and/or* Response references no task documents

	decision or solution is the best, given the presented evidence	why the proposed decision or solution is the best, given the presented evidence		supported by only one task document		

Appendix D5: Scoring Flow Chart (Task 2)

1a. Is the overall response illegible, incomprehensible or unrelated to the task question?

Yes => If yes, automatically score entire response as "0"	No => If no, continue scoring

1b. Does the response include at least one reference to a task document?

Yes => If yes, continue scoring	No => If no, automatically score entire response as "0"

2. Which documents are referenced in the response?

Used?	Doc	Relevant?	Credible?	Includes erroneous data?	Reliance on other national contexts?	Use of individual examples?	Missing info?
	A	Very relevant (Q1) Less relevant (Q2, Q3)	Yes	No	Yes	No	Yes (other variables that may contribute to trend)
	B	Less relevant (irrelevant for Q1)	No	N/A	No	No	N/A
	C	Very relevant (irrelevant for Q1)	Yes	Yes (questionable significance of reduction)	Yes	Yes	Yes (other variables that may impact reduction)
	D	Very relevant (irrelevant for Q1)	No	N/A	No	Yes	N/A
	E	Table 1 – Very relevant (irrelevant for Q1) Table 2 - Irrelevant	Yes	Yes (if use # instead of %)	No	No	Yes (regional factors)
	F	Very relevant (irrelevant for Q1)	Yes	Yes (% versus #)	No	No	No
	G	Very relevant (irrelevant for Q1)	Yes	No	Yes	No	No

3. Using this chart as a guide, continue to the scoring questions on page 2.

Scoring Questions

Skill A: Recognition of Personal Bias

Does the response acknowledge Ms. Nkurunziza's potential personal bias towards the use of mosquito nets?

If yes =>
Does the response also acknowledge that there may be a benefit to her husband's company (i.e. possibility for local production)?

If yes => Score "5"

If no (response rules out the use of Document B altogether) => Score "4"

If no =>
Does the respondent use Document B as evidence?

If yes =>
Does response use the existence of FABRWA as proof that mosquito nets are the best solution to the problem of malaria in Rwanda?

If yes => Score "1"

If no (simply uses document to show the potential of using nets as a solution, given the possibility of local production) => Score "2"

If no (Doc B not mentioned) => Score "3"

Skill B: Determination of Relevance

Does the response reference any irrelevant documents? (See chart on page 1)

If yes => Score

* For this skill, Documents E & F should not be counted individually (i.e. referencing both only counts as one relevant set of evidence)

If no =>
Does the response reference Document A in Question 1?

If yes =>
Does the response reference any other relevant documents?

If yes => How many other relevant documents?

If one => Score "3"

If two or three* => Score "4"

If more than three => Score "5"

If no => Score "2"

If no =>
Does the response reference any other relevant documents?

If yes =>
Does the response reference only one relevant document?

If yes => Score "2"

If no (response references multiple relevant documents aside from Document A) => Score "3"

If no => Score "1"

Skill C: Recognition of Source Credibility

Does the response reference any less credible documents? (Registered as "No" on the chart on page 1)

If yes =>
Does the response also reference credible documents (aside from the use of Document A in Question 1)?

If yes =>
Does the respondent discuss the relative credibility of the documents?

If yes => Score "4"

If no => Score "2"

If no => Score "1"

If no =>
Does the respondent discuss the relative credibility of the documents?

If yes => Score "5"

If no (selection of documents could be by chance) => Score "3"

Skill D: Identification of Methodological Errors in Information

Does the response argue against the claim in Question 1?

If yes =>
Does the respondent discuss the difference between causation and correlation by acknowledging that there may be other variables that could affect the graph?

If yes =>
Does the response rely on any erroneous data as evidence?*

If yes => Score "3"

If no =>
Does the response explicitly question erroneous evidence presented in at least one other document (aside from Doc A)?

If yes => Score "5"

If no => Score "4"

If no =>
Does the response rely on any erroneous data as evidence?*

If yes => Score "2"

If no =>
Does the response explicitly question erroneous evidence presented in at least one other document (aside from Doc A)?

If yes => Score "4"

If no => Score "3"

* Data refers to specific data within a document, not to the entire document

If no =>
Does the response rely on any erroneous data as evidence?*

If yes => Score "1"

If no => Does the response explicitly question erroneous evidence presented in at least one other document (aside from Doc A)?

If yes => Score "3"

If no => Score "2"

Skill E: Determination of Generalisability

Does the response use any documents that refer to other national contexts and/or that rely on individual cases? (See chart on page 1)

If yes =>

Does the response explicitly question the generalisability of either type of document?

If yes =>

Does the response question both the use of other contexts *and* the use of individual cases?

If yes => Score "5"

If no (only questions one) => Score "4"

If no =>

Does the response uncritically use both documents referencing other contexts *and* those relying on individual cases as evidence?

If yes => Score "1"

If no (only uses one) => Score "2"

If no => Score "3"

Skill F: Recognition of Missing Information

Does the response explicitly state that there is sufficient information available to adequately answer each question, including proposing an overall solution?

If yes =>
Score "1"

If no =>

Does the response acknowledge the need for more information overall?

If yes =>

Does the response also acknowledge the need for more information within particular task documents?

If yes =>

Does the response indicate that *multiple* documents are missing critical information?

If yes => Score "5"

If no (only mentions one doc) => Score "4"

If no => Score "3"

If no =>

Does the response acknowledge the need for more information within particular task documents?

If yes => Score "3"

If no => Score "2"

Skill G: Evaluation of How Information Connects

Does the response use more than one piece of evidence to respond to Questions 2 and 3?

If yes =>
Does the response use more than 4 pieces of evidence in the overall response (i.e. Questions 1-3)?

If yes =>
Does the response acknowledge that some of the evidence in the documents conflicts and some is complementary?*

If yes => Score "5"

If no (only discusses how evidence is complementary *or* how the evidence contradicts) => Score "3"

If no =>
Does the response acknowledge that some of the evidence in the documents conflicts and some is complementary?*

If yes => Score "4"

If no (only discusses how evidence is complementary *or* how the evidence contradicts) => Score "2"

If no =>
Score "1"

* For this skill, the respondent does not need to have identified errors, issues with credibility, etc, as discussed in other skills. Rather, this skill takes each piece of evidence at face value (i.e. does a given piece of evidence, *if taken to be correct as presented*, support or contradict the claims that the respondent says it supports or contradicts)?

Skill H: Evaluation of How Information Supports/Contradicts an Argument

Is all of the evidence from the documents used correctly to support or contradict claims* (i.e. is evidence used in support of a claim actually supportive of the claim)?

If yes =>
Does the response include evidence that supports *and* evidence that contradicts the claims presented in the questions?

If yes =>
Does the respondent discuss *why* the selected evidence supports or contradicts the claims?

If yes => Score "5"

If no (or sometimes no) => Score "4"

If no =>
Does the respondent discuss *why* the selected evidence supports or contradicts the claims?

If yes => Score "3"

If no (or sometimes no) => Score "2"

If no => Score "1"

* For this skill, the respondent does not need to have identified errors, issues with credibility, etc, as discussed in other skills. Rather, this skill takes each piece of evidence at face value (i.e. does a given piece of evidence, *if taken to be correct as presented*, support or contradict the claims that the respondent says it supports or contradicts)?

Skill I: Ability to Draw on Evidence when Formulating Decisions

Does the respondent propose a decision or an alternative solution?

If yes =>

Is the decision/solution based entirely on anecdotal evidence and/or personal experience?

If yes => Score "1"

If no =>

Does the response rely on more than one task document?

If yes =>

Does the evidence used to support the decision/solution actually support the decision/solution?

If yes =>

Does the response acknowledge differing perspectives and make a case for why the proposed decision/solution is the best, given the presented evidence?

If yes =>

Does the respondent make any use of anecdotal evidence or personal experience?

If yes => Score "4"

If no => Score "5"

If no => Score "3"

If no => Score "2"

If no => Score "2"

If no => Score

Appendix D6: Scoring Rubric (Task 2)

Cognitive Skill	Exemplary 5 points	Good 4 points	Average 3 points	Weak 2 points	Poor 1 point	Un-scorable 0 points
<u>Skill A</u> Respondent can recognise potential sources of personal bias	Response acknowledges Ms. Nkurunziza's potential personal bias Also acknowledges that her personal situation carries some advantages (i.e. nets can be produced locally in Rwanda)	Response acknowledges Ms. Nkurunziza's potential personal bias Respondent rules Document B out as a possible source, given the potential bias	Response does not mention Document B	Response uses Document B but not in support of Ms Nkurunziza's claim	Response actively uses Document B as evidence in support of Ms. Nkurunziza's claim	Response is illegible, incomprehensible or unrelated to the task question *and/or* Response references no task documents
<u>Skill B</u> Respondent can determine whether or not information is relevant to the situation	Responses references all of the most relevant documents, in addition to referencing Document A in Question 1	Response references two or three of the most relevant documents, in addition to referencing Document A in Question 1	Response references one relevant document, in addition to referencing Document A in Question 1	Response references one relevant document (*either* Document A in Question 1 or one relevant document in Questions 2/3)	Response references no relevant documents ---------- OR ---------- Response references at least one *irrelevant*	Response is illegible, incomprehensible or unrelated to the task question *and/or* Response references no task documents

			---------- OR ---------- Response references more than one relevant document but does not reference Document A in Question 1		document	
<u>Skill C</u> Respondent can recognise when a source of information is less credible or reliable	Response uses only credible sources of evidence Respondent explicitly discusses the credibility of the sources he or she uses	Response includes a mixture of credible and less credible documents (aside from the use of Document A in Question 1) Respondent explicitly discusses the credibility of the sources he or she uses	All documents included in the response are credible Less credible evidence is not used Respondent does not discuss the credibility of the documents he or she selected to include, indicating that documents may have been selected by chance	Response includes a mixture of credible and less credible documents (aside from the use of Document A in Question 1) Respondent does not discuss the credibility of the documents he or she selected to include, indicating that documents may have been selected by chance	Only less credible documents are included in response (aside from the use of Document A in Question 1)	Response is illegible, incomprehensible or unrelated to the task question *and/or* Response references no task documents

<u>Skill D</u> Respondent can identify statistical or methodological errors in presented information	Respondent argues against the claim in Question 1 by discussing the difference between causation and correlation Response does not use erroneous data as evidence Response explicitly questions erroneous evidence presented in at least one other document	Respondent argues against the claim in Question 1 by discussing the difference between causation and correlation Response does not use erroneous data as evidence Respondent does not explicitly question evidence in any other document ---------- OR Respondent argues against the claim in Question 1 but does not explain why the claim is problematic (or uses an untenable reason) Response does not use erroneous data as evidence Response explicitly	Respondent argues against the claim in Question 1 by discussing the difference between causation and correlation Response uses erroneous data as evidence ---------- OR ---------- Respondent argues against the claim in Question 1 but does not explain why the claim is problematic (or uses an untenable reason) Response does not use erroneous data as evidence, but also does not explicitly question evidence in any other document ---------- OR	Respondent argues against the claim in Question 1 but does not explain why the claim is problematic (or uses an untenable reason) Response uses erroneous data as evidence ----------OR ---------- Respondent accepts the claim made in Question 1 Response does not use erroneous data as evidence Respondent does not explicitly question evidence in any other document	Respondent accepts the claim made in Question 1 Response uses erroneous data as evidence	Response is illegible, incomprehensible or unrelated to the task question *and/or* Response references no task documents

		questions erroneous evidence presented in at least one other document	---------- Respondent accepts the claim made in Question 1 Response does not use erroneous data as evidence, but does explicitly question evidence in at least one other document			
<u>Skill E</u> Respondent can determine whether or not information can be generalised and/or applied to other situations	Response explicitly questions the generalisability of presented evidence from other contexts to Rwandan context *And* Response explicitly questions the use of individual examples as generalisable evidence	Response explicitly questions the generalisability of presented evidence from other contexts to Rwandan context *Or* Response explicitly questions the use of individual examples as generalisable evidence	Response does not use individual examples or other national examples as evidence Response includes no discussion of generalisability, indicating that the respondent may have avoided evidence by chance	Respondent uses evidence from other national contexts uncritically as evidence in the response *Or* Respondent uses evidence from individual examples uncritically as evidence in the response	Respondent uses evidence from other national contexts uncritically as evidence in the response *And* Respondent uses evidence from individual examples uncritically as evidence in the response	Response is illegible, incomprehensible or unrelated to the task question *and/or* Response references no task documents

<u>Skill F</u> Respondent can recognise when there is a lack of information	Response explicitly mentions the general need for more information in order to respond to the scenario *And* Response explicitly mentions that *multiple* documents are missing critical information	Response explicitly mentions the general need for more information in order to respond to the scenario *And* Response explicitly mentions that *one* document is missing critical information	Response explicitly mentions the general need for more information in order to respond to the scenario *Or* Response explicitly mentions that at least one document is missing critical information	Response does not acknowledge the need for additional information (either overall or within individual documents)	Respondent explicitly states that there is sufficient information (overall and within each document) to respond to the task	Response is illegible, incomprehensible or unrelated to the task question *and/or* Response references no task documents
<u>Skill G</u> Respondent can evaluate whether or not information is connected and, if so, whether the data is conflicting or complementary	Response draws on multiple pieces of evidence from the documents (>4) *And* Response acknowledges that some of the evidence in the documents conflicts and some of the evidence is complementary	Response draws on a few different pieces of evidence from the documents (2-4) *And* Response acknowledges that some of the evidence in the documents conflicts and some of the evidence is complementary	Response draws on multiple pieces of evidence (>4) *And* Response only discusses how the evidence is complementary *or* how the evidence contradicts (response does not discuss both areas of agreement and disagreement)	Response draws on a few different pieces of evidence from the documents (2-4) *And* Response only discusses how the evidence is complementary *or* how the evidence contradicts	Response only uses one piece of evidence to respond to each question	Response is illegible, incomprehensible or unrelated to the task question *and/or* Response references no task documents

<u>Skill H</u> Respondent can evaluate whether or not information supports or contradicts an argument	Response includes evidence that supports *and* evidence that contradicts the claims presented in the questions Evidence from the documents is used correctly to support or contradict the claims (i.e. evidence used in support of a claim actually does support the claim) Response includes an explicit discussion of why the evidence presented supports or contradicts the claim	Response includes evidence that supports *and* evidence that contradicts the claims presented in the questions Evidence from the documents is used correctly to support or contradict the claims Response includes no (or occasionally no) discussion of why evidence supports or contradicts the claim	Response only discusses *either* how evidence supports *or* contradicts the claims presented in the questions Evidence from the documents is used correctly to support or contradict the claims Response includes an explicit discussion of why the evidence presented supports or contradicts the claim	Response only discusses *either* how evidence supports *or* contradicts the claims presented in the questions Evidence from the documents is used correctly to support or contradict the claims Response includes no (or occasionally no) discussion of why evidence supports or contradicts the claim	At least some of the evidence included in the response is used incorrectly (i.e. it does not support an argument it is intended to support), reflecting an incorrect evaluation of how the evidence relates to the argument	Response is illegible, incomprehensible or unrelated to the task question *and/or* Response references no task documents

<u>Skill I</u> Respondent can draw on evidence when formulating a decision	Response proposes a decision – or an alternative solution Decision/alternative is entirely supported with data from multiple documents (No reference to anecdotal evidence or personal experience) Response acknowledges differing perspectives and makes a case for why the proposed decision or solution is the best, given the presented evidence	Response proposes a decision – or an alternative solution Decision/alternative is mostly supported with data from multiple documents (Occasional reference to anecdotal evidence/personal experience) Response acknowledges differing perspectives and makes a case for why the proposed decision or solution is the best, given the presented evidence	Response proposes a decision – or an alternative solution Decision/alternative is mostly supported with data from multiple documents (Occasional reference to anecdotal evidence or personal experience) Response ignores any conflicting evidence in the documents	Response proposes a decision – or an alternative solution --- Evidence used to support the decision or solution does not actually support the decision or solution (i.e. response uses conflicting evidence to support the proposed decision) OR Decision or solution is entirely supported by one task document	Response does not propose a decision or present an alternative solution --- OR --- Decision/alternative presented but either based entirely on anecdotal evidence/personal experience or formulated without reference to task documents	Response is illegible, incomprehensible or unrelated to the task question *and/or* Response references no task documents

Appendix E: Accompanying Questionnaire

Please answer the following demographic questions. The information will be used for statistical purposes during the analysis stage of the research study and will remain entirely confidential. Please do not hesitate to ask if you have any questions or concerns.

Demographic Information

1. Are you male or female? (Please check one) ☐ Male (2) ☐ Female (1)

2. In what year were you born? _____ 3. How old are you? _____ Years

---

Academic Background

University

4. Which university do you attend? ☐ NUR (1) ☐ KIST (2) ☐ SFB (3)

5. What is the name of your faculty? _____

6. What is the name of your department? _____

7. In which year are you at university? (Please check one)

☐ First (1) ☐ Fourth (2) ☐ Other (Please specify:_____) (3)

8. Do you study during the day or in the evening this year? ☐ Day (1) ☐ Evening (2)

9. Where do you live when university is in session (during the academic year)?

Province: _____ District: _____

Secondary School

10. We are interested in the secondary school where you *finished* your secondary studies. Where was your school located?

Province: _____ District: _____

11. What type of secondary school did you attend? (Please check one)

☐ Public school (1) ☐ Religious school (2) ☐ Private school (not religious) (3)

12. Was your secondary school located in an urban or rural area? (Please check one)

☐ Urban (2) ☐ Rural (1)

13. Was it located in a medium-sized city, large city, small town or village? (Please check one)

☐ Large City	☐ Medium-sized City	☐ Small Town	☐ Village
(More than 20,000 people)	(5,000-20,000)	(1,000-5,000)	(Less than 1,000)

14. What option did you do in secondary school? _____

15. What score did you receive on the National Exam? _____ out of _____

16. Did your National Exam score make you eligible for a Government Scholarship?

☐ Yes (2) ☐ No (1)

Family Background

17. Did you live with any adults when you were a child? (Please check one)

☐ Yes (1) ☐ No (2)

=> If you did not live with any adults, you may skip questions #18-20 and continue to question #21.

=> If you did live with adults, you may continue with question #18.

18. If you did live with adults, had *any of the adults in your household* ever attended school? (Please check one)

☐ Yes (1) ☐ No (2) ☐ Don't Know

=> If no adult in your household had ever attended school, you may skip questions #19-20 and continue to question #21.

=> If ONE adult in your household attended school, you may continue with question #19.

=> If MORE THAN ONE adult in your household attended school, you may skip question #19 and continue to question #20.

19. If ***only one adult*** in your household attended school, what was the **highest** level of education that he or she achieved? (Please check **one**)

☐ Primary (1)
☐ Vocational Training (2)
☐ Tronc Commun (3)
☐ Secondary (4)
☐ Technical College (5)
☐ University (6)
☐ Higher than University (7)
☐ Don't Know

=> If you answered this question, you may skip question #20 and continue to question #21.

20. If ***more than one adult*** in your household attended school, think of the person who attained the *highest level of education.* What level of education did he or she achieve? (Please check **one**)

☐ Primary (1)
☐ Vocational Training (2)
☐ Tronc Commun (3)
☐ Secondary (4)
☐ Technical College (5)
☐ University (6)
☐ Higher than University (7)
☐ Don't Know

=> If you answered this question, you may now continue to question #21.

Geographic Background

21. Do you have a family home where you regularly stay during the holidays?

☐ Yes (1) ☐ No (2)

=> If you answered "NO" to this question, you have finished the questionnaire. Please hand it to a member of the research team. Thank you!

=> If you answered "YES" to this question, please continue to question #22.

22. If yes, where is your family home located?

Province: _____ District: _____

23. Is your family home located in an urban or rural area? (Please check one)

☐ Urban (2) ☐ Rural (1)

24. Is your home located in a medium-sized city, large city, small town or village? (Please check one)

☐ Large City (More than 20,000 people) ☐ Medium-sized City (5,000-20,000) ☐ Small Town (1,000-5,000) ☐ Village (Less than 1,000)

**25. Does your family home have electricity? ☐ Yes (2) ☐ No (1)

26. A radio? ☐ Yes (2) ☐ No (1)

27. A television? ☐ Yes (2) ☐ No (1)

28. A refrigerator? ☐ Yes (2) ☐ No (1)

29. Does anybody in your household own a bicycle? ☐ Yes (2) ☐ No (1)

---

30. Does anybody in your household own a motorcycle/scooter? ☐ Yes (2) ☐ No (1)

31. Does anybody in your household own a car/truck? ☐ Yes (2) ☐ No (1)

32. Does anybody in your household own a phone? ☐ Yes (2) ☐ No (1)

---

33. At your family home, what type of fuel do you **primarily use for cooking**? (Please check **one**)

☐ Electricity (7)
☐ Natural Gas *(Gaze)* (6)
☐ Biogas (5)
☐ Kerosene (4)
☐ Charcoal (*Amakara)* (3)
☐ Firewood/Straw *(Inkwi/Ubwatsi)* (2)
☐ Dung *(Amase yumye)* (1)
☐ Don't Know
☐ Other (Please specify: _____)

34. What is currently the **main source of drinking water** for members of your household? (Please check **one**)

☐ Piped water to your dwelling (*Amazi ya robine mu nzu)* (13)
☐ Piped water to your yard/plot *(Amazi ya robine mu rugo)* (12)
☐ Public Tap *(Robine rusange)* (11)
☐ Open Well in your dwelling (*Iriba riri mu nzu)* (6)
☐ Open Well in your yard/plot *(Iriba riri mu rugo)* (5)
☐ Open Public Well *(Iriba rusange)* (4)
☐ Covered Well in your dwelling *(Iriba rifutse mu nzu)* (9)
☐ Covered Well in your yard/plot *(Iriba rifutse riri mu rugo)* (8)
☐ Covered Public Well *(Iriba rifutse rusange)* (7)
☐ Spring/River/Stream/Lake/Pond *(Isoko/Umugezi/Ikiyaga/Ikizenga)* (3)
☐ Dam (Barrage) (2)
☐ Rainwater *(Amazi y'imvura)* (1)
☐ Tanker Truck *(Amazi aza mu bigega by'amakamyo)* (10)
☐ Bottled Water (14)
☐ Don't Know
☐ Other (Please specify: _____)

35. What kind of toilet does your household use? (Please check **one**)

☐ Flush Toilet *(Umusarani wa Kijyambere)* (4)
☐ Traditional Pit Toilet *(Umusarani usanzwe)* (2)
☐ Ventilated Improved Pit Toilet (3)
☐ No Facility/Bush/Field/Beach (1)
☐ Don't Know
☐ Other (Please specify: _____)

36. What kind of material covers **most of the floors in your household**? (Please check **one**)

☐ Earth/Sand/Mud *(Itaka/Umucanga/Icyondo)* (2)
☐ Mud Mixed with Dung *(Icyondo kivanze n'amase)* (1)
☐ Wood Planks *(Imbaho)* (4)
☐ Palm/Bamboo *(Imigano/Imikenke)* (3)
☐ Parquet or Polished Wood *(Imbaho zisennye)* (5)
☐ Ceramic Tiles *(Amakaro)* (7)
☐ Cement *(Isima)* (6)
☐ Carpet *(Itapi)* (8)
☐ Don't Know
☐ Other (Please specify: _____)**

Thank you very much for your help with this research study! We will communicate with you when the study is complete. Murakoze cyane!

The questions between asterisks are based on the DHS Household Survey administered in Rwanda in 2010

Appendix F1: Sample Invitation Letter

February 2012

Dear _____:

I am contacting you today to ask if you would be willing to participate in a research study that I will be conducting on your university campus this month.

My name is Rebecca Schendel, and I am a PhD student at the Institute of Education in London. My doctoral research is focused on academic quality at Rwanda's public universities. As you know, Rwanda's universities are an important part of the government's development strategy. My research seeks to assess the ability of public university students to demonstrate critical thinking skills when faced with situations that they will likely face following graduation.

My study will begin with an assessment of these abilities. The assessment will be administered to students at three of the public universities in Rwanda. Following the assessment, I will be conducting a number of focus groups and interviews with students to discuss the university experience on your campus.

This research has been officially approved by the Rector of [University].

Your name was randomly selected from [University]'s enrolment list. Participation in the study is, of course, purely voluntary, but I would like to encourage you to participate if you are willing and able to do so.

I have attached a brief information sheet for your review. Please take the time to read the information before deciding whether or not to participate.

All selected students are asked to meet with me and a member of my research team on DATE to discuss participation in the study. Please come to XXX anytime between TIME and TIME pm on DATE. The meeting will only last about 15 minutes. Refreshments will be provided.

I look forward to meeting you soon.

Best wishes,

Rebecca Schendel

Appendix F2: Participant Information Sheet (Quantitative Phase)

Leading education and social research
Institute of Education
University of London

London International Development Centre

Participant Information Sheet

Title of Project: Critical thinking at Rwanda's public universities and implications for higher education reform

Name of Researcher: Rebecca Schendel

Thank you for your interest in this research study! I hope that this information sheet will give you all the information that you need about the project. If you find you have further questions or concerns, please do not hesitate to contact me or another member of the research team.

Why is this research being conducted?
I am a PhD student at the Institute of Education in London. This research study will be written up as my doctoral thesis. I also intend to share the results with university administrators and the Ministry of Education in Rwanda.

Who will be taking part?
Students from three public universities in Rwanda will be participating in the project. All participants will be selected randomly from university enrollment lists.

What will happen during the research?
Participants in the study will participate in an assessment activity in February 2012. The research team plans to hold a number of data collection sessions on your university campus, and we will ask that you attend one. During the session, you will be asked to complete a 90-minute assessment, as well as a short demographic questionnaire. Your assessment will be scored by a team of researchers and analyzed based on some of your responses to the demographic questionnaire.

I will be returning to Rwanda in a few months time to conduct some follow-up interviews. If you would be willing to be contacted to participate in this second phase of the study, please note your name and contact information on your consent form.

What questions will be asked?
The assessment takes the form of a "performance task". This means that you will be presented with a 'real-world' situation, a role to assume, and a number of related documents. You will be asked to read the documents and respond to a number of questions, imagining what you would do if you were presented with the situation in real life.

The demographic questionnaire will ask you some basic information about your family background, your academic background, and your secondary school.

What will happen to the research findings?
The findings will be analysed and used in my doctoral thesis. This means that I will present the thesis for examination at the Institute of Education. If I am successful, the thesis will remain available for public use in the Institute of Education library in London.

I will also produce a short report of the findings and related recommendations, which will be shared with university administrators and representatives from the Ministry of Education. Your participation will remain entirely confidential.

- - - - - - - - - -

All aspects of this project have been reviewed by the Research Ethics Committee at the Institute of Education, as well as by the Rwandan Ministry of Education.

Please do not hesitate to contact me with any questions or concerns. You may also contact any of the members of the research team. Our contact information is below:

[Contact details removed from Appendix for confidentiality purposes]

Murakoze cyane! Thank you again for your help!

Appendix F3: Participant Information Sheet (Qualitative Phase)

Leading education and social research
Institute of Education
University of London

London International Development Centre

Participant Information Sheet

Title of Project: Critical thinking at Rwanda's public universities and implications for higher education reform

Name of Researcher: Rebecca Schendel

Thank you for your interest in this research study! I hope that this information sheet will give you all the information that you need about the project. If you find you have further questions or concerns, please do not hesitate to contact me or another member of the research team.

Why is this research being conducted?
I am a PhD student at the Institute of Education in London. This research study will be written up as my doctoral thesis. I also intend to share the results with the KIST and NUR administration and with the Ministry of Education in Rwanda.

Who will be taking part?
Participants in this study include undergraduate students, faculty members and administrators from KIST and NUR. All contacted students participated in the first phase of the study in February 2012, at which time they volunteered to attend a follow-up interview. Contacted faculty members and administrators were identified by colleagues who have already participated in the research.

What will happen during the research?
There are two ways that participants may be involved. Some students may be asked to attend a focus group session with approximately ten other undergraduate students from their institution. The focus group will last approximately two hours and will be held on or near campus. Alternatively, participants may be asked to attend a one-hour interview with the researcher. All interviews may be conducted in either English or French, depending on participant preference. Students may also feel free to speak in Kinyarwanda, as translators will be present at all student interviews.

What questions will be asked?
Questions will focus on student experiences during university. Interviews with faculty members and administrators will also ask about institutional policies and the culture of the institution.

Will participants have the opportunity to review the final results?
Yes. Every participant will be given the opportunity to review the written transcript of his or her interview. The final results will also be disseminated on the university campus at the end of the project, and all participants are welcome to attend and give feedback.

What will happen to the research findings?
The findings will be analysed and used in my doctoral thesis. This means that I will present the thesis for examination at the Institute of Education. If I am successful, the thesis will remain available for public use in the Institute of Education library in London.

I will also produce a short report of the findings and related recommendations, which will be shared with university administrators and representatives from the Ministry of Education. *Your participation will remain entirely confidential.* Although the individual institutions will be named in the study, participant names will never be mentioned.

What is the benefit of participation?
Unfortunately, I am unable to provide any financial incentives for participation in this study. However, it is my hope that the results of this work will be very useful both for university administrators and for the Ministry of Education as they work together to improve higher education in Rwanda. You can be assured that your participation will help to achieve this goal!

- - - - - - - - - - -

All aspects of this project have been reviewed by the Research Ethics Committee at the Institute of Education, as well as by the Rwandan Ministry of Education.

Please do not hesitate to contact me with any questions or concerns. You may also contact any of the members of the research team. Our contact information is below:

[Contact details removed from Appendix for confidentiality purposes]

Murakoze cyane! Thank you for your help!

Appendix G: Informed Consent Form

Informed Consent Form

Title of Project: Critical thinking at Rwandan universities and implications for higher education reform

Name of Researcher: Rebecca Schendel

By signing below, I agree to the following statements:

1) I have read and understood the attached information sheet giving details of the project.
2) I have had the opportunity to ask the researcher any questions that I had about the project and my involvement in it, and I understand my role in the project.
3) My decision to consent is entirely voluntary, and I understand that I am free to withdraw at any time without giving a reason.
4) I understand that data gathered in this project may form the basis of a report or other form of publication or presentation.
5) I understand that my name will not be used in any report, publication or presentation and that every effort will be made to protect my confidentiality.

Participant's Signature: _____ Date: _____

Participant's name _____

Researcher's Signature: _____ Date: _____

Appendix H: Inventory of Decisions Made During Data Entry

Academic Field

- Aggregate fields determined by Faculty responses, not Department responses

Secondary School Type

- When determining typology, used "Urban" vs. "Rural" response, rather than estimate of size of town, even if there was a discrepancy between responses
    - Only exception was in instances when there was no response for "Urban" versus "Rural"; in this instance, used size of town response to generate "Urban" versus "Rural" response
- If respondent selected "Religious School", assumed to be private, unless also selected "Public School", in which case classified as Public

Parental Education Level

- If respondent said "no adults in school" or "did not know if adults in school" and then selected a level, changed response to "yes" for Adults in School
- If respondent entered a highest education level for both one and many adults, only entered response under Many Adults (and coded N/A under One Adult)
- If respondent selected two options for highest education level (either within Many Adults or between One and Many Adults), entered the highest
- One respondent said he did not live with adults, left question about education blank and then selected a level, counted this response as No Parents/No Education

Socio-Economic Status

- If respondent only ticked "Yes" for some of the binary asset variables and left others blank, assumed blanks should have been "No" and filled in accordingly
- If respondent entered more than one option for categorical asset variables, entered highest response (e.g. charcoal is "higher" than firewood/straw for cooking material)
- One respondent said he did not have a regular family residence but then responded to all of the demographic questions. Recoded him as having a regular family residence.

Appendix I: Evaluation of Assumptions and Additional Quantitative Results

A. Calculation of Wealth Quintiles

1) Missing Values Analysis for asset data

Table A1: Comparison of Means, Cases with and without Missing Data (n=220)

Skill	Missing or No Missing	Mean	Standard Deviation	Standard Error of Mean	Equal variances assumed? (based on Levene's test for Equality of Variances)	Results of *t*-test and Significance (*p*-value)
A: Bias	Missing (n=24) No Missing (n=196)	2.29 2.54	1.083 1.097	.221 .078	Yes	$t(218) = -1.030$ (p=.304)
B: Relevance	Missing (n=24) No Missing (n=196)	2.50 2.81	1.532 1.476	.313 .105	Yes	$t(218) = -.955$ (p=.340)
C: Credibility	Missing (n=24) No Missing (n=196)	2.29 2.23	.908 .942	.185 .067	Yes	$t(218) = .281$ (p=.779)
D: Errors	Missing (n=24) No Missing (n=196)	1.88 1.94	.900 .980	.184 .070	Yes	$t(218) = -.303$ (p=.762)
E: Generalisability	Missing (n=24) No Missing (n=196)	1.33 1.45	.868 .902	.177 .064	Yes	$t(218) = -.595$ (p=.552)
F: Missing Information	Missing (n=24) No Missing (n=196)	1.88 1.98	.680 .774	.139 .055	Yes	$t(218) = -.663$ (p=.508)
G: Evaluation of Connections	Missing (n=24) No Missing (n=196)	2.67 2.27	1.711 1.527	.349 .109	Yes	$t(218) = 1.184$ (p=.238)

H: Evaluation of Support	Missing (n=24)	2.79	1.793	.366	Yes	$t(218) = 2.175$ (p=.031)
	No Missing (n=196)	2.07	1.497	.107		
I: Use of Evidence	Missing (n=24)	2.13	1.424	.291	Yes	$t(218) = -.114$ (p=.910)
	No Missing (n=196)	2.16	1.340	.096		

** Significant at 1% level of significance*

2) Evaluation of Assumptions for PCA

- Sufficient sample size
  - n=195: This is very close to the 200 cases suggested as a fair sample size by Tabachnik and Fidell (2007)
- Missing Data
  - As discussed in Section 6.2.3.4, 40 cases were missing values for at least one of the 12 asset variables. Means could be substituted for the missing values for 15 of these cases. The remaining 25 were removed prior to analysis.
  - The PCA was also rerun without the substitutions for the 15 cases. There was no difference in the results of the PCA between the version with and the version without the substitutions.
- Normality/Linearity/ Homoscedasticity
  - Not relevant, given the binary nature of the variables

3) Results of PCA for wealth quintiles

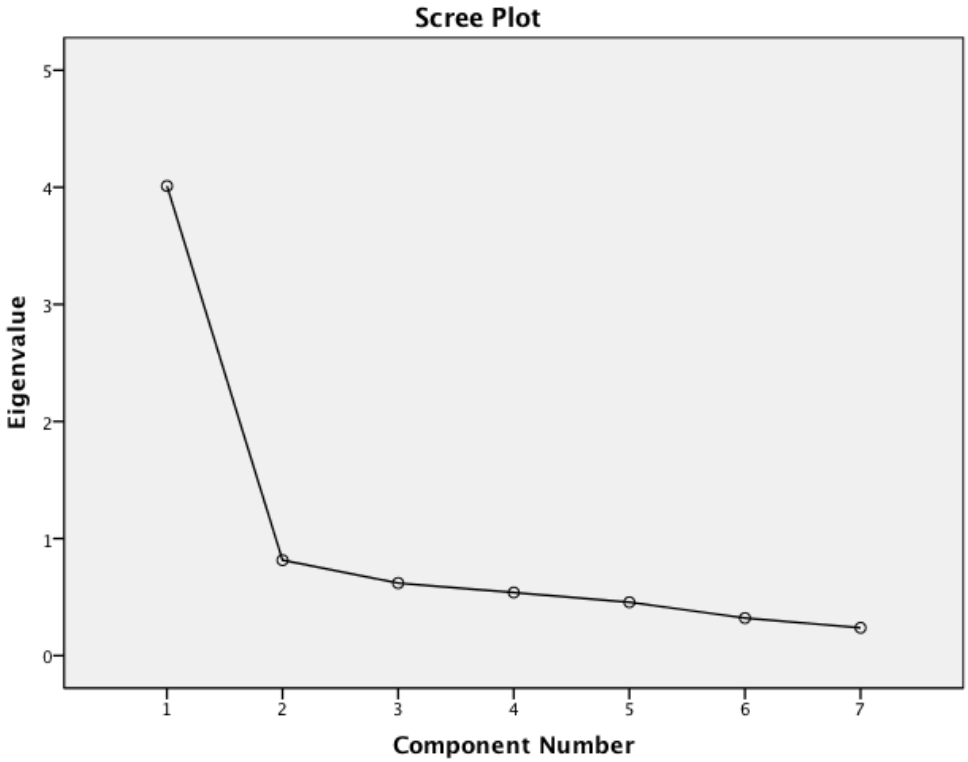
Table A2: Communalities, PCA for Wealth Quintiles

	Extraction
Dummy for Electricity	.598
Dummy for TV	.769
Dummy for Fridge	.508
Dummy for Non-Wood Cooking	.574
Dummy for Any Running Water	.591
Dummy for Flush Toilet	.496
Dummy for Manufactured Floor	.490

Table A3: Total Variance Explained, PCA for Wealth Quintiles

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance Explained	Cumulative %
1	4.026	57.513	57.513	4.026	57.513	57.513
2	.832	11.893	69.405			
3	.580	8.280	77.686			
4	.538	7.692	85.377			
5	.446	6.371	91.749			
6	.337	4.819	96.568			
7	.240	3.432	100.000			

Figure A1: Scree Plot for Wealth Quintiles



B. Principal Components Analysis: Assessment Scores

1) Evaluation of Assumptions

- Sufficient sample size
  - n=199: This is very close to the 200 cases suggested as a fair sample size by Tabachnik and Fidell (2007).
- Missing Data
  - As discussed in Section 6.3, 21 cases with missing data removed for analysis
- Normality/Linearity/Homoscedasticity
  - Although the individual score distributions all demonstrate significant skewness and kurtosis, the distribution in the overall population can be assumed to be normal, given that the sample size is significantly larger than the 30 cases required for the central limit theorem (Tolmie, Muijs & McAteer, 2011).
  - Scatterplots were inconclusive, given the ordinal nature of the scoring method.

2) *Results of PCA*

Table A4: Communalities, PCA for Main Sample

	Extraction
A: Bias	.994
B: Relevance	.895
C: Credibility	.997
D: Errors	.968
E: Generalisability	.981
F: Missing Information	.999
G: Evaluation of Connections	.819
H: Evaluation of Support	.830
I: Use of Evidence	.722

Table A5: Total Variance Explained, PCA for Main Sample

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
Component	Total	% of Variance	Cumulative %	Total	% of Variance Explained	Cumulative %	Total	% of Variance Explained	Cumulative %
1	2.756	30.619	30.619	2.756	30.619	30.619	1.887	20.97	20.97
2	1.271	14.123	44.742	1.271	14.123	44.742	1.22	13.55	34.52
3	1.148	12.759	57.501	1.148	12.759	57.501	1.05	11.661	46.182
4	.879	9.769	67.27	.879	9.769	67.27	1.021	11.34	57.522
5	.818	9.093	76.363	.818	9.093	76.363	1.015	11.281	68.804
6	.677	7.518	83.881	.677	7.518	83.881	1.006	11.179	79.982
7	.654	7.268	91.148	.654	7.268	91.148	1.005	11.166	91.148
8	.497	5.525	96.673						
9	.299	3.327	100						

C. Distribution of assessment scores

Figure A2: Histogram, Skill A Scores

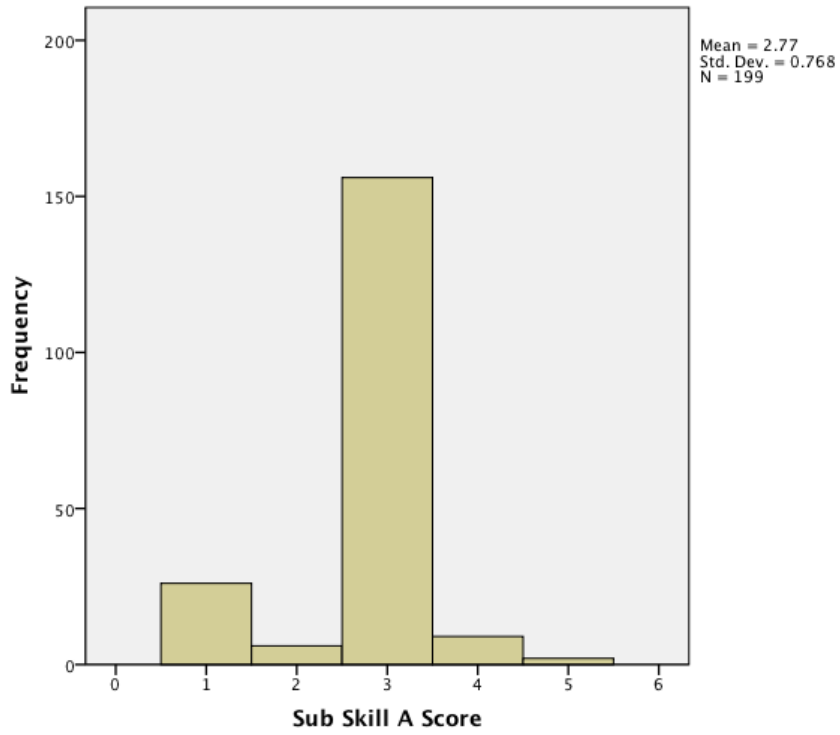


Figure A3: Histogram, Skill B Scores

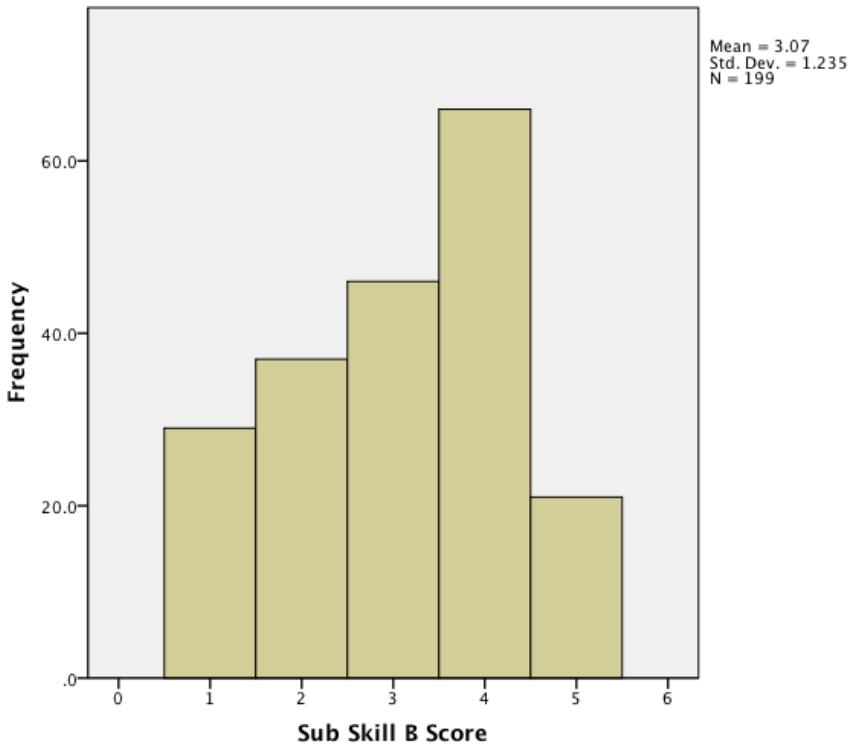


Figure A4: Histogram, Skill C Scores

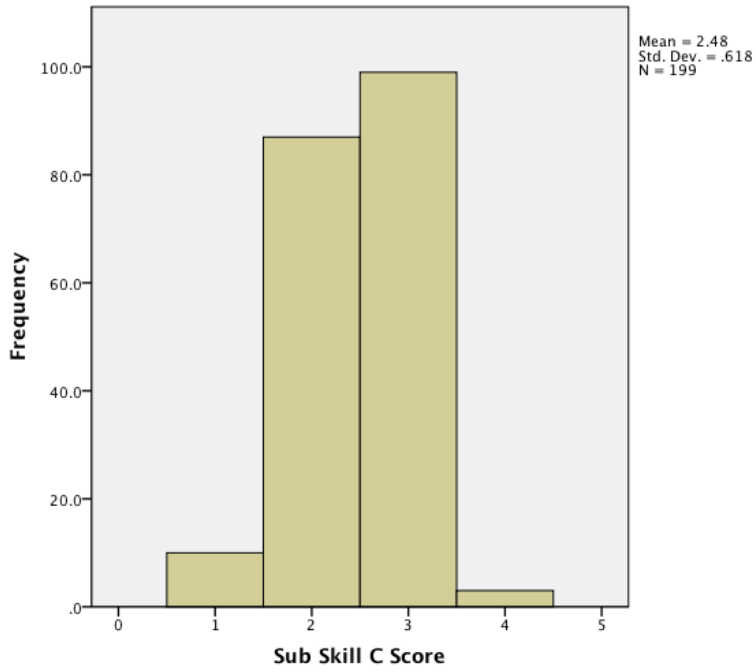


Figure A5: Histogram, Skill D Scores

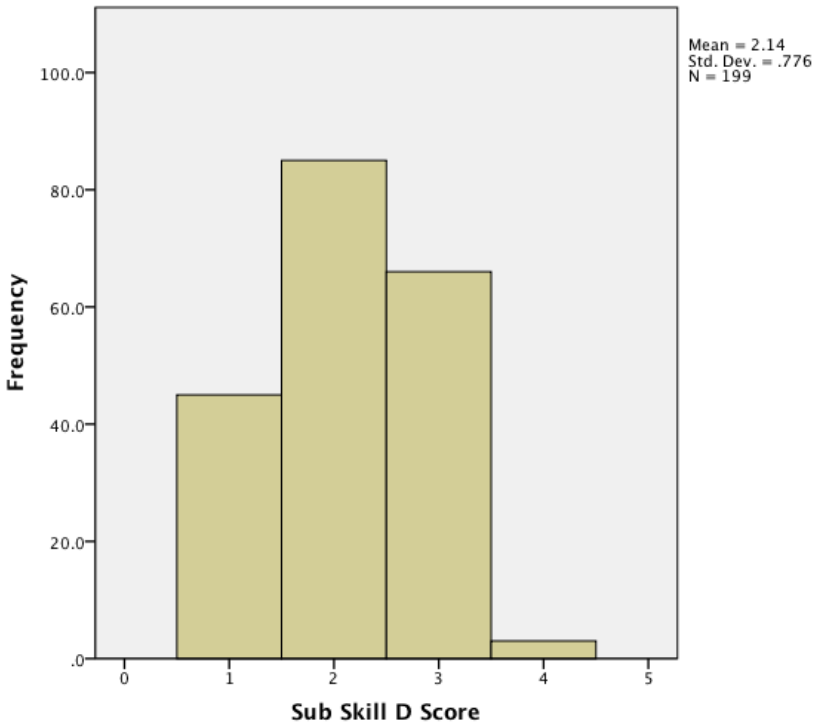


Figure A6: Histogram, Skill E Scores

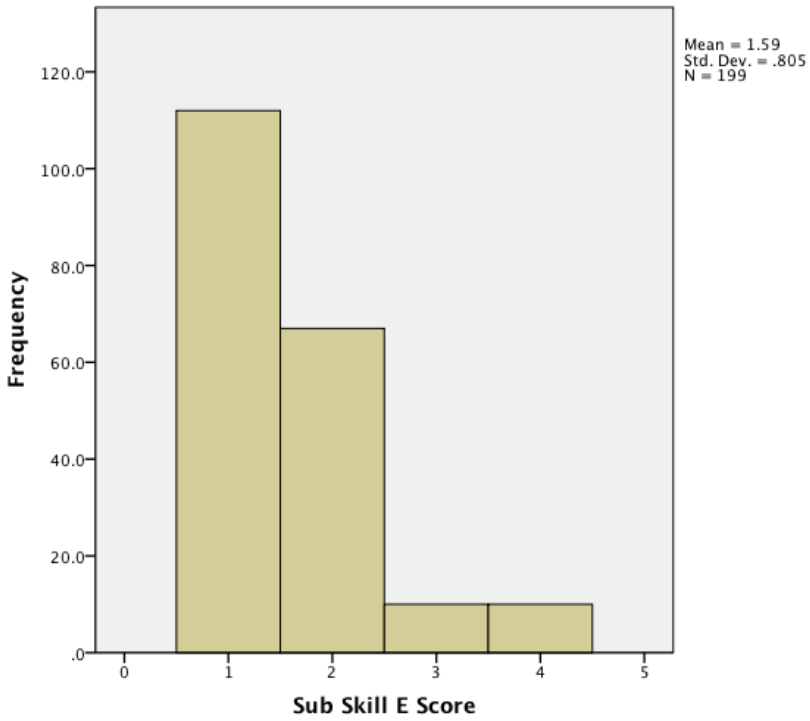


Figure A7: Histogram, Skill F Scores

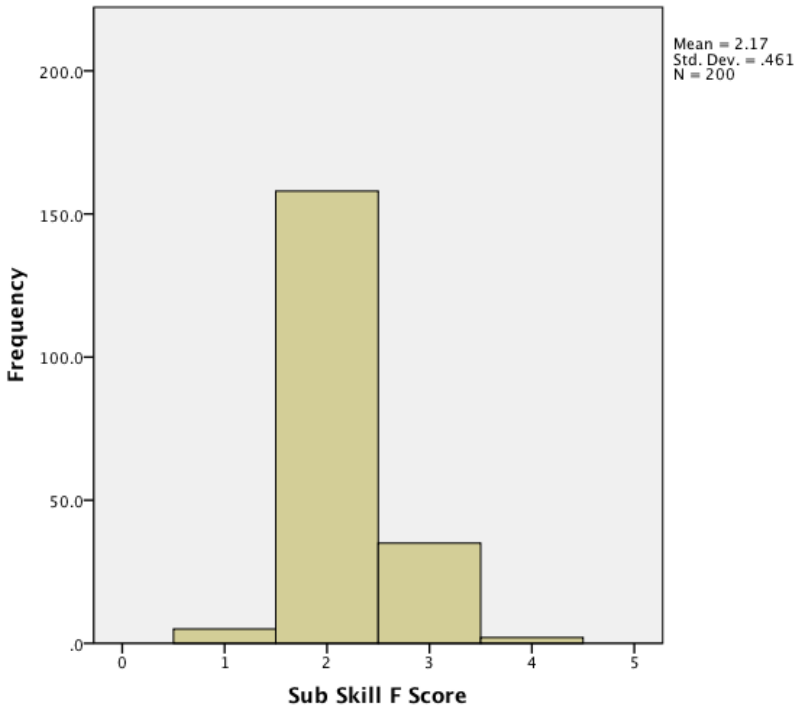
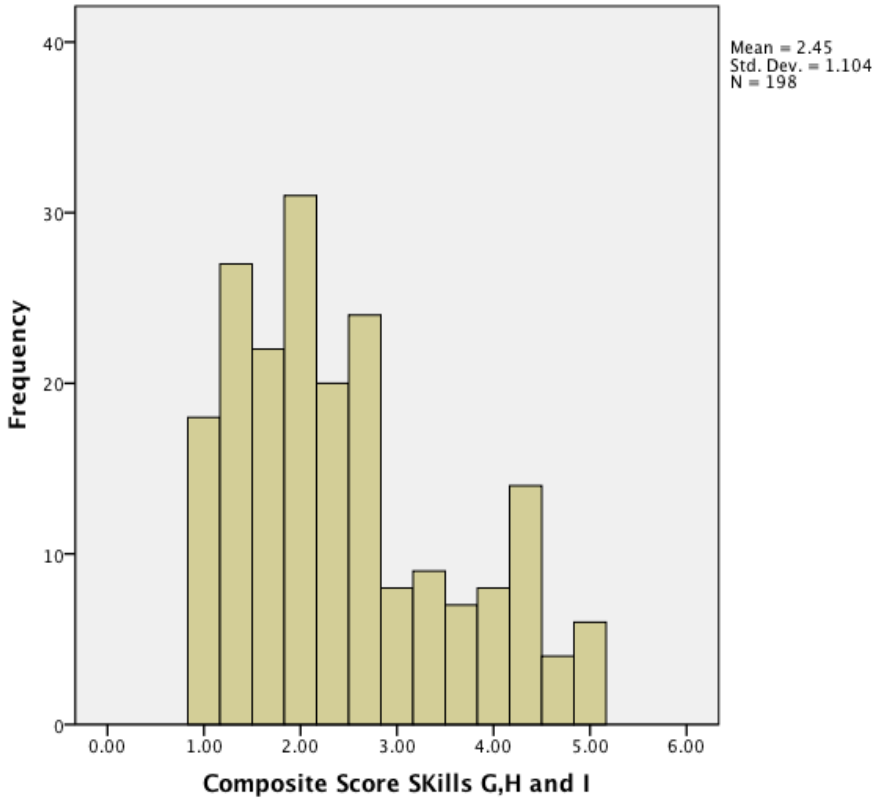


Figure A8: Histogram, Composite Skill Scores



D. Evaluation of Assumptions (MANOVA, ANOVA, T-Tests, Chi-square Tests)

1) Multiple Analysis of Variance (MANOVA) and Analysis of Variance (ANOVA)

- Task Comparison
  - Sufficient sample size
    - n=199
  - Missing Data
    - No missing data for this field
  - Sufficient cell size
    - More cases in each cell than variables (Task 1 = 101, Task 2 = 98)
  - Homogeneity of Variance
    - Box's M not relevant, as group sizes essentially equal (Tolmie, Muijs & McAteer, 2011)
    - Levene's test significant for Skills A (p<.001), F (p=.033), G (p=.018) and H (p<.001)
      - T-tests used to correct for homogeneity of variance

- Effect of Time Taken on Assessment
  - Sufficient sample size
    - n=195

    - Missing Data
        - Missing data for three cases, removed pairwise for analysis
        - One case with missing assessment scores (discussed in Section 6.3) also removed for MANOVA
    - Sufficient cell size
        - More cases in each cell than variables (Fewer than 60 minutes = 27, 60-89 minutes = 91, 90-119 minutes = 62, 120 minutes or More = 15)
    - Homogeneity of Variance
        - Box's test non-significant (p=.032)
        - Levene's test significant for Skills A (p=.007) and F (p=.003)

- Effect of Language of Response
    - Sufficient sample size
        - n=198
    - Missing Data
        - One case with missing assessment scores removed for MANOVA
    - Sufficient cell size
        - More cases in each cell than variables (English = 85, French = 84, Kinyarwanda/combination=29)
    - Homogeneity of Variance
        - Box's test *significant* for this analysis (p=.007), indicating a violation of the assumption
        - Levene's test significant for Skills A (p<.001), G (p=.011) and H (p=.010)

- Institutional Effects
    - Sufficient sample size
        - n=198
    - Missing Data
        - One case with missing assessment scores removed for MANOVA
    - Sufficient cell size
        - More cases in each cell than variables (NUR=93, KIST= 76, SFB=29)
    - Homogeneity of Variance
        - Box's test non-significant (p=.043)
        - Levene's test significant for Skill C (p=.048) and Skill F (p=.003)

- Gender
    - Sufficient sample size
        - n=198 for MANOVA
        - n=199 for *t*-tests
    - Missing Data
        - One case with missing assessment scores removed for MANOVA
    - Sufficient cell size
        - More cases in each cell than variables (M=157, F=41)
    - Homogeneity of Variance
        - Box's test non-significant (p=.195).
        - Levene's test significant for Skill A (p<.001) and marginal for Skill F (p=.051)
            - T-tests used to correct for homogeneity of variance

- Socio-economic Status
  - Sufficient sample size
    - n=182 for MANOVA
    - n=183 for ANOVAs
  - Missing Data
    - Missing data for 16 of the 199 cases with valid assessment scores
    - One case with missing assessment scores removed for MANOVA
    - Analysis run with missing data included and with missing data removed pairwise, and solutions remained the same
  - Sufficient cell size
    - More cases in each cell than variables (Lowest Quintile=86, Middle=21, Second-to-Highest=36, Highest=39)
  - Homogeneity of Variance
    - Box's test *significant* for this analysis (p<.001), indicating a violation of the assumption
    - Levene's test significant for Skill A (p=.003) and Composite Skill (p=.006)

- Parental Education Level
  - Sufficient sample size
    - n=192 for MANOVA
    - n=193 for ANOVAs
  - Missing Data
    - Missing data for six cases, removed pairwise for analysis
    - One case with missing assessment scores removed for MANOVA
  - Sufficient cell size
    - More cases in each cell than variables (No Adults/No Education=61, Some Primary=51, Some Secondary=45, Some Tertiary=35)
  - Homogeneity of Variance
    - Box's test non-significant (p=.156)
    - Levene's test significant for Composite Skill (p=.023)

- Secondary School Type
  - Sufficient sample size
    - n=198 for MANOVA
    - n=199 for ANOVAs
  - Missing Data
    - One case with missing assessment scores removed for MANOVA
  - Sufficient cell size
    - More cases in each cell than variables (Urban Public=45, Urban Private=45, Rural Public=56, Rural Private=52)
  - Homogeneity of Variance
    - Box's test non-significant (p=.567)
    - Levene's test not significant for any skill

- Academic Field
  - Sufficient sample size
    - n=198 for MANOVA
    - n=199 for *t*-tests
  - Missing Data
    - One case with missing assessment scores removed for MANOVA

    - Sufficient cell size
        - More cases in each cell than variables (Sciences=120, Social Sciences=78)
    - Homogeneity of Variance
        - Box's test non-significant (*p*=.025)
        - Levene's test significant for Skill C (*p*=.026)
            - T-tests used to correct for homogeneity of variance

- Year at University
    - Sufficient sample size
        - n=198 for MANOVA
        - n=199 for *t*-tests
    - Missing Data
        - One case with missing assessment scores removed for MANOVA
    - Sufficient cell size
        - More cases in each cell than variables (First Year=112, Fourth Year=86)
    - Homogeneity of Variance
        - Box's test non-significant (*p*=.844)
        - Levene's test significant for Skill F (*p*=.010)
            - T-tests used to correct for homogeneity of variance

2) *Independent Samples T-tests*

- Normal Distribution
    - Although the individual score distributions all demonstrate significant skewness and kurtosis, it can be assumed that the distribution in the overall population is normal, given that the sample size is significantly larger than the 30 cases required for the central limit theorem (Tolmie, Muijs & McAteer, 2011).
- Homogeneity of Variance
    - As outlined for each t-test in the relevant table of results. If Levene's test was violated, results for the corrected test are quoted.

3) Chi-square Tests

- Expected Frequencies >5 cases in each cell
- For Little's MCAR test on 0 responses, smallest cell size was 22
- For test of association between Time Taken on Assessment and Year at University, smallest cell size was 4
- For test of association between Language of Response and Year at University, smallest cell size was 8

E. Comparison of means and standard deviations, as a percentage of overall scale

Table A6: Means and Standard Deviations, in absolute value and as a percentage of the overall scale, Gender (n=199)

Skill	Standard Deviation, **Females** (n=42)	Standard Deviation, **Males** (n=157)	Difference in Means
A: Bias	.513 (10.26%)	.819 (16.38%)	0.196 (3.92%) (Female mean=2.93; Male mean=2.73)
B: Relevance	1.215 (24.3%)	1.227 (24.54%)	0.445 (8.9%) (Female mean=2.71; Male mean=3.16)
C: Credibility	.627 (12.54%)	.614 (12.28%)	0.149 (2.98%) (Female mean=2.6; Male mean=2.45)
D: Errors	.683 (13.66%)	.801 (16.02%)	0.009 (0.18%) (Female mean=2.14; Male mean=2.13)
E: Generalisability	.697 (13.94%)	.833 (16.66%)	0.039 (0.78%) (Female mean=1.62; Male mean=1.58)
F: Missing Information	.439 (8.78%)	.462 (9.24%)	0.155 (3.1%) (Female mean=2.05; Male mean=2.20)
Composite: Evaluation	.982 (19.64%)	1.116 (22.32%)	0.47035 (9.4%) (Female mean=2.07; Male mean=2.54)

Table A7: Means and Standard Deviations, in absolute value and as a percentage of the overall scale, Secondary School (n=199)

Skill	Standard Deviation, Urban Pub. (n=45)	Standard Deviation, Urban Private (n=46)	Standard Deviation, Rural Public (n=56)	Standard Deviation, Rural Private (n=52)	Largest Difference in Means
A: Bias	.650 (13%)	.853 (17.1%)	.773 (15.5%)	.785 (15.7%)	0.2 (4%) (Mean Rural Private = 2.83; Mean Urban Private =2.63)
B: Relevance	1.283 (25. 7%)	1.347 (26.9%)	1.168 (23.4%)	1.195 (23.9%)	0.09 (1.8%) (Mean Urban Public =3.11; Mean Rural Public =3.02)
C: Credibility	.583 (11.7%)	.657 (13.1%)	.630 (12.6%)	.609 (12.2%)	0.13 (2.6%) (Mean Rural Public = 2.55; Mean Urban Public = 2.42)
D: Errors	.714 (14.3%)	.705 (14.1%)	.869 (17.4%)	.791 (15.8%)	0.2 (4%) (Mean Urban Private = 2.24; Mean Rural Private = 2.04)
E: Generalisability	.757 (15.1%)	.834 (16.7%)	.759 (15.2%)	.871 (17.4%)	0.22 (4.4%) (Mean Rural Private = 1.71; Mean Urban Public = 1.49)
F: Missing Information	.387 (7.7%)	.416 (8.3%)	.508 (10.2%)	.513 (10.3%)	0.03 (0.6%) (Mean Rural Public = 2.18; Mean Urban Private = 2.15)
Composite: Evaluation	1.129 (22.6%)	1.084 (21.7%)	1.072 (21.4%)	1.163 (23.3%)	0.057 (1.1%) (Mean Urban Public = 2.474; Mean Rural Public = 2.417)

Table A8: Means and Standard Deviations, in absolute value and as a percentage of the overall scale, Parental Education (n=193)

Skill	Standard Deviation, Some Primary (n=52)	Standard Deviation, Some Secondary (n=45)	Standard Deviation, Some Tertiary (n=35)	Standard Deviation, No Education or No Adults (n=61)	Largest Difference in Means
A: Bias	.848 (17%)	.806 (16.1%)	.631 (12.6%)	.751 (15%)	0.18 (3.6%) (Mean Tertiary = 2.89; Mean Primary = 2.71)
B: Relevance	1.237 (24.7%)	1.242 (24.8%)	1.380 (27.6%)	1.152 (23%)	0.29 (5.8%) (Mean Secondary = 3.16; Mean Primary = 2.87)
C: Credibility	.576 (11.5%)	.548 (11%)	.710 (14.2%)	.644 (12.9%)	0.28 (5.6%) (Mean None = 2.57; Mean Tertiary = 2.29)
D: Errors	.760 (15.2%)	.804 (16.1%)	.874 (17.5%)	.726 (14.5%)	0.2 (4%) (Mean None = 2.2; Mean Tertiary = 2)
E: Generalisability	.841 (16.8%)	.837 (16.7%)	.780 (15.6%)	.797 (15.9%)	0.18 (3.6%) (Mean None=1.64; Mean Tertiary = 1.46)
F: Missing Information	.460 (9.2%)	.505 (10.1%)	.406 (8.1%)	.451 (9%)	0.05 (1%) (Mean Secondary & Tertiary = 2.2; Mean Primary = 2.15)
Composite: Evaluation	1.022 (20.4%)	1.227 (24.5%)	1.189 (23.8%)	.961 (19.2%)	0.726 (14.5%) (Mean Tertiary = 2.857; Mean None = 2.131)

Table A9: Means and Standard Deviations, in absolute value and as a percentage of the overall scale, Economic Background (n=183)

Skill	Standard Deviation, Lowest Quintile (n=86)	Standard Deviation, Middle Quintile (n=21)	Standard Deviation, Second-to-Highest Quintile (n=37)	Standard Deviation, Highest Quintile (n=39)	Largest Difference in Means
A: Bias	.7604 (15.2%)	.9035 (18.1%)	.5465 (10.9%)	.7334 (14.7%)	0.45 (9%) (Mean Highest=3.02; Mean Middle=2.57)
B: Relevance	1.1428 (22.9%)	1.2778 (25.6%)	1.3017 (26.0%)	1.2514 (25.0%)	0.44 (8.8%) (Mean Middle=3.29; Mean Highest= 2.85)
C: Credibility	.5434 (10.9%)	.5754 (11.5%)	.6866 (13.7%)	.6975 (13.9%)	0.17 (3.4%) (Mean Lowest=2.53; Mean Highest=2.36)
D: Errors	.7236 (14.5%)	.6529 (13.1%)	.8971 (17.9%)	.7828 (15.7%)	0.31 (6.2%) (Mean Highest=2.28; Mean Second-to-Highest=1.97)
E: Generalisability	.7818 (15.6%)	1.0648 (21.3%)	.6872 (13.7%)	.8317 (16.6%)	0.26 (5.2%) (Mean Middle=1.76; Mean Second-to-Highest=1.5)
F: Missing Information	.4869 (9.7%)	.4259 (8.5%)	.4410 (8.8%)	.4213 (8.4%)	0.14 (2.8%) (Mean Highest=2.23; Mean Middle=2.09)
Composite: Evaluation	.9422 (18.8%)	1.1746 (23.5%)	1.0395 (20.8%)	1.2575 (25.2%)	0.43 (8.6%) (Mean Second-to-Highest=2.65; Mean Lowest=2.22)

Table A10: Means and Standard Deviations, in absolute value and as a percentage of the overall scale, Academic Field (n=199)

Skill	Standard Deviation, Sciences (n=120)	Standard Deviation, Social Sciences (n=79)	Difference in Means
A: Bias	.822 (16.4%)	.681 (13.6%)	0.06 (1.2%) (Mean Social Science = 2.81; Mean Science = 2.75)
B: Relevance	1.207 (24.14%)	1.222 (24.44%)	0.507 (10.2%) (Mean Science = 3.27; Mean Social Science =2.76)
C: Credibility	.562 (11.2%)	.695 (13.9%)	0.09 (1.8%) (Mean Social Science = 2.53; Mean Science = 2.44)
D: Errors	.811 (16.2%)	.715 (14.3%)	0.153 (3.1%) (Mean Social Science = 2.23; Mean Science = 2.08)
E: Generalisability	.828 (16.6%)	.771 (15.4%)	0.075 (1.5%) (Mean Social Science = 1.63; Mean Science = 1.56)
F: Missing Information	.477 (9.5%)	.435 (8.7%)	0.054 (1.1%) (Mean Social Science = 2.2; Mean Science = 2.15)
Composite: Evaluation	1.089 (21.78%)	1.087 (21.74%)	0.41175 (8.23%) (Mean Science =2 .61; Mean Social Science =2.197)

Table A11: Means and Standard Deviations, in absolute value and as a percentage of the overall scale, Year at University (n=199)

Skill	Standard Deviation, First Years (n=112)	Standard Deviation, Fourth Years (n=87)	Difference in Means
A: Bias	.725 (14.5%)	.824 (16.5%)	0.048 (1%) (Mean First Years = 2.79; Mean Fourth Years = 2.75)
B: Relevance	1.234 (24.7%)	1.241 (24.8%)	0.129 (2.6%) (Mean Fourth Years = 3.14; Mean First Years = 3.01)
C: Credibility	.600 (12%)	.644 (12.9%)	0.011 (2.2%) (Mean First Years = 2.48; Mean Fourth Years = 2.47)
D: Errors	.780 (15.6%)	.776 (15.5%)	0.045 (0.9%) (Mean Fourth Years = 2.16; Mean First Years = 2.12)
E: Generalisability	.804 (16.1%)	.805 (16.1%)	0.126 (2.5%) (Mean First Years = 1.64; Mean Fourth Years = 1.52)
F: Missing Information	.470 (9.4%)	.437 (8.7%)	0.147 (2.9%) (Mean Fourth Years = 2.25; Mean First Years = 2.11)
Composite: Evaluation	1.0853 (21.7%)	1.1040 (22.1%)	0.3419 (6.8%) (Mean Fourth Years = 2.639; Mean First Years = 2.298)

Table A12: Table of Means, Background Characteristics and Field of Study (n=199)

	Mean (male)	**Mean** (fem)	**Mean** (high SES)	**Mean** (mid-to-high SES)	**Mean** (mid SES)	**Mean** (low SES)	**Mean** (urban public school)	**Mean** (urban private school)	**Mean** (rural public school)	**Mean** (rural private school)	**Mean** (parent ed tertiary)	**Mean** (parent ed sec)	**Mean** (parent ed prim)	**Mean** (no parent ed)	**Mean** (Sci)	**Mean** (Soc Sci)
A: Bias	2.73	2.93	3.02	2.92	2.57	2.71	2.82	2.63	2.80	2.83	2.89	2.82	2.71	2.74	2.75	2.81
B: Relevance	3.16	2.71	2.85	3.17	3.29	3.14	3.11	3.09	3.02	3.06	3.09	3.16	2.87	3.15	3.27	2.76
C: Credibility	2.45	2.6	2.36	2.47	2.38	2.53	2.42	2.46	2.55	2.46	2.29	2.53	2.46	2.57	2.44	2.53
D: Errors	2.13	2.14	2.28	1.97	2.05	2.19	2.11	2.24	2.16	2.04	2.00	2.11	2.17	2.20	2.08	2.23
E: Generalisability	1.58	1.62	1.64	1.50	1.76	1.61	1.49	1.57	1.57	1.71	1.46	1.60	1.63	1.64	1.56	1.63
F: Missing Information	2.20	2.05	2.23	2.17	2.09	2.17	2.18	2.15	2.18	2.17	2.20	2.20	2.15	2.16	2.15	2.2
Composite: Evaluation	2.54	2.07	2.61	2.65	2.55	2.22	2.47	2.43	2.42	2.47	2.86	2.59	2.40	2.13	2.61	2.20

Table A13: Table of Means, Year at University (n=199)

	Mean (First Years)	**Mean** (Fourth Years)
A: Bias	2.79	2.75
B: Relevance	3.01	3.14
C: Credibility	2.48	2.47
D: Errors	2.12	2.16
E: Generalisability	1.64	1.52
F: Missing Information	2.11	2.25
Composite: Evaluation	2.30	2.64

Appendix J: Assessment Results for KIST and NUR

Table A14: Distribution of Assessment Scores, KIST (n=76)

Skill	Mean	Standard Error of Mean	Standard Deviation	Range
A: Bias	2.74	.089	.772	Low 1; High 5
B: Relevance	3.18	.137	1.197	Low 1; High 5
C: Credibility	2.50	.066	.577	Low 1; High 4
D: Errors	2.00	.092	.800	Low 1; High 4
E: Generalisability	1.58	.092	.804	Low 1; High 4
F: Missing Information	2.09	.047	.406	Low 1; High 4
Composite: Evaluation	2.57	.121	1.052	Low 1; High 5

Table A15: Distribution of Assessment Scores, NUR (n=94)

Skill	Mean	Standard Error of Mean	Standard Deviation	Range
A: Bias	2.79	.081	.788	Low 1; High 5
B: Relevance	3.19	.128	1.238	Low 1; High 5
C: Credibility	2.45	.062	.598	Low 1; High 4
D: Errors	2.21	.080	.774	Low 1; High 4
E: Generalisability	1.57	.088	.849	Low 1; High 4
F: Missing Information	2.22	.050	.488	Low 1; High 3
Composite: Evaluation	2.47	.118	1.141	Low 1; High 5

Table A16: Analysis of Differences in Critical Thinking Ability by Year at University, KIST (n=76)

Skill	Year at University	Mean	Standard Deviation	Standard Error of Mean	Equal variances assumed? (based on Levene's test for Equality of Variances)	Results of *t*-test	Significance (*p*-value)
A: Bias	KIST First (n=42)	2.83	.537	.083	No	$t(48.54) = 1.146$	$p=.257$
	KIST Fourth (n=34)	2.62	.985	.169			
B: Relevance	KIST First (n=42)	3.24	1.122	.173	Yes	$t(74) = .434$	$p=.666$
	KIST Fourth (n=34)	3.12	1.297	.222			
C: Credibility	KIST First (n=42)	2.55	.550	.085	Yes	$t(74) = .797$	$p=.428$
	KIST Fourth (n=34)	2.44	.613	.105			
D: Errors	KIST First (n=42)	1.93	.838	.129	Yes	$t(74) = -.864$	$p=.391$
	KIST Fourth (n=34)	2.09	.753	.129			
E: Generalisability	KIST First (n=42)	1.55	.803	.124	Yes	$t(74) = -.375$	$p=.709$
	KIST Fourth (n=34)	1.62	.817	.140			
F: Missing Information	KIST First (n=42)	2.02	.412	.064	Yes	$t(74) = -1.649$	$p=.103$
	KIST Fourth (n=34)	2.18	.387	.066			
Composite: Evaluation	KIST First (n=42)	2.476	1.065	.164	Yes	$t(74) = -.905$	$p=.368$
	KIST Fourth (n=34)	2.696	1.039	.178			

** Significant at 5% level of significance*

Table A17: Analysis of Differences in Critical Thinking Ability by Year at University, NUR (n=94)

Skill	Year at University	Mean	Standard Deviation	Standard Error of Mean	Equal variances assumed? (based on Levene's test for Equality of Variances)	Results of *t*-test	Significance (*p*-value)
A: Bias	NUR First (n=51)	2.75	.868	.122	Yes	$t(92) = -.563$	$p=.575$
	NUR Fourth (n=43)	2.84	.688	.105			
B: Relevance	NUR First (n=51)	3.14	1.281	.179	Yes	$t(92) = -.461$	$p=.646$
	NUR Fourth (n=43)	3.26	1.197	.183			
C: Credibility	NUR First (n=51)	2.43	.575	.080	Yes	$t(92) = -.271$	$p=.787$
	NUR Fourth (n=43)	2.47	.631	.096			
D: Errors	NUR First (n=51)	2.20	.749	.105	Yes	$t(92) = -.226$	$p=.821$
	NUR Fourth (n=43)	2.23	.812	.124			
E: Generalisability	NUR First (n=51)	1.63	.848	.119	Yes	$t(92) = .657$	$p=.513$
	NUR Fourth (n=43)	1.51	.856	.130			
F: Missing Information	NUR First (n=52)	2.15	.500	.069	Yes	$t(93) = -1.487$	$p=.141$
	NUR Fourth (n=43)	2.30	.465	.071			
Composite: Evaluation	NUR First (n=51)	2.366	1.090	.153	Yes	$t(91) = -.997$	$p=.321$
	NUR Fourth (n=42)	2.603	1.201	.185			

** Significant at 5% level of significance*

Appendix K: Case Study Protocol[133]

A. Case Study Framework and Study Objectives

1. Case Study Objectives

The case study phase is guided by two overarching objectives.

First, findings from the case studies are intended to clarify some of the results obtained during the assessment phase. Specifically, investigation will focus on the following two questions:

1) Why do Rwandan students appear to differ in their ability to demonstrate the individual critical thinking skills included in the assessment?
2) Why does there appear to be no systematic relationship between student background characteristics and demonstrated critical thinking ability in Rwanda?

Second, the case study phase is intended to complement the results from the assessment phase by considering the potential role of additional inputs and institutional experiences not considered during the quantitative component of the study. Data collection will focus on the final research question governing the study: How are Rwandan universities helping (or hindering) the development of critical thinking skills in their students?

These questions will be considered within the context of two institutions – the National University of Rwanda and the Kigali Institute of Science & Technology. Analysis will be at the Faculty level, so six Faculties have been selected for investigation: the three Faculties at KIST, plus the Faculty of Science, the Faculty of Applied Sciences, and the Faculty of Economics & Management at NUR. Data collection will be completed at each institution independently, resulting in a final report for each institution. Cross-case comparison will then seek to illuminate similarities and differences between the six Faculties.

2. Conceptual framework

The conceptual framework underlying the study suggests that improvement in critical thinking ability during university depends on a number of factors, including inputs (i.e. student characteristics), elements of the university experience, and, potentially, elements of the student experience outside of the university context.

Past studies have demonstrated a link between the following variables and improvement in critical thinking ability:

- Certain student characteristics, including:
  - Demographics, particularly socio-economic status
  - Family background, particularly parental education level
  - Prior schooling experience
  - Incoming critical thinking ability
  - Motivations, attitudes and dispositions
- Aspects of the university experience that appear to influence critical thinking ability (particularly in terms of their impact on student engagement)
  - Progressive Curricula
  - Appropriate level of academic challenge in coursework
  - Pedagogy that emphasizes active and collaborative learning

[133] Format as suggested by Yin (2003).

- Enriching educational experiences
- Positive/enriching interactions with faculty
- Positive/enriching interactions with peers
- Enriching extracurricular activities

Although not yet well investigated empirically, there is also theoretical support for the idea that external influences, such as family pressures or work experiences, have a significant impact (both positive and negative) on student engagement and, therefore, on learning outcomes.

3. Case Study Questions

Results from the quantitative phase suggest that the case study component should be governed by the following questions:

- Why does student ability appear to differ on different critical thinking skills? Are some skills more incorporated than others into the Rwandan university experience?
- Why do the results suggest that there is no systematic relationship between demonstrated critical thinking ability and student characteristics, such as socio-economic status, parental education level and secondary school type?
- Are there other entering student characteristics that may impact demonstrated critical thinking ability?
- Why does it appear that Rwandan students do not significantly improve their critical thinking ability during university?
    - Is there a lack of appropriate "stimuli", in terms of pedagogy or course content?
    - Or is there a lack of student readiness, in terms of their level of engagement with the learning process?
        - If so, why?
        - If not, what other reasons could explain the lack of improvement?

B. Data Collection Procedures

1. Data collection plan

I intend to address these overarching questions through a series of interviews with stakeholders. Specifically, I aim to complete the following at each institution:

a. Focus groups with fourth-year students: intended to gain an understanding of student attitudes towards university and a sense of the university experience, in terms of course expectations, pedagogy and the university environment (target: four focus groups of 5-10 fourth-year students from differing academic backgrounds)

b. Semi-structured interviews with "positive outlier" students (i.e. those who did particularly well on the critical thinking assessment): intended to gain a deeper understanding of their individual experiences prior to and during university, in order to shed some light on potential factors that may influence demonstrated critical thinking ability (target: at least 5 interviews per institution)

c. Semi-structured interviews with faculty members: intended to gain an understanding of faculty attitudes and experiences with teaching and to understand faculty perceptions of critical thinking and its role in the university

curriculum (target: at least 10 interviews per institution, ideally from different backgrounds and representing all six Faculties)

d. Semi-structured interviews with university administrators: intended to gain an understanding of the university's official perspective on teaching and learning – and on the value of inculcating critical thinking skills as a part of the university curriculum (target: 3-5 interviews per case)

e. Review of official documents pertaining to teaching and learning on the campus: intended to shed light on official institutional policies regarding teaching methods, assessment and learning standards

It is anticipated that data collection at each institution will take one month to complete. After two weeks of data collection, an Interim Summary will be drafted and reviewed with members of the research team for initial feedback. The final report for each institution will be validated with research assistants who attend (or attended) the institution.

2. Data collection logistics

Focus groups will be facilitated by the lead researcher in English, along with two research assistants who will help with translation into Kinyarwanda, logistics and note taking. During the quantitative phase, participants were asked to note if they were willing to attend a follow-up interview. Fourth-year participants will be sampled purposively from this list of willing interviewees, in order to ensure a wide representation of academic backgrounds. Focus groups will be held in a confidential space on the university campus.

Interviews with students will be facilitated by the lead researcher in English or French, along with one research assistant to help with translation into Kinyarwanda if necessary. Positive outliers will be identified from the quantitative data set and contacted for participation, provided they agreed to attend a follow-up interview.

Interviews with faculty members and administrators will be facilitated by the lead researcher in English or French. Potential participants will be selected using a random purposive sampling technique.

All focus groups and interviews will be recorded and transcribed for analysis. Transcriptions will be shared with participants via email. If any portion of any interview is conducted in Kinyarwanda, the research team will translate the content using back-translation techniques.

3. Specific data collection tools

Focus group questions will use questions from the National Survey of Student Engagement (NSSE). Interview questions will reference focus group data and quantitative results when appropriate.

Complete interview schedules are attached to this protocol. All four schedules will be piloted prior to implementation.

C. Outline of Final Institutional Report

The intention is to organise the final report from each institution thematically, covering the following topics:

1. Connections between Institutional Data and the Conceptual Framework
2. Student Outcomes
    a. Emerging explanations for differences between scores on individual critical thinking skills
3. Student Backgrounds
    a. Emerging explanations for apparent lack of connection between background characteristics and critical thinking ability
    b. Relevance of additional "inputs", such as attitudes and motivations?
    c. Other important inputs in the Rwandan context?
4. Academic Experiences during University
    a. Are universities following "effective educational practices"? Why or why not?
5. Other Experiences during University: Extracurricular and External
    a. Other important factors in the Rwandan context?
6. Critical Thinking on Campus
    a. Stakeholder understandings of critical thinking
    b. The value of critical thinking in the university curriculum
    c. Lecturer preparation for teaching critical thinking
7. Conclusions

Appendix L1: Focus Group Guide

Preparing the Focus Group:

- Set up chairs in a circle
- Set up a welcome desk with pens and list of participants to tick off
- Prepare flip chart, handouts/dots, guidelines
- Set up recorder
- As participants come in, check them off and distribute pens

Welcome Procedures (5 minutes)

- Once everybody has arrived, introductions
- Remind participants about the overall study
    - What are the aims of the study?
    - What happens with the information?
    - Who is involved?
    - Basic procedures
- Will take between 1.5 and 2 hours with a break
- Will be audio-recorded
    - Confidentiality (remind them about consent forms)
- Ask if they have any questions
- Review Focus Group Ground rules
- Sign Informed Consent forms
- Turn on the recorder & begin

[Start by reminding participants about the purpose of the research and how we're hoping to work together to learn more about their experiences at university]

1. **Purpose of a University Education (10 minutes)**

[We're going to start off with a brainstorming activity. We're going to record your answers on the board.]

- What are some of the reasons why people decide to go to university in Rwanda?

(Listing activity)

- I'm going to give you three small dots each. Have a look at this list, and then I'd like you to come up and "vote" for the three reasons that were the most important for you personally. Which three reasons best reflect why *you* decided to attend university?

(Recap)

2. **Experiences at University (30 minutes)**[134]

[*Ok, thank you. That is very helpful. So, now that we know why you decided to come to university, we'd like to spend the rest of our time together talking about your experiences at university. I have a series of questions here. We're going to give you some small pieces of paper. Each piece of paper has the possible responses to my*

[134] The questions in Section 2 come from the NSSE

questions. For each question, I'd like you to record your answer. At the end, we will collect your responses and tally them during the break before discussing them all together. That way, your responses can remain anonymous, but we'll have a chance to discuss them as a group. Remember, there are NO WRONG ANSWERS. My research will be meaningless if you write down what you think you should say. Write down what you feel was your actual experience at university! To start, please fill in the name of your academic department on the top of the sheet. Do not write your name!]

2.1: During your time at KIST/NUR, *how often* have you done each of the following?
(Options: Often, Sometimes, Never)

1. Asked questions in class or contributed to a class discussion?
2. Gave a presentation in class?
3. Prepared more than one draft of an assignment before submitting it?
4. Worked on an assignment that required integrating ideas, information or perspectives from multiple sources?
5. Worked with other students on a project during or outside of class?
6. Went on a field trip as part of an academic course?
7. Approached an instructor outside of class to discuss your grades or assignments?
8. Approached an instructor outside of class to discuss the content of your class?
9. Approached an instructor outside of class to discuss your career plans?
10. Received prompt feedback (either written or oral) from an instructor on your academic performance?
11. Were surprised by how hard you worked or how much you achieved because of an instructor's encouragement?
12. Worked with faculty on activities other than coursework?
13. Discussed ideas from reading or classes with people outside of class?
14. Had serious conversations with students from very different backgrounds from your own?
15. Completed an internship for academic credit?
16. Did your own research for academic credit?

2.2: In an average year, about how much reading did you do?
(Options: 0, 1-4, 5-10, 11-20, 20+)
1. Approximately how many books or academic materials did you read each year for class?
2. Approximately how many books did you read each year for your own enjoyment or enrichment?

2.3: In an average year, approximately how many assignments did you do for each class?
1. Examinations or continuous assessment tests
2. Written papers or projects
3. Oral presentations

2.4: In a typical year, how often did you do the following activities?
(Options: Often, Sometimes, Never)

1. Attended an artistic performance
2. Exercised or played organized sport
3. Attending faith activities/services
4. Participated in community service or volunteer work

2.5: During your time at NUR/KIST, how often have you:
(Options: Often, Sometimes, Never)

1. Thought about your own views on a topic or issue
2. Tried to understand somebody else's perspective on a topic or issue
3. Learned something that changed the way you saw or understood a topic or issue

2.6 In a typical day, approximately how many hours did you spend doing the following?
(Options: 0, 0-1, 1-2, 2-3, 3-5, 5-7, 7+)

1. Preparing for class
2. Attending class
3. Working
4. Participating in extracurricular activities
5. Relaxing or socializing
6. Caring for your family or other dependents
7. Commuting to class

2.7 How much time and energy do you think the university staff and faculty spend on the following? *(Options: A lot of time and energy, Some time and energy, No time or energy)*

1. Encouraging students to spend a lot of time studying/on academic work
2. Providing students with academic support to help them succeed
3. Fostering contact between students from different backgrounds
4. Helping students to cope with non-academic responsibilities
5. Providing support to help students succeed socially
6. Organizing campus events and activities

2.8 How has your time at university contributed to your ability in the following areas?
(Options: A lot, Somewhat, Has not contributed)

1. Broad knowledge about the world
2. Knowledge and skills related to your field
3. Writing effectively
4. Speaking effectively
5. Working with others
6. Learning effectively on your own
7. Solving complex real-world problems

3. **BREAK with refreshments** **(10 minutes)**

4. Group Discussion (45-60 minutes)[135]

4.1: I want to start by looking at your responses to the questions about your classwork and academic experiences. *(Review tallies)* Are you surprised by any of these results? Do they look right for your institution?

- Possible follow-ups: Does everybody participate in group discussions in class? Why or why not?
- How do group assignments work? Do students like them? Why or why not?
- Do students feel that they can disagree with faculty members during class? Why do students feel that they can or cannot approach their instructors outside of class?

4.2: Now, let's talk about your assignments. What kinds of assignments did you have in a given year?

- For exams: What types of questions did you have to answer?
- What did you feel was the best way to prepare for your examinations?
- For written assignments or presentations: How long did they have to be?
- Did you feel that your assignments had a purpose? Could you understand why your instructors asked you to complete the assignments that they assigned?
- What happened if you failed an assignment?

4.3: I know that all of you are required to complete a final project. What did you think about your final project? Did you find it useful? Why or why not?

4.4: (For NUR only) I'd like to learn a little bit more about the new modular system at NUR. How did you think it worked? What are your opinions about the system?

4.5: If we look at your responses to the last question on the questionnaire, I can see that you thought your time at NUR/KIST had helped you to improve in the following abilities. Which of these abilities do you think it is important for a university graduate to have?

- Are there any areas where you wish you had improved but did not? Why do you think you did not?

5. Final Thoughts

Do you have anything that you would like to add or share with the group before we finish? Do you have any questions for me?

Wrap-Up Procedures

- Turn off the recorder
- Thank participants & remind them what will happen with information

[135] There will not be time to cover all of these questions. Use the tallied responses from the survey to guide which topics should be emphasised with particular groups.

Appendix L2: Focus Group Handouts

Name of your academic department: _____

1. During your time at university, *how often* have you done each of the following?

1. ☐ Often ☐ Sometimes ☐ Never
2. ☐ Often ☐ Sometimes ☐ Never
3. ☐ Often ☐ Sometimes ☐ Never
4. ☐ Often ☐ Sometimes ☐ Never
5. ☐ Often ☐ Sometimes ☐ Never
6. ☐ Often ☐ Sometimes ☐ Never
7. ☐ Often ☐ Sometimes ☐ Never
8. ☐ Often ☐ Sometimes ☐ Never
9. ☐ Often ☐ Sometimes ☐ Never
10. ☐ Often ☐ Sometimes ☐ Never
11. ☐ Often ☐ Sometimes ☐ Never
12. ☐ Often ☐ Sometimes ☐ Never
13. ☐ Often ☐ Sometimes ☐ Never
14. ☐ Often ☐ Sometimes ☐ Never
15. ☐ Often ☐ Sometimes ☐ Never
16. ☐ Often ☐ Sometimes ☐ Never

2. In an average year, about *how much* reading did you do?

1. ☐ 0 ☐ 1-4 ☐ 5-10 ☐ 11-14 ☐ 15-20 ☐ 20+

2. ☐ 0 ☐ 1-4 ☐ 5-10 ☐ 11-14 ☐ 15-20 ☐ 20+

3. Approximately how many assignments did you have to do for each class?

1. ☐ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ More than 4

2. ☐ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ More than 4

3. ☐ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ More than 4

4. In a typical year, *how often* did you do the following activities?

1. ☐ Often ☐ Sometimes ☐ Never

2. ☐ Often ☐ Sometimes ☐ Never

3. ☐ Often ☐ Sometimes ☐ Never

4. ☐ Often ☐ Sometimes ☐ Never

5. During your time at university, *how often* have you:

1. ☐ Often ☐ Sometimes ☐ Never

2. ☐ Often ☐ Sometimes ☐ Never

3. ☐ Often ☐ Sometimes ☐ Never

6. On a typical day, approximately *how many hours* did you spend on the following?

1. ☐ 0 ☐ 0-1 ☐ 1-2 ☐ 2-3 ☐ 3-5 ☐ 5-7 ☐ 7+

2. ☐ 0 ☐ 0-1 ☐ 1-2 ☐ 2-3 ☐ 3-5 ☐ 5-7 ☐ 7+

3. ☐ 0 ☐ 0-1 ☐ 1-2 ☐ 2-3 ☐ 3-5 ☐ 5-7 ☐ 7+

4. ☐ 0 ☐ 0-1 ☐ 1-2 ☐ 2-3 ☐ 3-5 ☐ 5-7 ☐ 7+

5. ☐ 0 ☐ 0-1 ☐ 1-2 ☐ 2-3 ☐ 3-5 ☐ 5-7 ☐ 7+

6. ☐ 0 ☐ 0-1 ☐ 1-2 ☐ 2-3 ☐ 3-5 ☐ 5-7 ☐ 7+

7. ☐ 0 ☐ 0-1 ☐ 1-2 ☐ 2-3 ☐ 3-5 ☐ 5-7 ☐ 7+

7. *How much time and energy* do you think the university staff and faculty spend on the following?

1. ☐ A Lot of Time and Energy ☐ Some Time and Energy ☐ No Time or Energy

2. ☐ A Lot of Time and Energy ☐ Some Time and Energy ☐ No Time or Energy

3. ☐ A Lot of Time and Energy ☐ Some Time and Energy ☐ No Time or Energy

4. ☐ A Lot of Time and Energy ☐ Some Time and Energy ☐ No Time or Energy

5. ☐ A Lot of Time and Energy ☐ Some Time and Energy ☐ No Time or Energy

6. ☐ A Lot of Time and Energy ☐ Some Time and Energy ☐ No Time or Energy

8. *How* has your time at university contributed to your ability in the following areas?

1. ☐ A Lot ☐ Somewhat ☐ Has Not Contributed

2. ☐ A Lot ☐ Somewhat ☐ Has Not Contributed

3. ☐ A Lot ☐ Somewhat ☐ Has Not Contributed

4. ☐ A Lot ☐ Somewhat ☐ Has Not Contributed

5. ☐ A Lot ☐ Somewhat ☐ Has Not Contributed

6. ☐ A Lot ☐ Somewhat ☐ Has Not Contributed

7. ☐ A Lot ☐ Somewhat ☐ Has Not Contributed

Appendix L3: Tallying Sheet for Focus Groups

Focus Group Code: _____

Question 1

	Often	Sometimes	Never
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			

Question 2

	0	1-4	5-10	11-14	15-20	20+
1						
2						

Question 3

	0	1	2	3	4	4+
1						
2						
3						

Question 4

	Often	Sometimes	Never
1			
2			
3			
4			

Question 5

	Often	Sometimes	Never
1			
2			
3			

Question 6

	0	0-1	1-2	2-3	3-5	5-7	7+
1							
2							
3							
4							
5							
6							
7							

Question 7

	A Lot	Some	Non
1			
2			
3			
4			
5			
6			

Question 8

	A Lot	Somewhat	Has Not
1			
2			
3			
4			
5			
6			
7			

Appendix L4: Questions from National Survey of Student Engagement (2012)

1. In your experience at your institution during the current school year, about how often have you done each of the following?
   (Options: Very Often, Often, Sometimes, Never)

   a. Asked questions in class or contributed to class discussions
   b. Made a class presentation
   c. Prepared two or more drafts of a paper or assignment before turning it in
   d. Worked on a paper or project that required integrating ideas or information from various sources
   e. Included diverse perspectives (different races, religions, genders, political beliefs, etc.) in class discussions or writing assignments
   f. Come to class without completing readings or assignments
   g. Worked with other students on projects during class
   h. Worked with classmates outside of class to prepare class assignments
   i. Put together ideas or concepts from different courses when completing assignments or during class discussions
   j. Tutored or taught other students (paid or voluntary)
   k. Participated in a community-based project (e.g. service learning) as part of a regular course
   l. Used an electronic medium (listserv, chat group, internet, instant messaging, etc.) to discuss or complete an assignment
   m. Used e-mail to communicate with an instructor
   n. Discussed grades or assignments with an instructor
   o. Talked about career plans with a faculty member or advisor
   p. Discussed ideas from your readings or classes with faculty members outside of class
   q. Received prompt written or oral feedback from faculty on your academic performance
   r. Worked harder than you thought you could to meet an instructor's standards or expectations
   s. Worked with faculty members on activities other than coursework (committees, orientation, student life activities, etc.)
   t. Discussed ideas from your readings or classes with others outside of class (students, family members, co-workers, etc)
   u. Had serious conversations with students of a different race or ethnicity than your own
   v. Had serious conversations with students who are very different from you in terms of their religious beliefs, political opinions, or personal values

2. During the current school year, how much has your coursework emphasised the following mental activities?
   (Options: Very much, Quite a bit, Some, Very little)

   a. Memorising facts, ideas or methods from your courses and readings, so that you can repeat them in pretty much the same form
   b. Analysing the basic elements of an idea, experience, or theory, such as examining a particular case or situation in depth and considering its components
   c. Synthesising and organising ideas, information, or experiences into new, more complex interpretations and relationships

d. Making judgments about the value of information, arguments, or methods, such as examining how others gathered and interpreted data and assessing the soundness of their conclusions
e. Applying theories or concepts to practical problems or in new situations

3. During the current school year, about how much reading and writing have you done?
(Options: None, 1-4, 5-10, 11-20, More than 20)

a. Number of assigned textbooks, books, or book-length packs of course readings
b. Number of books read on your own (not assigned) for personal enjoyment or academic enrichment
c. Number of written papers or reports of 20 pages or more
d. Number of written papers or reports between 5 and 19 pages
e. Number of written papers or reports of fewer than 5 pages

4. In a typical week, how many homework problem sets do you complete?
(Options: None, 1-2, 3-4, 5-6, More than 6)

a. Number of problem sets that take you more than an hour to complete
b. Number of problem sets that take you less than an hour to complete

5. Mark the box that represents the extent to which your examinations during the current school year have challenged you to do your best work
(Seven options between Very Little and Very Much)

6. During the current school year, about how often have you done each of the following?
(Options: Very Often, Often, Sometimes, Never)

a. Attended an art exhibit, play, dance, music, theatre, or other performance
b. Exercised or participated in physical fitness activities
c. Participated in activities to enhance your spirituality (worship, meditation, prayer, etc.)
d. Examined the strengths and weaknesses of your own views on a topic or issue
e. Tried to better understand someone else's views by imagining how an issue looks from his or her perspective
f. Learned something that changed the way you understand an issue or concept

7. Which of the following have you done or do you plan to do before you graduate from your institution?
(Options: Done, Plan to do, Do not plan to do, Have not decided)

a. Practicum, internship, field experience, co-op experience, or clinical assignment
b. Community service or volunteer work
c. Participate in a learning community or some other formal programme where groups of students take two or more classes together
d. Work on a research project with a faculty member outside of course or programme requirements
e. Foreign language coursework

f. Study abroad
g. Independent study or self-designed major
h. Culminating senior experience (capstone course, senior project or thesis, comprehensive exam, etc.)

8. Mark the box that best represents the quality of your relationship with people at your institution
    a. Relationships with other students
    (Seven options between Unfriendly/Unsupportive/Sense of alienation and Friendly/Supportive/Sense of belonging)
    b. Relationships with faculty members
    (Seven options between Unavailable/Unhelpful/Unsympathetic and Available/Helpful/Sympathetic)
    c. Relationships with administrative personnel and offices
    (Seven options between Unhelpful/Inconsiderate/Rigid and Helpful/Considerate/Flexible)

9. About how many hours do you spend in a typical 7-day week doing each of the following?
    (Options: 0, 1-5, 6-10, 11-15, 16-20, 21-25, 26-30, More than 30)

    a. Preparing for class (studying, reading, writing, doing homework or lab work, analysing data, rehearsing, and other academic activities)
    b. Working for pay on campus
    c. Working for pay off campus
    d. Participating in co-curricular activities (organisations, campus publications, student government, fraternity or sorority, intercollegiate or intramural sports, etc.)
    e. Relaxing and socialising (watching TV, partying, etc.)
    f. Providing care for dependents living with you (parents, children, spouse, etc.)
    g. Commuting to class (driving, walking, etc.)

10. To what extent does your institution emphasise each of the following?
    (Options: Very much, Quite a bit, Some, Very little)

    a. Spending significant amounts of time studying and on academic work
    b. Providing the support you need to help you succeed academically
    c. Encouraging contact among students from different economic, social, and racial or ethnic backgrounds
    d. Helping you cope with your non-academic responsibilities (work, family, etc.)
    e. Providing the support you need to thrive socially
    f. Attending campus events and activities (special speakers, cultural performances, athletic events, etc.)
    g. Using computers in academic work

11. To what extent has your experience at this institution contributed to your knowledge, skills, and personal development in the following areas?
    (Options: Very much, Quite a bit, Some, Very little)

    a. Acquiring a broad general education
    b. Acquiring job or work-related knowledge and skills
    c. Writing clearly and effectively

d. Speaking clearly and effectively
e. Thinking critically and analytically
f. Analysing quantitative problems
g. Using computing and information technology
h. Working effectively with others
i. Voting in local, state, or national elections
j. Learning effectively on your own
k. Understanding yourself
l. Understanding people of other racial and ethnic backgrounds
m. Solving complex real-world problems
n. Developing a personal code of values and ethics
o. Contributing to the welfare of your community
p. Developing a deepened sense of spirituality

12. Overall, how would you evaluate the quality of academic advising you have received at your institution?
    a. Excellent, Good, Fair, or Poor

13. How would you evaluate your entire educational experience at this institution?
    a. Excellent, Good, Fair, or Poor

14. If you could start over again, would you go to the same institution you are now attending?
    a. Definitely yes, Probably yes, Probably no, or Definitely No

Appendix M1: Student Interview Guide

Introductory Procedures (5 minutes)

- Introductions
- Check participant language of preference
- Review Information Sheet
- Review consent form
- Remind participants that there are no wrong answers, that interested in learning about their experiences – not what they think their experiences should have been – and that there is no way for any of the content of the interview to ever be connected to them – get verbal consent for recording
- Once consent given, turn on the recorder, CODE THE INTERVIEW and begin.

1. Reasons for attending university (5 minutes)

[Would like to start off by learning more about your motivations for coming to university.]

1.1: So, could you tell us why you decided to enrol in university?

1.2: When you started university, what did you hope you would gain during these four years? How did you think university would impact you and your future?

2. Student background (15 minutes)

2.1 Secondary School Background[136]

[Now, we'd like to move on to discuss your life before you came to university. So, we'd like to start by discussing your secondary school experience.]

2.1.1: Can you tell me where you went to secondary school?

- What type of school was it? (Private/public, religious/not-religious, single-sex/mixed, boarding/day)
- Where was it located?

2.1.2: What did you study in secondary school?

- Why did you choose that subject?

2.1.3: I'd like to hear more about your experiences in secondary school. What kinds of classes did you take? What subjects did you study?

2.1.4: Could you tell me a little bit about how your classes were taught?

- Follow-ups: How often did you ask questions in class or contribute to class discussions?
- How often did you give presentations in class?

---

[136] For fourth-year participants, only ask Questions 2.1.1, 2.1.2, 2.1.4, 2.1.5, and 2.1.7

- Did you ever work together with other students – in class or outside of class?

2.1.5: What kinds of assignments did you have?

- Did you have a chance to submit assignments for feedback before submitting final drafts?
- Did you get feedback on your final assignments?

2.1.6: What happened if you did not come to class or if you did not complete an assignment?

2.1.7: What kind of relationship did you have with your teachers? Could you talk to them about your grades or assignments? What about your future plans?

2.1.8: Did you get a government scholarship?

- Follow-ups: What do you think it takes to succeed on the national exam? Did your secondary school prepare you well for the exam? Why or why not?

2.2 Family Background

[Research has shown that there is often a connection between a student's family background and his or her experience at school, so would like to learn a little bit more about your family and your interactions with them.]

2.2.1: Would you mind telling me where you lived during secondary school?

- [If boarded] Where was your family home? Where did you go during the holidays?

2.2.2: Could you tell me whom you lived with during secondary school?

- Did any of your family members have jobs? If so, what were they?
- Did you have any books in your house? Did any of your family members like to read or listen to the news?
- Did you ever talk about school or what you were learning with your family members?

2.2.3: Did you get a scholarship to attend secondary school?

2.2.4: Did you have to work during secondary school? If so, what kind of work did you do?

3. Experiences at university (30 minutes)

[Thank you so much for sharing this with me. It's very useful to have a sense of where you came from before you arrived at university. Now, I'd like to talk about your experience at university.]

3.1 Experiences at University

3.1.1: What are you studying (or did you study) at university?

- Why did you choose that subject?

3.1.2: What types of courses did/do you take as part of that subject?

- Were/are you required to take classes in other departments?
- Did/do you have to take any general requirements? English?
- (For fourth years) Did you feel that all of your classes were relevant? Why or why not?
- (For NUR students) Could you tell me a little bit about the modular system and how it works? What is your opinion of the system?

3.1.3: Was there any induction program when you first arrived on campus?

- [If so] What was involved? Did you find it useful? Did you wish that they had included anything else?
- [If not] Do you think that would have been helpful? Why or why not? What would you have liked to see included?

3.1.4: Could you tell me a little bit about how your classes were/are taught?

- Follow-ups: How often did/do you ask questions in class or contribute to class discussions?
- What do you think I mean by class discussion? What was/is the purpose of discussions in your classes?
  - Did/do you feel that you could disagree with your instructor during class discussions? Why or why not?
- How often did/do you give presentations in class? What do presentations entail in your programme?
- Did/do you ever work together with other students – in class or outside of class? How do you feel that that worked/works? Did/do you enjoy group work? Why or why not?
- Did/do you ever go on field trips? What was/is your opinion of them?

3.1.5: What kinds of assignments did/do you have – and how many for each class?

- For exams: What types of questions did/do you have to answer?
- What did /do you feel was/is the best way to prepare for your examinations?
  - Prompt: What skills were/are you required to demonstrate?
- For written assignments or presentations: How long did/do they have to be?

- Did/do you have a chance to submit assignments for feedback before submitting final drafts? If not, did/do you ever write more than one draft of an assignment?
- Did/do you get feedback on your assignments (either positive or negative)?
- Did/do you feel that your assignments had a purpose? Could/can you understand why your instructors asked/ask you to complete the assignments that they assigned?
- What happened/happens if you failed an assignment?

3.1.6: What happened/happens if you did/do not come to class or if you did/do not complete an assignment?

3.1.7: In a typical year, how much reading were/are you required to do?

- About how many books did/do you have to read for your classes?
- About how many books did/do you read for your own enjoyment?

3.1.8: (For fourth years) Were you required to complete an internship or practical work as part of your degree?

- How did you feel that experience fit with the rest of your academic coursework?
- What was your opinion of the internship requirement?

3.1.9: [For fourth-year students] Did you have to complete a final year project? If so, what did you learn from the experience? Did you find it interesting or useful?

3.1.10: Now, I'm interested in learning about how you spent/spend your time during university. So, in a typical day or week, I want to know how many hours you spent/spend on different activities. Remember, there is no right answer! Every student is different!

- So, first, how much time in a day or in a week did/do you spend preparing for class?
- Attending class?
- Working?
- Participating in extracurricular activities?
- Relaxing or socialising?
- Caring for family?
- Commuting to class?
- Other (what?)

3.1.11: What kind of relationship did/do you have with your instructors? Could/can you talk to them about your grades or assignments? What about your future plans? Why or why not?

3.2 External Influences

[That's great. Thank you so much for sharing all of that with me. We're nearly done, but before we finish, I'd like to ask you a few questions about your life outside of university.]

3.2.1: Did you ever work during university? (Or do you currently have a job? Have you ever had one?) What kind of work did/do you do?

3.2.2: Were/are you involved with any other activities outside of work or school?

3.2.3: Did/do you have any family responsibilities?

3.2.4: What about friends outside of school? Did/do you socialise primarily with school friends, or did/do you spend time with other people away from campus?

3.2.5: Where did/do you live during the term? On campus or off campus?

3.2.6: How did/do you think your life outside of university impacted/impacts your university experience?

3.2.7: Did/do you find yourself ever talking about the content of your classes with people outside of university?

3.3 Graduate Skills

[Thank you so much for all of that. Now, to finish, I just have one more question for you, and it is about the skills you think you have learned – or will learn – during university.]

3.3.1: What are the most important abilities for a university graduate to have?

- Possible prompts: Broad knowledge about the world?
- Job-related knowledge and skills
- Writing ability
- Speaking ability
- Ability to work with others
- Ability to learn on your own
- Ability to solve complex real-world problems

3.3.2: What does it mean to be an intelligent person? What skills or abilities does an intelligent person have?

3.3.3: What does success at university mean? How do you know if you have been successful?

3.3.4: [For fourth-year students] At the beginning of the interview, we talked about your aspirations for university. Do you feel that you met your aspirations? Did you achieve what you hoped to achieve?

- Possible follow-up: Do you feel that you improved in all of these areas during your time in university? Refer back to skills list – did your time at university help you improve in any of these areas?
- [If not] What areas were missing? Why do you think you did not improve in these areas?

3.3.5: [For fourth-year students] Have you found a job?

- [If so] Is it related to what you studied? How did you find the job?
- [If not] Why do you think you're having trouble finding a job?
- Do employers value higher-level degrees? Why or why not?

4. Final Thoughts (5 minutes)

Thank you so much for sharing all of this with me. Before we finish, do you have any final thoughts you would like to share or comments you would like to make? Do you have any questions for me?

Appendix M2: Faculty Interview Guide

Introductory Procedures (5 minutes)

- Introductions
- Review Information Sheet
- Review & sign consent form
- Once the form is complete, turn on the recorder, CODE THE INTERVIEW and begin.

1. The purpose of university education (10 minutes)

[Would like to start off by talking about your views on university education. As you know, there are many opinions as to why university education is important, so I'd like to start by hearing your thoughts and opinions.]

1.1: What do you think is the purpose of a university education? Why is it important for young people to study at university? What should students gain from four years at KIST/NUR?

1.2: We met with a group of KIST/NUR students earlier this week, and they listed the following reasons as being the most common reasons why people go to university. Could you take a look at the list? Do any of these reasons surprise you? Do you agree with them?

2. The role of a faculty member (20 minutes)

[Thank you so much for that. Now, I'd like to switch topics a bit and talk a little about your own position at NUR/KIST.]

2.1 Primary responsibilities

2.1.1: So, to start off, can you tell me about your current position? What are your responsibilities?

2.1.2: Are you working anywhere else at the moment?

- [If so] What type of position do you hold elsewhere?

2.1.3: Of all of these competing priorities, which do you think are the most important? What do you think is the primary responsibility of a faculty member at KIST/NUR?

2.1.4: Do you think I would get a different answer if I talked to your colleagues in other departments? Why or why not?

2.1.5: In a typical week, how much time do you dedicate to your different responsibilities? *(List responsibilities they have mentioned and have them quantify the amount of time)*

2.2 Teaching Philosophy

2.2.1: Because my research is focused on the student experience, I'd like to concentrate on your teaching responsibilities. So, going back to this list, looking at the time you dedicate to teaching, can that time be broken down even further?

- How much time do you spend in the classroom?
- What about preparing for lessons?
- Talking to students outside of class?
- Scoring assessments and giving feedback?

2.2.2: Could you tell me a little bit about your teaching philosophy? What would a typical class of yours look like?

- Time spent lecturing?
- Discussion or class activities?
- What is the purpose of the different components? How do your classes fit with the rest of the departmental curriculum?
    - Possible prompt: Do your students go on field visits? If so, how do they connect to the rest of the curriculum?
    - How do internships fit with the rest of the programme curriculum in your department?

2.2.3: (For NUR participants) Could you tell me a bit more about the modular system?

- Why was it instituted? What was the purpose?
- How does it work logistically in your department?
- What is your opinion of the system?

2.2.4: How are your students assessed?

- What projects/papers/activities/exams do you assign?
- Do your students ever work in groups?
    - If so, how do you mark the individual members of a group?

2.2.5: I hear from many faculty members that it is difficult to cover all of the content during the semester. If you are not able to cover everything during class, how do students learn the rest of the content?

- [If readings] How do you choose the readings?
- [If learning on their own] Do you give them any guidance as to how to find materials?

2.2.6: What is your opinion of the final project? What is the value for students (if any)?

2.2.7: Have you ever participated in any pedagogical training?

- [If so] Where was that held? Who organized it? Was it a requirement that you attend?
- [If not] Would that be of interest to you, if it were offered?

2.3 Students

2.3.1: Can you tell me a little bit about your students?

- Do they come to university with a similar level of preparation?
  - If not, how do they differ?
- How do you manage to teach a classroom of students with differing language ability and differing levels of preparation?
- Has the level of ability changed over the past five years? Is there a difference between first-year students today and first-year students five years ago?

2.4 Incentives

2.4.1: How does the incentive structure work at your institution? How are decisions made about promotion?

3. Critical Thinking (20 minutes)

[That is all very helpful. Thank you so much. For the rest of the interview, I'd like to talk about the specific topic of my research, which is the concept of "critical thinking".]

3.1 Definitions

3.1.1: So, to start with, I wondered if you could tell me what you think I mean by "critical thinking"?

3.2 The importance of critical thinking/Validation of the tool

3.2.1: During the first phase of my research, I gave an assessment to students at KIST/NUR. *[Explain format of the assessment]* Do you think this is a realistic scenario that represents an experience that a student might have after graduation? Why or why not?

3.2.2: *[Explain skills that should have been assessed – handout]* Looking at this list, do you think that these are skills that students should be learning during university?

- [If do] Why do you think these skills are important?
  - Do you think that it is important for universities to encourage and improve such skills? Why or why not?
- [If not] Why not? Do you think the skills are not important, or do you think they are learned elsewhere (outside of university)?

3.2.3: How would you guess that your students would do on such an assessment? Would they perform better on some skills than others?

3.3 Discussion of Results

3.3.1: We're almost finished now, but before we end, I wanted to show you the results of the assessment. I have a handout showing the overall results of the whole national sample (students from NUR, KIST and SFB) and also results from your institution specifically. As you can see, students performed quite poorly overall. Are you surprised? Why or why not?

3.3.2: There also seems to be a difference between student ability on certain skills. Why do you think that might be? Do you emphasise any of these skills in the classes you teach? If so, which ones and how?

3.3.3: Why do you think we saw no gains between first- and fourth-year students?

4. Final Thoughts (5 minutes)

4.1: What are the biggest challenges that you face in terms of your role as a teacher?

Thank you so much for sharing all of this with me. Before we finish, do you have any final thoughts you would like to share or comments you would like to make? Any questions?

Appendix M3: Administrator Interview Guide

Introductory Procedures (5 minutes)

- Introductions
- Review Information Sheet
- Review & sign consent form
- Once the form is complete, turn on the recorder, CODE THE INTERVIEW and begin.

1. The purpose of university education (10 minutes)

[Would like to start off by talking about your views on university education. As you know, there are many opinions as to why university education is important, so I'd like to start by hearing your thoughts and opinions.]

1.1: What do you think is the purpose of a university education? Why is it important for young people to study at university? Do you think students come to KIST/NUR for these reasons?

1.2: In your opinion, what are the most important skills for a student to learn during university?

1.3: What do you think is the primary responsibility of a faculty member at KIST/NUR?

2. Quality in university education (20 minutes)

2.1 Role

2.1.1: Could you tell me a little bit about your role here? What are your primary responsibilities?

- How does this role interact with other academic roles on campus?

2.2 Teaching & Learning

2.2.1: Because my research is focused on the student experience, I'd like to concentrate on the teaching purpose of a university. There has been a lot of talk in recent years about improving the quality of higher education in Rwanda. How would you personally define a "quality" higher education?

2.2.2: Do you think that there is a teaching "philosophy" at KIST/NUR? How would you describe standard teaching practices at this institution?

- (For NUR) How do you feel that the new modular system is working?
- Is there an overarching Learning, Teaching and Assessment Strategy on campus?
  - If so, when was it written/implemented?

        - Was there any resistance to its implementation? Do you think it's being implemented in all of the Faculties? Why or why not?

2.2.3: What about the assessment culture?

- How are students assessed?
- Who determines the content of the main examinations?
- What happens to students that fail a class? Or multiple classes?
- Have there been any significant changes in the assessment requirements in recent years?

2.2.4: I hear from many faculty members that it is difficult to cover all of the content during the semester. Is there a standard way that professors are expected to teach the content that they cannot cover in class? Who sets those expectations?

2.2.5: What is your opinion of the final project? What is the value for students (if any)?

2.2.6: Does the university require the faculty to undergo any pedagogical training? Is there any pedagogical training available for those who wish to access it?

2.3 Students

2.3.1: Can you tell me a little bit about the KIST/NUR student body?

- Do they come to university with a similar level of preparation?
    - If not, how do they differ?
    - Does the university provide any training or support for lecturers to help them cope with differing student ability?
- Has the level of ability changed over the past five years? Is there a difference between first-year students today and first-year students five years ago?

2.4 Incentives

2.4.1: How does the incentive structure work at your institution? How are decisions made about promotion?

3. Critical Thinking **(20 minutes)**

[That is all very helpful. Thank you so much. For the rest of the interview, I'd like to talk about the specific topic of my research, which is the concept of "critical thinking".]

3.1 Definitions

3.1.1: So, to start with, I wondered if you could tell me what you think I mean by critical thinking?

3.2 The importance of critical thinking/Validation of the tool

3.2.1: During the first phase of my research, I gave an assessment to students at KIST/NUR. *[Explain format of the assessment]* Do you think this is a realistic scenario that represents an experience that a student might have after graduation? Why or why not?

3.2.2: *[Explain skills that should have been assessed – handout]* Looking at this list, do you think that these are skills that students should be learning during university?

- [If do] Why do you think these skills are important?
  - Do you think that it is important for universities to encourage and improve such skills? Why or why not?
- [If not] Why not? Do you think the skills are not important, or do you think they are learned elsewhere (outside of university)?

3.3 Discussion of Results

3.3.1: We're almost finished now, but before we end, I wanted to show you the results of the assessment. I have a handout showing the overall results of the whole national sample (students from NUR, KIST and SFB) and also results from your institution specifically. As you can see, students performed quite poorly overall. Are you surprised? Why or why not?

3.3.2: There also seems to be a difference between student ability on certain skills. Why do you think that might be?

3.3.3: Why do you think we saw no gains between first- and fourth-year students?

4. Final Thoughts **(5 minutes)**

4.1: What are the biggest barriers to "quality" currently facing the institution?

Thank you so much for sharing all of this with me. Before we finish, do you have any final thoughts you would like to share or comments you would like to make?

Do you have any questions for me?

Wrap-Up Procedures

- Turn off the recorder
- Thank participant & remind them what will happen with the information
- Get email for checking transcript and sending report

Works Cited

Africa-UK Engineering for Development Partnership (2012). *Engineers for Africa: Identifying engineering capacity needs in Sub-Saharan Africa*. London: Royal Academy of Engineering.

Ajayi, J. F. A., Goma, L. K. H., Johnson, G. A. and Association of African Universities. (1996). *The African experience with higher education*. Accra: Association of African Universities in association with James Currey and Ohio University Press.

Aldridge, J. M., Fraser, B. J. and Huang, T.-C. I. (1999). 'Investigating Classroom Environments in Taiwan and Australia with Multiple Research Methods'. *The Journal of Educational Research* 93 (1), 48-62.

Altbach, P. G. (1998). *Comparative higher education : knowledge, the university, and development*. Greenwich, CT ; London: Ablex.

Amutabi, M. N. and Oketch, M. O. (2003). 'Experimenting in Distance Education: The African Virtual University (AVU) and the Paradox of the World Bank in Kenya'. *International Journal of Educational Development,* 23 (1), 57-73.

Anderson, L. W. and Bloom, B. S. (2001). *A taxonomy for learning, teaching, and assessing : a revision of Bloom's Taxonomy of educational objectives*. (Abridged Ed.). New York, NY ; London: Longman.

Apentiik, C. R. A. and Parpart, J. L. (2006). 'Working in Different Cultures: Issues of Race, Ethnicity and Identity'. In V. Desai & R. B. Potter (Eds), *Doing Development Research* (pp. 34-43). London: Sage Publications.

Aronowitz, S. (2000). *The Knowledge Factory: Dismantling the Corporate University and Creating True Higher Learning*. Boston, MA: Beacon Press.

Arum, R. and Roksa, J. (2011). *Academically adrift : limited learning on college campuses*. Chicago, IL: University of Chicago Press.

Ashcroft, K. and Rayner, P. (2011). *Higher Education in Development: Lessons from sub-Saharan Africa*. Charlotte, NC: Information Age Publishing, Inc.

Assié-Lumumba, N. D. T. and CODESRIA. (2006). *Higher education in Africa : crises, reforms and transformation*. Dakar: CODESRIA.

Assiter, A. (1995). 'Transferable Skills: A Response to the Sceptics'. In A. Assiter (Ed.), *Transferable skills in higher education* (pp. 11-19). London: Kogan Page.

Association for the Development of Education in Africa Working Group on Higher Education and Association of African Universities. (2004). *Higher education in Sub-Saharan Africa, with specific reference to universities*. Accra: Association of African Universities.

Association of African Universities, Donors to African Education/Working Group on Higher Education and National University of Lesotho. (1995). *The university in Africa in the 1990's and beyond : summary report of a colloquium jointly organised by Association of*

African Universities, Donors to African Education/Working Group on Higher Education, [and] National University of Lesotho, January 16-20, 1995. Accra: Association of African Universities.

Astin, A. W. (1970). 'The Methodology of Research on College Impact, Part One'. *Sociology of Education,* 43 (3), 223-254.

Astin, A. W. (1984). 'Student Involvement: A Developmental Theory for Higher Education'. *Journal of College Student Personnel,* 25 (4), 297-308.

Astin, A. W. (1991). *Assessment for Excellence: The Philosophy and Practice of Assessment and Evaluation in Higher Education*. New York, NY: American Council on Education and Macmillan Publishing Company.

Astin, A. W. (1993a). 'Diversity and Multiculturalism on the Campus: How Are Students Affected?'. *Change: The Magaizine of Higher Learning,* March/April, 44-49.

Astin, A. W. (1993b). *What matters in college? Four critical years revisisted*. San Francisco: Jossey-Bass.

Astin, A. W. (2011). 'In 'Academically Adrift', Data Don't Back Up Sweeping Claim'. *The Chronicle of Higher Education,* February 14.

Bailin, S., Case, R., Coombs, J. R. and Daniels, L. B. (1999a). 'Common misperceptions of critical thinking'. *Journal of Curriculum Studies,* 31 (3), 269-283.

Bailin, S., Case, R., Coombs, J. R. and Daniels, L. B. (1999b). 'Conceptualizing critical thinking'. *Journal of Curriculum Studies,* 31 (3), 285-302.

Banta, T. W. (2008). 'Trying to Clothe the Emperor'. *Assessment Update,* 20 (2), 3-4 and 15-16.

Barnett, R. and Society for Research into Higher Education. (1997). *Higher education : a critical business*. Buckingham: Society for Research into Higher Education & Open University Press.

Bataineh, R. F. and Zghoul, L. H. (2006). 'Jordanian TEFL Graduate Students' Use of Critical Thinking Skills (as Measured by the Cornell Critical Thinking Test, Level Z)'. *International Journal of Bilingual Education and Bilingualism,* 9 (1), 33-50.

Baxter, G. P. and Glaser, R. (1998). 'Investigating the Cognitive Complexity of Science Assessments'. *Educational Measurement: Issues and Practice,* 17 (3), 37-45.

Baxter Magolda, M. (1990). 'Gender Differences in Epistemological Development'. *Journal of College Student Development,* 31 (6), 555-561.

Baxter Magolda, M. (1992). *Knowing and Reasoning in College Students: gender-related patterns in students' intellectual development*. San Francisco, CA: Jossey-Bass.

Baxter Magolda, M. B. (1999). *Creating Contexts for Learning and Self-Authorship: Constructive-Developmental Pedagogy*. Nashville, TN: Vanderbilt University Press.

Belenky, M. F., Clinchy, B. M., Goldberger, N. R. and Tarule, J. M. (1986). *Women's Ways of Knowing*. New York, NY: Basic Books.

Bell, J. (2005). *Evaluating Psychological Information: Sharpening Your Critical Thinking Skills*. Boston, MA: Allyn and Bacon.

Benjamin, R., Chun, M., Hardison, C., Hong, E., Jackson, C., Kugelmass, H., Nemeth, A. and Shavelson, R. (2009). *Returning to Learning in an Age of Assessment: Introducing the Rationale of the Collegiate Learning Assessment*. New York, NY: Council for Aid to Education.

Benjamin, R., Chun, M. and Shavelson, R. (2007). *White Paper. Holistic Tests in a Sub-Score World: The Diagnostic Logic of the College Learning Assessment*. New York, NY: Council for Aid to Education.

Bennell, P. (1996). 'Rates of Return to Education: Does the Conventional Pattern Prevail in sub-Saharan Africa?'. *World Development,* 24 (1), 183-199.

Bereiter, C. and Scardamalia, M. (1998). 'Beyond Bloom's Taxonomy: Rethinking Knowledge for the Knowledge Age'. In A. Hargreaves, A. Lieberman, M. Fullan & D. W. Hopkins (Eds), *International Handbook of Educational Change* (pp. 675-692). Dordrecht: Kluwer Academic Publishers.

Berg, B. L. (2007). *Qualitative Research Methods for the Social Sciences*. (6th Ed.). San Francisco, CA: Pearson Education Inc.

Biggs, J. (2001). 'Enhancing Learning: A Matter of Style or Approach?'. In R. J. Sternberg & L. F. Zhang (Eds), *Perspectives on Thinking, Learning and Cognitive Styles* (pp. 73-100). Mahwah, NJ: Lawrence Erlbaum Associates.

Binns, T. (2006). 'Doing Fieldwork in Developing Countries: Planning and Logistics'. In V. Desai & R. B. Potter (Eds), *Doing Development Research* (pp. 13-23). London: Sage Publications.

Birdsall, N. (1996). 'Public spending on higher education in developing countries: Too much or too little?'. *Economics of Education Review,* 15 (4), 407-419.

Blaich, C. and Wise, K. (2010). *Wabash National Study of Liberal Arts Education 2006-2009: Overview of Findings from the First Year*. [Online]. Available at: http://www.liberalarts.wabash.edu/study-4th-year-data/. [Last accessed March 6, 2011].

Blair, J. A. (1992). 'The Generalizability of Critical Thinking: The Evaluation of Sources'. In S. P. Norris (Ed.), *The Generalizability of Critical Thinking: Multiple Perspectives on an Educational Ideal* (pp. 125-137). New York, NY: Teachers College Press.

Bloom, B. S. and Association of College and University Examiners. (1956). *Taxonomy of educational objectives : the classification of educational goals*. New York, NY: David McKay.

Bloom, D., Canning, D. and Chan, K. (2006). *Higher Education and Economic Development in Africa*. Cambridge, MA: Harvard University.

Boeren, A. (2005). Mechanisms to support higher education and research in developing countries: The perspective of donors and programme administrators. In A. Boeren & G. Holtland (eds), *A Changing Landscape: Making support to higher education and research in developing countries more effective*. The Hague: Nuffic.

Bok, D. C. (2006). *Our underachieving colleges : a candid look at how much students learn and why they should be learning more*. Princeton, NJ ; Oxford: Princeton University Press.

Bowden, A., Fox-Rushby, J., Nyandieka, L. and Wanjau, J. (2002). 'Methods for pre-testing and piloting survey questions: illustrations from the KENQOL survey of health-related quality of life'. *Health Policy and Planning,* 17 (3), 322-330.

Bradby, H. (2002). 'Translating Culture and Language: A Research Note on Multilingual Settings'. *Sociology of Health & Illness,* 24 (6), 842-55.

Brannelly, L., Lewis, L. and Ndaruhutse, S. (2011). *Learning and Leadership: Exploring the linkages between higher education and developmental leadership*. Reading: CfBT Education Trust.

Bransford, J. D. (1979). *Human cognition : learning, understanding and remembering*. Belmont, CA: Wadsworth.

Bransford, J. D., Brown, A. L. and Cocking, R. R. (Eds) (1999), *How People Learn: Brain, Mind, Experience, and School*. Washington, DC: National Academy Press.

Braxton, J. M., Olsen, D. and Simmons, A. (1998). 'Affinity Disciplines and the Use of Principles of Good Practice for Undergraduate Education'. *Research in Higher Education,* 39 (3), 299-318.

Brislin, R. W. (1970). 'Back-translation for cross-cultural research'. *Journal of Cross-Cultural Psychology,* 1 (3), 186-216.

Brock-Utne, B. (2000). *Whose education for all? : the recolonization of the African mind*. New York, NY; London: Falmer Press.

Brown, K. (1998). *Education, Culture and Critical Thinking*. Aldershot: Ashgate Publishing.

Brown, P., Green, A. and Lauder, H. (2001). *High Skills: Globalization, Competitiveness, and Skill Formation*. Oxford: Oxford University Press.

Bryman, A. (2006). 'Integrating quantitative and qualitative research: how is it done?'. *Qualitative Research,* 6 (1), 97-113.

Bujra, J. (2006). 'Lost in Translation? The Use of Interpreters in Fieldwork'. In V. Desai & R. B. Potter (Eds), *Doing Development Research* (pp. 172-179). London: Sage Publications.

Burawoy, M. (1998). 'The Extended Case Method'. *Sociological Theory,* 16 (1), 4-33.

Caracelli, V. J. and Greene, J. C. (1993). 'Data Analysis Strategies for Mixed-Method Evaluation Designs'. *Educational Evaluation and Policy Analysis,* 15 (2), 195-207.

Carini, R. M., Kuh, G. D. and Klein, S. (2006). 'Student Engagement and Student Learning: Testing the Linkages'. *Research in Higher Education,* 47 (1), 1-32.

Case, R. (2009). 'Teaching and Assessing the 'Tools' for Thinking'. In J. Sobocan & L. Groarke (Eds), *Critical Thinking Education and Assessment: Can Higher Order Thinking Be Tested?* (pp. 197-214). London, Ont.: Althouse Press.

Castells, M. (1994). 'The University System: Engine of Development in the New World Economy'. In J. Salmi & A. M. Verspoor (Eds), *Revitalizing Higher Education* (pp. 14-40). Oxford: International Association of Universities.

Cattell, R. B. (1966). 'The meaning and strategic use of factor analysis'. In R. B. Cattell (Ed.), *Handbook of Multivariate Experimental Psychology* (pp. 174-243). Chicago, IL: Rand McNally & Company.

Chan, H. M. and Yan, H. K. T. (2008). 'Is There a Geography of Thought for East-West Differences? Why or why not?'. In M. Mason (Ed.), *Critical Thinking and Learning* (pp. 44-64): Blackwell Publishing.

Chi, M. T. H., Glaser, R. and Farr, M. J. (1988). *The Nature of Expertise*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Chi, M. T. H., Glaser, R. and Rees, E. (1982). 'Expertise in Problem Solving'. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 1, pp. 7-35). Hillsdale, NJ: Erlbaum.

Chu, J. (2009). 'Rwanda Rising: A New Model of Economic Development'. *Fast Company*, March 18.

Chun, M. (2008). *Introduction to Performance Tasks*. New York, NY: Council for Aid to Education.

Chun, M. (2010). 'Taking Teaching to (Performance) Task: Linking Pedagogical and Assessment Practices'. *Change: The Magaizine of Higher Learning,* 42 (2)*,* 22-29.

Clahsen, H. and Felser, C. (2006). 'Grammatical processing in language learners'. *Applied Psycholinguistics,* 27*,* 3-42.

Clark, D. (2006). *The Elgar companion to development studies*. Cheltenham: Edward Elgar.

Clark, H. H. and Wilkes-Gibbs, D. (1986). 'Referring as a collaborative process'. *Cognition,* 22*,* 1-39.

Cloete, N., Bailey, T., Pillay, P., Bunting, I. and Maassen, P. (2011). *Universities and Economic Development in Africa*. Wynberg: Centre for Higher Education Transformation.

Cloete, N. and Centre for Higher Education Transformation (South Africa). (1997). *Knowledge, identity and curriculum transformation in Africa*. Cape Town: Maskew Miller Longman.

Cohen, J. (1977). *Statistical Power Analysis for the Behavioral Sciences, Revised Edition*. London: Academic Press, Inc.

Cole, M. (1990). 'Cognitive development and formal schooling: The evidence from cross-cultural research'. In L. C. Moll (Ed.), *Vygotsky and education: Instructional implications and applications of sociohistorical psychology* (pp. 89-110). New York: Cambridge University Press.

Cole, M. and Scribner, S. (1974). *Culture and thought: a psychological introduction*. New York, NY: John Wiley.

Coleman, J. S. (1966). *Equality of Educational Opportunity*. Washington, DC: U.S. Department of Health, Education and Welfare.

Coleman, J. S. and Court, D. (1993). *University development in the Third World : the Rockefeller Foundation experience*. Oxford: Pergamon.

Collins, C. S. (2011). *Higher education and global poverty: university partnerships and the World Bank in developing countries*. Amherst, NY: Cambria Press.

Cook, P., Johnson, R., Moore, P., Myers, P., Pauly, S., Pendarvis, F., Prus, J. and Ulmer-Sottong, L. (1996). *Critical Thinking Assessment: Measuring a Moving Target. Report & Recommendations of the South Carolina Higher Education Assessment Network Critical Thinking Task Force*: South Carolina Higher Education Assessment Network, Rock Hill.

Coombe, T., Ford Foundation and Rockefeller Foundation. (1991). *A consultation on higher education in Africa : a report to the Ford Foundation and the Rockefeller Foundation*. London: Institute of Education, University of London.

Cottrell, S. (2003). *The Study Skills Handbook*. (2nd Ed.). Basingstoke: Palgrace Macmillan.

Council for Aid to Education (2008). *CLA Frequently Asked Technical Questions*. New York, NY.

Council for Aid to Education (2011). *Architecture of the CLA Tasks*. New York, NY.

Council of Independent Colleges (2008). *Evidence of Learning: Applying the Collegiate Learning Assessment to Improve Teaching and Learning in the Liberal Arts College Experience*. Washington, DC.

Court, D. (1995). 'The challenge to the liberal vision of universities in Africa'. In L. Buchert & K. King (Eds), *Learning from Experience: policy and practice in aid to higher education* (pp. 109-121). The Hague: Centre for the Study of Education in Developing Countries.

Court, D. and Kinyanjui, K. (1986). 'African Education: Problems in a High-growth Sector'. In R. J. Berg & J. S. Whitaker (Eds), *Strategies for African Development: A Study for the Committee on African Development Strategies* (pp. 361-392). Berkeley, CA: University of California Press.

Cowan, J. (1994). 'Research into student learning - Yes, but by whom?'. In S. Tornkvist (Ed.), *Teaching Science and Technology at Tertiary Level* (pp. 51-59). Proceedings of Conference of Royal Swedish Academy of Engineering Sciences, Stockholm, Sweden, 6-9 June.

Creswell, J. W. and Plano Clark, V. L. (2007). *Designing and Conducting Mixed Methods Research*. Thousand Oaks, CA: Sage Publications.

Crossley, M. and Watson, K. (2003). *Comparative and international research in education : globalisation, context and difference*. London: Routledge Falmer.

Crossman, P. (2004). 'Perceptions of 'Africanization' or 'endogenization' at African universities: Issues and recommendations'. In P. T. Zeleza & A. O. Olukoshi (Eds), *African Universities in the Twenty-First Century. Volume 2: Knowledge and Society.* (pp. 321-340). Dakar: CODESRIA.

Daneman, M. and Carpenter, P. A. (1983). 'Individual Differences in Integrating Information Between and Within Sentences'. *Journal of Experimental Psychology,* 9 (4), 561-584.

Daneman, M. and Tardif, T. (1987). 'Working Memory and Reading Skill Re-examined'. In M. Coltheart (Ed.), *Attention and Performance XII: The Psychology of Reading* (pp. 491-508). Hove and London: Lawrence Erlbaum Associates.

Davies, W. M. (2006). 'An 'infusion' approach to critical thinking: Moore on the critical thinking debate'. *Higher Education Research and Development,* 25 (2), 179-193.

de Corte, E. (2000). 'Marrying theory building and the improvement of school practice: a permanent challenge for instructional psychology'. *Learning and Instruction,* 10, 249-266.

de Vaus, D. (2002). *Surveys in Social Research*. (5th Ed.). London: Routledge.

Denzin, N. K. and Lincoln, Y. S. (1994). *Handbook of qualitative research*. Thousand Oaks, CA: Sage Publications.

Deregowski, J. B. (1970). 'Effect of Cultural Value of Time upon Recall'. *British Journal of Social and Clinical Psychology,* 9, 37-41.

Dewey, J. (1933 (renewed 1960)). *How We Think: A Restatement of the Relation of Reflective Thinking to the Educative Process*. Lexington, MA: D.C. Heath and Company.

Dey, E. (1991). Community service and critical thinking: An exploratory analysis of collegiate influences. Paper presented at the meeting of the conference "Setting the Agenda for an Effective Research Strategy for Combining Service and Learning in the 1990s". Racine, WI.

Dias, M. A. R. (1992). 'Improving Co-operation to Ameliorate Quality and Relevance in Higher Education'. *International Journal of Educational Development,* 12 (2), 123-129.

Dowling, P. and Brown, A. (2010). *Doing research/reading research : re-interrogating education*. (2nd Ed.). London: Routledge.

Drew, S. and Staff and Educational Development Association. (1998). *Key skills in higher education : background and rationale*. Birmingham, AL: Staff & Educational Development Association.

Dunbar, K., Fugelsang, J. and Stein, C. (2007). 'Do naive theories ever go away?'. In M. Lovett & P. Shah (Eds), *Thinking with Data: 33rd Cambridge Symposium on Cognition* (pp. 193-205). Mahwah, NJ: Erlbaum.

Durkin, K. (2008). 'The adaptation of East Asian masters students to western norms of critical thinking and argumentation in the UK'. *Intercultural Education,* 19 (1), 15-27.

Eckstein, H. (1975). 'Case Studies and Theory in Political Science'. In F. I. Greenstein & N. W. Polsby (Eds), *Handbook of Political Science* (Vol. 7, pp. 94-137). Reading, MA: Addison-Wesley.

Educational Consultants India Limited (2001). *Study of Technical and Vocational Training Needs for Rwanda's Reconstruction and Role of KIST. Project supported by the African Development Bank and MINEDUC. Final Report, August 2001.*

Edwards, R. (1998). 'A Critical Examination of the Use of Interpreters in the Qualitative Research Process'. *Journal of Ethnic and Migration Studies,* 24, 197-208.

Eide, P. and Allen, C. B. (2005). 'Recruiting Transcultural Qualitative Research Participants: A Conceptual Model'. *International Journal of Qualitative Methods,* 4 (2), 1-10.

Eisemon, T. O. and Kourouma, M. (1994). 'Foreign Assistance for University Development in sub-Saharan Africa and Asia'. In J. Salmi & A. M. Verspoor (Eds), *Revitalizing Higher Education* (pp. 274-304). Oxford: International Association of Universities.

Ennis, R. H. (1962). 'A concept of critical thinking'. *Harvard Educational Review,* 32, 81-111.

Ennis, R. H. (1964). 'A Definition of Critical Thinking'. *The Reading Teacher,* 17, 599-612.

Ennis, R. H. (1985). 'A Logical Basis for Measuring Critical Thinking'. *Educational Leadership,* 43 (2), 45-48.

Ennis, R. H. (1993). 'Critical Thinking Assessment'. *Theory Into Practice,* 32 (3), 179-186.

Ennis, R. H. (2009). 'Investigating and Assessing Multiple-Choice Critical Thinking Tests'. In J. Sobocan & L. Groarke (Eds), *Critical Thinking Education and Assessment: Can Higher Order Thinking Be Tested?* (pp. 75-97). London, Ont.: Althouse Press.

Ennis, R. H. and Weir, E. (1985). *The Ennis-Weir Critical Thinking Essay Test: Test Manual*. Pacific Grove, CA: Midwest Publications.

Entwistle, N. (1997). 'Contrasting Perspectives on Learning'. In F. Marton, D. Hounsell & N. Entwistle (Eds), *The Experience of Learning: Implications for Teaching and Studying in Higher Education* (2nd ed., pp. 3-22). Edinburgh: Scottish Academic Press.

Entwistle, N. and Entwistle, A. (1997). 'Revision and the Experience of Understanding'. In F. Marton, D. Hounsell & N. Entwistle (Eds), *The Experience of Learning: Implications for Teaching and Studying in Higher Education* (2nd ed., pp. 145-155). Edinburgh: Scottish Academic Press.

Entwistle, N. and Peterson, E. R. (2004). 'Conceptions of learning and knowledge in higher education: Relationships with study behaviour and influences of learning environments'. *International Journal of Educational Research,* 41, 407-428.

Ericsson, K. A. and Simon, H. A. (1993). *Protocol analysis : verbal reports as data*. (Revised Ed.). Cambridge, Mass. ; London: MIT Press.

Esposito, N. (2001). 'From Meaning to Meaning: The Influence of Translation Techniques on Non-English Focus Group Research'. *Qualitative Health Research,* 11 (4), 568-579.

European University Association (2010). *Outcomes and Recommendations of the Project: "Access to Success: Fostering Trust and Exchange between Europe and Africa" (2008-2010)*. Brussels.

Everett, M. D. and Zinser, O. (1998). 'Interdisciplinary Social Science Courses: Using a Critical Thinking Approach'. *The Journal of General Education,* 47 (3), 225-241.

Facione, N. C. and Facione, P. A. (1997). *Critical Thinking Assessment in Nursing Education Programs: An Aggregate Data Analysis*. Millbrae, CA: The California Academic Press.

Facione, P. A. (1990). *Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction. Executive Summary.* Millbrae, CA: California Academic Press.

Facione, P. A. (1991). *Using the California Critical Thinking Skills Test in Research, Evaluation, and Assessment*. Millbrae, CA: California Academic Press.

Fehnel, R. (2003). 'Massification and Future Trends in African Higher Education'. In D. Teferra & P. G. Altbach (Eds), *African Higher Education: An International Reference Handbook* (pp. 73-81). Bloomington, IN: Indiana University Press.

Feldman, K. A. (1987). 'Research Productivity and Scholarly Accomplishment of College Teachers as Related to their Instructional Effectiveness: A Review and Exploration'. *Research in Higher Education,* 26 (3), 227-282.

Field, A. (2009). *Discovering Statistics Using SPSS*. (Third Ed.). Thousand Oaks, CA: Sage Publications.

Filmer, D. and Pritchett, L. H. (2001). 'Estimating Wealth Effects without Expenditure Data - or Tears: An Application to Educational Enrollments in States of India'. *Demography,* 38 (1), 115-132.

Fisher, A. and Scriven, M. (1997). *Critical Thinking: Its Definition and Assessment*. Norwich: Centre for Research in Critical Thinking.

Flavell, J. H. (1976). 'Metacognitive Aspects of Problem Solving'. In L. B. Resnick (Ed.), *The Nature of Intelligence* (pp. 231-235). Hillsdale, NJ: Lawrence Erlbaum Associates.

Floyd, C. B. (2011). 'Critical thinking in a second language'. *Higher Education Research and Development,* 30 (3), 289-302.

Fonlon, B. N. (2009). *Genuine intellectuals : academic and social responsibilities of universities in Africa*. Bamenda, Cameroon: Langaa Research & Publishing Common Initiative Group.

Forsyth, B. H. and Lessler, J. T. (1991). 'Cognitive laboratory methods: a taxonomy'. In P. P. Biemer, S. M. Groves, L. E. Lyberg, N. A. Mathiowetz & S. Sudan (Eds), *Measurement errors in surveys* (pp. 393-418). New York, NY: John Wiley.

Fraser, A. (2009). 'Aid-Recipient Sovereignty in Historical Context'. In L. Whitfield (Ed.), *The Politics of Aid: African Strategies for Dealing with Donors* (pp. 45-73). Oxford: Oxford University Press.

Freedman, S. W., Weinstein, H. M., Murphy, K. L. and Longman, T. (2011). 'Teaching History in Post-Genocide Rwanda'. In S. Straus & L. Waldorf (Eds), *Remaking Rwanda: State Building and Human Rights after Mass Violence* (pp. 297-315). Madison, WI: University of Wisconsin Press.

Fulton, O., Gordon, A. and Williams, G. (1980). 'Higher Education and Manpower Planning: A Comparison of Planned and Market Economies'. *Education Policy Bulletin,* 8 (1), 83-113.

Gahutu, J. B. (2010). 'Physiology teaching and learning experience in a new modular curriculum at the National University of Rwanda'. *Advances in Physiology Education,* 34 (1)*,* 11-14.

Gardner, M. K. and Clark, E. (1992). 'The psychometric perspective on intellectual development in childhood and adolescence'. In R. J. Sternberg & C. A. Berg (Eds), *Intellectual development* (pp. 16-43). Cambridge: Cambridge University Press.

Giancarlo, C. A. and Facione, P. A. (2001). 'A Look Across Four Years at the Disposition Toward Critical Thinking Among Undergraduate Students'. *The Journal of General Education,* 50 (1)*,* 29-55.

Giancarlo-Gittens, C. A. (2009). 'Assessing Critical Thinking Dispositions in an Era of High-Stakes Standardized Testing'. In J. Sobocan & L. Groarke (Eds), *Critical Thinking Education and Assessment: Can Higher Order Thinking Be Tested?* (pp. 17-34). London, Ont.: Althouse Press.

Gmelin, W., King, K. and McGrath, S. (Eds) (2001), *Development Knowledge, National Research and International Cooperation*. Edinburgh: Centre of African Studies, University of Edinburgh, in conjunction with the German Foundation for International Development and the Northern Policy Research Review and Advisory Network for Education and Training.

Goma, L. H. K. and Tembo, L. P. (1984). *The African university : issues and perspectives*. Lusaka: National Educational Company of Zambia for the Institute for African Studies, University of Zambia.

Goodnow, J. J. (1976). 'The Nature of Intelligent Behavior: Questions Raised by Cross-Cultural Studies'. In L. B. Resnick (Ed.), *The Nature of Intelligence* (pp. 169-188). Hillsdale, NJ: Lawrence Erlbaum Associates.

Gourevitch, P. (1998). *We Wish to Inform You That Tomorrow We Will Be Killed With Our Families*. New York, NY: Farrar, Straus & Giroux.

Gouveia, T. F. P.-B. (2010). Students' views of higher education in their transitions to work in Portugal. PhD thesis. London: Institute of Education.

Green, A. and Sakamoto, A. (2001). 'Models of High Skills in National Competition Strategies'. In P. Brown, A. Green & H. Lauder (Eds), *High Skills: Globalisation, Competitiveness, and Skill Formation* (pp. 56-160). Oxford: Oxford University Press.

Greene, J. C. (2007). *Mixed methods in social inquiry*. San Francisco, CA: Jossey-Bass.

Greene, J. C., Caracelli, V. J. and Graham, W. F. (1989). 'Toward a Conceptual Framework for Mixed-Method Evaluation Designs'. *Educational Evaluation and Policy Analysis,* 11 (3)*,* 255-274.

Greenfield, P. M. and Bruner, J. S. (1969). 'Culture and Cognitive Growth'. In D. Goslin (Ed.), *Handbook of Socialization Theory and Research* (pp. 633-657). Chicago, IL: Rand McNally.

Grice, P. (1975). 'Logic and Conversation'. In P. Cole & J. L. Morgan (Eds), *Syntax and Semantics* (Vol. 3, Speech Acts, pp. 41-58). New York, NY: Academic Press.

Groarke, L. (2009). 'What's Wrong with the California Critical Thinking Skills Test? CT Testing and Accountability'. In J. Sobocan & L. Groarke (Eds), *Critical Thinking Education and Assessment: Can Higher Order Thinking Be Tested?* (pp. 35-54). London, Ont.: Althouse Press.

Guile, D. (2006). 'What is Distinctive About the Knowledge Economy? Implications for Education'. In H. Lauder, P. Brown, J.-A. Dillabough & A. H. Halsey (Eds), *Education, Globalization, and Social Change* (pp. 355-366). Oxford: Oxford University Press.

Gwatkin, D. R., Rutstein, S., Johnson, K., Suliman, E., Wagstaff, A. and Amouzou, A. (2007). *Socio-Economic Differences in Health, Nutrition and Population: Rwanda*. Washington, DC: World Bank.

Haas, P. F. and Keeley, S. M. (1998). 'Coping with Faculty Resistance to Teaching Critical Thinking'. *College Teaching,* 46 (2)*,* 63-67.

Haertel, E. H. (2006). 'Reliability'. In R. L. Brennan (Ed.), *Educational Measurement* (Fourth ed., pp. 65-110). Westport, CT: American Council on Education and Praeger Publishers.

Haertel, E. H. and Linn, R. L. (1996). Comparability. In G. W. Phillips (ed), *Technical Issues in Large-Scale Performance Assessment* (pp. 59-78). Washington, DC: U.S. Department of Education Office of Educational Research and Improvement.

Hagedorn, L. S., Pascarella, E. T., Edison, M., Braxton, J., Nora, A. and Terenzini, P. T. (1999). 'Institutional Context and the Development of Critical Thinking: A Research Note'. *The Review of Higher Education,* 22 (3)*,* 265-285.

Halpern, D. F. (1996). *Thought and knowledge : an introduction to critical thinking*. (3rd Ed.). Mahwah, N.J: L. Erlbaum Associates.

Hambleton, R. K. (1994). 'Guidelines for Adapting Educational and Psychological Tests: A Progress Report'. *European Journal of Psychological Assessment,* 10 (3)*,* 229-244.

Hambleton, R. K. and Kanjee, A. (1995). 'Increasing the Validity of Cross-Cultural Assessments: Use of Improved Methods for Test Adaptations'. *European Journal of Psychological Assessment,* 11 (3)*,* 147-157.

Hammersley, M. and Atkinson, P. (1995). *Ethnography : principles in practice*. (2nd Ed.). London: Routledge.

Hammond, J. and Gibbons, P. (2001). 'What is scaffolding?'. In J. Hammond & P. Gibbons (Eds), *Scaffolding teaching and learning in language and literacy education* (pp. 1-14). Sydney: The Primary English Teaching Association (PETA).

Hardison, C. and Vilamovska, A.-M. (2008). *Critical thinking performance tasks: Setting and applying standards for college-level performance*. Santa Monica, CA: Rand.

Harkness, S. and Super, C. M. (1982). 'Why African Children Are So Hard To Test'. In L. L. Adler (Ed.), *Cross-Cultural Research at Issue* (pp. 145-152). New York, NY: Academic Press, Inc.

Hatcher, D. L. (2009). 'The Institutional Assessment of Critical Thinking: A Fifteen-Year Perspective'. In J. Sobocan & L. Groarke (Eds), *Critical Thinking Education and Assessment: Can Higher Order Thinking Be Tested?* (pp. 229-246). London, Ont.: Althouse Press.

Hayman, R. (2005). *The Contribution of Post-Basic Education and Training (PBET) to Poverty Reduction in Rwanda: balancing short-term goals and long-term visions in the face of capacity constraints*. Edinburgh: Centre for African Studies, University of Edinburgh.

Hayman, R. (2007). 'Are the MDGs enough? Donor perspectives and recipient visions of education and poverty reduction in Rwanda'. *International Journal of Educational Development,* 27 (4), 371-382.

Hayman, R. (2009a). 'Going in the "Right" Direction? Promotion of Democracy in Rwanda since 1990'. *Taiwan Journal of Democracy,* 5 (1), 51-76.

Hayman, R. (2009b). 'Milking the Cow. Creating Policy Space in Spite of Aid Dependence'. In L. Whitfield (Ed.), *The Politics of Aid: African Strategies for Dealing with Donors* (pp. 156-184). Oxford: Oxford University Press.

Holm, J. D. (2012). 'The Brain Drain Within Africa'. *The Chronicle of Higher Education,* October 9.

Howe, C., Tolmie, A. and Rodgers, C. (1992). 'The acquisition of conceptual knowledge in science by primary school children: Group interaction and the understanding of motion down an incline'. *British Journal of Developmental Psychology,* 10, 113-130.

Hyland, T. and Johnson, S. (1998). 'Of Cabbages and Key Skills: exploding the mythology of core transferable skills in post-school education'. *Journal of Further and Higher Education,* 22 (2), 163-172.

Inman, P. and Pascarella, E. T. (1998). 'The impact of college residence on the development of critical thinking skills in college freshmen'. *Journal of College Student Development,* 39 (6), 557-568.

Inter-University Council for East Africa (2010). *A Road Map to Quality: Handbook for Quality Assurance in Higher Education*. Kampala: The Inter-University Council for East Africa/DAAD.

International Association of Universities. (1979). *The role of the university in developing countries : its responsibility toward the natural and cultural environment*. Paris: International Association of Universities.

Irwin, M., Klein, R. E. and Townsend, J. W. (1982). 'Indigenous versus Construct Validity in Cross-Cultural Research'. In L. L. Adler (Ed.), *Cross-Cultural Research at Issue* (pp. 23-29). New York: Academic Press, Inc.

Israel, G. D. (1992). *Determining Sample Size* (Fact Sheet PEOD-6, Florida Cooperative Extension Service, Institute of Food and Agricultural Sciences). Gainesville, FL: University of Florida.

Jobe, J. B. and Mingay, D. J. (1989). 'Cognitive research improves questionnaires'. *American Journal of Public Health,* 79 (8), 1053-1055.

Johnson, D. W., Johnson, R. T. and Smith, K. A. (1996). *Academic Controversy: Enriching College Instruction through Intellectual Conflict*. Washington, DC: The George Washington University, Graduate School of Education and Human Development.

Johnson, R. B., Onwuegbuzie, A. J. and Turner, L. A. (2007). 'Toward a definition of mixed methods research'. *Journal of Mixed Methods Research,* 1 (2), 112-133.

Johnson, R. H. (1992). 'The Problem of Defining Critical Thinking'. In S. P. Norris (Ed.), *The Generalizability of Critical Thinking: Multiple Perspectives on an Educational Ideal* (pp. 38-53). New York, NY: Teachers College Press.

Johnson, R. H. (2009). 'The Implications of the Dialectical Tier for Critical Thinking'. In J. Sobocan & L. Groarke (Eds), *Critical Thinking Education and Assessment: Can Higher Order Thinking Be Tested?* (pp. 55-74). London, Ont.: Althouse Press.

Just, M. A. and Carpenter, P. A. (1992). 'A Capacity Theory of Comprehension: Individual Differences in Working Memory'. *Psychological Review,* 99 (1), 122-149.

Kagame, P. (2011). President Kagame addresses NUR students (May 16). Huye. Transcribed from podcast released by the Government of Rwanda, May 17.

Kajubi, W. S. (1984). 'Higher Education and the Dilemma of Nation-Building in Africa: A Retrospective and Prospective View (Lecture delivered January 28, 1982, at Teachers College, Columbia University)'. In A. Taylor (Ed.), *Insights into African Education: The Karl W. Bigelow Memorial Lectures* (pp. 17-48). New York, NY: Columbia University.

Kalyuga, S. (2006). *Instructing and testing advanced learners : a cognitive load approach*. New York, NY: Nova Science Publishers.

Kamali, A. (2006). National examinations in Rwanda : on the right track? MA Dissertation in Curriculum, Pedagogy and Assessment. London: Institute of Education.

Kane, M. T. (2006). 'Validation'. In R. L. Brennan (Ed.), *Educational Measurement* (Fourth ed., pp. 17-64). Westport, CT: American Council on Education and Praeger Publishers.

Kapur, D. and Crowley, M. (2008). *Beyond the ABCs: Higher Education and Developing Countries*. Washington, DC: Center for Global Development.

Karabenick, S. A. and Collins-Eaglin, J. (1996). *Relation of perceived instructional goals and incentives to college students' use of learning strategies*. Paper presented at the American Educational Research Association. New York, NY.

Kember, D. and Gow, L. (1990). 'Cultural Specificity of Approaches to Study'. *British Journal of Educational Psychology,* 60, 356-363.

Kember, D. and Gow, L. (1994). 'Orientations to Teaching and Their Effect on the Quality of Student Learning'. *The Journal of Higher Education,* 65 (1), 58-74.

Kember, D. and Leung, D. Y. P. (2005). 'The Influence of the Teaching and Learning Environment on the Development of Generic Capabilities Needed for a Knowledge-Based Society'. *Learning Environments Research,* 8, 245-266.

Kember, D., Leung, D. Y. P., Jones, A., Yuen Loke, A., McKay, J., Sinclair, K., Tse, H., Webb, C., Yuet Wong, F. K., Wong, M. and Yeung, E. (2000). 'Development of a Questionnaire to Measure the Level of Reflective Thinking'. *Assessment & Evaluation in Higher Education,* 25 (4), 381-395.

Kemper, E. A., Stringfield, S. and Teddlie, C. (2003). 'Mixed Methods Sampling Strategies in Social Science Research'. In A. Tashakkori & C. Teddlie (Eds), *Handbook of mixed methods in social and behavioral research* (pp. 273-296). Thousand Oaks, CA: Sage Publications.

King, K. (2009). 'Higher Education and International Cooperation: the role of academic collaboration in the developing world'. In D. Stephens (Ed.), *Higher Education and International Capacity Building: twenty-five years of higher education links* (pp. 33-50). Oxford: Symposium Books.

King, P. M. and Kitchener, K. S. (1994). *Developing Reflective Judgment: Understanding and Promoting Intellectual Growth and Critical Thinking in Adolescents and Adults*. San Francisco, CA: Jossey-Bass Inc.

KIST Directorate of Planning & Development (2007). *KIST Strategic Plan 2007-2010*. Kigali: Kigali Institute of Science & Technology.

KIST Directorate of Quality Assurance (2012). *Learning, Teaching and Assessment Strategy*. Kigali: Kigali Institute of Science and Technology.

KIST Office of Planning & Development (2012). *KIST Statistics*. Kigali: Kigali Institute of Science & Technology.

Kitchener, K. S. and Fischer, K. W. (1990). 'A Skill Approach to the Development of Reflective Thinking'. In D. Kuhn (Ed.), *Developmental Perspectives on Teaching and Learning Thinking Skills* (pp. 48-62). Basel: Karger.

Klees, S. J. (2002). 'World Bank education policy: new rhetoric, old ideology'. *International Journal of Educational Development,* 22, 451-474.

Klein, S., Benjamin, R., Shavelson, R. and Bolus, R. (2007). 'The Collegiate Learning Assessment: Facts and Fantasies'. *Evaluation Review,* 31, 415-439.

Klein, S., Kuh, G., Chun, M., Hamilton, L. and Shavelson, R. (2005). 'An Approach to Measuring Cognitive Outcomes Across Higher Education Institutions'. *Research in Higher Education,* 46 (3), 251-276.

Klein, S., Liu, O. L. and Sconing, J. (2009). *Test Validity Study (TVS) Report*. New York, NY: Council for Aid to Education with support from the Fund for the Improvement of Postsecondary Education (FIPSE).

Kneale, P. E. (1999). *Study Skills for Geography Students: A Practical Guide*. London: Arnold.

Koretz, D. M. (2008). *Measuring up : what educational testing really tells us*. Cambridge, MA ; London: Harvard University Press.

Kozulin, A. (1998). *Psychological tools : a sociocultural approach to education*. Cambridge, Mass.: Harvard University Press.

Krueger, R. A. (1993). 'Quality Control in Focus Group Research'. In D. L. Morgan (Ed.), *Successful Focus Groups: Advancing the State of the Art* (pp. 65-88). Newbury Park, CA: Sage Publications.

Kuh, G. (1995). 'The Other Curriculum: Out-of-Class Experiences Associated with Student Learning and Personal Development'. *The Journal of Higher Education,* 66 (2), 123-155.

Kuh, G. D. and the Documenting Effective Educational Practice Project. (2005). *Student success in college : creating conditions that matter*. San Francisco, CA: Jossey-Bass.

Kuhn, D. (1991). *The skills of argument*. Cambridge: Cambridge University Press.

Kuhn, D. (1995). 'Microgenetic study of change: What has it told us?'. *Psychological Science,* 6, 133-139.

Kuhn, D. (1999). 'A Developmental Model of Critical Thinking'. *Educational Researcher,* 28 (2), 16-25+46.

Kuhn, D. (2005). *Education for Thinking*. Cambridge, MA: Harvard University Press.

Lane, S. and Stone, C. A. (2006). 'Performance Assessment'. In R. L. Brennan (Ed.), *Educational Measurement* (Fourth ed., pp. 387-432). Westport, CT: American Council on Education and Praeger Publishers.

Lauder, H., Brown, P., Dillabough, J.-A. and Halsey, A. H. (2006). 'The Prospects for Education: Individualization, Globalization, and Social Change'. In H. Lauder, P. Brown, J.-A. Dillabough & A. H. Halsey (Eds), *Education, Globalization and Social Change* (First ed.). Oxford: Oxford University Press.

Lave, J. (1988). *Cognition in practice : mind, mathematics and culture in everyday life*. Cambridge: Cambridge University Press.

Leach, F. (1994). 'Expatriates as Agents of Cross-Cultural Transmission'. *Compare: A Journal of Comparative and International Education,* 24 (3), 217-231.

Lebeau, Y. (2008). 'Universities and social transformation in sub-Saharan Africa: global rhetoric and local contradictions'. *Compare: A Journal of Comparative and International Education,* 38 (2), 139-153.

Lehman, D. R. and Nisbett, R. E. (1990). 'A Longitudinal Study of the Effects of Undergraduate Training on Reasoning'. *Developmental Psychology,* 26 (6), 952-960.

Liamputtong, P. (2010). *Performing Qualitative Cross-Cultural Research*. Cambridge: Cambridge University Press.

Lim, D. (1999). 'Quality Assurance in Higher Education in Developing Countries'. *Assessment & Evaluation in Higher Education,* 24 (4), 379-390.

Lindow, M. (2011). *Weaving Success: Voices of Change in African Higher Education*. New York, NY: Institute of International Education.

Lipman, M. (1988). 'Critical Thinking - What Can It Be?'. *Educational Leadership,* 46, 38-43.

Lipman, M. (2003). *Thinking in Education*. (2nd Ed.). Cambridge: Cambridge University Press.

Lloyd-Evans, S. (2006). 'Focus Groups'. In V. Desai & R. B. Potter (Eds), *Doing Development Research* (pp. 153-161). London: Sage Publications.

Loevinger, J. (1957). 'Objective tests as instruments of psychological theory'. *Psychological Reports, Monograph Supplement,* 3, 635-694.

Lombard, K. and Grosser, M. (2004). 'Critical thinking abilities among prospective educators: ideals versus realities'. *South African Journal of Education,* 24 (3), 212-216.

Lombard, K. and Grosser, M. (2008). 'Critical thinking: are the ideals of OBE failing us or are we failing the ideals of OBE?'. *South African Journal of Education,* 28, 561-579.

Long, J. B. and Harding-Esch, E. (1977). 'Summary and recall of text in first and second languages: some factors contributing to performance differences'. In H. W. Sinaiko & D. Gerver (Eds), *Proceedings of the NATO Symposium on Language Interpretation and Communication* New York, NY: Plenum Press.

Lonka, K. and Ahola, K. (1995). 'Activating instruction: How to foster study and thinking skills in higher education'. *European Journal of Psychology of Education,* 10 (4), 351-368.

Lulat, Y. G. M. (2005). *A history of African higher education from antiquity to the present : a critical synthesis*. Westport, CT ; London: Praeger.

Lun, V. M.-C., Fischer, R. and Ward, C. (2010). 'Exploring cultural differences in critical thinking: Is it about my thinking style or the language I speak?'. *Learning and Individual Differences,* 20 (6), 604-616.

Maassen, P., Pinheiro, R. and Cloete, N. (2007). *Bilateral Country Investments and Foundations Partnership Projects to Support Higher Education Across Africa (2006/07 Report)*. Wynberg: Centre for Higher Education Transformation.

Maliyamkono, T. L., Ishumi, A. G. M., Wells, S. J. and Eastern African Universities Research Project. (1982). *Higher education and development in Eastern Africa : an Eastern African Universities Research Project Report on the impact of overseas training on development*. London: Heinemann Educational.

Mamdani, M. (2001). *When victims become killers : colonialism, nativism, and the genocide in Rwanda*. Princeton, NJ: Princeton University Press.

Mamdani, M. (2007). *Scholars in the Marketplace: The Dilemmas of Neo-Liberal Reform at Makerere University, 1989-2005*. Dakar: CODESRIA.

Marsh, H. W. and Hattie, J. (2002). 'The Relation between Research Productivity and Teaching Effectiveness: Complementary, Antagonistic, or Independent Constructs?'. *The Journal of Higher Education,* 73 (5), 603-641.

Marshall, C. and Rossman, G. B. (2006). *Designing Qualitative Research*. (Fourth Ed.). Thousand Oaks, CA: Sage Publications.

Marton, F., Hounsell, D. and Entwistle, N. (Eds) (1997), *The Experience of Learning: Implications for Teaching and Studying in Higher Education* (2nd ed.). Edinburgh: Scottish Academic Press.

Marton, F. and Saljo, R. (1976). 'On Qualitative Differences in Learning: 1 - Outcome and Process'. *British Journal of Educational Psychology,* 46, 4-11.

Mason, J. (1994). 'Linking qualitative and quantitative analysis'. In A. Bryman & R. G. Burgess (Eds), *Analyzing Qualitative Data*. London: Routledge.

Materu, P. (2007). *Higher education quality assurance in Sub-Saharan Africa : status, challenges, opportunities and promising practices*. Washington, D.C.: World Bank.

Matiru, B., Zentralstelle für Erziehung Wissenschaft und Dokumentation and Deutsche Stiftung für internationale Entwicklung. (1991). *Towards academic and professional excellence in higher education*. Bonn: Zentralstelle für Erziehung Wissenschaft und Dokumentation.

Mayer, R. E. and Wittrock, M. C. (1996). 'Problem-solving transfer'. In D. C. Berliner & R. C. Calfee (Eds), *Handbook of educational psychology* (pp. 47-62). New York, NY: Prentice Hall International.

Mayoux, L. (2006). 'Quantitative, Qualitative or Participatory? Which Method, for What and When?'. In V. Desai & R. B. Potter (Eds), *Doing Development Research* (pp. 115-129). London: Sage Publications.

Mazimhaka, J. and Daniel, G. F. (2003). 'Rwanda'. In D. Teferra & P. G. Altbach (Eds), *African higher education : an international reference handbook* (pp. 500-511). Bloomington, IN: Indiana University Press.

Mazimpaka, J., Daniel, G. F. and Association of African Universities. (2000). *Post-genocide restructuring of higher education in Rwanda : an overview*. Accra: Association of African Universities.

Mbabazi Bamwesiga, P., Dahlgren, L.-O. and Fejes, A. (2012). 'Students as learners through the eyes of their teachers in Rwandan higher education'. *International Journal of Lifelong Education,* 31 (4), 503-521.

McKeachie, W. J., Pintrich, P. R., Lin, Y.-G. and Smith, D. A. F. (1986). *Teaching and Learning in the College Classroom: A Review of the Research Literature*. Ann Arbor, MI: Regents of The University of Michigan.

McKenzie, D. J. (2003). *Measuring Inequality with Asset Indicators*. Durham, NC: Bureau for Research in Economic Analysis of Development.

McLean Hilker, L. (2011). 'The role of education in driving conflict and building peace: The case of Rwanda'. *Prospects,* 41, 267-282.

McPeck, J. E. (1981). *Critical thinking and education*. Oxford: Martin Robertson.

McPeck, J. E. (1992). 'Thoughts on Subject Specificity'. In S. P. Norris (Ed.), *The Generalizability of Critical Thinking: Multiple Perspectives on an Educational Ideal* (pp. 198-205). New York: Teachers College Press.

Messick, S. (1989). 'Meaning and Values in Test Validation: The Science and Ethics of Assessment'. *Educational Researcher,* 18 (2), 5-11.

Messick, S. (1994). 'The Interplay of Evidence and Consequences in the Validation of Performance Assessments'. *Educational Researcher,* 23 (2), 13-23.

Meyers, C. (1986). *Teaching Students to Think Critically*. San Francisco, CA: Jossey-Bass.

Miles, M. B. and Huberman, A. M. (1994). *Qualitative data analysis : an expanded sourcebook*. (2nd Ed.). Thousand Oaks, CA ; London: Sage Publications.

MINEDUC (2007). *Rwandan National Qualifications Framework for Higher Education*. Kigali: Republic of Rwanda Ministry of Education (MINEDUC).

MINEDUC (2008). *Education Sector Strategic Plan 2008-2012*. Kigali: Republic of Rwanda Ministry of Education (MINEDUC).

MINEDUC (2010). *Education Sector Strategic Plan 2010-2015*. Kigali: Republic of Rwanda Ministry of Education (MINEDUC).

MINEDUC (2012). *Rwanda Education Statistics*. Kigali: Republic of Rwanda Ministry of Education (MINEDUC).

Mines, R. A., King, P. M., Hood, A. B. and Wood, P. K. (1990). 'Stages of Intellectual Development and Associated Critical Thinking Skills in College Students'. *Journal of College Student Development,* 31 (6), 538-547.

Ministry of Health (MOH) [Rwanda], National Institute of Statistics of Rwanda (NISR) and ICF Macro (2009). *Rwanda Interim Demographic and Health Survey 2007-08*. Calverton, MD: MOH, NISR and ICF Macro.

Mkandawire, P. T. (2005). 'African intellectuals and nationalism'. In P. T. Mkandawire (Ed.), *African Intellectuals: Rethinking Politics, Language, Gender and Development* (pp. 10-55). Dakar, Senegal and London: CODESRIA Books with Zed Books.

Mohamedbhai, G. (2008). *The Effects of Massification on Higher Education in Africa*. Accra: Association of African Universities.

Moja, T. (2004). 'Policy responses to global transformation by African higher education systems'. In P. T. Zeleza & A. O. Olukoshi (Eds), *African Universities in the Twenty-First Century: Liberalisation and Internationalisation* (Vol. 1, pp. 21-41). Dakar: CODESRIA.

Moon, J. A. (2008). *Critical thinking : an exploration of theory and practice*. London: Routledge.

Moore, T. (2004). 'The Critical Thinking Debate: How general are general thinking skills?'. *Higher Education Research and Development,* 23 (1), 3-18.

Moore, T. (2011). 'Critical thinking and disciplinary thinking: a continuing debate'. *Higher Education Research and Development,* 30 (3), 261-274.

Moseley, D. (2005). *Frameworks for thinking : a handbook for teaching and learning*. Cambridge: Cambridge University Press.

Mosha, H. J. (1986). 'The Role of African Universities in National Development: A Critical Analysis'. *Comparative Education,* 22 (2), 93-109.

Mugisha, I. S. (2010). Assessment and Study Strategies: A Study Among Rwandan Students in Higher Education. PhD thesis. Linkoping: Linkoping University.

Murenzi, R. and Hughes, M. (2006). 'Building a prosperous knowledge economy in Africa: Rwanda as a case study'. *International Journal of Technology and Globalisation,* 2 (3/4)*,* 254-269.

Musaazi, J. C. S. (2011). *Planning and development in education : African perspectives*. London: Routledge.

Mutebi, F. G., Stone, S. and Thin, N. (2003). 'Rwanda'. In D. Booth (Ed.), *Fighting Poverty in Africa: Are PRSPs Making a Difference?* (pp. 217-245). London: Overseas Development Institute.

National Survey of Student Engagement (2012). *National Survey of Student Engagement 2012: The College Student Report*. Bloomington, IN: Indiana University.

Newell, A. and Simon, H. A. (1972). *Human Problem Solving*. Englewood Cliffs, NJ: Prentice-Hall.

Ng'ethe, N., Subotzky, G., Afeti, G. and World Bank. (2008). *Differentiation and articulation in tertiary education systems : a study of twelve African countries*. Washington, D.C.: World Bank.

Ngara, C. (2012). 'African Ways of Knowing: Rethinking Pedagogy in Africa'. In H. K. Wright & A. A. Abdi (Eds), *The Dialectics of African Education and Western Discourses: Counter-Hegemonic Perspectives* (Vol. 21, pp. 129-145). New York, NY: Peter Lang.

Nicholls, J. G. (1983). 'Conceptions of Ability and Achievement Motivation: A Theory and its Implications for Education'. In S. G. Paris, G. M. Olson & H. W. Stevenson (Eds), *Learning and Motivation in the Classroom* (pp. 211-238). Hillsdale, NJ: Lawrence Erlbaum Associates.

Nisbett, R. E., Choi, I., Peng, K. and Norenzayan, A. (2001). 'Culture and Systems of Thought: Holistic Versus Analytic Cognition'. *Psychological Review,* 108 (2)*,* 291-310.

Norenzayan, A., Smith, E. E., Kim, B. J. and Nisbett, R. E. (2002). 'Cultural preferences for formal versus intuitive reasoning'. *Cognitive Science,* 26*,* 653-684.

Norris, S. P. (Ed.) (1992), *The Generalizability of Critical Thinking: Multiple Perspectives on an Educational Ideal*. New York, NY: Teachers College Press.

NUR Department of Planning & Development (2008). *The National University of Rwanda Strategic Plan (NURSP) and Business Plan (NURBP) 2008-2012*. Huye: National University of Rwanda.

Nwaboku, N. and UNESCO. (1996). *Enhancing instructional competence in higher education : a study of pedagogic needs in some African universities*. Dakar: UNESCO Regional Office for Education in Africa, Higher Education Unit.

Nyerere, J. K. (1968). *Freedom and socialism= Uhuru na ujamaa : a selection from writings and speeches, 1965-1967*. Dar es Salaam ; London: Oxford University Press.

Obura, A. P., UNESCO and International Institute for Educational Planning. (2003). *Never again : educational reconstruction in Rwanda*. Paris: International Institute for Educational Planning.

Official Gazette of the Republic of Rwanda (2006). *Law governing the organisation and functioning of Higher Education*. Kigali: Republic of Rwanda.

Okagaki, L. and Sternberg, R. J. (1990). 'Teaching Thinking Skills: We're Getting the Context Wrong'. In D. Kuhn (Ed.), *Developmental Perspectives on Teaching and Learning Thinking Skills* (pp. 63-78). Basel: Karger.

Oketch, M. O. (2003). 'Affording the Unaffordable: Cost Sharing in Higher Education in Sub-Saharan Africa'. *Peabody Journal of Education,* 78 (3)*,* 88-106.

Omokhodion, J. O. (1989). 'Classroom Observed: The Hidden Curriculum in Lagos, Nigeria'. *International Journal of Educational Development,* 9 (2)*,* 99-110.

Onwuegbuzie, A. J. and Johnson, R. B. (2006). 'The Validity Issue in Mixed Research'. *Research in the Schools,* 13 (1)*,* 48-63.

Osborne, J. (2010). 'Arguing to Learn in Science: The Role of Collaborative, Critical Discourse'. *Science* 328*,* 463-466.

Osman, R. M. and Githua, B. N. (2009). 'Analysis of Students' Critical Thinking Skills by Location, Gender and the Skills Relationship to Motivation to Learn Academic Disciplines in Amoud University, Somaliland'. *Journal of Technology and Education in Nigeria,* 14 (1 & 2)*,* 1-10.

Palmer, B. and Marra, R. M. (2004). 'College student epistemological perspectives across knowledge domains: a proposed grounded theory'. *Higher Education,* 47*,* 311-335.

Palmer, R., Wedgwood, R., Hayman, R., with King, K. and Thin, N. (2007). *Educating out of poverty? : a synthesis report on Ghana, India, Kenya, Rwanda, Tanzania and South Africa*. London: Great Britain Department for International Development and Centre of African Studies, University of Edinburgh.

Parker, R. and Moore, B. N. (1989). *Critical Thinking: Evaluating Claims and Arguments in Everyday Life. Second Edition*. Mountain View, CA: Mayfield Publishing Company.

Parker Webster, J. and John, T. A. (2010). 'Preserving a space for cross-cultural collaborations: an account of insider/outsider issues'. *Ethnography and Education,* 5 (2)*,* 175-191.

Pascarella, E. T. (1985). 'College Environmental Influences on Learning and Cognitive Development: A Critical Review and Synthesis'. In J. C. Smart (Ed.), *Higher Education: Handbook of Theory and Research* (Vol. 1, pp. 1-62). New York, NY: Agathon Press, Inc.

Pascarella, E. T., Blaich, C., Martin, G. L. and Hanson, J. M. (2011). 'How robust are the findings of Academically Adrift?'. *Change: The Magaizine of Higher Learning,* 43 (3)*,* 20-24.

Pascarella, E. T., Cruce, T., Umbach, P. D., Wolniak, G. C., Kuh, G. D., Carini, R. M., Hayek, J. C., Gonyea, R. M. and Zhao, C.-M. (2006). 'Institutional Selectivity and Good Practices in Undergraduate Education: How Strong is the Link?'. *The Journal of Higher Education,* 77 (2)*,* 251-285.

Pascarella, E. T., Seifert, T. A. and Blaich, C. (2010). 'How Effective are the NSSE Benchmarks in Predicting Important Educational Outcomes?'. *Change: The Magaizine of Higher Learning,* January/February, Accessed online at:

http://www.changemag.org/Archives/Back%20Issues/January-February%202010/full-how-effective.html.

Pascarella, E. T. and Terenzini, P. T. (2005). *How college affects students*. (2nd Ed.). San Francisco, CA: Jossey-Bass.

Paul, R. (1981). 'Teaching Critical Thinking in the "Strong" Sense: A Focus on Self-Deception, World Views, and a Dialectical Mode of Analysis'. *Informal Logic: Reasoning and Argumentation in Theory and Practice,* 4 (2), 2-7.

Paul, R. W. (1987). 'Dialogical Thinking: Critical Thought Essential to the Acquisition of Rational Knowledge and Passions'. In J. B. Baron & R. J. Sternberg (Eds), *Teaching Thinking Skills: Theory and Practice* (pp. 127-148). New York, NY: W.N. Freeman and Company.

Peck, A. (2012). 'Overloaded with information, students need critical thinking skills'. *University World News,* 26 February.

Pellegrino, J. W., Chudowsky, N., Glaser, R. and National Research Council (U.S.) Division of Behavioral and Social Sciences and Education Committee on the Foundations of Assessment. (2001). *Knowing what students know : the science and design of educational assessment*. Washington, DC: National Academy Press.

Penfield, R. D. and Lam, T. C. M. (2000). 'Assessing Differential Item Functioning in Performance Assessment: Review and Recommendations'. *Educational Measurement: Issues and Practice,* 19 (3), 5-15.

Perkins, D. N. and Salomon, G. (1994). 'Transfer of Learning'. In T. Husen & T. N. Postlethwaite (Eds), *The International Encyclopedia of Education* (2nd ed., Vol. 11, pp. 6452-6457). Oxford: Elsevier/Pergamon.

Perry, W. G. and Harvard University Bureau of Study Counsel. (1970). *Forms of intellectual and ethical development in the college years : a scheme*. New York, NY ; London: Holt Rinehart & Winston.

Phan, H. P. (2011). 'Deep Processing Strategies and Critical Thinking: Developmental Trajectories Using Latent Growth Analyses'. *The Journal of Educational Research,* 104 (4), 283-294.

Philips, S. U. (1972). 'Participant Structures and Communicative Competence: Warm Springs Children in Community and Classroom'. In C. B. Cazden, V. P. John & D. Hymes (Eds), *Functions of Language in the Classroom* (pp. 370-394). New York, NY: Teachers College Press.

Phillips, H. (1960). 'Problems of Translation and Meaning in Fieldwork'. In R. N. Adams & J. J. Preiss (Eds), *Human Organisation Research: Field Relations and Techniques*. Homewood, IL: Dorsey Press Inc.

Piaget, J. (1975). *The Development of Thought: Equilibration of Cognitive Structures*. (A. Rosin, Trans.). Oxford: Basil Blackwell.

Pinto, L. E. and Portelli, J. P. (2009). 'The Role and Impact of Critical Thinking in Democratic Education: Challenges and Possibilities'. In J. Sobocan & L. Groarke (Eds), *Critical Thinking Education and Assessment: Can Higher Order Thinking Be Tested?* (pp. 299-318). London, Ont.: Althouse Press.

Pithers, R. T. and Soden, R. (1999). 'Assessing Vocational Tutors' Thinking Skills'. *Journal of Vocational Education and Training,* 51 (1), 23-37.

Pithers, R. T. and Soden, R. (2000). 'Critical thinking in education: a review'. *Educational Research,* 42 (3), 237-249.

Plass, J. L., Moreno, R. and Brünken, R. (2010). *Cognitive load theory*. Cambridge: Cambridge University Press.

Pottier, J. (2002). *Re-Imagining Rwanda: Conflict, Survival and Disinformation in the Late Twentieth Century*. Cambridge: Cambridge University Press.

Prunier, G. (1995). *The Rwanda crisis, 1959-1994 : history of a genocide*. Kampala: Fountain Publishers.

Psacharopoulos, G., Tan, J.-P., Jimenez, E. and World Bank Education and Training Department. (1986). *Financing education in developing countries : an exploration of policy options*. Washington, DC: World Bank.

Ramsden, P. (1979). 'Student Learning and Perceptions of the Academic Environment'. *Higher Education,* 8, 411-427.

Ramsden, P. (1997). 'The Context of Learning in Academic Departments'. In F. Marton, D. Hounsell & N. Entwistle (Eds), *The Experience of Learning: Implications for Teaching and Studying in Higher Education* (pp. 198-216). Edinburgh: Scottish Academic Press.

Renner, J. and Lawson, A. (1975). 'Intellectual development in preservice elementary school teachers: An evaluation'. *Journal of College Science Teaching,* 5, 89-92.

Renner, J. and Paske, W. (1977). 'Comparing two forms of instruction in college physics'. *American Journal of Physics,* 45, 851-859.

Republic of Rwanda Ministry of Finance and Economic Planning (2000). *Rwanda Vision 2020*. Kigali: Republic of Rwanda.

Republic of Rwanda National Council for Higher Education (2009). *Strategic Planning Guidelines for Public and Private Sector Higher Education Institutions: Achieving the 2020 Vision*. Kigali: National Council for Higher Education.

Republic of Rwanda National Council for Higher Education (2011). *National Skills Survey 2010*. Kigali: National Council for Higher Education.

Richardson, J. T. E. (1994). 'Cultural Specificity of Approaches to Studying in Higher Education: A Literature Survey'. *Higher Education,* 27 (4), 449-468.

Robson, C. (2011). *Real world research : a resource for users of social research methods in applied settings*. (3rd Ed.). Chichester: Wiley.

Rogoff, B. (2003). *The cultural nature of human development*. Oxford: Oxford University Press.

Rogoff, B. and Chavajay, P. (1995). 'What's Become of Research on the Cultural Basis of Cognitive Development?'. *American Psychologist,* 50 (10), 859-877.

Rosenshine, B. and Meister, C. (1992). 'The Use of Scaffolds for Teaching Higher-Level Cognitive Strategies'. *Educational Leadership,* 49 (7), 26-33.

Rostow, W. W. (1960). *The Stages of Economic Growth: A Non-Communist Manifesto*. Cambridge: Cambridge University Press.

Rummel, R. J. (1970). *Applied Factor Analysis*. Evanston, IL: Northwestern University Press.

Rutayisire, J., Kabano, J. and Rubagiza, J. (2004). 'Redefining Rwanda's Future: The Role of Curriculum in Social Reconstruction'. In S. Tawil & A. Harley (Eds), *Education, Conflict and Social Cohesion* (pp. 315-374). Geneva: UNESCO International Bureau of Education.

Rutstein, S. and Johnson, K. (2004). *The DHS Wealth Index*. Calverton, MD: ORC Macro.

Ryan, J. (1992). 'Finding Generalizable Strategies in Scientific Theory Debates'. In S. P. Norris (Ed.), *The Generalizability of Critical Thinking: Multiple Perspectives on an Educational Ideal* (pp. 66-80). New York, NY: Teachers College Press.

Ryan, J. and Louie, K. (2008). 'False Dichotomy? 'Western' and 'Confucian' concepts of scholarship and learning'. In M. Mason (Ed.), *Critical Thinking and Learning* (pp. 65-78). Oxford: Blackwell Publishing.

Saavedra, A. R. and Saavedra, J. E. (2011). 'Do colleges cultivate critical thinking, problem solving, writing and interpersonal skills?'. *Economics of Education Review,* 30, 1516-1526.

Salmi, J., Verspoor, A. M. and International Association of Universities (IAU). (1994). *Revitalizing higher education*. Oxford: Pergamon for IAU Press.

Salomon, G. and Perkins, D. N. (1989). 'Rocky Roads to Transfer: Rethinking Mechanisms of a Neglected Phenomenon'. *Educational Psychologist,* 24 (2), 113-142.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Salsana, M. and Tarantola, S. (2008). *Global Sensitivity Analysis: The Primer*. Hoboken, NJ: John Wiley and Sons.

Samuelson, B. L. and Freedman, S. W. (2010). 'Language policy, multilingual education, and power in Rwanda'. *Language Policy,* 9, 191-215.

Schafer, J. L. and Graham, J. W. (2002). 'Missing Data: Our View of the State of the Art'. *Psychological Methods,* 7 (2), 147-177.

Scheaffer, R. L., Mendenhall, W. and Ott, L. (2006). *Elementary survey sampling*. (6th Ed.). Belmont, CA; London: Duxbury.

Schilling, K. (1991). *Assessing Models of Liberal Education: An Empirical Comparison*. Washington, DC: Fund for the Improvement of Postsecondary Education.

Schmeiser, C. B. and Welch, C. J. (2006). 'Test Development'. In R. L. Brennan (Ed.), *Educational Measurement* (Fourth ed., pp. 307-353). Westport, CT: American Council on Education and Praeger Publishers.

Schommer, M. and Walker, K. (1995). 'Are Epistemological Beliefs Similar Across Domains?'. *Journal of Educational Psychology,* 87 (3), 424-432.

Schultz, T. W. (1961). 'Investment in Human Capital'. *The American Economic Review,* 51 (1)*,* 1-17.

Schweisfurth, M. (2006). 'Global and cross-national influences on education in post-genocide Rwanda'. *Oxford Review of Education,* 32 (5)*,* 697-709.

Scriven, M. (2009). 'Foreword'. In J. Sobocan & L. Groarke (Eds), *Critical Thinking Education and Assessment: Can Higher Order Thinking Be Tested?* (pp. xiii-xv). London, Ont.: Althouse Press.

Senghor, L. S. (1962). 'De la Negritude. Psychologie du Negro-Africain'. In J. Reed & C. Wake (Eds), *Leopold Sedar Senghor: Prose and Poetry* (pp. 29-34). London: Heinemann.

Serageldin, I. (2000). University Governance and the "Stakeholder Society". Keynote speech given at the Eleventh General Conference of the International Association of Universities. Durban, South Africa.

Serpell, R. (2007). 'Bridging between orthodox western higher educational practices and an African sociocultural context'. *Comparative Education,* 43 (1)*,* 23-51.

Shabani, J. (1995). 'Higher Education in Sub-saharan Africa: strategies for the improvement of the quality of training'. *Quality in Higher Education,* 1 (2)*,* 173-178.

Shavelson, R. (2010). *Measuring College Learning Responsibly: Accountability in a New Era*. Stanford, CA: Stanford University Press.

Shermis, M. D. (2008). 'The Collegiate Learning Assessment: A Critical Perspective'. *Assessment Update,* 20 (2)*,* 10-12.

Shizha, E. (2012). 'Linguistic Independence and African Education and Development'. In H. K. Wright & A. A. Abdi (Eds), *The Dialectics of African Education and Western Discourses: Counter-Hegemonic Perspectives* (Vol. 21, pp. 148-161). New York, NY: Peter Lang.

Shrout, P. E. and Fleiss, J. L. (1979). 'Intraclass Correlations: Uses in Assessing Rater Reliability'. *Psychological Bulletin,* 86 (2)*,* 420-428.

Siegel, H. (1992). 'The Generalizability of Critical Thinking Skills, Dispositions, and Epistemology'. In S. P. Norris (Ed.), *The Generalizability of Critical Thinking: Multiple Perspectives on an Educational Ideal* (pp. 97-108). New York: Teachers College Press.

Sindayigaya, B. S. (2010). *Statistical Information on Higher Learning Institutions in Rwanda*. Kigali: Republic of Rwanda National Council for Higher Education.

Skinner, B. F. (1954). 'The Science of Learning and the Art of Teaching'. In B. F. Skinner (Ed.), *Cumulative Record*. London: Methuen & Co Ltd.

Skon, L., Johnson, D. W. and Johnson, R. T. (1981). 'Cooperative Peer Interaction Versus Individual Competition and Individualistic Efforts: Effects on the Acquisition of Cognitive Reasoning Strategies'. *Journal of Educational Psychology,* 73 (1)*,* 83-92.

Smith, D. G. (1977). 'College Classroom Interactions and Critical Thinking'. *Journal of Educational Psychology,* 69 (2)*,* 180-190.

Smith, D. G. (1981). *Instruction and outcomes in an undergraduate setting*. Paper presented at the meeting of the American Educational Research Association. Los Angeles, CA.

Sobocan, J. (2009). 'The Ontario Secondary School Literacy Test: Creative Higher-Order Thinking?'. In J. Sobocan & L. Groarke (Eds), *Critical Thinking Education and Assessment: Can Higher Order Thinking Be Tested?* (pp. 177-192). London, Ont.: Althouse Press.

Spady, W. G. (1994). *Outcomes-based Education: Critical issues and answers*. Arlington, VA: American Association of School Administrators.

Stake, R. E. (1995). *The art of case study research*. Thousand Oaks, CA ; London: Sage Publications.

Star, C. and Hammer, S. (2008). 'Teaching generic skills: eroding the higher purpose of universities, or an opportunity for renewal?'. *Oxford Review of Education,* 34 (2)*,* 237-251.

Stark, J. S., Lowther, M. A., Bentley, R. J., Ryan, M. P., Genthon, M. L., Martens, G. G. and Wren, P. A. (1990). *Planning Introductory College Courses: Influences on Faculty*. Ann Arbor, MI: University of Michigan, National Center for Research to Improve Postsecondary Teaching and Learning.

Steedle, J., Kugelmass, H. and Nemeth, A. (2010). 'What Do They Measure? Comparing Three Learning Outcomes Assessments'. *Change: The Magaizine of Higher Learning,* 42 (4)*,* 33-37.

Stein, B. S., Haynes, A. F. and Unterstein, J. (2003). *Assessing Critical Thinking Skills*. Paper presented at the SACS/COC Annual Meeting. Nashville, TN.

Stephens, D. (Ed.) (2009), *Higher Education and International Capacity Building: twenty-five years of higher education links*. Oxford: Symposium Books Ltd.

Sternberg, R. J. (1985). 'Critical Thinking: Its Nature, Measurement, and Improvement'. In F. R. Link (Ed.), *Essays on the Intellect* (pp. 45-66). Alexandria, VA: Association for Supervision and Curriculum Development.

Sternberg, R. J. and Berg, C. A. (1992). *Intellectual development*. Cambridge: Cambridge University Press.

Stiggins, R. J. (1987). 'The design and development of performance assessments'. *Educational Measurement: Issues and Practice,* 6*,* 33-42.

Strauss, A. L. and Corbin, J. M. (1998). *Basics of qualitative research : techniques and procedures for developing grounded theory*. (2nd Ed.). Thousand Oaks, Calif. ; London: Sage.

Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I. and Taggart, B. (2012). *Effective Pre-School, Primary and Secondary Education Project (EPPSE 3-14). Final Report from the Key Stage 3 Phase: Influences on Students' Development from age 11 - 14*. London: Institute of Education, University of London, and Birkbeck, University of London.

Tabachnick, B. G. and Fidell, L. S. (2007). *Using multivariate statistics*. (5th Ed.). Boston, MA ; London: Pearson/Allyn & Bacon.

Takano, Y. and Noda, A. (1993). 'A Temporary Decline of Thinking Ability during Foreign Language Processing'. *Journal of Cross-Cultural Psychology,* 24 (4)*,* 445-462.

Task Force on Higher Education and Society, World Bank and UNESCO. (2000). *Higher education in developing countries : peril and promise*. Washington, DC: World Bank.

Teferra, D. and Altbach, P. G. (2004). 'African higher education: Challenges for the 21st century'. *Higher Education,* 47, 21-50.

Teferra, D. and Knight, J. (2008). *Higher education in Africa : the international dimension*. Boston, MA: Center for International Higher Education, Lynch School of Education, Boston College, and Accra: Association of African Universities.

Temple, B. and Young, A. (2004). 'Qualitative Research and Translation Dilemmas'. *Qualitative Research,* 4 (2), 161-178.

Terenzini, P. T., Springer, L., Pascarella, E. T. and Nora, A. (1995). 'Influences Affecting the Development of Students' Critical Thinking Skills'. *Research in Higher Education,* 36 (1), 23-39.

Terenzini, P. T., Springer, L., Yaeger, P., Pascarella, E. T. and Nora, A. (1994). *The multiple influences on students' critical thinking skills*. Paper presented at the Association for the Study of Higher Education. Orlando, FL.

Tettey, W. J. and Partnership for Higher Education in Africa. (2009). *Deficits in academic staff capacity in Africa and challenges of developing and retaining the next generation of academics*. [Online]. Available at: http://www.foundation-partnership.org/pubs/pdf/Tettey_NGA_Indicators_Dec2009.pdf.

Thorndike, E. L. (1913). *Educational Psychology (Vols. 1 and 2)*. New York, NY: Columbia University Press.

Thorndike, E. L. and Woodworth, R. S. (1901). 'The influence of improvement in one mental function upon the efficiency of other functions'. *Psychological Review,* 8, 247-261.

Tikly, L. and Great Britain Department for International Development. (2003). *Globalisation and skills for development in Rwanda and Tanzania*. London: Department for International Development.

Tilley, S. A. (2003). '"Challenging" Research Practices: Turning a Critical Lens on the Work of Transcription'. *Qualitative Inquiry,* 9 (5), 750-773.

Tinto, V. (1975). 'Dropout from Higher Education: A Theoretical Synthesis of Recent Research'. *Review of Educational Research,* 45 (1), 89-125.

Tolmie, A., Muijs, D. and McAteer, E. (2011). *Quantitative methods in educational and social research using SPSS*. Maidenhead: Open University Press.

Tsui, L. (1998). Fostering critical thinking in college students: A mixed-methods study of influences inside and outside of the classroom. PhD thesis. Los Angeles, CA: University of California, Los Angeles.

Tsui, L. (2000). 'Effects of Campus Culture on Students' Critical Thinking'. *The Review of Higher Education,* 23 (4), 421-441.

Tsui, L. (2001). 'Faculty Attitudes and the Development of Students' Critical Thinking'. *The Journal of General Education,* 50 (1), 1-28.

Tsui, L. (2002). 'Fostering Critical Thinking through Effective Pedagogy: Evidence from Four Institutional Case Studies'. *The Journal of Higher Education,* 73 (6), 740-763.

Twale, D. and Sanders, C. S. (1999). 'Impact of Non-Classroom Experiences on Critical Thinking Ability'. *NASPA Journal,* 36 (2), 133-146.

Twinn, S. (1998). 'An analysis of the effectiveness of focus groups as a method of qualitative data collection with Chinese populations in nursing research'. *Journal of Advanced Nursing,* 28 (3), 654-661.

UNESCO (2009). *Final Report: World Conference on Higher Education*. Paris: UNESCO.

UNESCO, Association of African Universities and Priority Africa Programme. (1993). *Development of higher education in Africa : the African University: into the new millenium*. Paris: UNESCO.

UNESCO Regional Office for Education in Africa. (1992). *Higher education in Africa : trends and challenges for the 21st century*. Dakar: UNESCO.

Uvin, P. (1998). *Aiding violence : the development enterprise in Rwanda*. West Hartford, CT: Kumarian Press.

van den Bor, W. and Shute, J. C. M. (1991). 'Higher education in the Third World: status symbol or instrument for development?'. *Higher Education,* 22, 1-15.

Vogel, I. (2012). *Research Capacity Strengthening: Learning from Experience*. London: UK Collaborative on Development Sciences.

Vyas, S. and Kumaranayake, L. (2006). 'Constructing socio-economic status indices: how to use principal components analysis'. *Health Policy and Planning,* 21 (6), 459-468.

Vygotsky, L. S., Cole, M. and Luria, A. R. (1978). *Mind in society : the development of higher psychological processes*. Cambridge, MA ; London: Harvard University Press.

Wainer, H. and Thissen, D. (1994). 'On Examinee Choice in Educational Testing'. *Review of Educational Research,* 64 (1), 159-195.

Walker-Keleher, J. (2006). 'Reconceptualizing the Relationship between Conflict and Education: The Case of Rwanda'. *Praxis: The Fletcher Journal of Human Security,* 21, 35-53.

Wandira, A. (1977). *The African university in development*. Johannesburg: Ravan Press.

Wengraf, T. (2001). *Qualitative research interviewing: Biographic narrative and semi-structured methods*. London: Sage Publications.

Werner, P. H. (1991). 'The Ennis-Weir Critical Thinking Essay Test: An instrument for testing and teaching'. *Journal of Reading,* 34 (6), 494-495.

Wiggins, G. P. (1998). *Educative Assessment: Designing Assessments to Inform and Improve Student Performance*. San Francisco, CA: Jossey-Bass.

Williams, K., Wise, S. L. and West, R. F. (2001). *Multifaceted Measurement of Critical Thinking Skills in College Students*. Paper presented at the Annual Meeting of the American Educational Research Association. Seattle, WA.

Willis, K. (2006). 'Interviewing'. In V. Desai & R. B. Potter (Eds), *Doing Development Research* (pp. 144-152). London: Sage Publications.

Wiredu, K. (1980). *Philosophy and an African Culture*. London: Cambridge University Press.

Wober, M. (1972). 'Culture and the Concept of Intelligence: A Case in Uganda'. *Journal of Cross-Cultural Psychology,* 3 (3)*,* 327-328.

Wood, D., Bruner, J. S. and Ross, G. (1976). 'The Role of Tutoring in Problem Solving'. *Journal of Child Psychology and Psychiatry,* 17, 89-100.

World Bank. (1988). *Education in Sub-Saharan Africa : policies for adjustment, revitalization, and expansion*. Washington, DC: World Bank.

World Bank. (2002). *Constructing knowledge societies : new challenges for tertiary education*. Washington, DC: World Bank.

World Bank. (2004). *Education in Rwanda : rebalancing resources to accelerate post-conflict development and poverty reduction. A World Bank country study*. Washington, DC: World Bank.

World Bank. (2009). *Accelerating Catch-up: Tertiary Education for Growth in Sub-Saharan Africa*. Washington, DC: World Bank.

World Bank. (2010). *Financing higher education in Africa*. Washington, DC: World Bank.

World Bank (2011a). *The Little Data Book on Africa*. Washington, DC: World Bank.

World Bank. (2011b). World Bank EduStats.

World Bank. (2013). World Development Indicators.

World Bank and Salmi, J. (1994). *Higher education : the lessons of experience*. Washington, DC: World Bank.

Wright, S. P. (1992). 'Fostering Intellectual Development of Students in Professional Schools Through Interdisciplinary Coursework'. *Innovative Higher Education,* 16 (4)*,* 251-261.

Yaffee, R. A. (1998). *Enhancement of Reliability Analysis: Application of Intraclass Correlations with SPSS/Windows v.8*. New York, NY: New York University.

Yang, K.-S. (1988). 'Will Societal Modernization Eventually Eliminate Cross-Cultural Psychological Differences?'. In M. H. Bond (Ed.), *The Cross-Cultural Challenge to Social Psychology* (pp. 67-85). Newbury Park, CA: Sage Publications.

Yesufu, T. M. and Association of African Universities. (1973). *Creating the African university : emerging issues in the 1970's*. Ibadan ; London: Oxford University Press for the Association of African Universities.

Yin, R. K. (2003). *Case study research : design and methods*. (3rd Ed.). Thousand Oaks, CA ; London: Sage Publications.

Yin, R. K. (2009). *Case study research : design and methods*. (4th Ed.). Los Angeles, CA ; London: Sage Publications.

Zeleza, P. T. (2004). 'Neo-liberalism and academic freedom'. In P. T. Zeleza & A. O. Olukoshi (Eds), *African Universities in the Twenty-First Century. Volume 1: Liberalisation and Internationalisation.* (Vol. 1, pp. 42-68). Dakar: CODESRIA.

Zoyem, J.-P. (2010). *Employer and Graduate Survey: Final Report*. Huye: National University of Rwanda.