Explaining students' deep and surface approaches to studying through their interactions in a digital learning environment for mathematics

Maria Margeti

UCL

PhD thesis submitted for the degree of Doctor of Philosophy at University College London

Declaration of Authenticity

I, Maria Margeti, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed……………………………..

Acknowledgments

I wouldn't have been able to complete this piece of work without the great support and help of my husband Chris and the 'patience' of my son Alexander.

I would like to thank the module leader and the module team at the university where the study took place for their collaboration and for being receptive to new pedagogical ideas and applications. I would also like to thank: the team behind ActiveMath at DFKI, University of Saarbrücken, for their technical support and collaboration when conducting the study; Dr Peter Monthienvichienchai for great guidance and support when setting up and conducting the study with the DFKI team; and Dr Harvey Mellar for seeing through the MPhil/PhD transfer.

Finally, I would like to thank my supervisor, Dr Manolis Mavrikis, for taking over after a long PhD interruption, steering me towards the right direction and helping me with his constructive feedback to finish this piece of work. I will always appreciate his encouragement and invaluable support, which he gave me throughout the last crucial stages of this "Odyssey".

Abstract

This thesis presents the results of a study that embraces and tests Entwistle's theory of deep and surface approaches in relation to students’ interaction with a digital learning environment for mathematics, in real conditions, during tutorial sessions. In contrast to most of the work in the field that seeks ways of adapting a system to students’ specific learning styles, the aim is to find ways to support tutors and researchers to identify students’ prominent approach in order to ultimately encourage the adoption of a deep approach to studying while discouraging a surface approach.

To achieve this aim there is an in-depth examination of the relationship between the various scales and subscales of the Approaches and Study Skills Inventory for Students (ASSIST) and metrics occurring from the interaction in the digital learning environment ActiveMath. Furthermore, the potential influence of students’ prior knowledge in mathematics in “deep” and “surface” models is discussed. The results point to insights for tutors regarding identifying students’ deep and surface approaches from their interaction with the digital learning environment; suggestions regarding the design of features that encourage a deep approach to studying; and methodological recommendations for researchers regarding future studies which can help to distinguish further deep and surface approaches and to examine them in similar or different educational settings.

Table of Contents

List of Tables

Table of Figures

Chapter 1 - Introduction

Mathematics has a very important role in the curriculum of computer science courses. First-year computer science students need to know and use a considerable amount of mathematics. Osmon (2009) notes that the struggle with mathematics during the first-year of study in computer science courses is a widespread problem. His investigation of the mathematics qualifications on entry to university computer science courses showed that many students have no qualification beyond the GCSE level, despite the need for a better mathematical foundation in this type of course. He argues that these students typically disliked mathematics at school and tend to have a negative attitude to studying it at university when their interest is really in studying their major. However, computer science courses do require a higher foundation than that provided by GCSE, and so it falls to the universities to deal with the poor level of prior knowledge and the negative attitude towards mathematics of these students. Furthermore, at the computer science department of the university where the author works, there is an ongoing concern with failure rates in the mathematics module and a need to look for ways of providing greater support for first-year students with their mathematics.

Educators have given a lot of attention to finding ways to support first-year undergraduate students in their mathematics classes, turn their learning of mathematics into a positive experience, and improve students' mathematical understanding and performance (Sangwin, 2004; Perkin et al., 2013; Freeman et al., 2014). The author's own professional background lies in teaching in the areas of design and implementation of digital learning environments for a variety of subjects, and therefore an obvious line of enquiry for the author is how interactive learning environments might be used to address the aforementioned concerns.

The role of interactive learning environments in higher education -and in mathematical education- has been widely researched. As discussed in section 2.1.7, some authors have questioned the effectiveness of their use, and others have pointed towards moderate effectiveness, whereas others have argued that these systems can be effective by assisting towards creating a positive attitude, increasing students' engagement, encouraging the exploration of mathematical patterns and relationships, and generally improving performance. Those who argue that these systems are effective would nevertheless be clear that there remains much room for improvement. This improvement should start from increasing our understanding of how these systems are used, in order ultimately to design them better and increase their effectiveness.

The investigation into the use of interactive learning environments in mathematical education has generated research in relation to a wide range of theories and methodologies (see Hoyles and Noss, 2003). An area, which has been widely researched, is concerned with students' individual characteristics such as prior knowledge and goals, attitudes, beliefs, emotions, preferences, "styles" and motivation towards learning and mathematics. As discussed later on in sections 2.1.3-2.1.7, studies in this area have investigated whether there are connections between students' individual characteristics and their use of mathematical interactive learning environments (henceforth ILEs). This is, in essence, an effort to understand better how learners interact with them, and to identify ways to support students as well as provide information about their interaction to their tutors.

In order to understand students' interaction and support them based on their individual characteristics, one of the areas which has been involved in this type of investigation is the theory of "styles".

In the 1990's, there was a gradual increase in studies that examine various "style" constructs in relation to the way students use ILEs. The main aim of most studies was to eventually propose ILEs which can be adapted to suit students' "style"; a term which really depends on the context of the tradition is found, but in the majority of theoretical frameworks describes a behaviour that is sustained and thus repeated over time. However, in recent years there has been a waning in the popularity of applying "styles" in this context, as well as in the broader educational context. As discussed in detail in 2.6.1, this is because of certain criticisms that arise from such an application: whether the proposed measurements capture subtleties and complexities of individual human behaviour in real educational settings; whether they measure what they claim to measure; and whether their psychometric standards are sound. In addition, the wider debate on the wisdom of matching or mismatching style to instruction and the lack of solid empirical evidence about the effect on performance, led to the need for a different perspective in the context of ILEs and mathematical education.

The current investigation does not seek ways of "pigeonholing" the way students learn based on controlled experimental conditions, neither adapting content or interventions according to learning styles to increase students' performance. It rather attempts to explain students' interaction when using ILEs, in real learning conditions, in a way that can potentially help our understanding of how their approach to studying can improve. It ultimately aims to examine the potential of helping tutors identifying a prominent approach towards studying through students' interactions, and improving an ILE's design, so as to encourage students' adoption of a deep approach to studying and/or discourage a surface approach, based on the theoretical framework behind Entwistle's ASSIST construct. This choice is reinforced by evidence that adoption of a deep approach to studying by first-year undergraduate students correlates positively with their performance; whereas a surface approach correlates negatively to their performance (Entwistle and Ramsden, 1983; Tait and Entwistle, 1996). Also, the ideas of discouraging a surface approach and encouraging a deep approach towards studying are very relevant in the context of mathematical education. The issue of discouraging rote memorization of formulas and rules (as well as tendencies to treat concepts and methods as unrelated bits of knowledge) and encouraging a deep understanding of mathematics is quite prominent in particular in mathematics education (Crowe and Zand, 2000b; Saha et al., 2015). It has been observed that students often carry out mathematical procedures without any understanding of the concepts involved and they tend to follow a surface rather than a deep approach towards studying (Liston and O'Donoghue, 2009). So, investigating deep and surface approaches towards studying has the potential to address the issue of discouraging a surface approach while encouraging a deep one in university courses which involve mathematics.

This new perspective requires conducting the investigation for interactions in ILEs for mathematics in real learning conditions. In this way, it is possible to reveal subtleties and complexities of interactions in ILEs with regards to deep and surface approaches, which can only manifest themselves in real educational settings. It can ultimately offer recommendations for more supportive learning environments inside the class for tutors. Finally, it can be a starting point for future studies in this context as it has the potential to reveal what sort of methodological recommendations can be made, and what sort of improvements can further take place in terms of designing ILEs and capturing data.

Chapter 2 examines the literature on the effectiveness of systems for mathematical education and students' interactions in using these systems. It also provides a critical examination of the three main traditions of “style” theory: the learning, the cognitive and the personality tradition, as it is generally considered important to have a full understanding of the theory, before proceeding with any choices in terms of theoretical framework (Ali et al., 2014). There is also discussion on the existing empirical evidence about the relationships between “styles” and students’ performance and interaction when practising in tutoring systems; and the influence of individual characteristics, which are not included within the concept of “style”, such as prior knowledge. Finally, there is an in-depth analysis on how methodological issues and the need for a new perspective, when examining individual differences in the context of students’ interaction in ILEs in mathematical education, led to the involvement of Entwistle’s (1997) deep and surface approaches towards studying in the current investigation.

Chapter 3 sets out the research questions and the methodology of this study. In particular, it provides a closer examination of the ASSIST instrument, outlines the thinking behind the selection of appropriate and relevant “interaction” metrics with regards to the students’ deep and surface approaches towards studying. The chapter also describes the study’s design, data collection techniques and the procedures followed as well as the strategy for the development of multiple regression models for both “deep” and “surface” scales.

Chapter 4 presents an analysis of the data which involves both the correlational and “multiple regression” statistics of the “deep” and “surface” scales and the “interaction” metrics; and an individual interpretation for the suggested regression models representing the “deep” and “surface” scales.

Chapter 5 presents a general discussion with comparisons between the “deep” and “surface” models; and a detailed discussion on the influence of prior knowledge in all “deep” and “surface” models.

Chapter 6 presents a summary of the principal arguments of the thesis and discusses the contribution to knowledge of the thesis. Finally, in Chapter 7, there is discussion as to the implications of the work and potential areas for future studies.

Chapter 2 - Literature

This chapter first examines the literature in relation to the impact of interactive learning environments (ILEs) for mathematics education and their use and common interactions observed. The second section looks at the learning, cognitive and personality styles as possible choices in terms of expressing individual differences. The third section critically reviews mathematical ILEs and examines students' performance and interactions in interactive learning environments. The fourth section discusses the relevance of "styles" in the context of interactive learning environments. While the intension is not to provide an exhaustive coverage of the vast field of "styles" and their use in interactive learning environments, relevant aspects to the current investigation are discussed in a way that will help shaping the aim and research questions. The fifth section examines the influence of non-style factors on the relationship between "styles" and students' interactions. Lastly, the sixth and seven sections present how the insights of the literature shaped the research inquiry of the current thesis.

2.1 Students' interactions in systems for mathematical education

This section (2.1) examines a number of computer systems currently used to support teaching and learning in undergraduate mathematical classes (with specific focus on those that are used in algebra as this is central to the computer-science undergraduate curriculum); and the existing research on students' interactions when using these systems.

2.1.1 Introduction – Systems for mathematical education

Literature in ILEs does not clearly indicate a set of criteria that an interactive learning environment should adhere to. Part of the reason is that there is a great variety of ILEs, even from the earlier days of their use in classroom, which offer a different combination of capabilities and affordances. Another reason is that there does not seem to be a common consensus amongst researchers, tutors, and practitioners in the field of mathematical education, as they concentrate on different learning needs and preferences, individual research interests, and educational needs.

However, the author has identified certain common patterns of criteria across a number of reviews dating from the time the data collection took place. These criteria have their roots on ILE aspects or characteristics discussed in early reviews, such the one by Crowe and Zand (2000b) (a review that was at the time this research started perhaps the most exhaustive and comprehensive review and taxonomy on mathematical ILEs). As indicated later on, these criteria with regards to mathematical ILEs still carry weight in more recent critiques, and the reason for looking into them is not only to understand what different types of ILEs can offer but ultimately to inform the decision of which ILE to chose for the current research.

Furthermore, these criteria mainly stem from pedagogical theories. For example, they stem from the idea that a mathematical ILE has the potential to help towards re-construction of knowledge, encouraging a constructivist approach by providing ways for further investigation and organisation of concepts (Crowe and Zand, 2000b; Thomas and Holton, 2003; McDonald, 2016). Another idea is that mathematical ILEs have also the potential to provide communication and collaboration amongst students, hence encouraging a more social-cultural approach (Crowe and Zand, 2000b). As indicated later on, a variety of elements in an ILE which support collaboration and communication are also suggested across different types of ILEs and in more recent critiques (Borba et al., 2013; Joshi, 2017; Artigue, 2013).

Finally, the purpose of this section is not to provide an exhaustive review of related ILEs, but rather a more ‘inclusive’ discussion based on criteria, which will help towards clarifying the final choice of ILE for the current research enquiry. It is worth mentioning, however, that the following criteria are not a panacea and there are practical issues to consider which can lead tutors and practitioners to be more pragmatic with regards to their ILE choices.

2.1.1.1 Criteria

1) Multimodality with particular emphasis on visualisation

Multimodality, and particularly visualisation, is a characteristic which is consistently discussed or mentioned across reviews and critiques in mathematical ILEs.

According to Borba et al. (2013) multimodality can be defined with regards to a mathematical ILE as a combination of media such as text, hyper links, video, images/graphics, diagrams, audio recordings, animations and graph plotters. Taleb et al. (2015) also support multimodality in “m-learning tools” with audio and visual mediums as a way to increase students’ comprehension, motivation and self-confidence. At the same time researchers on the field of mathematical

ILEs point out that multimodal content can add a level of complexity when designing ILEs (Borba et al., 2013). They further argue that “the biggest conceptual transition for e-learning designers is to envision the content and learning objectives through graphical imagery and user interactions rather than explaining the content through text”. Part of this challenge is providing a high quality in terms of mathematical symbols and graphics which are consistent with any textbook, but also to avoid potential excessive use of multi-media elements so they add value to the learning context without distracting (Engelbrecht and Harding, 2005a; Engelbrecht and Harding, 2005b; Borba et al., 2013). Further practical issues with regards to the implementation of ILEs are discussed further later on.

The emphasis in most reviews and critiques is on visualisation, as a desirable element in a variety of mathematical ILEs, such as web-based courses, m-learning tools and simulations, and through features such as graphics, diagrams, graph plotters, animated images and java applets (Engelbrecht and Harding, 2005a; Engelbrecht and Harding, 2005b; Taleb et al., 2015; Borba et al., 2013; Joshi, 2017; Juan et al., 2008; McDonald, 2016). Such “visualisation” features can enhance comprehension and motivation (Engelbrecht and Harding, 2005a; 2005b). Finally, one of the reasons visualisation is such an essential element, is because it contributes in presenting a mathematical content through multiple representations in variety of ILEs, an issue discussed in the following section.

2) Multiple representations of mathematical content in ILEs

According to Crowe and Zand (2000b), ILEs can serve towards creating representations of mental problems and facilitate their manipulation, as they have the capability, for example, to “transform symbolic-algebraic situations into spatial-geometric ones”. This is the case of graphing a function, where there is linking between numerical and graphical representations. Furthermore, the ability to make connections and translate properties across different representations is referred to as “representational versatility” or “representational fluency” (e.g. making connections amongst tables, graphs and expressions with regards to the concept of “function”) (Thomas and Holton, 2003).

The support for multiple representations and what they can offer across different mathematical ILEs spreads throughout older and more recent critiques.

Multiple representations, especially those involving visualisation, can complement the formal definitions and it is a way to “organise” mathematical concepts” or “re-organise” them, hence reconstructing knowledge (Crowe and Zand, 2000b; Artigue, 2013). Mathematical ILEs such as spreadsheets, interactive graph plotters, CAS, and geometric visualisation tools are good examples, where there is effective linking of multiple representations, as different views of same concept and their links can be viewed and investigated simultaneously, and understood further (Crowe and Zand, 2000b; Lagrange et al., 2003; Galbraith, 2006).

Lagrange et al. (2003) also emphasise the possibilities in the use of multiple representations involving visualisation from simple calculators (i.e. with numerical, graphical and symbolic capabilities) to CAS, web-based tutorials and more intelligent environments (i.e. with symbolic, geometric and graphical capabilities). They cautioned, however, with regards to the modifications, which “technological modifications” can bring to mathematical notions, pointing out that visualisation, for example does not make necessarily the learning easier but it can make it certainly richer by revealing the complexities of the symbolic aspects.

Finally, Thomas and Holton (2003) argue that ILEs, such as CAS, graph plotters, geometric visualisation tool, and web-based systems can support representational fluency or versatility and reduce fragmentation and compartmentalisation of knowledge, which are so common with the traditional use of algebraic symbolic language. An ILE can engage students in higher education with multiple representations, and enhance their mathematical understanding, however, it can be still difficult for students to think of concepts in the same versatile way as tutors do. There can be also other obstacles as an ILE can be still used in higher education in a way that focuses on the procedural aspects of mathematics rather than the conceptual ones. This points to the issue that the curriculum or learning content has to be designed or redesigned in a way which supports conceptual and multiple representational aspects before integrating it in an ILE; an issue which is discussed further later on.

3) Organisation/Structure of content in ILEs with regards to scaffolding

There are a number of suggestions and cautions as to how the learning content can be organised in an ILE in a way that supports and engages students. Some of these are related to the fact that mathematics is a tall subject. This means that it is relatively linear in that a concept is built upon prior concepts and its natural structure is more like scaffolding with many interconnected conceptual structures (Crowe and Zand, 2000b). The idea is that we can build the structure higher with new concepts as long as there are previous concepts reliably and solidly placed as “previous layers”.

ILEs should support effective scaffolding (McDonald, 2016). At the same time this can be challenging as ILEs can “liberate” the order in which topics are presented and students have the freedom to explore the concept in any order (without necessarily mastering the previous ones) (Crowe and Zand, 2000b).

A suggestion is that the tutor should give clear guidance with regards to the order in which topics are accessed in ILEs when using them in classroom, taking into consideration students’ characteristics such as prior knowledge; in essence relating scaffolding to students’ characteristics (Crowe and Zand, 2000b). This is also a capability that intelligent ILEs can do, without tutors’ intervention, with hints, directions and time-sensitive prompts (Kulik and Fletcher, 2016; VanLEHN, 2011; Scandura, 2012). However, non intelligent ILEs such as Khan’s academy can also achieve this by informing students’ guidance with self-diagnosed activities and tests, as shown in 6.4.2 (see also Appendix 6.3).

A structure with effective scaffolding should also facilitate the linking between new knowledge and prior knowledge, by allowing students to review their work regularly, and facilitating their revision (Galbraith and Haines, 1998; Joshi, 2017). There are features in ILEs, discussed later on, which facilitate students’ interaction with regards to these aspects.

Finally, effective scaffolding is also related to features in an ILE, which allow further investigation of concepts, as discussed in following section.

4) Investigation

Facilitating further investigation of mathematical concepts and procedures is an aspect in ILEs which is considered important in various reviews (Wong, 2003; Engelbrecht and Harding, 2005a; Engelbrecht and Harding, 2005b; Borba et al.; 2013; Joshi, 2017), because it can assist towards “reconstructing knowledge” (Crowe and Zand, 2000b) or “construct their own narrative” (Laurillard, 2002).

There are suggestions for a number of ways to apply the “investigative” aspect in mathematical ILEs.

First of all the “investigative” aspect is related to multiple representations. More specifically, features which allow the manipulation of representations (e.g.

graph plotters and CAS) facilitate students' investigation and understanding of the nature of their connections (Crowe and Zand, 2000b; Joshi, 2017).

Secondly, this aspect can be applied by facilitating the investigation of both concepts and procedures and how they are linked, helping in this way towards both procedural and conceptual understanding (i.e. proceptual) (Crowe and Zand, 2000b; Thomas and Holton, 2003; Engelbrecht and Harding, 2005a; Engelbrecht and Harding, 2005b).

Thirdly, this aspect can be applied by allowing the search of concepts and procedures both within the learning content of an ILE and beyond, using "search" type of features, hyper links within the learning material to relate ideas, as well as external links to internet resources for mathematics (Engelbrecht and Harding, 2005a; Engelbrecht and Harding, 2005b; Joshi, 2017). According to Borba et al. (2013), exploring internet resources beyond the ILE, will require first for a student to do "a good deal of scaffolding" in order to have a focused search and find relevant data. They also argue that there is an issue with regards to the authenticity and the veracity of the search results. However, it seems that an Internet search is inevitable during a learning situation as a problem or question sparks most of the times an Internet search. This trend is identified by experts in the field of mathematical ILEs, such as Artigue (2013), as "Google reflex".

5) Amplifier of calculations

ILEs can play an enabling role by performing calculations which would take time and effort to do them by hand (Crowe and Zand, 2000b). However, there is no really common agreement as to whether ILEs should play an enabling role (Crowe and Zand, 2000b), also this is not a characteristic which all ILEs have. Part of the problem is that there is no common agreement as to what "necessary algorithmic skills" mean and therefore there is lack of consensus as to what calculations should be done by hand and what calculations can be left to the ILE (Crowe and Zand, 2000b).

Those in favour of using ILEs with this characteristic, from simple calculators to CAS, argue that it allows students more time to concentrate on more important tasks such as: formulating a problem; deeper understanding of mathematical concepts; and solving a complicated problem (e.g. non-linear algebraic calculations) without tracing an algorithmic solution and surpassing in this way the need for rote learning (Crowe and Zand, 2000b; Juan et al., 2008).

There are those however, who are more cautious about leaving the calculations to an ILE, pointing to the issue of "black box". Crowe and Zand (2000b) argue that if students leave the calculations to an ILE, then they may apply the results without checking whether the answer makes sense. They also support evidence which shows that: some student do not progress beyond the "black box level of utilisation"; and from a cognitive aspect, the use of such ILEs can make the "mind rely heavily on tools" and can hinder problem solving skills. Taking these concerns into consideration, it is the author's opinion that the issue of the black box relates ultimately to the way students approach their studying and the use of ILEs. As Artigue (2013) indicated students can use technology as a white box when they ask technologies to carry out mathematical processes they are familiar with or they have mastered; however, they can also use it as a black box when they ask technologies to carry out for them mathematical processes which they do not know. It is reasonable to also indicate that the tutor's intervention and guidance and the way the ILE is allowed to be used in class plays also a role as to the "black box" or "white box" utilisation of an ILE.

6) Communication and Collaboration

Finding ways to facilitate communication and collaboration amongst students and between tutors and students through ILEs can encourage a social-cultural approach in a learning situation; an approach which is valued by experts in the field mathematical ILEs (such as Crowe and Zand, 2000b; Lagrange et al., 2003; Engelbrecht and Harding, 2005a; Engelbrecht and Harding, 2005b; Borba et al., 2013; Joshi, 2017; Taleb et al., 2015). In these reviews, experts argue that integrating collaborative aspects in ILEs can help engagement and deeper understanding, as students are actively involved in discussing maths, giving and receiving feedback, and explaining their thinking process with regards to problem solving strategies. However, advocates of the social aspect such as Engelbrecht and Harding (2005a) at the same time caution that students need clear guidance with regards to their collaboration and communication and cannot simply generate their own ways.

There seems to be a variety ILEs in which this aspect is incorporated either specifically for mathematics, or for a variety of subjects such as Moodle and Blackboard (Juan et al., 2008; Artigue, 2013). There is particular emphasis on features that can be offered in web-based mathematical ILEs in order to encourage collaboration and/or communication. Reviews by Borba et al. (2013), Galbraith (2006), and Thomas and Holton (2003) indicate "collaborative" features that enable sharing annotations with regards to ideas, questions, and feedback; and direct communication (e.g. chatrooms).

Finally, it is worth mentioning that latest developments in mathematical ILEs such as m-learning tools facilitate the collaborative aspects (Joshi, 2017).

7) Providing mathematical activity with support.

There is a great variety of suggestions and opinions with regards to how activities in mathematical ILEs can be more supportive of students' learning. According to Crowe and Zand (2000b), supportive mathematical activity requires first a certain degree of acquaintance with symbolic notation, hence there should be an effort to provide relevant support. Juan et al. (2008) note that this is especially important for adult students who may have not used mathematical notation for a long time, so there should be facilitation of revision of previous concepts and notations they are supposed to know. The degree of acquaintance differs across different ILEs. As discussed later, effective activity when using CAS, for example, requires much more than a minimum level of acquaintance with symbolic notation (Galbraith and Haines, 1998). Experts in mathematical ILEs, such as Artigue (2013) and Keady et al. (2006) also emphasise the need for guidance as to how the ILE works but also for creating intentional and organised activity.

Furthermore, there are propositions with regards to web-based ILEs to facilitate a range of activities which are built on existing knowledge and which encourage, motivate, and shape interactivity with a learning purpose (Engelbrecht and Harding, 2005a; 2005b). More recent reviews such as those by Joshi (2017) and Taleb (2015) emphasise that this can be achieved by including ILE features which: allow individual preferences to be considered, give students control and flexibility with regards to learning activities, and allow third party recognition such as rewards.

Depending on the type of ILE there is a variety of activities suggested for supporting and engaging students when practicing exercises such as: interactive quizzes based on a pool of questions which facilitates unlimited practice and revision; automated feedback; model answers which allow students to compare against their own solutions; editable notepad to encourage students to articulate their findings (Crowe and Zand, 2000b; Laurillard, 2002; Engelbrecht and Harding, 2005a; Engelbrecht and Harding, 2005b; Galbraith and Haines, 1998; Sangwin, 2004). In addition, as discussed earlier on, some features support activities related to the characteristics of visualisation, multiple representations, investigation, and communication and collaboration, mentioned earlier on (e.g. mathematical forums, activities involving graph plotters, “search” features) (Sun et al., 2018; Wong, 2003; Joshi, 2017; Thomas and Holton, 2003). Designing activities which support the aforementioned ILE characteristics can lead towards a more reflective approach when students are practicing, rather than adopting “trial and error strategies” or “fishing behaviour” Wong (2003).

Finally, there are criticisms regarding the way mathematical activities are designed and delivered in ILEs. For example, a common criticism is related to the integration of mathematical notation in an ILE. Engelbrecht and Harding (2005a) and Juan et al. (2008) note that in online type of assessments typing mathematical symbols and interpreting mathematical symbolism in students' answers can be problematic. They further argue that poor integration of mathematical notation in an ILE can cause students to find the communication of mathematical concepts tedious. Hoyles and Noss (2003) also point out that syntax and semantics of an ILE is also something that requires time for students to find out how it works. However, as indicated later on, in ILEs such as CAS, there are solutions which can address these issues.

2.1.1.2 Practical Issues - Being pragmatic

The above criteria are commonly observed across a number of reviews; however, they are no guarantee for resulting in effective learning and teaching. As a recent meta-analyisis on studies, conducted before and after 2010 on the use of ILEs by Drijvers (2016), shows, it is not easy to prove the beneficial use of ILEs, as there are many factors involved, and when there is a positive impact, the effects are small. It makes sense to assume that digital technology is not a panacea and its effectiveness can depend on particular implementations and situations, such as educational setting, orchestration by the tutors, use of digital tools based on pedagogical theories and practices (Drijvers, 2016).

At this point, it is also worth making a clarification. It is not only a matter of the ILS to support the aforementioned criteria; the curriculum of a course (i.e. learning content) should support them as well. Thomas and Holton (2003) argue that, for example, tertiary mathematics curriculum is not always designed to engage students with inter-representational thinking, and it is usually procedural in nature. When integration of an ILE is considered at classroom, they point out to the importance of rethinking the curriculum in terms of content and structure and even suggesting changing tutors' teaching style. However, this is not always feasible and it cannot be always fully implemented. Thomas and Holton (2003) gave an example of integrating first-year curriculum in CAS for a linear algebra, where to fully integrate it would require changing the order, trivialising or making redundant topics, and adding new topics; something that at the end could not be fully implemented due to due to "lack of congruency and faculty support".

It is worth mentioning that it is not always the case that the curriculum has to drastically change in order to be fully integrated in an ILE. It is possible to have an ILE which has features supporting the aforementioned criteria, while at the same time follows a course's curriculum. This can be a reasonable solution given that a university's policy (i.e. regulations) does not always allow drastic changes in a course's curriculum, or it is not always practically feasible to re-design the curriculum of a course in order to integrate it in an ILE.

There are a number of experts in the field pointing to practical issues preventing from integration or full integration across different types of ILEs, such as CAS and web-based courses. Reviews by Keady et al. (2006), Hoyles and Noss (2003), Juan et al. (2008), Thomas and Holton (2003), and Joshi (2017) point to financial and time-related issues which require: human resources for development and sustaining ILEs; time and money for training staff on using ILEs in classrooms; and time and money for setting up properly the technical infrastructure. Thomas and Holton (2003) in particular discuss how both students and staff should be allowed time to built "familiarity, suitability and expertise" when working with a medium. Other reviews by Joshi (2017) and Lavicza (2010) also discuss policy-related issues, for example, integration of ILEs may require changes in a university department's policy and regulations, hence management or teaching or administrative staff may be unsupportive or unwilling of such changes.

Finally, while it is essential to perform further research when integrating the curriculum in an ILE, as Artigue (2013) points out, it is still a challenge to have "well-developed and tested curriculum" in an ILE. Holyles and Noss (2003) also supports that the aforementioned practical issues may prohibit research in ILEs.

2.1.1.3 ILEs

At the time the current research was conducted, a variety of ILE systems for learning and teaching was used in undergraduate mathematical classes (see reviews by: Crowe and Zand (2000a) and (2000b), Crowe and Zand (2001), Handal and Herrington (2003), Surowiec (2004), Engelbrecht and Harding (2005a), Keady et al. (2006), and Thomas and Holton (2003)).

In this section, there was specific focus on the systems that seemed to be most commonly used in algebra, as this was central in the mathematics undergraduate curriculum and the most relevant to the current investigation at the time the data collection was taking place. These systems were: computer algebra systems (CAS), geometric visualisation environments and graph plotters, assessment systems, and interactive tutorials[1]. More specifically, the following sections discuss common activities and interactions related to the aforementioned criteria in these systems. It is worth clarifying that, intelligent tutoring systems are not discussed. The intention was simply to explore systems in which students' behaviour was not altered by intelligent help or guidance, as this was thought to serve better the research purposes of the current investigation[2].

1) Algebraic manipulation system - Computer Algebra System (CAS)

Computer Algebra Systems have been quite widely used, and the most commonly used[3] CAS are: Maxima, Mathematica and Maple (Crowe and Zand, 2001; Keady et al., 2006; Thomas and Holton, 2003).

The primary function of CAS is the manipulation of mathematical expressions in symbolic form. More specifically, a typical CAS interface provides users with a type of text-editor in which they can type commands, variables and operators (Pratap, 2006). In this way users can perform symbolic algebraic computations on linear and non-linear equations, multiplication of matrices etc. (Daku, 2006). A secondary function of CAS is graphing. Besides plotting graphs, tutors can also create more interactive environments for visualisation such as animated graphs and graph plotters (Thomas and Holton, 2003). Hence, with regards to the aforementioned characteristics, CAS can be used as an amplifier for calculations, but also serves as a visualisation tool and can facilitate the manipulation of multiple representations (Thomas and Holton, 2003). From this aspect it also serves the "investigative" aspect as it facilitates students' investigation and understanding of the nature of the connections amongst these

[1] At this point it is worth mentioning that the aforementioned reviews also show that these systems can be combined.

[2] As indicated in the methodology (see sections 3.6. and 3.7), the study takes place in real learning conditions, and we are after the participants genuine interactions. In this way both deep and surface approaches may manifest themselves in a genuine manner and this will serve better the aim which is establishing connections between approaches to studying and students' interactions (see section 2.7.5).

[3] At least that was the case at the time the current research was taking place.

multiple representations (Crowe and Zand, 2000b; Crowe and Zand, 2001).

However, there are mainly issues regarding its use, which may influence the effectiveness of mathematical activities and require further support from tutors and institutions. Starting with practical issues, CAS does not simply require a minimum level of knowledge of symbolic notation but also symbolic input in order to perform activities and internal processes (Crowe and Zand, 2000b). Galbraith (2006) discusses in depth a case where students' performance was influenced by how well students know the syntax of Maple. Hence, there is the aforementioned need to allow time for familiarity and expertise of working in the specific medium. In addition, Engelbrecht and Harding (2005a) also note that students typing mathematical symbols and CAS can be also problematic, however according to them there is a further issue with the interpretation of mathematical symbolism in students' answers. However, they cite a solution suggested by Sangwin (2004), according to which it is possible to interpret the students' input in a test[4].

Furthermore, it requires for the curriculum to be redesigned which may cause policy-related issues and lack of institutional support (Lavicza, 2010; Thomas and Holton, 2003). For example, content favoured by CAS may be trivial in nature and of limited educational value, so new topics are included, or topics should be taught in a different order (Wong, 2003; Thomas and Holton, 2003; Crowe and Zand, 2000b).

Finally, it is reasonable that these aforementioned practical issues may require a certain degree of financial investment in terms of human resources for development and training. Another thing to consider is the cost surrounding the use of CAS (Keady et al., 2006), as with the exception of Maxima, CAS such as Maple and Mathematica are not open source.

2) Geometric visualisation environments and graph plotters

In this section, the intention is to discuss ILEs whose primary characteristic is visualisation, and which they have been primarily used to support teaching in geometry and graphing. At the time the current research was taking place the

---

[4] The solution which is currently offered is via the computer aided assessment package (CAA) called STACK. While it provides questions for the quiz of learning management system Moodle, it makes use of a CAS to evaluate the mathematical expressions (see: https://moodle.org/plugins/qtype_stack). The logic and 'mechanics' behind CAA underpinned by CA packages is discussed thoroughly in the papers by Keady (2006) and Sangwin et al. (2010).

most commonly discussed ILEs of this type were: Dynamic Geometry (DG) environment such as Cabri Geometry and GeoGebra, and graph plotters (Crowe and Zand, 2001; Thomas and Holton, 2003; Hoyles and Noss, 2003).

DGS has been used increasingly in classroom to support teaching and learning of geometry (Hoyles and Noss, 2003). However, DG such as GeoGebra along with geometry has also incorporated a number of other topics and aspects related to higher education such as graphing, calculus, and algebra[5]; hence it has been used for all levels of education, including tertiary-level mathematics (Abdulwahed et al., 2012; Lavicza, 2010). From a practical aspect, GeoGebra also provides an authoring environment which supports tutors towards creating interactive material.

A graph plotter is a typical example of a multiple representation tool and it is widely used in undergraduate math courses. Graph plotters can represent functions in a graphical, numerical and symbolic form; and typical examples of graph plotters are provided by Kleitman (2010b) and Waner (2010).

Both DG environments and graph plotters are primarily used to enhance visualisation (Lagrange et al., 2003), and they may assist the students in their learning by showing them different representations of the same concept simultaneously . Students, for example, can manipulate those representations and hopefully come to understand the connections between graphic and numerical representation (Engelbrecht and Harding, 2005a; Crowe and Zand, 2000a).

In this regard, they also serve the “investigative” aspect as they facilitate students’ investigation and understanding of the nature of the connections amongst these multiple representations (Borba et al., 2013; Abdulwahed et al., 2012; Thomas and Holton, 2003; Joshi, 2017). Borba et al. (2013) specifically refers to observations in class which indicate that students understand better equations by exploring their graphs and trying to make sense of the relationships between equations and graphs. For this to occur successfully, there should be also supportive interactive activities involving a number of representations which expose students to conceptual processes and objects, as well as procedural skills (Thomas and Holton, 2003). In this respect, DG environments can provide direct feedback that can serve as a basis for reflection and further support in activities (Thomas and Holton, 2003; Hoyles

[5] (see: https://www.geogebra.org/about)

and Noss, 2003). At the same time, both Hoyles and Noss (2003) and Thomas and Holton (2003) discuss that students may encounter difficulties in terms of interpretation and construction. These issues however, can be resolved in classroom with tutors' intervention. In addition, the combination of CAA system and DG environment can provide further support in activities, for example, by not just comparing students' answer to model answers, but by also indicating "in what aspects and to what extend a wrong answer deviates from the correct one" (Sangwin et al., 2010).

3) Assessment

There are a number of online assessments, exercises and quizzes for undergraduates in mathematics, as shown by Saab (1999), which they are used to assist students with their preparation for an undergraduate course or with their undergraduate studies (without being part of one of the ILEs mentioned in this section). The ones with diagnostic purpose can assist tutors to test students' prior knowledge (Crowe and Zand, 2001). A typical example of a diagnostic test is the one provided by the University of California[6]. Furthermore, revision type of tests in previous concepts and notations at the start of a course, may help towards more effective mathematical activity as it requires first a certain degree of acquaintance with symbolic notation. This may help, for example , adults who may have not used mathematical notation in a long time, and as discussed in section 2.1.1 *7)* , and who they may need further support. Overall in this type of assessments, students can usually respond by clicking on multiple-choice and multiple selections, or typing the answer in a fill-in-the-blank with numeric and numeric-plus-strings (e.g. algebraic expressions) type of response (Crowe and Zand, 2001).

In relation to feedback, there are tests which communicate feedback (or hints about the correct answer) when answering each question; and tests which communicate the feedback for all questions after submitting the whole test (Stroppel et al., 2007). Depending on what learning objective they serve, feedback can be of formative or summative nature, and while the degree of support and guidance provided in feedback may differ according to that, the idea of giving feedback which does not only allow for comparisons to model answer but also reveals how far away students' answer is from model answer (see Sangwin et. al., 2010), is a way towards a more constructive and helpful

[6] (see: https://www.math.ucla.edu/ugrad/diagnostic)

approach.

Finally, there are tests which allow them to try a specific number of times and tests which allow them to try until they get the correct answer (Narciss, 2007) or improve their score (Crowe and Zand, 2001). From a practical perspective this means generating a test from a ‘pool’ of questions and providing automated feedback (Crowe and Zand, 2001). To support further tutors with practical aspects of implementation, there are suggestions such as sharing databases with questions and CAA modules (Keady et al., 2006).

4) Interactive tutorials

Based on the definitions of tutorial systems by Alesi and Trollip (2001) and Handal and Herrington (2003), “tutorials”[7] refer to a system which assists the learning of mathematical concepts and procedures by offering instruction and practice with exercises.

As discussed in 2.1.1.1 *3)*, maths is a tall subject and effective scaffolding should be supported in ILEs. In interactive tutorials, the order of topics is evident through menus and content pages, which underline the linear structure of the math curriculum. Simple examples are provided by Kleitman (2010a) and Seward and Puckett (2009).

At the same time, there are interactive tutorials which can “liberate” the order on which the topics are presented, encouraging a more “investigative” approach, as discussed in section 2.1.1.1 *4)*. This can be achieved both by providing “search” type of features, hyper links within the learning material to relate ideas, but also by providing features which allow the manipulation of multiple representations through graph plotters and function evaluators (see the interactive tutorial provided by Waner (2007)).

Furthermore, the characteristics of visualisation and multimodality are served through animations, graphics, graph plotters, and function evaluators (see interactive tutorials by Waner (2007) and Husch (2001a, 2001b, 2001c)).

Despite that the above examples serve the aforementioned criteria, students might feel overwhelmed by the sheer volume of material provided and also let their investigation into the material go too far. As discussed in 2.1.1.1 *3)*, students need to do first a good deal of scaffolding, and also as discussed in

---

[7] However, to avoid confusion with the tutorials which are taking place in class this type of ILEs will be referred as “interactive tutorials” in the thesis.

2.1.1.1 *6)* and 2.1.1.1 *7)*, they need a good deal of support and guidance. If an ILE is used in class, then tutors need to point towards the right direction in terms of: visiting or revising a previous concept or process; communicating the intentional activity for the session and how to approach it; giving further feedback; directing the students' investigation, etc. However, if this is not the case, then the ILE should provide this support during activities in class.

A current interactive tutorial, which according to author's opinion has managed to serve well the aforementioned criteria of "investigation", "visualization", "multimodality" with a variety of features and activities, while at the same time there is effective scaffolding, communication, guidance and support, is Khan's Academy. However, this is discussed further in section 6.4.2 as it has been evolved the last years from an interactive tutorial to an open course type of environment (i.e. MOOC).

2.1.1.4 ActiveMath (AM) – Justification

The empirical aspects of the present study will be carried out using the ActiveMath system, which is based on the work of the DFKI team of the University of Saabrucken (Melis et al., 2006). It is essentially the type of interactive learning environment (ILE) which is a web-based "interactive tutorial".

At the time the data collection was taking place, ActiveMath was considered a good choice because:

1. It had or allowed elements and features which served the criteria mentioned in section 2.1.1.1
2. It was a pragmatic solution when considering practical aspects discussed in section 2.1.1.2
3. It could serve methodological aspects (keeps logs of users actions and serve theoretical assumptions in relation to the chosen pedagogical theory).

Starting from the first point, it had the potential to serve reasonably well the criteria of:

- Multiple representation (e.g. connecting numerical, symbolic and graphical aspects of a function) and visualisation via graphics and a graph plotter
- Investigation through a search-type of feature and hyper links for further exploration of mathematical concepts

- Organisation as it provides consistently a table of contents (TOC) and previous/next buttons, which shows clearly the order of topics and allows to follow a linear structure, hence facilitating scaffolding
- Communication and collaboration as it provides the feature of "interactive notes", similarly to the idea of sharing annotations discussed in section 2.1.1.1 *6)* (i.e. the notes can be made public allowing in this way a possible collaboration amongst students and communication with tutor)
- Mathematical activity with a reasonable degree of support, as it is possible to integrate a course's curriculum (e.g. definitions, examples of mathematical concepts and procedures, model answers, and different types of mathematical exercises for which students can get feedback). Furthermore, students can annotate private or public notes for each exercise, similarly to the idea of "editable notepad" mentioned in 2.1.1.1 *7)*. Also, it is possible to combine exercises with the use of graph plotter, encouraging the further exploration of multiple representational aspects of an exercise.

  There will be more detailed discussion in section 3.10.1.4 with regards to its interface and its features.

With regards to the second point of practical considerations, ActiveMath had the potential to serve as an "empty shell", with the aforementioned features included; that is an "empty shell" which could be adapted to the given curriculum and the learning outcomes of a maths module. Based on what was discussed in 2.1.1.2, and given that that its use in classroom would be in real conditions (during the actual tutorial sessions) and was destined for research purposes, it was deemed important, that its integration would not trigger changes at a university department's policy and regulations; hence causing management or teaching or administrative staff being unsupportive or unwilling of its use. In other words, it was important that its integration would not require severe re-design of existing curriculum (such as changing the order of topics, trivialising or making redundant topics, and/or adding new topics).

In terms of financial and time related issues, although the author would have to contribute a significant amount of time adopting AM to a given curriculum (see section 3.10.1.3 and 3.10.1.4), there would be no financial cost for its licence, or for its further development and its maintenance, as these aspects would be supported by the AM team. Furthermore, it was deemed important that training the teaching staff or students for its use would not require a significant amount of time, hence AM was deemed appropriate as it provided the familiar interface of a website with typical, and relatively simple, navigational and “activity” features, as discussed in more detail in 3.10.1.4.

At the same time, ILEs such as CAS were not deemed as a good solution. It was thought, for example, that their integration would interfere with the given structure and content of a course’s curriculum to the point of requiring changes in a university’s processes or regulations. It would also require more time to train staff and students for their use in classroom (e.g. with regards to its syntax), as indicated in 2.1.1.3. It would also require to rely on the university’s human resources and technical infrastructure, for its development and maintenance. In addition, it was thought at the time of the current investigation that the integration of a graph plotter, designed and implemented specifically for the needs of the specific learning material would cover well the criterion of multiple representations; hence the use of a DG environment was not deemed necessary. Finally, other interactive tutorials described in 2.1.1.3 could not offer the methodological and practical aspects discussed below.

With regards to the third point of methodological aspects, AM was also deemed appropriate. As discussed later on in 3.10.1.4, the use of features such as “notes”, “search”, “hyper links”, “TOC”, and “previous/next” contributes towards forming theoretical associations with aspects of a specific pedagogical theory (i.e. studying approaches), hence forming relevant assumptions. These are just some examples, however, other elements of AM (such as allowing students not to follow the given structure or allowing them to try a specific number of times to solve exercises or cancel them) also allowed to form relevant assumptions, as shown later on in sections 4.1.1 - 4.1.10. Furthermore, the implementation of AM was based on an XML mark-up which could reveal semantic information with regards to the content of the pages visited (i.e. definitions, examples, theory, and exercises), hence allowing the calculation of more refined “interaction” metrics and offering the potential of making associations between, for example, time spent on a specific type of pages and aspects of the chosen pedagogical theory. Also, from a practical perspective, the fact that the team behind AM had the infrastructure, expertise and human resources to help with the implementation and do the action analysis (i.e. recording of web logs and calculation of required metrics) free of charge, influenced also this choice.

Finally as discussed in section 3.10.3.11, the choice of AM served ethical issues. For example, one of the main reasons AM was chosen was because it could help collect data in an unobtrusive way (e.g. recording web logs). Another ethical issue was avoiding setting tasks that would disrupt the learning process, hence it was reasonable to choose an ILE such as AM, which would not require significant amount of training for both students and staff.

2.1.2 Students' interactions when using mathematics education systems

From the description of the above systems, it is evident that students need to make a number of decisions when studying in these systems. As supported by Heid and Todd (2001), Weigand and Weller (2001), and Borba (2009), the way students work in mathematics is likely to change when using computer systems. The next section presents research on students' interactions in mathematical ILEs. However, because there are various empirical findings as to whether the use of mathematical ILEs actually benefits achievement in mathematics, the discussion starts with an examination of relevant empirical investigations.

Starting from the general effect of the use of technology on achievement in higher education, a meta-analysis by Schmid et al. (2009) reveal that there is an overall average low to moderate statistically significant effect size of 0.28. In general, their method seems to follow what is required in order to obtain internal and external validity. Their result is based on the results of a representative sample of 231 studies conducted in 1990 and onwards in higher education (and after reviewing 491 studies and selecting the ones that provide sufficient statistical information).

The selected studies vary in terms of their research design, the type of technology used (e.g. systems with and without "cognitive support"), and the "intensity of their use" (e.g. whether there is use of more than one systems; the number of their features and functions; the duration and frequency of their use). Schmid et al. (2009) conduct further analysis to examine the impact of each of the above factors. Their results indicate that for systems which provide cognitive support (that is, systems that "guide, and extend the thinking processes of their users" (Jonassen, 1994)) there is a moderate effect size of 0.41; whereas the intensity of use has a low effect size.

The effect of technology, therefore, depends heavily on the way and the conditions in which the technology is used in a higher education class; consequently the results of the effect on achievement may vary. This seems to be also true for the effect of the use of systems for mathematical education on achievement in a higher education mathematics class.

There is a number of studies that suggest that the use of systems for mathematical education does not make any difference in relation to students' performance (Jenks and Springer, 2002; Buteau and Muller, 2006). On the other hand, there are more positive views and empirical evidence about the effect of systems for mathematical education on performance. Khoju et al. (2005), Hagerty and Smith (2005), Tokpah (2008), and Saha et al. (2015) argue that the use of systems for mathematical education influence significantly the students' performance in mathematics compared to the use of traditional (non-computer) learning material. In terms of the use of graph plotters in class, researchers such as Crowe and Zand (2000b), Goulding and Kyriacou (2007), and Hong and Thomas (2015) argue that it may be possible to enhance understanding by manipulating and exploring further the output (e.g. by scaling, zooming, moving, rotating, and reading data from it) and that this should be encouraged in the class.

Furthermore, meta-analyses conducted by Schenker (2007), and Rosen and Salomon (2007) found effect sizes in the region of 0.23 and 0.46, respectively, on achievement. This difference can be explained if we consider that the meta-analysis conducted by Schenker (2007) simply compares the use or non-use of mathematical systems, whereas in the meta-analysis by Rosen and Salomon (2007) it is the type of instructional method rather than the sheer existence of computer use that resulted in a moderate effect size on performance. The way and the conditions, therefore, in which the mathematical systems are used can make a difference; a conclusion which is in accordance with the findings of Schmid et al. (2009) about the general effect of the use of technology in education.

These findings indicate a mixed image on the effectiveness of mathematical ILEs, so this leaves room for improvement. Later on in section 2.1.7, it is discussed that to improve the effectiveness of mathematical ILEs, it is worth looking at how students interact with the ILEs and how their individual characteristics affect their interactions.

The following sections examine literature describing some common student interactions when they use mathematical systems such as: interactions in low and high level control mathematical systems, interactions when they practise exercises, and degree of computer and mathematics interaction.

2.1.3 Students' interactions and low and high degrees of control

The degree of freedom a system allows learners when interacting with it varies. The degree of freedom in a system can be in relation to the choice of content, the order in which learners view the topics, or the choice to follow or ignore the advice given by the system (Lunts, 1997).

Lunt (1997) reviews a number of studies which examine the effectiveness of systems with various degrees and types of control on students' performance. She concludes that there is lack of significant difference.

According to Lunt (1997) the lack of significant difference may be due to the fact that the studies she reviewed on low and high control systems do not consider individual characteristics. She argues that students' personalities, prior knowledge, abilities, goals, and preferences differ, and that these differences may affect the degree of control with which students are able to perform well.

In the context of ILE for mathematical education Lunt's argument seems reasonable. For example, students without prior knowledge may simply learn best when maths topics are viewed in the given order; especially since one of the characteristics of undergraduate mathematics is that it builds on previous concepts (Thurston, 1990). Differences in ability may also have an effect on performance depending on the degree of control an ILE provides. Li and Edmonds (2005), in their discussion, argue that when dealing with at-risk adult students in mathematics only low control systems with small sequences and drill and repetition can help improve their performance.

Finally, students' interaction with a low and a high control system may differ. Henry (1995) compared the way students interact in a low and high control ILE for mathematics. More specifically, he compared the way they move around the system in terms of the number of pages, the number of revisitations and the time spent on the system. His results reveal that between the low and high control systems, students who use the low control system spend significantly more time interacting with it compared to students who use the high control system. He claims that research into learner's characteristics may explain these differences further.

2.1.4 Students' interactions when carrying out practice on exercises

Empirical research into the way students work while practising on exercises in systems for mathematical education reveals that there are some common patterns of interactions. The reason behind this type of research is usually to examine whether the way the students work with these exercises actually differs compared to the way they work in a traditional session, and whether the way students practise in these systems leads to bad or good performance. This knowledge is important in order to improve the system's design or give appropriate guidance in class.

A common behaviour when practising exercises is "trial and error" where students try different solutions, but without using a systematic approach which would help them to improve with each attempt (Cazes et al., 2006; Berry et al., 2006). Berry et al. (2006) also observe some students using a more systematic approach, commonly called "trial and improvement", which eventually leads to better results.

Another common behaviour is "gaming the system" when students try to achieve good results by taking advantage of the system's feedback (Baker et al. 2008). "Gaming the system" has been linked to poor performance and has prompted researchers such as Cazes et al. (2006) to conduct relevant studies. They conduct a case study and use log files to record students' behaviour when practising their exercises in an ILE, they find that some students tend to click every checkbox within a set of multiple-choice answers until the system identifies a correct answer and allows students to advance. Based on these findings, Cazes et al. (2008) make suggestions about how tutors may deal with such behaviour in class.

Cazes et al. (2008) find patterns of interactions through a detailed recording of students' interaction with the ILE when practising; however, they do not attempt to explain this behaviour by linking it, for example, to students' individual characteristics. A reason for this may be simply that "gaming the systems" may be a natural part of the learning process and not necessarily something that can be dealt with. It is also possible that learners may apply compensatory strategies to simplify a task; a tactic which, according to Niederhauser (2007), is frequently applied in ILEs. On the other hand, the relationship between "gaming the system" and individual characteristics is examined thoroughly in the study of Baker et al (2008). Although their study concerns intelligent ILEs, which are not examined in this report, it is worth mentioning it briefly because they look into a fairly large number of individual characteristics such as goals, attitudes, beliefs, emotions and preferences towards learning and mathematics. More specifically, their results reveal that there are significant positive and fairly moderate correlations between "gaming" behaviour and the students' characteristics of dislike for maths, frustration with the software, and lack of educational self-drive.

2.1.5 Computer and mathematics interaction

According to Galbraith and Haines (1998) students indicating a high degree of computer and mathematics interaction (that is the extent to which mathematical thinking interacts with computer medium) have specific beliefs and attitudes towards computers and mathematics. For example, they believe that computers help their learning mainly through examples and by linking algebraic and geometric ideas, and that note-taking can complement the information given by the system.

Gomez-Chacon and Haines (2008) conduct a study in which the aim is to gain a better understanding of the affective and behavioural factors of undergraduate students in mathematical courses, in which systems such as graph plotters are used. In their discussion, they suggest that research on students' reactions and characteristics may provide information on how to best use mathematics tools and real world interfaces, evaluate computer courses and develop computer based curricula. Initially, they test the hypothesis that undergraduate students with high computer and mathematics interaction tend to link algebraic and geometric ideas, take notes, and review their activities within the system. Their results confirm that students with high interaction scores tend to link algebraic and geometric ideas.

2.1.6 Discussion on students' behaviour in systems for mathematical education and individual characteristics

Section 2.1.3 discusses that the effects of the use of these systems on performance are real but of medium strength (with effect sizes in the region of 0.3-0.4). In addition, based on what is discussed in 2.1.3 and 2.1.5, the systems on their own do not improve performance and it is rather the way they are used in class that can make a difference by providing additional and appropriate instructions, activities in the class (e.g. to prevent "gaming the system" etc.). Most importantly, this area of literature shows that the issue of students' characteristics resurfaces in most aspects of students' behaviour in using ILEs for mathematics, in an attempt to explain the behaviour and consequently make a system more effective. In previous sections, for example, we saw that students' interaction and practising in mathematical systems have been linked to their attitudes, beliefs, goals, preferences and strategies as a way to improve the design of the system.

Given that these ILEs are at best only moderately effective it is natural to ask how we might increase the effectiveness of these systems, and the impact of the individual differences on the effectiveness of these systems suggests that the study of individual differences may provide some answers to this question. However, there is a large amount of research in the area of students' characteristics, and it is reasonable to ask which research area of students' characteristics might be most useful in addressing this issue.

A research area that has a serious claim on offering further understanding on students' individual differences when using ILEs and that could possibly help make ILEs more effective is the area of learning styles. As shown, later on, in section 2.3, they have played a significant role in the context of more general research on ILEs. This is not necessarily the case in the context of interactions in mathematical ILEs, however, there can be potential connections between learning styles and students' interactions in mathematical ILEs. For example:

- Moreno (2002) shows empirically students with a reflective learning style can cope better with the cognitive load generated by the interaction with a mathematical ILE compared to those with an impulsive learning style.
- As shown in Henry's study (1995) there are temporal and revisitation metrics which can indicate differences in students' interaction in high and low control mathematical ILEs. This connection, between this type of metrics and learning styles, has been explored empirically in high and low level control ILEs in a number of studies as shown in the review by Chen and Macredie (2002).
- Connections between the "gaming" behaviour and learning styles are possible. More specifically, according to Entwistle et al. (1979) and Entwistle (1997b) students with a high degree of a surface approach to studying may well exhibit "gaming the system" behaviour because it allows them to appear to succeed without risking failure. In the context of students' interactions in an ILE, the behaviour of "gaming the system" has been linked empirically to anxiety about failing in the study of Baker et al. (2008), and it may have connections to Entwistle's learning style which is linked to anxiety.
- There can be a connection between learning styles and level of interaction in mathematical ILEs. According to the empirical observations of McCune (1998) students with a deep approach to studying tend to go vigorously through the learning material by making notes, and relating ideas in order to achieve personal understanding. It is possible that students intending to follow a deep approach show a similar type of active engagement with the learning material in a mathematical ILE, leading to a high degree of interaction. As shown in section 2.1.6, the study by Galbraith and Haines (1998) indicates that a high degree of

what they refer to as "computer and mathematics interaction" is related to linking algebraic and geometric ideas, and note-making.

These are some examples of possible connections between learning styles and the way students interact in systems for mathematical education. However, the area of learning styles is large with a number of theoretical frameworks, definitions and measurements. Before proceeding, therefore, with a more detailed account of existing empirical evidence of the relationship between learning style and students' interaction in ILEs, the intention is to examine the approaches taken by learning style theorists to explain individual differences through a critical examination of its main traditions.

2.2 'Style' traditions

One of the common criticisms of researchers who apply the 'style' theory is that they concentrate on applying a specific theoretical framework and measurement without really having a full understanding and overall view of the theory and its variety of theoretical frameworks, measurements, and definitions. This type of criticism one may find even in overall positive reviews with regards to applicability and usefulness the 'style' theory in education, such as the one by Ali et al. (2014). At this point, therefore, it is essential to acquire a more thorough understanding of the theory before making choices with regards to the theoretical framework the author intends to concentrate on.

According to the reviews of researchers such as Rayner and Riding (1997), Sadler-Smith (1997), Curry (1987), and Coffield et al. (2004), there are three main traditions in theorising about styles: the personality tradition, the cognitive tradition, and the learning tradition. These traditions derive from three different movements arising from the historical development of style. Behind these categorisations of styles, learning style theorists attempt to rationalise, clarify and therefore facilitate the application of style theory by identifying common characteristics and differences among existing style constructs. However, as the literature on style is vast, the intention is to concentrate only on the cognitive style tradition and the learning style tradition, as they seem to be the most empirically relevant in the context of students' interaction in ILEs (hence the personality tradition is excluded from the current literature review).

2.2.1 Cognitive tradition

The systematic development of styles in the cognitive tradition starts with a 30-

year period beginning in the 1940s where studies in cognition and perception are conducted by cognitive psychologists (Rayner and Riding, 1997) . It has its roots in early 20th-century German typological theories (Sternberg and Grigorenko, 2001). In particular the Gestalt school's research in perceptual differences inspired the first experiments on perceptual functioning in individuals when locating an upright object in space (Rayner and Riding, 1997).

Sternberg and Grigorenko (1997) argue that the interest of cognitive psychologists in style started when research in intelligence could not really explain the processes of generating individual differences. In other words, there was a need to understand cognitive functioning beyond the measurement of intelligence. Research then started to focus on identifying styles and dimensions of cognitive processing (processing and organising information, problem solving) and perceptual functioning (Sternberg and Grigorenko, 2001).

The cognitive tradition has produced a variety of style constructs in an attempt to capture individual differences in cognitive and perceptual functioning, generating at the same time a number of criticisms.

One of the main criticisms is the resemblance of cognitive tradition styles to abilities, while there should be a clear distinction. One of the reasons for this resemblance is that certain measurements, such as Witkin's field independence vs. field dependence, assess the maximum performance in the same way as the abilities tests (Sternberg and Grigorenko, 1997).

Another criticism is that in this great variety of constructs, different labels are used to represent the same dimensions. According to Riding (1997), the reason behind this phenomenon is that between the 1940s and 1970s researchers worked in isolation, developing fragmented constructs and measurements. It is not until the 1990s that efforts for organisation of styles in this tradition take place by researchers such as Riding and Cheema (1991).

The cognitive tradition starts as a way of expressing individual differences in the cognitive processes. However, in the 1970s and especially during the 1980s cognitive measurements begin to be used for educational purposes. The on-going desire to apply cognitive theory in educational context leads to the creation of new measurements and the birth of the learning tradition.

2.2.2 Learning tradition

The systematic development of this approach begins in the 1970s and involves theories of individual differences in learning processes (Riding and Cheema, 1991). Unlike the cognitive tradition, the constructs and measurements in the learning tradition have been created to serve exclusively educational purposes. Therefore, this tradition focuses on the impact of individual differences during learning activities and the development of new constructs that are based on pedagogical theories (Rayner and Riding, 1997; Sternberg and Grigorenko, 2001).

According to Sternberg and Grigorenko (2001), an influential pedagogical theory that formed conceptualisations for the styles of this tradition is the one proposed by Kolb in (1984). Coffield et al. (2004) also point out the pedagogical significance of Entwistle's theoretical framework in relation to approaches to studying.

Kolb's theory proposes a four-stage experiential learning cycle. It consists of concrete experience (being involved in experiences and dealing with real situations); reflective observation (understanding the meaning of ideas and situations by observing and reflecting on them); abstracting conceptualisation (using logic in order to form general theories, build concepts, manipulate abstract symbols); and active experimentation (testing of concepts in new situations with an emphasis in practical applications (Kolb, 1984). Kolb's theory formed styles and generated the Learning Style Inventory (LSI) measurement that proves to be extremely popular in various educational studies and applications.

Entwistle's construct is based mainly on the model of Marton and Säljö (1976). The learning style construct of Marton and Säljö (1976) arises from qualitative research. They conducted a qualitative analysis which indicated two terms: deep and surface. Initially, these terms represented approaches to reading (reading for meaning, and reading which focuses on the words and facts); however, these terms have been broadened to describe more general approaches to studying (Entwistle et al., 2001). Deep approach represents the intention to understand and analyse concepts; whereas surface approach represents the intention to complete the task with little personal engagement and memorising without reflecting (Entwistle et al., 2001). These two approaches are measured as separate scales in the Approaches and Study Skills Inventory for Students (ASSIST). Entwistle (1997) added also a third scale in ASSIST which measures a strategic approach to studying. The strategic approach represents the intention to achieve high academic performance through organised studying, time management, alertness to assessment demands and monitoring effectiveness.

Styles in the learning tradition have bonds with cognitive dimensions, at the same time they are strongly related, by definition, to the learning process. This leads to constructs that are influenced directly by the learning environment and which are very susceptible to change (Rayner and Riding, 1997).

There are various criticisms concerning the constructs of the learning tradition. The first criticism concerns the ambiguity in the terms “learning strategy” and “learning style” (Rayner and Riding, 1997). Strategy refers to an organised series of tactics based on attitudes and motives (where tactics are the procedure of deciding which processes to apply, in which order, and using what skills) (Eysenck, 1994). Therefore, strategies may be learned and developed in order to cope with learning tasks, compared to learning styles that are relatively quantitative (Riding, 1997). Similar observations are made by Pask (1976b, p.133) who notes that strategies “are induced by systematic enforcement”, whereas learning styles manifest themselves in more “relaxed” learning environments. So, learners can apply the learning strategy of utilising knowledge about their learning style in order to improve their performance by matching style to instruction; or style flexibility can be applied when choosing or developing an appropriate strategy and employing different tactics in a novel situation (Curry, 1999).

The popularity of the “styles” of the learning tradition in education has gradually increased due to their strong relevance to learning processes. Most of these theories start from the point of view of the student, rather than that of the researcher or the teacher and therefore, they offer a better understanding of the reality of student learning. Their ultimate goal also is to provide a theoretical and empirical rationale for practical solutions which may improve learning and teaching.

Furthermore, in learning style theories such as the one proposed by Entwistle, there is emphasis on the way content and context influences students’ approach to learning. Characterisation of a student’s approach to learning is linked to a specific subject, department, school, thus providing a more realistic view on what is going on in a real learning environment.

After reviewing the two traditions, it is evident that clarification of terminology is necessary. In the next section, the intention is to examine the differences between the terms “learning style” and “cognitive style”, what they represent, and their interrelationships.

2.2.3 Interrelationships and comparisons between terms and traditions

It is evident from the above reviews of the two traditions that clarification of terminology in style theory is not straightforward. In addition, there is not really a common agreement as to how the styles from the two traditions influence each other and how they differ.

Starting from the definition of the learning tradition, "learning style" is the pattern of behaviour related to acquiring, manipulating, and using knowledge in a learning environment or through a learning procedure. A learning environment or procedure consists of learning strategies and instructional preferences. This definition is derived from the definitions of Eysenck (1994) and Pask (1988), who relate learning style to habitual preferences and strategies.

"Cognitive style" has a broader definition; it is defined as the pattern of acquiring knowledge in general and not specifically in a learning environment. The acquisition of knowledge derives from internal cognitive processes such as processing, organising, representing, perceiving information, and problem solving (Riding, 1997; Liu and Ginther, 1999). Cognitive style differs from ability in that performance will increase as ability increases, whereas the same style can influence both negatively and positively the performance depending on the nature of the task (Curry, 1999; Lemire, 2000; Riding, 1997).

The key difference between learning style and cognitive style is that learning style, unlike cognitive style, is influenced by the learning environment. Cognitive style is fairly quantitative, i.e. more stable over time, compared to learning style the interaction of which with the learning environment makes it less stable (Riding and Cheema, 1991; Liu and Ginther, 1999). In addition, this key difference between cognitive and learning style seems to be supported by most researchers in the Delphi study conducted by Armstrong et al. (2011) where there is an effort to establish consensus amongst researchers with regards to the definitions and the differences between cognitive and learning style (although total consensus does not occur as it shows that 27% of the researchers still believe that a student's learning style is relatively stable).

There is a range of perspectives on the interrelationships of these two traditions. According to Sadler-Smith (1997), there is empirical evidence showing that there is no relationship between cognitive style and learning style. However, there are those who argue that conceptually learning style constructs may arise from aspects of cognition (Sternberg and Grigorenko, 1997).

So, these two traditions may interact, influence and even relate to each other, however they are very distinct and they should be represented accurately in the literature of styles by using the appropriate terminology. Awareness of the tradition behind a construct plays an important role in understanding the conceptual relationship between a style construct and other concepts such as students' behaviour and interactions in learning environments. It can also help in selecting the appropriate tradition and eventually the appropriate style measurement from that tradition for a study.

The next section will use these ideas in order to support a discussion of studies looking at students' interactions in ILEs.

2.3 Empirical evidence about styles and students' interaction when using ILEs

The concept of students' interaction in the studies discussed in this section, can refer to a number of things, for example:

- type of navigational options students select in order to move around in the ILE;
- how long they spent on the content of the ILE;
- how many pages and what type of pages (e.g. with more or less degree of detail) they visit and revisit;
- whether they follow the given structure of the content in the ILE or impose their own organisation.

A methodological design that has been commonly used in the context of styles and ILEs is to compare differences in students' performance across different matching conditions. This type of study follows an experimental quantitative design (e.g. true or quasi experiments) where researchers investigate the effect of matching condition on learning performance. Learners are divided into groups according to their style, then each group is assigned to different treatments (matched or mismatched) and at the end each group's learning performance is assessed. For example:

- Students' styles are matched to ILEs with different levels of control (different types of structures which allow more or less freedom of choice). Typical examples of such studies were conducted by: Lin and Davidson-Shivers (1996) who used Witkin's cognitive style construct (field dependence vs. field independence); Graff (2005) who used Allinson and Hayes' Cognitive Style Index (intuitive vs. analytical); Rasmussen and Davidson-Shivers (1998) who used Kolb's learning style construct LSI; and Dunser and Jirasko (2005) who used Felder and Silverman’s learning style construct (global vs. sequential dimension).

- Students’ styles are matched to either an ILE that provides an overview of the available topics before entering in to the details of each topic (breadth-first order), or to an ILE that allows students to go directly into the details of each topic (depth-first order). A typical example of such a study was conducted by Ford and Chen (2001) who used Riding’s cognitive style construct CSA (analytic vs. holist).

However, there is an ongoing debate as to whether matching or mismatching “style” to instruction is beneficial for the students. In a more general educational context, there is ongoing attention to the implications of matching style to instruction on performance. Researchers such as Witkin et al (1977), Dunn (2000), Riding (1997), and Myers-Briggs (1986) consider matching style to instruction an important factor for achievement. Another advocate of matching to instruction, Felder (1993) (cited in Coffield et al., 2014, p.122) claims based on empirical evidence in science courses that mismatching instruction to students’ “style” can result in lower grades and lack of intrinsic interest in the course material. However, Coffield et al. (2004), and Rohrer and Pasher (2012) (cited in Ali et al, 2014, p.83) argue that there is no empirical support for matching instruction to style. There are also researchers in the field of styles such as Curry (1990) and Moran (1991) who are critical of both matching and mismatching, and state that optimal results are not achieved when styles are systematically mismatched or matched to instructional methods.

The debate of the matching hypothesis does not only concern the evidence about immediate performance gains, but it also extends to wider pedagogical concerns. For example, some researchers support deliberate mismatch as a way to create more versatile learners (Coffield et al., 2004; Curry, 1990). A more recent review by Ali et al. (2014) in the field of “styles” also discusses that, while there is support for matching, there is also support for mismatching student’s “style” and instruction, as it can challenge students so they can overcome their weaknesses and prepare them better for real life. This position resonates also with researchers in higher education such as Entwistle and Peterson (2004) who urge educators not to hesitate to design an environment which make students “somewhat uncomfortable yet provide enough support for new strategies to be developed without undue anxiety”. On the other hand, there are other researchers who consider deliberate mismatch unethical and even catastrophic for students whose personality does not favour versatility (Miller, 1991).

Because of these pedagogical concerns and mixed image on the performance gains of matching instruction to style, it was decided that a “matching hypothesis” type of design would not be appropriate for the current investigation. Hence, there will not be further detailed discussion with regards to “matching hypothesis” studies.

However, there is type of methodological design which seems to be more relevant and appropriate for the current investigation. Relational design is also a widely used design, through which it is possible to investigate the relationship between style and students’ interaction. In this type of design, students’ interaction is related to their style as identified by style measurements. Since this type of design is more relevant to the current investigation, a detailed accounted is provided in the following sections, but it is restricted to university courses of any subject.

2.3.1 Style and students’ interactions in terms of navigational metrics

This section focuses on the empirical evidence that associates “style” to the use of navigational options presented in an ILE.

2.3.1.1 Maps and menus and their use

Use of navigational options has been also related to “styles” and particularly to the cognitive style constructs of Witkin’s GEFT (1971) (field dependence vs. field independence) and Riding’s CSA (2001) (analytic vs. holist).

More specifically, a study by Ford and Chen (2000) indicates that there is a weak, positive, and statistically significant relationship between the CSA analytic-holist dimension and the use of the index. This means that the closer the learners are to the analytic end of the dimension the more they tend to use the index (which allows them to navigate quickly to the exact piece of information they prefer to focus on). It was found, however, that the magnitude of the shared variance accounted for is very small (r^2=0.08). In relation to the use of the map, the results indicate a moderate, negative, and significant relationship between the CSA Analytic-Holist dimension and the use of the map. This means that the closer the learners are to the holist end of the dimension the more they tend to use the map (which allows them a conceptual overview of the learning content). The magnitude of the shared variance accounted for is small (r^2= 0.10). In terms of generalisability, there are limitations as the sample is a self-selected convenience sample and is relatively small (N=65).

Finally, a review by Chen and Macredie (2002) present studies where the use of navigational options is related to Witkin's cognitive styles (e.g. field independent tend to use more "search" and "index" options to find specific details and procedures in the content of the ILE). Again though, most studies rely on relatively small samples ranging from 5 to 63 participants.

2.3.1.2 Linearity and the use of previous/next buttons

A key aspect of a hypermedia design which is frequently investigated in relation to style is linearity. According to Botafogo et al. (1992), linearity indicates to what extent the hypermedia developer imposes order of reading, that is the trail of pages that must be viewed consecutively. There have been a number of studies which examine the relationship between cognitive style and path linearity when using an ILE. A review by Chen and Macredie (2002) presents studies which indicate, for example, that field dependent learners tend to take a more linear path in an ILE than field independent learners at the early and middle stages of the learning process.

Furthermore, there are parts of ILEs that can offer linear navigation with the help of navigational options such as previous/next buttons; thus, linearity in users' navigation is expressed through the use of these options. Chen and Ford (1998) examine in their study the relationship between Riding's cognitive styles and use of the previous/next buttons. The results show that there is a statistically significant, negative, moderate relationship between the values of analytic/non-analytic variable and the use of previous/next buttons. The magnitude of the shared variance accounted for is moderate (r^2=0.239). These results seem to support the hypothesis that learners classified as analytics demonstrate a less linear navigational behaviour than those classified as non-analytics; however, given the small size of the above study (N=20), definite conclusions cannot be drawn.

2.3.2 Style and students' interactions

Students' interactions when using ILEs also differs in terms of path length, how often pages are revisited, and how long they spend on different pages. This section looks at some of the studies that have examined the relationship of these interactions to style.

2.3.2.1 Path length, visits and revisits

Connections have been found between "path length" (total number of pages visited) and revisitation metrics (the number of times the same page was visited) and "styles" such as Riding's analytic vs. holist and Entwistle's learning style ASSIST.

A study by Chen and Ford (1998) indicates that there is a statistically significant, negative, high relationship between the analytic/non-analytic variable and the metric of path length. This means that learners classified as analytics are likely to perform a shorter path than those classified as non-analytic. The magnitude of the shared variance accounted for is moderate (r^2=0.31). In the same study, there are also some findings in relation to a revisitation metric called "pages duplicated". The findings show that there is a statistically significant, negative, high relationship between the analytic/non-analytic variable and the metric of revisitation. This means that learners classified as analytics are less likely to revisit than those classified as non-analytic. The magnitude of the shared variance accounted for is moderate (r^2=0.31).

Chen and Ford (1998) say that it is not clear how individual differences become evident in an ILE. They interpret the navigational behaviour of the analytic learners as a sign of more efficient learning with fewer page duplications and navigational moves, although this interpretation would be more convincing if it was found that analytic learners had visited more distinct pages compared to the non-analytic learners which was not the case in their study.

Measurements from the learning tradition have been also related to this aspect of students' interaction. In the study of Mimirinis and Dafoulas (2008), the path length and the number of pages visited in specific sections of the ILE are examined in relation to the deep and surface scales of ASSIST (see section 2.2.2). They conduct this study with final year computer-science students and their results indicate that:

- there are no correlations between the above metrics and the deep scale and its subscales.
- the higher students score on the surface subscale of unrelated memorising the more likely they are to visit the Content section (the theoretical section of the system). The magnitude of the shared variance accounted for is fairly moderate (r^2=0.16).

The size of their study is relatively small (N=31). However, their findings can give an indication of possible connections between path length and visitation metrics and Entwistle's surface scales and subscales. For example, students with high values on the unrelated memorising surface subscale simply follow the route of rote learning without the intention of personal understanding (Entwistle, 1998); hence, they may simply try to get through as much learning content as possible without necessarily reflecting on what needs to be understood further or what is required of them.

In an earlier study by Allinson (1992), it has been found that students (from art and science courses) with high scores on deep scale and low scores on the surface scale tend to go to more pages compared to those with low scores on the deep scale and high scores on the surface scale.

The results from these two studies about the association between deep and surface approaches and students' interaction are in disagreement. In the study by Miriminis and Dafoulas (2008) the higher students score in unrelated memorising –a surface subscale – the more pages they visit; whereas in the study of Allinson (1992) students with the high scores in the surface scale tend to visit fewer pages. This difference may be due to the educational context since the ASSIST scale is intended to be context specific – according to Entwistle (1997a) students should think of their answers in terms of the specific course. However, students may visit a lot of pages for different reasons, some because they tend to skim through the learning material in order to memorise and some because they want to elicit deeper meaning and understanding thus manifesting the same behaviour in ILEs but for different reasons.

The above findings and interpretations should be treated with caution given the small size of all three studies and the differences in context. However, they do suggest that the number of page visits and revisits to a certain section of an ILE may be associated with cognitive and learning styles.

2.3.2.2 Use of time

Measuring the time that students spend doing various activities in the system may help achieve more accurate interpretations of interaction. This will be particularly the case when viewing times are recorded in relation to sections or pages with specific characteristics that reveal the degree of depth in subject hierarchy, degree of detail of learning material, or type of learning material (e.g. exercise or theory).

Based on the notion that analytic learners tend to focus on details and procedures in a field (Sadler-Smith, 1997; Entwistle and Hanley, 1977), it might be expected that analytics are likely to spend more time on pages, sections or levels of the ILE that provide detailed learning material. This hypothesis is tested in the experimental correlational study of Ford and Chen (2000).

Ford and Chen (2000) examine the correlation between the scores on Riding's CSA Analytic-Holist dimension and measures of the time spent on the sections of the ILE with more and less detailed information about a subject. The results indicate that:

- There is a weak, positive, and statistically significant relationship between the score on the Analytic-Holist dimension and the overall time spent on the section "Detailed Techniques" (which means that the closer the learners are to the analytic end of the dimension the more time they are likely to spend on the "Detailed Techniques" section). The magnitude of the shared variance accounted for is very small (r^2=0.07).
- There is a moderate, negative, and statistically significant relationship between the score on the Analytic-Holist dimension and the score of the overall time spent on one of the higher levels of the subject hierarchy (which means that the closer the learners are to the analytic end of the dimension the less likely they are to use sections with less detailed information about the subject). The magnitude of the shared variance accounted for is very small (r^2=0.09).
- There is a moderate, positive, and statistically significant relationship between the score on the Analytic-Holist dimension and the score of the overall time spent on one of the lower levels of the subject hierarchy (which means that the closer the learners are to the analytic end of the dimension the more likely they are to use sections with more detailed information about the subject). The magnitude of the shared variance accounted for is small (r^2=0.14).

The direction and statistical significance of the above relationships supports to a certain degree the hypothesis that analytic learners are likely to spend more time on sections that cover more details about the learning material; although more empirical evidence would be needed to support the practical significance of the relationship.

The findings of this section indicate how cognitive and learning style can be associated with path length, visitation, use of navigational tools, revisitation and temporal metrics; although there should be careful interpretation of the results since practical significance ranges from weak to moderate. The empirical evidence, presented in this section, also shows that cognitive style constructs seem to be more widely used in this type of studies than learning styles.

2.4 The relevance of the traditions to a study of interaction in an ILE

There has been a range of empirical work, therefore, looking at the relationships between styles and student interaction when using ILEs. At this point, it is worth examining whether any of the traditions of work on style have particular relevance or appropriateness in the context of students' interaction in an ILE.

- Cognitive style tradition

We have seen there are strong conceptual connections between cognitive style measurements and certain aspects of students' interaction, and the empirical evidence in section 2.3 indicates the domination of cognitive style constructs in studies where different aspects of interaction in ILEs is involved. The theories behind the cognitive style constructs seem to fit in well conceptually with the different aspects of students' behaviour in a variety of ILEs. As shown in section 2.3, cognitive style measurements such as GEFT and CSA have been used widely in a variety of studies of different aspects of students' interaction in ILEs. There is also the Allinson and Hayes' CSI which, although it has not been used as extensively as the other two, has received positive criticism psychometrically (Coffield et al., 2004).

Finally, it seems that this type of early empirical work, discussed in section 2.3, has formed the foundation for further empirical research and application in the field of adaptive ILEs and cognitive style measurements. As shown in the reviews by Al-Azawei and Badii (2014) and Ali et al. (2014), this seems to be particularly the case for all three cognitive style measurements: Witkin's GEFT, Riding's CSA, and Allison and Haye's CSI.

- Learning style tradition

The learning tradition has been concerned both empirically and conceptually with the examination of the students' interactions in ILEs through the learning style measurement of Kolb and Entwistle.

With regards to Silverman's ILS, as mentioned in 2.3, an early study has indicated a possible connection to linearity. In Coffield et al. (2004) review, its potential impact as a leading learning style construct is not supported; however, recent reviews in the field of interactive learning environments show that ILS is amongst the most frequently used learning style measurements because it is considered easy to apply with regards adaptive interactive learning environments (Deborah et al., 2014; Al-Azawei and Badii, 2014).

There is a possible conceptual connection between Kolb's LSI and the level of control in an ILE, as mentioned in section 2.3. This early empirical work seems to have paved the way for further application. Kolb's construct is used in the field of adaptive ILEs (Drissi and Abdelkirim, 2012). However, LSI has been criticised for its lack of psychometric rigour (Ruble and Stout, 1994; Coffield et al., 2004).

Entwistle's learning style measurement ASSIST may be able to offer a better theoretical framework for examining styles in relation to students' behaviour in ILEs. This measurement has been involved in studies of relational design with regards to visitation as shown in 2.3.2.1 and has received a positive critique by sceptics of the style theory such as Coffield et al. (2004).

According to Curry (1987), Coffield et al. (2004), and Eysenck (1994) the ASSIST deep subscales "relating ideas" and "use of evidence" are based on Pask's learning styles holist and serialist, respectively. Students with a "relating ideas" approach tend to build up a broad view of the learning task and impose their own organisation on content (Entwistle, 1997a; Entwistle, 1981; Entwistle et al., 1979). On the other hand, students with a "use of evidence" approach tend to build up meaning from the details and prefer a linear sequence when studying the learning content (Entwistle, 1997a; Entwistle et al., 1979).

The ASSIST learning style measurement is based on an extensive research on the way students approach their studying in real conditions. Entwistle (1981) and Entwistle and Ramsden (1983) through the discussion of their data (mainly interviews) provide a rich insight into issues such as: the amount of time that students with certain approaches are likely to spend on their studying; their intention to rehearse and repeat in order to memorise; their intention to follow the given structure in the learning content as defined by the lecturer.

To give some examples, approaches of the ways in which ASSIST scales may connect to behaviour when using ILEs:

- Students with an “unrelated memorising” approach to studying tend to repeat the learning content in order to memorise (Entwistle, 1981), and this could be linked to revisitation of pages in an ILE.
- Students with a “fear of failure” approach to studying –which relates to anxiety- tend to work slowly (Entwistle, 1981), so it is possible that this can be connected to time spent on the ILE.
- Students with a “seeking meaning” approach - intention to understand the learning content for oneself - tend to go through the learning content fairly slowly, and impose their own structure on the content they are given (Entwistle and Ramsden, 1983). It is possible that the “seeking meaning” approach can be linked to the time spent on an ILE, to the degree of linearity in a hypermedia design and to students’ navigation in the system.

Entwistle (1981) and Entwistle and Ramsden (1983) have connected their data and discussion on the way students approach studying to the subject being studied. They discuss and comment, for example, on the way science students approach problem solving or the way psychology students approach the writing of their essays. ASSIST considers how style may affect students’ behaviour in a specific learning environment, in a specific subject area and university course, making the interpretations of the findings difficult to generalise but more accurate.

The above discussion indicates that from the learning tradition ASSIST is a sophisticated measurement which has been well researched.

2.5 The influence of non-style factors on student interaction in ILEs

Coffield et al. (2004) point out that style should not be considered as the only predictor of performance and behaviour. They suggest that “learners are not alike nor do they live out their lives in psychological laboratories…[and that factors such as] age, gender and race and class, all interact to influence their attitudes to learning” (Coffield et al., 2004, p.142). Furthermore, in the field of navigational behaviour, Scherly et al. (2000) suggest that users’ characteristics should be taken into account; and Ford and Chen (2001) also argue that the failure to take into account background variables may result in findings appearing to be significant when in fact they are not.

In the majority of studies discussed in 2.3, the background variables of age, gender, and prior knowledge are the ones that are most often examined as likely additional potential factors of influence on the relationship between style and interaction. However, with regards to age the population of the current investigation concerns first-year undergraduates, so the age of participants is not likely to be varied. With regards to gender, there are indications of a relationship between gender and interaction in ILEs (e.g. temporal metrics) (Chen and Ford, 1998; Fiorina et al., 2007). There seem to be, however, few findings from which concrete conclusions about the connection between gender and interaction in ILEs can be drawn.

Amongst these background variables, therefore it makes sense to focus on prior knowledge, which seems to have stronger influence and relevance in the context of the current investigation. First, the involvement of prior knowledge is highly relevant in the context of mathematics education. There is a notable struggle in mathematics classes during the first-year of study in computer science courses, because of the poor level of entry qualifications of some students in those courses (Osmon, 2009). Secondly, it can influence the way students interact in ILEs, as shown in the following section.

2.5.1 Prior knowledge of subject area

Prior knowledge may enable a faster engagement with and understanding of a subject and this can be reflected on the way students interact with ILEs. For example:

- Chen and Paul (2003) suggest that more knowledgeable learners may impose their own structure simply because they understand better the conceptual structure of the subject matter. In terms of empirical evidence, there are only few studies, in which the main focus of the investigation is the relationship of style and navigation, and where prior knowledge is also investigated in detail.
- Chen and Ford (1998) and Chen and Ford (2000) examine correlations between prior knowledge in subject area to interaction metrics such as temporal ones, revisitation ones, and path length. The strength of most of these correlations varies from low to moderate. There is, also, a fairly high and significant correlation between prior knowledge of subject area and depth (i.e. the more knowledge of the subject area learners have, the deeper they are likely to navigate into the hypermedia hierarchy).

Furthermore, these early studies have formed the foundation for further application of prior knowledge in the field of adaptive ILEs. Mampadi and Mokodedi (2012) present a good overview of prior knowledge’s involvement in this type of application and empirical work. In their study, they adapt the ILE for experts and novices with regards to link hiding, provision of tips, annotated tips and navigational tools. Their results indicate that novices benefit more compared to experts in terms of the difference between the pre-test and post-test scores from this adaptation.

The above empirical findings indicate that when examining the relationship between style and students' interaction in ILEs, prior knowledge seems to be a factor which is worth considering for further investigation. A student with a tendency towards a specific "style" can interact with an ILE in a specific way, albeit because of it, but its interactions are likely to be also influenced by the prior knowledge on a subject.

2.6 How the Literature Review helps shaping the research enquiry

The literature in this chapter has indicated possible connections between "style" and students' use of ILEs. At this stage, it is discussed how addressing certain mainly methodological issues and criticisms in the field of "style" helps in shaping further the research enquiry in the current investigation.

2.6.1 Issues in the measurement of "style"

If one is to use a quantitative research design to address the research aims, similarly to studies discussed in 2.3, then we need to be able to measure style. In the literature review, the general term "style" is used to refer to a behaviour that is sustained and thus repeated over time and has cognitive and learning aspects. Two issues arise at this point. First whether these aspects can be measured, and second whether they measure what they claim to measure.

In relation to constructs coming from the cognitive tradition, well-known reviews such as those of Coffield et al. (2004), Curry (1983), De Bello (1990), Riding and Cheema (1991), and Ehrman (1990) argue that the cognitive aspect can be measured. The main debate on measuring this aspect is whether existing instruments measure the cognitive aspect of style or whether they measure ability; an issue that is briefly discussed in 2.2.1 and is analysed further below.

In relation to the learning tradition, there is a greater debate on whether style can be measured. According to Coffield et al. (2004), the opponents of the idea of measuring the learning aspect of style dispute the objectivity of such results, because of the subjective judgements that students make about themselves.

However, a review by Al-Azawei et al. (2014) indicates that the alternative automated, data-driven, methods of identifying students' style also has disadvantages as there are difficulties when measuring and interpreting students' interactions in ILEs. To conclude, although self-report measurements have their limitations both in general educational context and in that of ILEs, alternative methods have not provided a significant advantage. A self-reported measurement which occurs from solid educational background can be an appropriate and satisfactory solution as a quantitative data collection method, if it has been designed in a way that can be linked to subtleties in students' interaction without oversimplifying.

Another general criticism is that learning style measurements cannot capture the subtleties and complexities of the individual human behaviour (Coffield et al., 2004). Prominent theorists in the learning style tradition, such as Entwistle et al. (2001), not only encounter the debate between quantitative and qualitative methods when identifying approaches towards studying, but they also find ways to address it.

Entwistle et al. (2001) notice that qualitative research methods, such as interviews and observation into everyday studying, were needed to complement or even counter the way that measurements oversimplify the complexity of studying in different environments. On the other hand, they also notice that by simplifying these complexities, they could identify constructs that provide a precise language to describe and discuss everyday studying. Such a construct may also be easier to link to educational practice. As a result, Entwistle's learning style construct encompasses both qualitative and quantitative research methods by drawing ideas from two models: Holist-Serialist of Pask (1976) and deep-surface of Marton and Säljö (1976b). Pask's construct is based on quantitative methods and demonstrates students' consistency in experimental situations and normal studying from which his measurement derived. Marton and Säljö use qualitative studies to show evidence of variability, where students adapt their approaches according to the demands of the specific task. Based on this qualitative evidence Marton and Säljö produce eventually a theoretical construct.

The above insights raise the issue of consistency versus variability of the learning aspect of style (as represented respectively by the constructs of Pask and Marton and Säljö). One of the questions in the field of learning styles is whether it is possible to measure the learning aspect of style if it is not stable. Entwistle et al. (1979, p.367) argue that "students will exhibit sufficient consistency in intention and process across broadly similar academic tasks to justify measuring it as a dimension.". At the same time, there is certainly an interaction between an individual's learning style, the nature of the academic task, and the whole learning environment, so consistency and variation in learning approaches can manifest simultaneously, (Entwistle et al, 2001). Based on these ideas, Entwistle and his team produced the Approaches and Study Skills Inventory for Students (ASSIST) (1997).

To return to the initial argument, a learning style construct that proposes a qualitative method is a possible way of identifying the learning aspect of style. The qualitative methods used in the model of Marton and Säljö are well documented and can be reproduced. However, quantitative methods can offer a precise language to discuss everyday learning and at the same time offer an easier way to link to educational practice.

The above criticisms of quantitative methods of identifying style should not be ignored. Similarly to Entwistle's approach, finding ways to endorse and address the above conclusions and criticisms on quantitative methods can be constructive for the current investigation**.** Whilst "style" is defined as a behaviour that is sustained and thus repeated over time, its sustainability may be a characteristic of the cognitive aspect of style, but, in relation to its learning aspect, there are certain constraints. The learning aspect of style may be sustained and repeated, but this is likely to happen across the same academic tasks, subject area, and in general similar learning environments, as suggested by Entwistle (2001).

2.6.2 "Style" measurements and psychometric rigour

In section 2.4, it is argued that in both traditions there are measurements which can have empirical relevance or conceptual relevance or both to students' behaviour in ILEs. The empirical account in section 2.3 and the discussion in section 2.4 gave an insight into representative examples of style measurements from each tradition that:

- are most frequently selected for studies where students' interaction in learning environments is examined, such as Witkin's, and Riding's cognitive style measurements, Kolb's LSI, and Felder and Silverman's ILS learning style measurements;

- are used less frequently but may offer a more well-grounded and realistic learning perspective in this type of studies, such as the Entwistle's learning style measurement;
- are used less frequently in this type of studies but have sound psychometric properties, such as the Allinson and Hayes' CSI;

However, independently of their frequent use and possible links in terms of student's interaction in ILEs, the psychometric rigour of learning style measurements was (when the current investigation started) and still is a major concern as shown in reviews such as the one by Coffield et al. (2004) and Ali et al. (2014). Therefore, besides the empirical and conceptual relevance, it is also important to select measurements with reasonable psychometric properties.

2.6.2.1 Validity and reliability issues

According to Curry (1990), Lemire (2000), Coffield et al. (2004), and Ali et al. (2014) the lack of psychometric rigour is one of the main criticisms the style constructs face. In these reviews, the most frequently proposed types of validity and reliability are: construct validity, concurrent validity, test-retest reliability and internal consistency. In terms of validity, construct validity is supported by all the reviews and critiques on styles as the one criterion which researchers should examine before selecting a style measurement for any type of study. In addition, concurrent validity is used in detailed reviews, such as the one by Coffield et al. (2004), to support the meaning of the construct based on the comparison and correlation with another one. In terms of reliability, both test-retest and internal consistency are used and it is argued that they are necessary criteria for the selection of style measurements.

2.6.2.2 Practical Considerations

Another criterion for selection is whether a style measurement is practically manageable. In other words, whether it can be easily administered, scored, interpreted and also appropriately documented, as suggested in a review by James and Blank (1993). Practical considerations such as the above may not seem as important as the psychometric rigour of a measurement; however, they should be considered, since they may influence the organisation and the results of a study and therefore its validity and reliability.

2.6.2.3 A brief critique on psychometric rigour

With regards to the psychometric rigour and practical considerations of the aforementioned “style” measurements, the main issues for each measurement are discussed briefly here:

- Witkin’s GEFT (1971): There are no really issues with its internal consistency, reliability, and administration (Curry, 1987). However, there are serious issues with its construct validity, as it has been found to have correlations to ability, so its unclear relationship with intelligence and ability has generated a lot of controversy (Sternberg and Grigorenko, 1997; Coffield et al., 2004).
- Riding’s CSA (2001): While its construct validity is acceptable and its administration is relatively easy, there are major concerns with its test-retest reliability (Coffield et al., 2004). Issues with its reliability also raised in the rather “positive” for “styles” review by Al-Azawei et al. (2014).
- Allinson and Hayes’ (CSI) (1997): Its validity and reliability are considered solid and its administration can be conducted easily (Coffield et al., 2004).
- Kolb’s Learning Style Inventory (LSI) (1985; 1999): There are mixed reviews with regards to its psychometric rigour. Its construct validity and its reliability, both internal consistency and test-retest, are considered poor by certain reviews (Ruble and Stout, 1994); while its administration is considered easy (James and Blank, 1993). Issues with its construct validity are also raised by more recent review such as the one by Ali et al. (2014); whereas in the review by Al-Azawei et al. (2014) its validity and reliability is supported.
- Felder and Silverman’s (ILS) (1996): There are mixed reviews with regards to its validity and reliability as shown in the review by Al-Azawei and Badii (2014). It is relatively easy to administer (Deborah et al., 2014).
- Entwistle’s ASSIST (1997a): its construct validity and internal consistency are considered good based on independent evaluations, but there is a need for independent evaluation of its test-retest reliability (Coffield et al., 2004).

It seems that amongst potential “candidates” for involvement in studies of students’ interaction in learning environments, the relatively more frequently used “style” measurements such as Witkin’s GEFT, Riding’s CSA, Kolb’s LSI, and Felder and Silverman’s ILS have received either poor or mixed evaluations with regards to their psychometric rigour. With regards to the relatively less frequently used measurements: Entwistle’s ASSIST has received a relatively good evaluation and Allinson and Hayes’ (CSI) a particularly good one.

2.6.3 The choice of measurement

In current review, there are examples of style measurements which were prominent when the current investigation started in the field of students' interaction in learning environments (albeit some more than others). In choosing a "style" measurement to explore students' interaction in interactive learning environments for mathematics in higher education, there is a greater need to find one which:

- Addresses, relatively well, criticisms and issues raised in the field of "styles"
- Show promise of conceptual and/or empirical relevance in the field of students' interactions in interactive learning environments

A "style" measurement which can address the above issues reasonably well is Entwistle's ASSIST.

ASSIST can be relevant, appropriate and promising for investigating students' interactions in a learning environment, particularly when this takes place in a real learning context. It is a measurement which is based on a construct which is well-grounded in an educational context, as it is based on longitudinal educational research, spanning over four decades, and which has been conducted in a realistic learning context in higher education. The extensive qualitative research conducted to develop ASSIST provides a rich account of students' interactions in class: the way students approach their studying, in terms of the amount of time that students are likely to spend on their studying, their intention to rehearse the learning content, the amount of learning content they go through, their intention to follow the given structure of the learning content, and the way they practise their exercises. This type of research has been conducted through observations, interviews and eventually through the earlier and current versions of ASSIST, and in a variety of courses across different institutions in higher education.

It is reasonable to say that a similar research with the specific construct has rarely been applied in the context of interactive learning environments. ASSIST has not really been used frequently in a digital learning context for a specific course. As discussed in 2.3 and 2.4, in comparison there are more prominent and frequently used constructs and measurements in the field of students' interaction in learning environments from both the cognitive and learning tradition, such as Witkin's GEFT, Riding's CSA, and Kolb's LSI, and Felder and Silverman's ILS. In addition, as shown in 2.6.2.3, in comparison to Entwistle's ASSIST, Allison and Hayes' CSI seems to have received better evaluation in terms of psychometric rigour (specifically with regards to test-retest reliability). However, there are specific reasons for which ASSIST is deemed a better choice for the current investigation compared to the aforementioned measurements:

- There is no issue of a close relationship with ability like with some "style" measurements from the cognitive tradition (see 2.6.2.3).
- As discussed in 2.6.1, its varied educational research background shows that it can capture the complexity of studying in different environments, without oversimplifying by simply labelling students independently of the learning environment, task or subject area. It addresses reasonably, therefore the issue of consistency versus variability by acknowledging that the sustainability of the learning style depends on the academic task, subject area, and learning environment.
- The rationale behind the development of ASSIST, and in general its use in any educational context, does not concern or serve matching/mismatching style to instruction. ASSIST is about identifying deep and surface approaches towards studying and encouraging and discouraging them respectively; which can bring more valuable insights in the context of the current investigation compared to the much-debated match/mismatch approach (an issue which is discussed further in section 2.7). On the contrary, most of the aforementioned "style" measurements, based on the early empirical evidence briefly mentioned in 2.3 and the more recent developments discussed in 2.4, seem to serve mainly matching/mismatching scenarios. More specifically, Witkin's GEFT, Kolb's LSI and particularly Felder and Silverman's ILS, are becoming increasingly more and more heavily involved in applications and studies regarding intelligent adaptive learning environments as indicated in reviews by Al-Azawei and Badii (2014). This is not a surprising development since the authors of most of these "style" measurements are advocates of matching style to instruction (with the exception of Allison and Hayes) and consider it an important factor for achievement, as mentioned in section 2.3.1. However, by using ASSIST, the current investigation moves away from the matching/mismatching approach and chooses to examine and apply a different pedagogical perspective.
- As conducting a study in real learning conditions has become increasingly important in the context of students' interactions in a learning environment (see section 2.6.4), ASSIST seems more appropriate mainly because it has been developed based on observations, interviews and surveys which explored the

variability and consistency of the learning approaches within specific contexts in higher education. ASSIST, therefore, gives an authentic account of learning context in higher education and provides the potential to capture and identify any complexities involved in how students interact in a learning environment in natural settings. Such examples of complex and authentic students' interactions are discussed further in section 2.7.

- If the only criterion for selection was psychometric rigour, one could argue that Allison and Hayes CSI is the best choice because of its solid validity and reliability which commonly acknowledged by reviews such as Coffield et al. (2004). However, CSI has been developed in the context of decision-making and work performance, and although it has been used in educational context, the pedagogical implications of the model have not been fully explored and more empirical evidence in educational context is needed, as suggested by the authors of CSI and Coffield et al. (2004). On the other hand ASSIST has a more solid educational research background which can reflect better the reality of how students learn in interactive learning environments, as discussed earlier. This in combination with reasonably sound psychometric properties make ASSIST more appropriate for the current investigation compared to CSI.

2.6.4 Methodological weaknesses of existing studies

The studies discussed in section 2.3 indicate two main methodological issues. More specifically:

- In terms of empirical settings and sample sizes, they were usually carried out in experimental contexts rather in naturalistic teaching settings, and with relatively small or medium size samples.
- In terms of practical significance, in studies where it was possible to calculate it, mainly weak and medium magnitudes of the shared variance were found.

The intention is to address the above issues in the methodology of the current investigation (e.g. by planning a study appropriately so it can be carried out in natural teaching settings and obtaining a sample which is large enough so the findings are not limited in terms of generalisability). Finally, given that most of these studies encountered in the literature were conducted in experimental conditions and the intention is to carry out the current one in genuine learning conditions, it would be reasonably successful to find moderate correlation coefficients in the region of 0.3-0.4 and shared variance in the region of 9%-16%.

2.7 Developing the aims and research questions

This section discusses how the literature has shaped the aims and research questions of the thesis.

2.7.1 Choosing a different pedagogical perspective

The unresolved issues in studies that investigate the matching hypothesis in the context of style and ILEs, as shown in 2.3, led to the conclusion that there is a need to focus on researching the relationship between style and students' interaction in ILEs similarly to the studies discussed in sections 2.3.1 and 2.3.2. Most importantly, the current literature led to also look at constructs with different pedagogical philosophy which examine students' interactions in a learning environment from a different perspective.

The aim of most studies that investigate how learners with different learning styles behave in ILEs is to help to design a system that adapts to learners' preferences or needs in order to improve their performance. As discussed in section 2.3, there is an ongoing debate on the wisdom of matching or mismatching style to instruction and doubts have been expressed about whether it is pedagogically correct or whether there is enough empirical evidence about the effect of matching style and instruction on performance. On the other hand, Entwistle and his team showed when developing ASSIST that there are cases where both contrasting learning styles are used. For example, Entwistle and Ramsden (1983) find that a deep approach can be followed with a holist way of studying and a serialist way of studying; and as a result, Entwistle (1997a) includes in the deep scale of the ASSIST measurement both ways of studying as subscales. Entwistle (1981) also shows that achieving high grades can be attained by using a combination of approaches to studying, including the less appropriate studying approach of unrelated memorising. More specifically, students, with an intention to seek the meaning of the learning material of what they study, may later on try to commit facts and formulas to memory in order to deal with demands of a forthcoming closed-book examination. So, ASSIST acknowledges that students' approaches to learning is something which can change or become more versatile depending on the academic task and in general on the demands of a learning environment. The pedagogical perspective which ASSIST construct represents, although more complex, strikes one as a more authentic account of what actually occurs in a class in terms of how students approach their studying. In addition, Entwistle's observations on the versatility and sustainability of students approaches towards studying resonate with the author's experience as a tutor. Students in undergraduate courses during their practical sessions, for example, may exhibit a deep approach towards studying by trying to seek the meaning of what they learn and perform their one independent research on topic they are interested in, but occasionally the same students may show, for example, complete reliance on the given learning material (i.e. "syllabus boundness") or memorise a concept without really understanding its meaning (especially when coursework and examination deadlines are close-by).

Furthermore, it is reasonable to say that students with combined approaches to studying are not always encountered. There are students who may approach their studying in a deep or surface manner rather consistently, for example, during a specific module in a course; it is mentioned in 2.6.1 that students can exhibit sufficient consistency in intention and process across broadly similar academic tasks. Entwistle (1997a) describes again realistically through his construct what this means in a learning environment. Students with a surface approach, for example, tend to treat the learning material as unrelated bits of knowledge, concentrate narrowly on the minimum that the course requires, and feeling excessive anxiety about their course (Entwistle, 1997a; Entwistle and Peterson, 2004).There are also students who tend to understand concepts for themselves and explore them further concepts independently (Entwistle, 1997a). First-year undergraduates especially can demonstrate different approaches which have their origins in their schooling experiences (Entwistle and Peterson, 2004). Again, this resonates with the author's experience as a tutor. Challenging students preconceived ideas about their way of approaching learning, and their preferred way of teaching during their first year of their studies can be a challenge in itself (e.g. a habitual approach to treat learning material as unrelated bits of knowledge rather than seeking its meaning). It is a challenge, however, which as Entwistle and Peterson (2004) emphasises, should rather be embraced.

2.7.2 How is this pedagogical perspective beneficial?

So, this study will not seek ways of adapting a system to a specific learning style, but rather try to explain the students' interactions in a learning environment. This eventually can serve towards creating ILEs which assist tutors during practical sessions in class to identify, quickly and effectively, students in particular with surface approaches. Considering the aforementioned challenges, this can be quite valuable for tutors, especially in large undergraduate classes where identifying and keeping track of students' approaches towards studying can be difficult. Besides identifying approaches towards studying, tutors can be also helped to shift students' surface approaches towards studying by the design of features of interactive learning environments which could encourage, for example, a deep approach towards studying and discourage a surface one. Entwistle and Peterson (2004) argue with regards to the importance of designing learning environments which override students existing approaches towards studying while providing enough support to develop new ones. So, the current investigation may ultimately serve to find ways of helping tutors identifying a prominent approach based on students' interactions in an ILE and/or improving the design/features of an ILE, and thus encourage a deep approach to studying and/or discourage a surface approach to studying when used in classroom.

Finally, this pedagogical perspective can eventually also help students' academic performance. Entwistle and Ramsden (1983) and Tait and Entwistle (1996) find that the adoption of a deep approach to studying by first-year undergraduate students correlates positively to their performance; whereas a surface approach to studying correlates negatively to their performance.

2.7.3 The relevance and value of the new pedagogical perspective to mathematics education

However, at this stage, the question is how relevant and valuable can this perspective be for mathematical education? In particular in mathematics education, the issue of discouraging rote memorisation of formulas and rules and encouraging a deep understanding of mathematical procedures is quite prominent (Ladson-Billings, 1997; Crowe and Zand, 2000b). Similar concerns have been raised by Liston and O'Donoghue (2009), who point to research which indicates that students often carry mathematical procedures without really understanding the concepts involved, and that they focus on each procedure separately rather than trying to find connections between different parts of mathematics. Saha et al. (2015) also give examples in algebra in which students cannot relate methods and concepts. What Liston and O'Donoghue (2009) and Saha et al. (2015) point to can be considered examples of surface approaches or rather the lack of deep ones in mathematics learning. It shows that in mathematical education, there are indications that some students may not seek the meaning of what they apply, and that they lack the versatility of combining both serialist and a holist approach towards studying (i.e. they just concentrate on one mathematical procedure at a time without relating it to other concepts and procedures). To avoid such tendencies, Sangwin (2004) recommends moving students out of their comfort zone, for instance, by asking them to produce examples which reveal an “awareness of global criteria”, rather than asking them to apply steps of a particular mathematical procedure or simply applying a procedure after looking into a worked solution.

Therefore, there are issues regarding deep and surface understanding in mathematical education which need to be addressed. An investigation on how undergraduate students approach their studying during their mathematical practical sessions can help us understand these issues further, and can eventually indicate how they can be resolved. As discussed in 2.1, just the use of mathematical ILE in class is not enough. What can eventually make a difference is for tutors to use them in a way that considers students’ individual approaches towards studying and accordingly support them with appropriate instructions and activities. Finally, it is also particularly valuable if the proposed investigation targets first-year students, for two main reasons: it can help towards the aforementioned challenge of teaching students who come with preconceived ideas about teaching and learning because of their school experiences, and it can help tutors to deal with students’ varied (or even poor) levels of prior knowledge in mathematics, based on their entry levels.

2.7.4 New pedagogical perspective and some “natural” questions

Despite the aforementioned benefits, the pedagogical perspective the author intends to explore in current investigation is not without issues and concerns.

In order to discourage or encourage an approach, it will have to be identified. So, a crucial question is whether students’ deep and surface approaches towards studying can be identified through their interaction with a digital learning environment, given that there is currently little empirical evidence, as shown in section 2.3. It has been discussed previously how ASSIST is a construct which shows promise in capturing authentic complexities of students’ interactions in a learning environment used in natural settings in class. In its educational background research, one can find a rich account of students’ interactions in traditional learning environments which can be translated into interactions in digital ones. In section 2.4, there are examples on possible interactions with regards to deep and surface approaches: the surface approach of unrelated memorising can link to page revisitation; the surface approach of fear of failure can link to time spent on learning material; the deep approach of seeking meaning can link to time spent on learning material and the linearity of the path followed when going through the learning material. In the methodology chapter, it will be shown that, in similar way, theoretical assumptions with regards to interactions in the learning environment can be formed for all deep and surface approaches.

Another valid question is whether the adoption of a deep approach is something that can be induced. Fransson (1978) discusses how the efforts of Marton (1974), Dahlgren (1975) and Säljö (1975) to induce a deep approach by providing content neutral questions, different types of questions and content-oriented guidance, were simply unsuccessful. A deep approach cannot be induced in order to improve students' performance, however, it can be encouraged or stimulated (Entwistle and Ramsden, 1983). Entwistle (1997a) finds linkages suggesting that the approach to studying can be affected by the type of teaching experienced. For example, Fransson (1978) argues, that although it may not be possible to induce the deep approach of intrinsic interest in content for learning (a state where the relevance of the content of the learning material is the learner's main reason for learning) it is possible that a deep approach to studying, such as interest in ideas, can be encouraged and stimulated. For example, Beaty et al. (1997), Vockell (2006) and Martens et al. (2004) suggest in their discussion that intrinsic interest in learning content can be promoted or encouraged by allowing learners freely to choose what they want to learn and how they want to learn it; and by stimulating their curiosity in the learning environment which then encourages exploration of the content.

An investigation of how students with a deep approach interact in a learning environment might help us understand whether there is a need for features that might promote a deep approach (e.g. freedom of choice, exploration) or whether certain features need to be designed in a way that stimulates curiosity in order to encourage or maintain an intrinsic interest approach towards the content of mathematics. In addition, an investigation into how students with a surface approach interact in an interactive learning environment can help us to understand whether there are features of the system that encourage, for example, unrelated memorising instead of discouraging such an approach.

In a similar way, another question is: can a surface approach to studying being discouraged? According to Entwistle (1981) a surface approach to studying relies primarily on unrelated memorising (that is rote memorisation through overlearning). He argues that by explaining new concepts in terms of familiar ones and promoting meaningful associations, it is possible to discourage unrelated memorising.

Another question is whether the strategic approach, which is also part of the ASSIST measurement, is relevant in the context of the current investigation. The empirical research conducted by Entwistle and Ramsden (1983) indicates that the strategic approach is more likely to be relevant in more formal learning situations where formal assessment is taking place (e.g. preparation and revision for exams, implementation of coursework, implementation of final year project). As Entwistle (1997b) notes the distinction of deep and surface approaches to learning can be found in natural settings where the outcome of learning does not count, whereas the strategic approach expresses behaviour in learning situations where formal assessment is involved. The current investigation focuses on deep and surface approaches and examines students' interactions when they are engaged in practical sessions where they are not being formally assessed. As such the strategic scale is not as relevant as the deep and surface scales.

2.7.5 Research aims

It is now possible to refine the basic aim of this study as to investigate how first-year undergraduate students with a deep approach and those with a surface approach to studying interact when using an ILE such as ActiveMath (AM) for mathematics in tutorial sessions, in the classroom, in real learning conditions. This can potentially serve towards identifying deep and surface approaches based on their interaction in AM.

From this general aim, specific testable questions will be formed, as shown later on in the methodology chapter, about the relationship between the deep and surface approaches, as measured through ASSIST, and the students' interactions in AM when practicing in tutorial sessions in the classroom. Chapter 3 also discusses which metrics have the potential to express students' interactions in the chosen interactive learning environment for mathematics AM. More specifically, it discusses the way in which the measurement of the students' interactions is operationalized through specific "interaction" metrics. After looking at the possible theoretical connections between deep and surface approaches and "interaction" metrics, the intention is to examine:

- The empirical associations between "deep" and "surface" approaches towards studying and "interaction" metrics

- Whether it is possible to predict surface and deep approaches towards studying from the combined knowledge of “interaction metrics” (i.e. there is an effort statistically to explain the variance of “deep” and “surface” scales by using a set of “interaction” metrics)
- Whether it is possible to determine which “interaction” metrics are better predictors for deep and surface approaches towards studying (i.e. whether there are distinguishable aspects and tendencies in students interaction in an ILE such as AM which can identify deep and surface approaches)

Furthermore, part of this aim, but of complementary and secondary value, is to investigate the influence of the non-style factor of prior knowledge on the relationship between deep and surface approaches towards studying and “interaction” metrics.

Finally, the aforementioned examination can ultimately contribute towards:

- Providing pedagogical insights in understanding the connections between deep and surface approaches towards studying and students interaction in an ILE such as AM
- Helping a tutor to identify deep and surface approaches during students’ interaction in an ILE in order to facilitate intervention in class by providing certain recommendations
- Providing a good starting point in terms of methodological recommendations for future studies in similar or different educational settings (i.e. “interaction metrics” which can be recommended as predictors of deep and surface approaches)
- Providing recommendations with regards to future improvements in data capture and an ILE’s interface design and features

Chapter 3 - Methodology

The aim is to look at the relationship between deep and surface approaches to studying (as measured by Entwistle's ASSIST) and students' interactions when using interactive learning environments (ILEs) in higher education, and in a specific context. This relationship is examined in the context of an ILE for mathematics; a subject area which can benefit from this type of investigation, as shown in the background research. In particular, the author intends to carry out the study involving the ILE for mathematics ActiveMath (AM), and for undergraduate first-year students; a sample to which the author has access. Given the aim and the context in which this aim is examined, the intention is to look at which "interaction" metrics can be relevant to identify deep and surface approaches to studying.

3.1 Research Questions

The primary research questions are:

1. What is the relationship between students' interactions in the learning environment AM and deep and surface approaches towards studying when learning mathematics at tutorial sessions in the classroom?
2. To what extent do students' interactions in the learning environment AM explain deep and surface approaches when learning mathematics at tutorial sessions in classrooms?

The secondary research question is:

3. Is the relationship between students' interactions in the learning environment AM and the deep and surfaces approaches towards studying influenced by different levels of prior knowledge in mathematics, when learning mathematics at tutorial sessions in the classroom, and to what extent?

3.2 Why quantitative and qualitative research design

There are existing theoretical frameworks for deep and surface approaches and students' interactions in learning environments that can provide a satisfactory degree of understanding in terms of what is the students' intention and approach towards their studying, and what pages and features students access, use and spend time on, respectively. As discussed in 2.2.2 and 2.4 the theoretical framework for deep and surface dimensions give a satisfactory conceptual understanding of the phenomenon of how students approach their studying. In 2.6.1, it is also discussed how students exhibit enough consistency in intention and process across similar academic tasks to justify measuring their approaches towards studying as dimensions.

In addition, research on students' interactions in learning environments produces a relatively well-articulated theory for the specific phenomenon. The theoretical frameworks of these two concepts (students' approach towards studying and students' interactions) can provide, therefore, a substantial amount of conceptual understanding, before planning an empirical study to investigate their relationship. According to Robson (2002), these characteristics fit well with the theory-driven methodology of quantitative design.

What also derives from the research questions is that the phenomena (approaches towards studying and students' interactions in AM) and their relationships are examined in the classroom. As discussed in section 3.10, one of the intentions is to plan ahead in order to avoid disruptions in the classroom as much as possible. According to Robson (2002), when a substantial amount of pre-specification is required, then this is an indication that a quantitative design methodology is appropriate. Furthermore, the literature review indicates that students interaction can be identified based on: "interaction" metrics that derive from web logs which is a frequently used quantitative method[8]; and "interaction" metrics that derive from dynamic observation techniques such as "think aloud", which is considered a widely accepted qualitative method (Protopsaltis, 2006; Cockburn and McKenie, 2001). The use of web logs is an automated way to record users' navigational behaviour that occurs in real time and, more importantly, without interfering with the learning process (Juvina and Van Oostendorp, 2004; Barab et al., 1997). Therefore, in comparison to the "think aloud" technique, web logs are better suited for the requirement of maintaining the learning process as undisrupted as possible.

Furthermore, a quantitative design allows investigating the research questions of the current study for larger samples, and thus supporting more the generalisation of the results to a larger population, an issue that constitutes one of the major criticisms of studies of the field.

A quantitative design methodology, therefore, seems an appropriate choice and

[8] The logs provide quantitative data.

will constitute the primary methodology in order to answer the research questions and fulfil the aim of the current investigation. However, as the research questions indicate that the phenomena and relationships are examined in the classroom, there may be interactions and incidents that may influence them. A weakness of the quantitative design is that it cannot capture the subtleties and complexities of individual human behaviour in the way that qualitative design can. Ford and Chen (2001) argue that there is a need for complementary qualitative research in the "hypermedia" field, which may assist in developing a deeper understanding of the interactions between learners and the learning material. Robson (2002) suggests that a pragmatic approach where both qualitative and quantitative designs are combined is feasible in some studies and may cover the aims of a study in a more satisfactory way. According to Robson (2002), an observation technique may be used as a supportive or supplementary data collection method to complement data obtained by other techniques that constitute the primary method.

In the current study, ASSIST measurement shows great promise in capturing subtleties and complexities in students' interactions in a learning environment, as discussed in 2.7. However, also conducting observations in class, as a complementary task, could aid towards recording expected and unexpected events, which may influence the quantitative data, and even aid the interpretation of empirical findings. For example, conducting observations can help towards "cleaning the data" (i.e. deciding whether to exclude cases). It can aid towards obtaining complementary information about incidents or interactions that may influence the relationship between students' studying approaches and "interaction" metrics such as: tutor's tasks and instructions; incomplete sessions due to technical difficulties; students' absences from the tutorial sessions; and students' short attendance. It can also serve towards monitoring processes (i.e. students' registration on the interactive learning environment).

Entering into an investigation that involves other people is necessarily a complex and sensitive undertaking; especially when it involves real situations such as the current one. A quantitative design can assist in specifying in advance the majority of details required and a qualitative design can give complementary information to enhance further the capture of potential complexities of such an undertaking. The details for both designs are discussed further in section 3.10.

3.3 Approaches and Study Skills Inventory for Students (ASSIST)

3.3.1 The ASSIST instrument

The ASSIST instrument initially provides an explanation of its purpose, and a general instruction that encourages students to answer truthfully and "work their way through the questionnaire quite quickly". It consists of three parts. In all three parts, each statement can be ranked from 1 to 5 on a Likert scale, and in the second and third part, students are encouraged to avoid selecting the unsure response in the middle of the scale.

The first part is called "What is learning?" According to Entwistle (1997a), it is based on the conceptions of learning as described by Marton and Säljö (1976b) and extended by Hattie in 1996. It consists of six items and requires the students to consider general statements in relation to learning and rate them according to how close they are to their own way of thinking. This first part does not provide scores to indicate different styles or approaches.

The second part is called "Approaches to studying" and consists of 52 items. It derives from ideas as described by Marton and Säljö (1976) and Pask (1976) (especially in relation to deep and surface approaches) and from work by Entwistle and Ramsden (1983), especially in relation to the strategic approach.

The students are required to express their agreement or disagreement with statements about studying made by other students. The instructions for this part state that students should give their answers with a specific course in mind; that they have to answer all the questions and give their immediate responses. The scoring process of ASSIST for this part is carried out by adding the scores for statements corresponding to each of the three approaches (deep, surface and strategic) and their 13 subscales. There are 16 questions for each of the deep and surface scales (giving scores from 16 to 80); 20 questions for the strategic scale range (giving a score from 20 to 100); and four questions for each subscale (giving a score from 4 to 20).

The third part of the questionnaire is called "Preferences for different types of course and teaching" and consists of eight items. It requires the students to express their agreement or disagreement with specific statements in relation to organisation and design of a course. The third part of the questionnaire does not provide scores to indicate different styles or approaches, so the results deriving from the third part of the questionnaire are not used in the current study.

ASSIST is considered an ordinal-level measurement. In order to measure the learning styles of the students in this study only the second part of the questionnaire will be used, since this is the element of the instrument that is concerned with learning styles.

3.3.2 The ASSIST scales and subscales

This section discusses further the deep and surface scales and subscales[9].

- **The "deep" scale** measures the extent to which there is an intention to understand for oneself; there is an interest in the subject, students relate ideas to previous knowledge and experience, and students check evidence and relate it to conclusions (Entwistle, 1997b; McCune, 1998). It is based on the research conducted by Marton and Säljö (1976b) who find that some students have an intention for personal understanding by interacting with ideas and evidence in order to draw their own conclusions. Further research reveals that their behaviour when studying is characterised by their intention to spend time on studying and impose their own structure. Svensson (cited in Entwistle and Ramsden, 1983, p.19) finds that students adopting a deep approach tend to spend longer in studying. In relation to their overall performance in a course, Entwistle and Ramsden (1983) also find positive correlations between score on the deep scale and performance.

- **Deep subscale "seeking meaning":** This measures the extent to which there is an intention to understand the learning content for oneself (McCune, 1998; Entwistle, 1997a). Entwistle (1998) finds based on students' interviews that understanding the learning content for oneself means: how much material students bring together, how much effort had been used in forming connections between concepts, and the extent to which students impose their own structure on the lecturer's learning content. Based on the interviews conducted by Entwistle and Ramsden (1983), students who follow this approach when studying do not skim through the learning content, but they tend to go through it fairly slowly (Entwistle and Ramsden, 1983).

- **Deep subscale "interest in ideas":** This measures the intrinsic interest of students in the content of a course they are taking (Entwistle, 1997b). According to the ASSIST manual, the subscale of "interest in ideas" is a correlate of intrinsic motivation (Entwistle, 1997a). Entwistle et al. (1979) find

[9] As discussed in section 2.7.4, the strategic approach is not relevant to the current investigation, so the strategic scale and subscales are not discussed here.

that a distinct type of intrinsic motivation is the one that stems from the interest of the students in the subject matter.

- **Deep subscale "relating ideas":** This measures the extent to which students study by building up a broad view of the learning task and looking into relationships between ideas (Entwistle, 1997a; Entwistle, 1981; Entwistle et al., 1979). Entwistle based this approach to studying on research conducted by Pask (1976b) on the holist learning style. Pask (1976) notes the preference of students with such a style for a broad view and a personal organisation of the learning content. Students with a holist learning style tend to explore several topics of what may be known in the area and tend to adopt a top-down approach, in which they first concentrate on establishing an overview of what is learned before attending to the low level procedural information (Ford and Chen, 2001).

- **Deep subscale "use of evidence":** This measures the extent to which students study by building up meaning from the details, checking evidence to relate it to conclusions, and their preference for a linear sequence in their learning (Entwistle, 1997a, Entwistle et al., 1979). Entwistle based this approach to studying on research conducted by Pask (1976b) on the serialist learning style. Pask (1976) notes the preference of students with such style for step-by-step and tightly structured learning. Students with a serialist learning style tend to master one topic at a time and adopt a bottom-up approach in which they pay attention to the low level detail, building an overview at a later stage (Ford and Chen, 2001).

- **The "surface" scale** measures the extent to which there is an intention to memorise and treat the content as unrelated bits of knowledge; there is motivation to avoid failure; there is an overall negative attitude to studying and an intention to cope minimally with the course requirement; and an intention to follow strictly the instructions and the structure of the learning content, and focus on the minimal requirements of the course (Entwistle, 1997a; Entwistle and Ramsden, 1983). It is based on the research conducted by Marton and Säljö (1976b) who find that some students have an intention to complete the task with very little engagement and with unreflective memorisation. In relation to their overall performance in a course, Entwistle and Ramsden (1983) find negative correlations between the surface scale and performance.

- **Surface subscale "unrelated memorising":** This measures the extent to which students intend to memorise facts and procedures and treat the learning content as unrelated bits of knowledge (Entwistle et al., 2001, Entwistle, 1997a).

- **Surface subscale "fear of failure":** This measures the extent to which students are motivated to avoid failure (Entwistle, 1997b; Tait et al., 1998; Entwistle, 1981). Students with "fear of failure" have an over-anxious concern about possible failure (Entwistle, 1981). Fransson (1978) finds that it is not so much that a threatening learning environment or situation causes this concern; it is rather that students perceive the situation as threatening. This type of anxiety affects the way in which students tackle their study, resulting most of the time in rote memorisation (Entwistle, 1981).
- **Surface subscale "lack of purpose":** This measures the lack of interest in the subject (Entwistle, 1997a; Entwistle and Ramsden, 1983). It expresses an overall negative attitude to studying, and so when studying students tend to cope with the course requirements in a minimal way which is likely to result in ineffective studying (Entwistle and Ramsden, 1983; Entwistle et al., 2001).
- **Surface subscale "syllabus boundness":** This measures the extent to which students prefer clear instructions, deadlines, and defined course materials with clear structure (Entwistle, 1997a; Entwistle et al., 1979; Entwistle and Ramsden, 1983). Students who follow this approach are not autonomous when studying and they study little beyond what is required to pass (they simply focus on the course's minimum requirements) (Entwistle, 1997a).

Based on what is discussed in sections 3.3.1 and 3.3.2, it is worth mentioning that it is possible for a student to have a relatively high score in both the deep and surface scales and subscales. Entwistle et al. (2001) argue that whilst students have a tendency to adopt a particular approach to studying, their approach may vary in reaction to specific circumstances. For example, empirical evidence by Entwistle (1981) and Entwistle and Ramsden (1983) suggest that particularly in science subjects personal understanding may involve acquiring basic knowledge and a certain amount of rote learning especially in the early stages of learning. Thus, a deep approach may involve rote memorisation on occasion, at the same time these variations in approach do not minimise the crucial difference in intention (to achieve personal understanding of a topic, versus satisfying the teacher in a minimal way). This has been found across different subjects and is expressed through the ASSIST scales and subscales.

3.4 Examining students' interactions when using an ILE – "Interaction" metrics

Previous studies looking at the relationship between styles and students' interactions in ILEs have examined interactions such as: the number of pages visited and revisited, the order in which students move around the learning material (e.g. following the given linear order of the material), the time they spend on the ILE; and the use of navigational options that allow them to have a more global or detailed view of the learning material.

The present study will be carried out using AM, and so the students' behaviour is to some degree specific to that ILE. Metrics will be chosen that enable us to capture how students move around and what students do in AM. We also wish to examine the link to the deep and surface scales and subscales of the ASSIST measurement, and so the design and choice of metrics will also be based on their relevance to examining this link.

It is evident from the students' interaction in an ILE for mathematical education, shown in section 2.1, that the more interactivity ILEs offer, the more choices and complex decisions students have to make. AM is a fairly interactive learning environment for mathematics in which students have to decide on: the pages (theory and exercises) they visit or revisit, the order in which they will go through the learning material, how long they spend on each reading or exercise page; how many exercises they do, whether to research a mathematical concept further (e.g. in relation to other concepts) using the search option, whether to make notes using the notes tool, whether to use the previous/next button to follow the given structure of the content.

3.4.1 Path length and visitation metrics

Path length is defined as the number of pages the user visits during the navigation session including revisits (in accordance with: Herder and Juvina, 2004). Berendt and Brenstein (2001) and Cockburn and McKenie (2001) describe the number of pages visited by a user as one of the typical measures used to examine the way users navigate in a system. Cockburn and McKenie (2001) define the term "visit" in the context of navigation as the act of displaying a page regardless of the options used to reach the page.

As shown in section 2.3.2.1 of the literature review, a path length metric has been involved in studies where there is an attempt to interpret students' behaviour in ILEs and in relation to styles. A path length metric may give an overall idea of the amount of use of AM, as it may show the amount of activity which takes place during the tutorial session.

Path length can be analysed further to show the amount of activity in specific parts of an ILE. As shown in the study of Mimirinis and Dafoulas (2008) in section 2.3.2.1, the *number of page visits* is calculated separately for the practical and theoretical part of the ILE. In this way, a general “interaction” metric such as the *number of page visits* can be placed in a learning context and provide further information about what students do in an ILE. The creation of visitation metrics for the practical and theoretical part can reflect what students do in AM during their tutorial session - where they need both to visit exercises and practise and visit the mathematical theory, concepts and working examples.

The question is how path length and visitation metrics can be related to the deep and surface scales and subscales and how path length and visitation metrics can be used in the context of an ILE such as AM.

In relation to Entwistle’s studying approaches and visitation metrics, the study conducted by Mimirinis and Dafoulas (2008), as discussed in 2.3.2.1, indicates that there is a possible positive correlation between the *number of visits* in the theoretical section of the ILE and the surface subscale of “unrelated memorising”. However, path length and visitation metrics in current investigation such as the *number of reading pages visited* and the *number of exercises pages visited* may have a positive relationship to “unrelated memorising” subscale due to revisitation (as discussed in 3.4.3). So, a path length type of metric, such as the *number of distinct pages* (which does not include revisits) may express perhaps more accurately the interactions of students with a high “unrelated memorising” score, who tend to perform activities in a more repetitive manner compared to those with a lower score. More specifically, it is also possible that students with higher scores in the “unrelated memorising” scale may perform a more limited amount of activities in tutorial sessions by not accessing as many exercises, compared to those with the lower scores; or by visiting more of the same AM pages and less a variety of distinct AM pages, compared to those with the lower scores. So, high values on the “unrelated memorising” scale may result in low values in metrics such as: *number of exercises accessed*, and *number of distinct pages visited*.

Interviews conducted by Entwistle and Ramsden (1983) show that students with “fear of failure” tend to feel that they are drowning in the sheer amount of material they have to cope with. It is possible that students with a high “fear of failure" score may feel overwhelmed by the amount of material, and so they may access fewer AM reading pages and exercises pages, resulting in a shorter path than students with a low "fear of failure" score.

Another possible link is between the deep approach and visits to the AM feature "notes". According to the empirical observations of McCune (1998) students with deep approach to studying tend to interact vigorously with the learning material by making notes, selecting and organising learning material based on their own interests, relating ideas in order to achieve personal understanding**.** So, a metric with regards to visits to the AM "notes" feature can show an intention to create notes.

The visitation metrics involved in this study are:

- ***Number of reading (content) pages visited***: The total number of reading pages visited. The reading pages [10] present the theory of mathematical concepts, procedures and the mathematical examples (see Appendices 3.4.1 and 3.4.10). They can be requested by clicking on the links in the table of contents (TOC) (See Appendix 3.4.1), previous and next buttons (see Appendix 3.4.2), or by using the back and forward browser buttons.
- ***Number of exercise pages visited*****:** The total number of exercise pages visited. The exercise pages present a list with exercise links (see Appendix 3.4.2). They can be requested by clicking on the links in the table of contents (TOC), previous and next buttons, or by using back and forward browser buttons.
- ***Number of exercises accessed based on tries*****:** The number of times the student accesses an individual exercise by clicking on an exercise link on the exercise pages (see Appendices 3.4.4-3.4.7)
- ***Number of distinct pages visited*****:** The number of pages a student visits at least once.
- ***Number of times "notes" link is clicked*****:** the number of times a student visits the "notes" option by clicking on the notes link (see Appendices 3.4.8.a-3.4.8.c).
- ***Path length*****:** The sum of
  - *the number of times exercise and reading (content) pages are visited*

[10] Throughout the thesis the AM pages which contain the theoretical learning material will be referred to as "content" or "reading" pages.

- *the number of times an individual exercise is accessed by clicking on an exercise link on an exercise page*
- *the number of times a student requests help by clicking on the help link*
- *the number of times a student visits the notes option by clicking on the notes link*
- *the number of times a student visits the search option by clicking on the search link*
- *the number of search results clicked on the search option*
- *the number of submitted queries in the search option*
- *the number of times student follows hyperlinks in the text of reading and exercise pages.*

3.4.2 Metrics related to use of hyperlinks and the search option

Students can view the learning content in AM using hyperlinks in the text of the reading and exercise pages or by using the search option (in which students can submit queries for mathematical terms).

The hyperlinks provided in AM can help towards relational linking between mathematical terms. Scherly et al. (2000) argue that the frequent use of the hyperlinks in the text and search option is an indicator of active reading. Furthermore, the use of hyperlinks for mathematical terms in the text may be connected conceptually to Entwistle's "relating ideas" and "interest in ideas" subscales. Entwistle (1981) describes students with an "interest in ideas" approach as having active interest in the course content, so it is likely that they would use the option of the hypertext links to follow up on ideas in which they are interested in.

Students with a "relating ideas" approach tend to look into relationships between concepts (Entwistle, 1997a; Entwistle, 1981; Entwistle et al., 1979). This approach may lead them to use the AM hyperlinks that provide a relational linking between mathematical terms.

In relation to the AM search option, students can use it to submit queries for any mathematical term; the search returns all relevant AM reading pages in which the specific term is involved and it gives the student also the option to search beyond AM using standard search engines (see Appendix 3.4.3). After a query is submitted in AM, the search results show all the related theory, examples and exercises, facilitating the exploration of relationships between mathematical concepts and their application. This allows the students to explore and select their reading based on their interest.

The use of the search option may be connected conceptually to Entwistle's "interest in ideas" subscales. Entwistle (1981) refers to the empirical work of Fransson (1978) on intrinsic motivation, who found that interest in content is likely to encourage a more exploratory search when students retrieve information. The search option can offer a more exploratory and active retrieval of the mathematical concepts which suits students with an "interest in ideas" approach, who according to Beaty et al. (1997), tend to have an exploratory and active approach towards the learning content.

The use of the search option can also facilitate looking into relationships between mathematical concepts for students with a "relating ideas" approach.

The use of hyperlinks can be expressed with the metric of number of hyperlinks (i.e. "concept links" linking to mathematical terms) of reading and exercise pages (see Appendix 3.4.11); whereas the use of search option can be expressed through the number of submitted queries and the number of search results visited.

So, based on what is discussed in this section the metrics involved in this study are:

- ***Number of hyperlinks (concept links) visited in reading and exercise pages*****:** The number of hyperlinks to mathematical concepts visited followed in reading and exercise pages.
- ***Number of times search option is clicked*****:** The number of times a student visits the search option by clicking on the search link.
- ***Number of submitted queries in search option*****:** The number of queries submitted in the search option.
- ***Number of search results visited in search option*****:** The number of search results visited in the search option after the submission of queries.

3.4.3 Revisitation metric

As shown in section 2.3.2.1, revisitation behaviour has been examined in studies where there is an attempt to interpret students' behaviour in ILEs in relation to styles. In the study conducted by Chen and Ford (1998) revisiting pages is interpreted as a sign of less efficient learning with the ILE.

It is likely that there is a connection between revisitation behaviour and the surface subscale of “unrelated memorising”. Students with an “unrelated memorising” approach tend to rehearse and overlearn the learning content (Entwistle et al., 2001; Entwistle, 1997a; Entwistle and Ramsden, 1983). It is possible, of course, that students independently of their studying approach may revisit, for example, a reading page in AM in order to review the mathematical concepts, procedures and working examples to help them with their exercises. However, it is also likely that revisitation will be used as a way to rehearse and overlearn the learning material by students with an “unrelated memorising” approach; and as a result students with high “unrelated memorising” scores may more often revisit pages than students with low “unrelated memorising” scores.

A formula which can express how often students are likely to revisit pages has been suggested by researchers in the field of revisitation behaviour such as (Obendorf et al., 2007; Cockburn and McKenie, 2001; Tauscher and Greenberg, 1997; Herder, 2003). It is given as the recurrence rate R or relative amount of revisits. Tauscher and Greenberg (1997) calculate the formula for the recurrence rate as:

R = ((“total URLs visited” – “different URLs visited”)/”total URLs visited”)*100%

where “total URLs visited” is the number of pages visited (including revisits) and “different URLs visited” is the number of distinct pages.

Based on what is discussed in this section the “revisitation” metric involved in this study is:

- ***Relative amount of revisits*****:** The probability that the visit to any URL is a repeat of a previous visit.

3.4.4 Temporal metrics

Time spent by a user viewing a page is defined as the time spent on each page (excluding the time spent on loading the page) until the moment they select to leave that page by clicking on a link (in accordance with Shahabi et al., 1997). As shown in 2.3.2.2, time spent on pages of an ILE has been examined empirically in relation to styles. In relation to the ASSIST measurement, there are possible connections to: the “deep” scale, “interest in ideas” subscale, and “fear of failure” subscale.

In relation to the ASSIST measurement, there are possible connections to the two subscales of “interest in ideas” and “fear of failure”. Entwistle (1981) finds that students with relevant "interest in ideas" tend to "get hooked" on topics and keep studying them, and in general explore the content in a more leisurely manner. Regarding the "fear of failure" subscale, Entwistle (1981) argues that students with a higher score tend to work more slowly, putting more effort into tasks and persisting longer to solve problems.

Therefore, temporal metrics could be related to those two subscales. It should, however, be noted that the tutorial sessions in the current study have a relatively fixed duration and therefore all students are likely to spend around the same overall time on the AM reading and exercise pages. What may differ, though, is the time a student chooses to spend on each AM reading or exercise page as a student's intrinsic interest in a topic or persistence in tasks due to anxiety may increase the time spent on specific pages.

According to Berendt and Brenstein (2001), average view time is a typical measure which can be used to formalise temporal behaviour per page in an ILE. Herder and Juvina (2004) note that average view time is a temporal metric which has been involved in studies examining behaviour in hypermedia environments in relation to individual differences. Shahabi et al. (1997) and Herder and Juvina (2004) also find that average view time is an indicator of users' interest in the content of a hypermedia environment.

In the context of AM, students with higher scores in "interest in ideas" subscale focus and spend more time on average on a reading or exercise page than ones with lower scores do. Time per page may also assist in showing that students with higher scores in the "fear of failure" subscale are likely to persist in their tasks and spent more time on a reading or exercise page than the ones with lower scores. What we have here are a deep and a surface approach, whose relationship to the temporal metrics of *average view time* is assumed to have the same direction.

Besides the *average view time*, there are also other temporal metrics which have the potential to enlighten the current investigation in terms of students' interaction according to their studying approach (i.e. they may contribute to a more enriching complete profile for a scale or subscale). For example, increasing the *maximum view time* on an AM page (relating to an increasing score on a deep scale or a surface scale) can indicate some sort of "extreme interaction" or "imbalance" in terms of how students allocate their time with regards to reading and exercise pages in AM. This "imbalance" in terms of time can be due to an effort to seek and research further the meaning of a concept or process it further by relating it to other concepts and processes with regards to a deep approach. Or it can be due to experiencing difficulties with solving the exercises related to a specific mathematical concept or process, because they tend to treat it as unrelated bits of knowledge with regards to a surface approach. A similar theoretical assumption can be made for *minimum view time* on an AM page: decreasing *minimum view time* on an AM page (relating to an increasing score on "lack of purpose" and "syllabus boundness" subscales) can indicate some sort of "imbalance" in terms of how students engage with the learning process or learning material. This "imbalance" in terms of time can be due to the intention of engaging minimally with the learning process or the learning material with regards to these two surface subscales.

So, the temporal metrics involved in this study are:

- ***Average View Time on Exercise Pages*:** The average time during which an exercise page in AM is viewed in a tutorial session.
- ***Average View Time on Content (Reading) Pages*:** The average time during which a reading page in AM is viewed in a tutorial session.
- ***Maximum View Time on Exercise Page*:** The longest time a student has viewed an exercise page in AM during a tutorial session.
- ***Minimum View Time on Exercise Page*:** The shortest time a student has viewed an exercise page in AM during a tutorial session.
- ***Maximum View Time on Content (Reading) Page*:** The longest time a student has viewed a reading page in AM during a tutorial session.
- ***Minimum View Time on Content (Reading) Page*:** The shortest time a student has viewed a reading page in AM during a tutorial session.

3.4.5 Path metrics: stratum and compactness

These "interaction" metrics are called "path" metrics, and their purpose is to characterise (or quantify) user's navigational path in a hypermedia environment (McEneaney, 2001). Herder and Juvina (2004) also consider them as metrics of navigational "complexity"[11].

In order to look at users' paths in hypertext systems McEneaney (2001) developed the concepts of *path stratum* and *path compactness*, which he

[11] This type of metrics may reveal the "navigational complexity" of a user which can be defined as "any form of navigation that is not strictly linear" (Herder and Juvina, 2004).

calculated by adapting the formulae of Botafogo et al. (1992), using the *number of distinct pages* of the user's path in place of the pages in the network. In a communication with the DFKI team and the author, MacEneaney clarified that in his formulae for *stratum* and *compactness*, he used the distinct pages of the user's path instead of the number of nodes (pages) the hypertext consists of because "this seemed a reasonable choice on both theoretical and practical grounds". More specifically:

- Path Stratum

  Stratum refers to the degree of linearity of a network, as indicated "by the extent to which a network is organised so that certain pages must be read before others" (McEneaney, 2001, p.765). It captures the extent to which pages can be identified as departure, destination and in-between points based on the structure of the network. The value of the *stratum* is closer to one when the network is more linear; whereas it is closer to zero when it is less linear (which means that almost every page is connected to every other page and there is no structural basis for distinguishing points of departure and destination) (McEneaney, 2001). To make this distinction, a "prestige" value is assigned to each page which indicates the "pecking order" of the page in relation to the rest of the pages in the user's path (McEneaney, 2001). This calculation of *stratum* was initially introduced by Botafogo et al. (1992) to assess and improve hypertext systems. Finally, *stratum* has been used to express linearity in the more general context of users' navigation. Juvina and van Oostendorp (2006) suggest that *stratum* is a fairly common measure for computing linearity by assessing the user's path.

  Linearity path is the extent to which the user follows the given structure of the content in a hypermedia environment. It is a concept which has been examined frequently in relation to styles, as shown in section 2.3, to indicate the extent to which students with a certain style prefer to impose their own structure or simply follow the given structure of the ILE. In the more general context of user's navigation it has been also used to investigate whether users choose a more or less passive way to retrieve information (McEneaney, 2001).

  There may be a possible connection between the surface subscale of "unrelated memorising" and linearity. Based on interviews conducted by Entwistle and Ramsden (1983), it is found that science students with an intention for "unrelated memorising" are unlikely to be autonomous in their learning. They appear to transfer lecture information to the memory without thinking about it, without reassessing it or reorganising it, so in a rather "passive" way. This may suggest that students with high "unrelated memorising" scores are more likely to follow the given structure of the content in AM and navigate in a more passive linear way.

There may also be a connection between the deep scales and linearity. Entwistle (1998) finds that students who make an effort to seek an individual interpretation tend to impose their own structure on their learning material. This suggests that students with high scores on "seeking meaning", for example, are less likely to follow a linear path in order to retrieve the AM pages, compared to those with low score.

In the current context, students who tend to seek the meaning of what they learn may follow a less linear path when materials using materials, thus producing smaller *stratum*. Students with a surface "unrelated memorising" approach, who tend to rely exclusively on the structure of the given content, may follow the pecking order of the pages in AM thus producing a higher path *stratum*.

- Path Compactness

According to McEneaney (2001), "compactness refers to the connectedness of a network". A network's *compactness* is closed to zero when "it is sparsely linked"; whereas it is closer to one when "it is more densely connected" (McEneaney, 2001). So, the value of the connectedness is closer to one when the network is more compact; whereas it is closer to zero when it is less compact.

In the current investigation, there can be links to the studying approaches. According to Entwistle and Ramsden (1983), students with an intention for "unrelated memorising" tend to overlearn and repeat what they learn. So, students with a tendency for repetitive overlearning are likely to interact more closely around a certain set of pages. So, high values on surface subscale of "unrelated memorising" may result in high values in *compactness*.

The metrics discussed in this section have the potential to express the extent to which students with a specific approach to studying follow a linear and compact path in AM. The intention is therefore to involve the following metrics:

- ***Path Stratum*:** The degree of linearity of students' path, calculated according to the formula of McEneaney (2001).
- ***Path Compactness*:** The degree of compactness of students' path, calculated

according to the formula of McEneaney (2001).

3.4.6 Metrics related to use of AM navigational options

In the context of AM, students have the option of visiting the reading and exercise pages via the use of the following navigational options:

- previous/next options

The use of previous/next options offers a linear way to explore the learning content. Linearity of path with the use of previous/next options, is one of the aspects of students' interactions that has been examined in relation to styles, as shown in section 2.3.1.2. Furthermore, McEneaney (2001) found that passive readers were more likely to rely on the use of previous/next buttons to move through hypermedia materials. So, similarly to what is discussed in section 3.4.5 with regards to "unrelated memorising" subscale and linearity, it is possible that there is a positive relationship between the "unrelated memorising" subscale and the use of previous/next buttons. The involvement of the use of previous/next options may be also valuable.

- table of content (TOC)

The use of TOC in AM allow students to navigate quickly to the exact piece of information they prefer to focus on (like when using an index or a menu), however it can also indicate the structure of the learning material and it being used in a linear way. So, the TOC can be used to navigate through the learning material in both a linear and non-linear way. As shown in 2.3.1.1, there are empirical findings which indicate the preference of certain "styles" for specific navigational options. In addition, Chen et al. (2016) refer to empirical studies which indicate that serialists, based on Pask's construct, prefer both back/foreword buttons and index. In the current context, there are no clear indications that the use of TOC is the likely interaction of a specific studying approach. The use of TOC can be relevant to the "use of evidence" and "relating ideas" approaches to studying, since they are based on Pask's construct serialist/holist (two learning styles which have been involved empirically in relation to the use of navigational options as indicated by Chen et al. (2016)). However, as has been shown earlier, serialists can prefer both back/forward and index options, and TOC in AM is designed to serve flexibly linear and non-linear preferences in terms of navigation, so it is not possible to make associations with a positive or negative direction. Nevertheless, the inclusion of use of TOC as a metric is deemed valuable in the current investigation. First, because it may enlighten us as to the navigational preferences of the studying approaches, and second, because it may offer a sort of contrast to the use of previous/next options. So, the metrics involved in this study are:

- ***Number of pages visited using the previous/next buttons*:** The number pages visited from the previous and next option (see Appendix 3.4.10).
- ***Number of pages visited using the TOC*:** The number pages visited from TOC option (see Appendix 3.4.1).

3.4.7 Performance-related metrics related to number of tries when practising exercises

As shown in 2.1.5, practising exercises is an important activity in a mathematical ILE. AM provides exercises for students to practise in their tutorial sessions. The activity of practising exercises is designed so that it allows students to evaluate their answers and receive feedback. If the answer is incorrect, they are encouraged to try again, and are allowed three attempts in all (see Appendices 3.4.4-3.4.7).

Capturing data on the students' responses to these exercises can indicate whether they get the correct answer, and also whether they get it correct on the first, second or third attempt, or whether they do not manage to solve the question at all. This data may indicate whether students are clicking every checkbox until they get the correct answer; a "gaming" behaviour or "trial and error" behaviour where students try different solutions without a systematic approach. These are interactions which have been commonly observed in mathematical ILEs, as discussed in section 2.1.7.

Which approaches to studying are likely to be associated with such interactions? As discussed in section 2.1.7, anxiety has been linked to "gaming" behaviour, hence such behaviour can manifest in students with high "fear of failure" scores. In the context of the current study, it is likely that the higher students score in the "fear of failure" subscale, the lower the number of exercises they are likely to solve on the first attempt and the higher the number of exercises they are likely to solve on the second or third attempt, or fail to solve at all.

Interviews conducted by Entwistle and Ramsden (1983) suggest that science students with an "unrelated memorising" approach seem to be just "grinding the numbers, getting some kind of solution", and then if the result is not correct "go back and pick different values". This is a "trial and error" approach trying different values in the formula in an unsystematic way until they get correct answer. In the context of the current study, it is likely that the higher students score on the "unrelated memorising" subscale, the lower the number of exercises they are likely to solve on the first attempt, and the higher the number of exercises they solve on the second or third attempt, or fail to solve altogether.

In addition, there may be connections between the subscale of "syllabus-boundness" and quitting (or cancelling) exercises. Mavrikis et al. (2003) observe during the pilot testing of WALLIS (an ILE for mathematics) that a proportion of students quit their activities. Based on the data collected in the interviews, they find that one of the reasons behind this behaviour is that students do not see the relevance of the activities to their assessment. This behaviour seems to fit the profile of students who tend to be bound to the syllabus and who, according to Entwistle and Ramsden (1983), tend to gear their studying closely to just what seems relevant to their assessment and in general do very little beyond what is required.

Furthermore, these metrics may give indications as to whether the students do well (or perform well) while practising their exercises during their tutorial sessions. Entwistle and Ramsden (1983) find that scores on the surface scale had a negative correlation with performance, whereas score on the deep scale has a positive correlation to performance. So, it is possible that there are positive correlations between the deep scale and subscales and the *number of exercises solved on first try* metric; and negative correlations between the deep scale and subscales and the metrics related to the *number of exercises solved on second try*, the *number of exercise solved on third try*, or the *number of exercises finished but not solved*. Also, in a similar way and in line with what is discussed earlier, it is possible that there are negative correlations between surface scales and subscales and the *number of exercises solved on first try* metric; and positive correlations between the surface scale and subscales and the metrics related to the *number of exercises solved on second try*, the *number of exercise solved on third try*, or the *number of exercises finished but not solved*.

Therefore, the metrics related to the number of tries on the exercises, used in the study are:

- ***Number of Exercises Solved on First Try*:** number of exercises solved on the first attempt

- ***Number of Exercises Solved on Second Try*:** number of exercises solved on the second attempt
- ***Number of Exercises Solved on Third Try*:** number of exercises solved on the third attempt
- ***Number of Exercises Finished but Not Solved*:** number of exercises that were completed but not solved on the third attempt
- ***Number of Exercises Cancelled*:** number of exercises that were cancelled

3.4.8 Metrics related to average number of links followed per page

This group of metrics is in relation to the *average number of links followed per page*. According to Herder and Juvina (2004), this type of so-called "frequency per page" is represented by calculating the ratio between the *number of links followed* and the *number of distinct pages visited*, which is expressed as: *number of links followed / number of distinct pages*. The formula is used for the calculation of metrics in relation to AM features such as "search" and "notes" to give a more proportional image of the intention to use these features. So, the intention is to involve:

- ***Average number a "notes" link is clicked per page*:** It is calculated as: number of times "notes" link is clicked / number of distinct pages
- ***Average number a "search" link is clicked per page*:** It is calculated as: number of times "search" link is clicked / number of distinct pages

As, these metrics are likely to highly correlate with *number of times "notes" link is clicked* and *number of times "search" link is clicked* respectively, it will be shown later on in strategy how the choice between them will occur in terms of their involvement in the statistical analysis.

3.5 Development of specific hypotheses for each ASSIST scale and subscale

As shown in section 3.4, the development of hypotheses is based on existing literature about the way students with a specific approach to studying interact in traditional learning environments, and on the limited empirical research which has examined the associations between the ASSIST scales and subscales and students' interactions in learning environments. In section 3.4, the author has already shown examples of connections between students' "interaction" metrics and the ASSIST scales and subscales, to justify the selection of metrics with regards to their relevance to the scales. To help the reader, Chapter 4 presents all the theoretical assumptions concerning the associations between each of the aforementioned metrics and each deep and surface ASSIST scale and subscale.

3.6 Correlational design

The hypotheses require the investigation of the relationship between ASSIST deep and surface scales and subscales and students' "interaction" metrics. This is a relationship in which each factor has a number of variables, and where these variables may need to be examined simultaneously. On one hand, there are the ten ASSIST scales and subscales; while in terms of students' interaction in AM, there are selected metrics, which may be associated with those scales. Therefore, the type of quantitative design to be applied should allow for the examination of a plethora of variables and their relationships simultaneously. According to Cohen (2000) and Robson (2002), a correlational or relational design is the type of non-experimental quantitative strategy which can meet this requirement.

In addition, there is no intention deliberately to alter the interactions of the participants. Conducting the study in real learning conditions assumes certain considerations about treating all participants the same without involving treatment in the form of different ILEs. A non-experimental type of design can fulfil these requirements. This type of design allows the study of learning behaviour in a more realistic setting (Cohen et al., 2000, Robson, 2002); and it can be applied when modifying variables is not feasible or should be avoided (Robson, 2002).

3.7 Real learning conditions

As discussed in 2.6.4, one of the main methodological issues in the existing studies is that they are carried out in experimental conditions rather than in naturalistic settings. However, conducting the current investigation in real learning conditions can be particularly valuable. Ford (2000) suggests that research and real course delivery can complement each other. This argument is supported by Robson (2002) who suggests that there is much to be gained by transferring an enquiry from the laboratory to the real world since the emphasis tends to be on the practical significance rather than assessing the statistical significance.

Real settings may also favour the validity of a study, as in more experimental conditions participants tend to do what they think the researcher wants them to do (Robson, 2001). On the other hand, the same author argues that the validity may be threatened by "compensatory equalisation of treatments" (this means that if one group receives special treatment, then there may be pressures by the organisation or institution for a control group to receive it). Furthermore, in experimental conditions (where students know that they are not really in a teaching-learning situation, that they are observed and that certain things are demanded of them), they are not likely to engage with the learning material in the same way they would in a real teaching-learning situation. This is because of what Robson (2002) calls "demand characteristics". Such biases and constraints make the applicability and generality of the results to the real world questionable (Entwistle, 2008).

Similarly, generalisability seems to be favoured by the fact that the study may be replicated in real conditions. However, there are constraints because of the choice of the subject area and it certainly depends on the representativeness of the sample. Therefore, conducting a study in real life settings is not an absolute guarantee for better generalisability.

Robson (2002) argues that proper access to study people in real life settings can be obtained only if researchers use their skills to provide some sort of service. Conducting a real world enquiry, therefore, may help in obtaining more genuine interactions.

As discussed also in section 2.7, the choice of ASSIST as a construct shows promise in capturing authentic complexities of students' interactions in a learning environment used in natural settings in class. So, it is an opportunity to examine whether it can indeed give valuable insights into students' interactions in a learning environment when these take place in the actual tutorial sessions of mathematics in a course.

So, conducting the study in real learning conditions allows us to combine addressing the research aims of the current study with the requirements of the course examined. Carrying out a study in real learning conditions serves the research questions better in that it makes it possible to answer questions and address potential criticisms related to the impact of the students' studying approaches in real situations and in the context of mathematical interactive learning environments in higher education. Also, conducting a study in real life settings does not require students to perform a task outside the scope of the course, and it may even provide solutions to help improve the course delivery and facilitate students' learning.

Ford (2000) and Ford and Chen (2001) highlight the importance of conducting studies in real world contexts but do not give a clear definition about what this actually means. However, what they seem to mean is that the tasks of students carried out using the hypermedia learning environment are related directly to students' coursework, and that there is consultation with teaching staff about students' coursework, and about students' study needs and the course's outcome. Robson (2002) suggest that what also characterises a real world enquiry is that it takes place in a real environment such as a school and not in a research institution. In this study, real learning conditions will be taken to mean that:

- The study is conducted in the classroom
- ActiveMath as well as any learning material provided are part of the course delivery
- ActiveMath and its learning material is structured and designed according to the course's outcomes and syllabus and the students' study needs and there is consultation about it with teaching staff
- Students are assessed on the learning material delivered via ActiveMath

This carried a number of implications for the study. Since students are assessed on learning material delivered via ActiveMath, there is an ethical consideration that none of the participants should lose out and all participants should be treated equally by using the same version of the interactive learning environment as well as completing the ASSIST questionnaire[12].

It is also important that the study runs smoothly in parallel with the learning process and without disrupting it. This can be achieved by:

- Planning the procedures at an early stage and in a detailed way.
- Seeking to apply relatively unobtrusive methods of data collection where possible, as shown in Ford's review (2000).
- Designing the study so that students' behaviour is not directed by tasks that are set by the researcher when the actual learning process takes place.

[12] This was one of the main ethical issues that were discussed with a member of the ethics committee of the University of Westminster.

- Planning a pre-session with AM activities, with written and verbal instructions about AM, so students are familiar with the interface, navigational aids and AM before using it for learning purposes.
- Collaborating closely with teaching staff for the integration of the current study into the learning process.

3.8 Sampling strategy

The use of probability samples allows for generalisation from the sample to the population, however, choosing a random mathematics course or modules among UK universities did not seem a feasible strategy for the current study, as getting permission to access the sample would require time-consuming negotiations. In addition, planning the study in real conditions required close collaboration with teaching, administrative and technical staff as well as awareness of the regulations and policies of the chosen institution. It was certainly an advantage to be familiar with environment, people and regulations of the chosen institution, before even starting the planning of a study of this type. Furthermore, this was a "cold-start" empirical investigation, in that at the time the current investigation started there were no really prior similar empirical findings [13] to support methodological choices, theoretical assumptions and interpretation of the findings[14]. One of the reasons for using a "convenient" sample, therefore, was that the author's familiarity and experience as a tutor with the general learning environment and the type of students recruited in the specific institution would aid towards making appropriate choices and ultimately to a better understanding of the findings.

In addition, a crucial criterion for selecting a sample is its size. One of the main criticisms of most studies in the field is the lack of generalisability because of relatively small sample size, and the use of non-representative samples (arising from a combination of convenience sampling and self-selection). Later on in the strategy, it will be clarified exactly how the sample size relates to the statistical power and how this influences the development of regression models in terms number of explanatory variables (i.e. predictors) included in the regression

[13] By similar studies, the author means studies whose aim was to identify deep and surface approaches towards studying through students' interactions in ILEs when practising exercises during tutorial sessions in the classroom.

[14] "Cold-start" is a term borrowed from the context of automated systems whose purpose is to detect profiles based on users' actions. Al-Azawei and Badii (2014) discuss that in a "cold start" there is no initially data to detect profiles and initialize models.

models. However, it is worth emphasising here that the bigger the sample size the more statistical power is ensured, and the greater the number of predictors allowed to be included in the models (i.e. the bigger the sample size the more metrics can be included in the models to explain the interactions in terms of studying approaches). One of the reasons, therefore, for choosing a sample of convenience was that it offered the potential to work with a big size course, which would ultimately result in a big size sample.

In the current study, therefore, convenience sampling could not be avoided and in that respect the generalisability of the study was compromised to a certain degree.

More specifically, the sample used in the current study occurred from a combination of sampling of convenience and self-selecting (since the sample depends on students' attendance). The study was carried out at the School of Computer Science, where the author is a member of staff and so is familiar with the environment, regulations and procedures and as well as staff. The course selected was the mathematics course "Information Fundamentals". This was a core course, taken by all first-year undergraduates in the specific school during the first semester. It offered the potential of working with a big size sample, as previous records showed that there were approximately 240 to 300 first-year undergraduate students registering each year and that there was also good attendance on this course.

3.9 Sample

The initial sample consisted of 276 first-year undergraduate students from a London-based university who registered for the common core course of mathematics in the School of Computer Science. However, 38 students withdrew from the course and another 5 second-year students, repeating the module, never attended the tutorial sessions or registered in AM. This resulted in working with a sample size of 233 students who registered in the course and in AM as users. The specific course takes place during the first semester of the first year of their computer science studies. In the sample, there are 190 males and 43 females aged between 18 and 46.

3.10 Description of the study

3.10.1 Planning the study

Planning the study required a great number of details to be pre-specified in order for the study to run smoothly and in parallel with the real learning process, which was one of the main considerations set out in section 3.7.

Planning of the study started in January 2005, when the author introduced the idea of developing an ILE for the undergraduate computer-science students to the leader of the maths course and the teaching and learning coordinator of the computer science department.

Around the same time the author approached the team of DFKI from the University of Saarbrücken in Germany, regarding the provision of ActiveMath and because of DFKI's experience in action analysis; an experience that was valuable for the current study since the data for navigational behaviour would be collected through web-logs. Discussions with both parties continued over the following months.

3.10.1.1 Planning with the maths teaching team

The author discussed with the leader of the maths course issues relating to:

- How the course is organised
- The usual level of attendance in the course
- Which chapters from the existing handbook would be most useful to integrate in AM
- When and how AM would be introduced and used during the tutorial slots; conducting a pilot study in order to test the process and AM
- How the learning material could be structured and designed for an ILE
- How the study would be designed in a way that would not disrupt but help the learning process and the learning outcomes of the course

Eventually, it was agreed that:

- A pilot study would take place during the summer school in August 2005 with undergraduate students who were deferred or referred in maths over the academic year 2004-2005.
- Chapters with the most demanding graphs (such as Functions, Graphs and Matrices[15]) would benefit from an ILE in which improved pictorial representation and a graph plotter would be provided.

[15] There were exercises in the subchapter of "solving simultaneous equations using the inverse matrix" which would benefit from the use of the graph plotter.

- Storyboards indicating the instructional and interface design[16] would be checked and approved first by the leader of the maths course before proceeding with the implementation.

AM would be involved in:

- All tutorial sessions which, based on the records of previous years, were estimated to be around 15 sessions.
- The first teaching week (from the 26th to the 30th of September) during which students would register and get familiar with AM before using it for learning purposes (in order to avoid any disruptions during the learning process).
- The third teaching week (from the 10th to the 14th of October) for the learning of chapters “Functions” and “Graphs”.
- The fourth teaching week (from the 17th to the 21st of October) for the learning of chapter “Matrices”.

It was also agreed that:

- Due to overlapping tutorial sessions in the timetable, a research assistant would help with the observational tasks and registration process.

Administration of the ASSIST questionnaire would take place during the tutorial sessions of the revision week, for the following reasons:

- The instructions at the start of the ASSIST questionnaire state that students should answer the questions in terms of a particular course
- Judging from previous years, there would be enough time during the revision week to administer the questionnaire without disrupting the learning process; in previous years, the revision week also had very good attendance.

The author discussed with the six tutors involved in the maths course the scope and settings of the study, the use of AM, and in general how the process would be conducted in parallel with the learning process each week. In general, the tutors seemed keen on integrating AM in their tutorials and agreed with the research process.

As none of the tutors had encouraged collaboration between students in previous years (since the tutorial exercises have not been designed for group

[16] The decisions in relation to interface design (what type of features are needed, what type of features exist, and how they would be customised) were also influenced by previous research on styles and navigational behaviour.

work), it was agreed the students would continue to work on their own. This point was considered important for the study since collaboration between students could alter a student's own behaviour in AM.

Tutors' intervention [17] could also alter a student's own behaviour in AM. However, asking tutors to intervene less was not a possibility as this would have meant severe interference with the learning process.

3.10.1.2 Planning with the AM team

A series of meetings with DFKI in Germany was arranged at the start of June 2005. During the first meeting the author presented the proposed study and discussed the following issues: what it was feasible to implement considering the human resources and time constraints; details on how the study would be conducted; and how the application could be customised to suit the research questions of the study and the learning outcomes of the maths course.

In terms of the process, it was concluded that:

- The AM team could calculate most of the proposed "interaction" metrics[18]
- Learning material would be broken down into more pages as this would provide more detail in relation to what type of learning material the subject examines and for how long[19]
- There should be close supervision and strict instructions during the registration process to avoid double registration
- There should be a demonstration of AM to the students and written instructions should be provided before the registration process in order to avoid mistakes
- There should be an administration account for the author in order to access accounts, create accounts, delete potential double registrations, and help the students create a new password in case they forgot their existing one

In terms of the application, it was concluded that:

[17] This was raised when some tutors explained that they tend to give specific tasks or conduct revisions.

[18] Certain metrics could not be calculated. For example, the use of the browser buttons or the number of external links explored through the search engine could not be calculated.

[19] Another solution was the use of the AM eye-tracker feature. The eye-tracker highlighted exactly what the users were pointing at, while the rest of the content became blurred. However, it was decided that its use would be disruptive for the learning process and it was removed from AM.

- AM features such as search engine, help and hyperlinks would be customised to serve the research questions of the current investigation and the learning outcomes of the course
- AM features such as the different visual presentation for each element (that is theory, example, and exercise) would be maintained
- AM features required by the maths course leader such as the graph plotter did not exist in AM and could not be implemented by the AM team
- The AM team would provide software development training to the author and they would also help with the integration of the learning material in AM (and specifically the exercises)
- Given the time constraints it was only feasible to implement the three chapters mentioned above

The above conclusions helped to further shape the process and the application.

3.10.1.3 Implementation of AM

Following the meetings with maths course leader and DFKI, the design and implementation of the application was intensified. The author designed (according to the given feedback) and completed the storyboards on the 19th July 2005, and the author designed a graph plotter to fit the requirements of the learning material.

The integration of the three chapters in AM took place from mid-June until mid-September. The implementation consisted of image processing (e.g. creation of diagrams) and XML programming. The final testing of AM took place from the 14th September to the 22nd September 2005 in various computer labs using a range of browsers. Technical errors and typos were corrected.

It is estimated that for the design of storyboards, implementation, and testing of AM the author spent approximately 1000 hours.

3.10.1.4 Design of AM

The design version of AM used in the current study was based on the work of DFKI team (Melis et al., 2006). However, its interface and its features were changed to suit the study, the learning outcomes of the maths course, and the requirements set by the module team.

The AM interface and features were customised to integrate the learning material, as defined in the syllabus, and serve the learning outcomes of the course. The interactive learning environment was therefore changed to integrate concepts, procedures, examples and exercises in verbal (text and numerical) and visual form with regards to Functions, Graphs and Matrices.

In addition, AM allows students control in terms of what theory example and exercise they will view and practise on and what options they will use to access them. Initially, students login with their own username and password. On the home page, students have to choose between the three available chapters: Functions, Graphs, and Matrices (see Appendix 3.4.9). After choosing a chapter, students can navigate through two main options:

- The table of contents (TOC), which is consistently placed on the left-hand side of the screen (see Appendix 3.4.1). It is placed in a position which can draw students' attention, as it starts from the top-left corner of the interface, which is considered a primal optical area (Lidwell et al., 2010), and expands to the bottom-left corner of the interface. The items in the TOC are ordered and structured according to the topics of the curriculum. Each topic may be a theory, an example or an exercise and is presented on a link to either a superordinate category (e.g. "Linear Graphs") or a subordinate category (e.g. "The meaning of gradient"). In the TOC, students can follow the specified order, or they can jump to a specific topic.
- Students can also navigate in a linear way using the Previous/Next buttons at the bottom of the page (see Appendix 3.4.2).

These were navigational options which were already provided in AM and it was considered reasonable to retain both, as providing different ways for the users to navigate through the material adds more flexibility to the way the information is accessed in a system (Issa and Isaias, 2015; Dix et al., 1998). When selecting a topic from the TOC or clicking on the Previous/Next button, the relevant content is displayed on the right-hand side of the screen. There are two different types of learning content presented on the right-hand side of the screen:

- Reading page: students can view the mathematical concepts, theory, and working examples for mathematical procedures (see Appendices 3.4.1 and 3.4.10). The reading pages also have hyperlinks which allow the student to visit new concepts or revisit previous concepts by opening a pop-up window with the relevant information.

- Exercise page: students can view on each exercise page a group of exercises (see Appendix 3.4.2). For each exercise, there is a short description and a “Start exercise” link which opens a pop-up window, allowing students to work on the exercise. There are three types of exercises: multiple choice, multiple selections and fill-in-the-blank. In the fill-in-the-blank exercises the type of response can be numeric or numeric-plus-strings. After selecting (or typing), students receive feedback which tells them whether their answer is correct. If the answer is incorrect, they are encouraged to try again up to three times. Throughout their three attempts, the previous wrong answers are retained and highlighted by AM (see Appendices 3.4.4-3.4.7). After the third attempt, they receive the correct answer (see Appendix 3.4.4). Students can also cancel an exercise at any point.

  Furthermore, the decision to implement three different type of exercise was based on the way the exercises were presented on the textbook of the module. It was agreed with the module team that it was important to keep the way the exercises were presented consistent between the textbook and AM as much as possible (as students would also be allowed to consult the textbook during the tutorial sessions). The decision to allow the students to try three times for the correct answer before getting the right answer from AM was also something which was decided by the module team.

  Students have constant access to Help, Home Page and Logout, and the interactive and investigative features of Notes, Graph Plotter, and Search (see Appendix 3.4.1). These options are grouped in a horizontal bar placed at the top-right corner of the screen. Although they do not occupy the primal optical area, they are placed in the “strong fallow area” which is the next area in the interface to attract the users’ attention (Lidwell et al., 2010). The students can interact with these tools as follows:

- Students can choose two different types of graph plotter, depending on the exercise. For example, the “two lines” graph plotter can be used in simultaneous solutions for two linear equations and the “parabola and line” graph plotter can be used in simultaneous solutions for quadratic and linear equations. Each graph plotter has two panels. Students can change the graphs in the right-hand panel by moving the sliders in the left-hand panel for changing the quadratic coefficient, the linear coefficient, the constant (for the quadratic equation), and the gradient and intercept for the linear equation. Students can see the change of function graphically and numerically.

- Students can write notes by clicking on the Notes option. This action opens a pop-up window in which they can type and save a note (see Appendices 3.4.8.a-3.4.8.c).
- Students can enter a maths term in the text field of the Search option (see Appendix 3.4.3). After submitting their query, there is a list of search results from which they can choose. There is also an option to expand their search beyond the AM tutorial by searching for the term on other websites such as Google, Wikipedia and MathWorld.

The decision to include the above tools was based on what was considered by the module team as beneficial for the students, and it was also based on the needs of the study (i.e. forming associations between the use of features and studying approaches). More specifically, with regards to the use of graph plotter, it was a request made by the module team as it was thought that it would contribute to the process of solving specific exercises. The beneficial role of the graph plotter was also supported in the literature, as indicated in section 2.1.1 it could help students to understand the connections between graphic and numerical representation.

Furthermore, the “notes” feature was an existing feature in AM which was retained for the current investigation. This was because making notes was found capable of being linked to deep approaches towards studying, as discussed in 3.4.1. So, its use contributes towards forming associations with the studying approaches, but also at the same time it may encourage a deep approach towards studying.

Similarly, the “search” feature was also an existing feature in AM which was retained for the current investigation. This was because its use can offer a more exploratory and active retrieval of the mathematical concepts which suits students with an “interest in ideas” approach, as discussed in 3.4.2. So, its use contributes towards forming associations with the studying approaches, but also at the same time it may encourage a deep approach towards studying.

Finally, there was a decision to retain the hyperlinks (concept links) in the text of the AM pages. As discussed in 3.4.2, the use of hyperlinks for mathematical terms in the text may be connected to the “relating ideas” and “interest in ideas” approaches towards studying. Besides contributing towards forming associations with the studying approaches, it may also facilitate students in relating mathematical concepts.

3.10.2 The pilot study

The pilot study was conducted on 15th August 2005 during the summer school[20]. Previously, the author had prepared AM accounts for all students in order to help them with their registration, prepared the materials for ASSIST questionnaire, and created an observation sheet.

3.10.2.1 The process

It was expected that 24 deferred and referred students in maths would participate, however, summer school was not compulsory and only three students attended.

The session with AM followed after a revision lecture by the maths course leader where the author was present. At the start, the author presented AM to students and helped them to log in. The students used the usernames and passwords set by the author. The above tasks took 15 minutes. The maths course leader instructed them to start from Linear Graphs without specifying, though, whether they should start from exercises, examples or theory. They used AM for the duration of 1.5 hours.

During this time, the author tried to keep her intervention minimal. However, because of the students' slow progress (and after the course leader's indication), the author suggested (half an hour before the end of the session) exploring a different subchapter.

At the end of the session, the ASSIST questionnaire was administered. The completion of the questionnaire took an average time of 15 minutes.

3.10.2.2 Conclusions from pilot study

Because of the poor attendance, the study did not have any value in terms of collecting data but only in terms of testing the process and the AM application. In terms of the AM application, the pilot indicated one technical problem in relation to the graph plotter which was addressed in time for the main study. In terms of process, it was observed that students initially needed time to familiarise themselves with the interface, structure and navigation of AM. This fact reinforced the idea of planning a session with AM activities before the actual use of AM for learning purposes in the main study.

---

[20] The purpose of the summer school was to prepare the deferred and referred students for their August exams.

3.10.3 Main study

After conducting the pilot study and until the commencement of the main study - on 19th September - the author continued with the preparation of materials and the refinement of the process for the main study. During the registration week, the number of tutorial sessions was finalised to 15.

3.10.3.1 Preparatory task for observation

Starting from the observation task, the forms used in the pilot study seemed to cover well the requirements for recording incidents and interactions that may influence the relationship. Observations would give complementary information about incidents or interactions that may influence the relationship between students' learning styles and students' behaviour in AM such as tutor's tasks, incomplete sessions due to technical difficulties, students' absences from the tutorial sessions, reasons for short attendance; and collaboration between students.

The observation task was planned to be as unobtrusive as possible. The observer would record the number of the computer used by each student, and match it to the student ID number in the attendance list.

3.10.3.2 Preparatory task for registration

In relation to the registration process, it was decided that the students would have to register themselves. The author designed the online registration form so that all the form elements were mandatory and there were simple and concise instructions for its completion (see Appendix 3.10.4). Additional instructions would be given in a written form (see Appendix 3.10.2), along with the manual for AM, to avoid overloading the interface of the online form with more details.

3.10.3.3 First week of study - registration and introduction to AM

During the first teaching week, the students were introduced to AM. The presentation lasted an average time of 15 minutes. During the presentation, the author demonstrated the features and interface of AM, highlighting the advantages of its use. The author also explained that users' AM interactions would be recorded for research purposes and asked the students' permission. Furthermore, it was highlighted that the students could contact the author via email for any kind of technical assistance with AM. After the presentation, the author distributed the instructions for registration (see Appendix 3.10.2) and explained the registration process.

The main problem encountered during the first day was that a number of students double-registered. The students' IDs of these students were recorded on the observation sheet and double-registrations were tracked down and deleted through the administration account (see Appendix 3.10.5). Double-registrations decreased after highlighting more frequently the importance of checking the form with the tutor before submitting it.

After the registration session, the students were given the AM manual (see Appendix 3.10.1) and a sheet with AM navigational exercises (see Appendix 3.10.3). These exercises had the purpose of helping the students become familiar with AM in order to avoid disruptions during the learning process (as suggested in 3.10).

At the end of the first week, 195 students had registered and spent time getting familiar with AM. In 10 out of the 15 tutorial sessions, the students used AM approximately for 1 hour. There were, however, 3 tutorial sessions that were taking place at the same time. As a result, three tutorial groups, which were introduced to AM only during the second hour of the tutorial session, had an average time of use of 20 minutes.

3.10.3.4 Second week of study - registration of remaining students

By the end of the first week, it was noticed that not all expected students had attended the first week's tutorial sessions. It was estimated that there were approximately 43 students who had not registered in AM[21], because they did not show up to the tutorial sessions. It was decided with the maths course leader that during the second teaching week (from 3rd to 7th October) the author would visit the classes and help register those students who had not attended during the first week. The registration took place during breaks, at the very start or the very end of the two-hour tutorial session to avoid any disruptions. By the end of the second week, 13 more students had registered and there were 208 students in total in the AM registry.

3.10.3.5 Conclusions from first and second week

The events and observations of the first two weeks of the study revealed that it

[21] During the first three weeks, there were students who did not show up, changed courses etc. As a result, it was not feasible at that point to know the exact number of those who were registered on the course.

is essential to have a pre-session where students register and test the ILE. This way, any potential disruptions during the learning process can be avoided; but also the collection of web log data will not be influenced heavily by technical and administrative problems, which frequently occur during the first week of a big size course, as well as by the inexperience of users with the ILE.

3.10.3.6 Third week of study - Use of AM for learning purposes

During the third week, from 10th until 14th October, AM was used for the first time for learning purposes, specifically for the chapters of “Functions” and “Graphs”. The tutorial sessions ran smoothly since most students were familiar by now with the environment and the login process. There were 15 more students who registered because they had not attended the course in the two previous weeks. This raised the number of AM users to 223.

The author recorded observations for each tutorial session with the help of a research assistant who covered the overlapping sessions. The observers recorded incidents and interactions.

Finally, during the third week, 170 students attended the tutorial sessions and participated in AM. According to the web logs, the average time for AM use during the two-hour tutorial session was 1 hour and 27 minutes.

3.10.3.7 Fourth week of study - use of AM for learning purposes

During the fourth week, from 17th until 21st October, AM was used for the learning of the “Matrices” chapter. There were 7 new students who registered in AM raising the number of possible AM users to 230. During the fourth week, there were 154 students who attended and used AM. According to the web-logs, the average time for AM use during the two-hour tutorial session was 1 hour and 26 minutes. In general, sessions ran smoothly and the observation process was conducted similarly to the third week.

3.10.3.8 Fifth week of study – use of AM for revision

During the fifth week, there was unexpected use of AM by some tutors who used it for revision purposes. The tutorials of the fifth week, from 24th to 28th October, were dedicated to the revision for an in-class test that would take place during the sixth teaching week. Three tutors planned their revision in a way that included AM, so it was used by 90 students in 7 tutorial slots and the average time of use was 42 minutes[22].

During the fifth week 5 more students, who attended for the first time, registered in AM. This raised the number of possible AM users to 233 out of the 238 who had registered for the course[23].

3.10.3.9 Fifth week of study – administration of ASSIST

ASSIST was administered in all 15 sessions during the tutorials of the revision week. Out of the 184 students who attended the tutorial sessions, 174 students completed ASSIST voluntarily. There were 10 students who did not complete it because they had to leave class early for personal reasons. In most tutorials, the administration of ASSIST took place at the end of the session; however, in three sessions which were taking place at the same time the process took place at the start of the session. The author initially explained the purpose of the questionnaire and the study. She also repeated the instructions written on the questionnaire, emphasising that students should think of their answers in terms of the maths course. The process lasted an average time of 20 minutes per tutorial session, and the actual completion of the questionnaire lasted an average of 15 minutes per tutorial session.

After collecting the data, ASSIST was examined statistically in terms of internal reliability (i.e. internal consistency) using the statistical method of Cronbach's α in SPSS. According to Field (2009), the generally accepted value is 0.8, however the value of α depends also on the number of items, so a value of 0.5 and above can be also respectable when there are a few items in a scale. The "deep" and "surface" main scales appeared to have good internal consistency, with α =0.804 and α=0.837, respectively. With regards to the "deep" subscales of "seeking meaning", "interest in ideas", "use of evidence" and "relating ideas" the value of α ranges between 0.518 and 0.593, which is considered respectable given that they consist of only 4 items. Finally, regarding the "surface" subscales of "lack of purpose", "unrelated memorising", "syllabus boundness" and "fear of failure", the value of α ranges between 0.563 and 0.737, which is considered respectable given that each subscale consists of only 4 items. So, overall, the scales of ASSIST appeared to have good internal consistency.

---

[22] The average time was less compared to previous weeks because the tutors also spent time revising chapters not included in AM.

[23] There were 5 students in total who did not attend any tutorial sessions and never registered in AM, possibly because they were retaking the module.

3.10.3.10 Data collection of secondary background variable – prior knowledge

Data in relation to the secondary background variable, prior knowledge (that is, their maths level based on their entry qualifications), was collected together with the completion of ASSIST measurement. On the same way data was also collected in relation to the secondary background variables age and gender, however, as stated in a previous chapter, the intention is to focus on the influence of the prior knowledge.

The level of maths was based on the students' university entry qualifications and it was discussed and assessed in collaboration with the module leader of the maths module team. The great disparity in students' prior knowledge in mathematics due to accepting students (especially through the clearing process) with a variety of entry level qualifications and the way it affects the students' learning, teaching and performance has been always a concern for the maths module team of the university in which the current investigation took place, and the specific year the study took place it was no exception: the students had again great disparity with regards to prior knowledge. This reinforced the argument made in the literature about the influence of prior knowledge. In addition, measuring prior knowledge based on their entry qualifications, although not a straightforward and completely unbiased process, seemed the most reasonable way of collecting such data at the time, since logistically a pre-test on their maths knowledge could not be implemented and administered.

For the purpose of the current study, as will be shown later on, it was essential to distinguish between two groups.[24] So, after identifying the different entry qualifications, then two groups were formed. The first group consisted of those students, who were considered to have low prior knowledge and it was called "low prior knowledge" group. This group consists of students with GCSE maths qualifications graded D-G[25]. The second group consisted of those students, who were considered to have medium to high prior knowledge and it was called "high prior knowledge". This group consists of students with the following maths qualifications: GCSE graded A-C and diplomas or certificates (e.g. BTEC) in

[24] As is shown in 2.11.3, this distinction between the two prior knowledge groups will be used to split the sample for each multiple regression model in each ASSIST scale and subscale.

[25] Those were students who at the time had certificates in AVCE ICT or BTEC National Diplomas (relevant to computer science and multimedia subjects) and in which either there were no maths units or they were optional and were not taken. Hence, their GCSE maths qualifications were considered.

which maths units were taken; Foundation course in Computer Science (where maths units are part of the course's diet); European/International Baccalaureate (where maths units with standard or high level were taken); AS-level Maths (graded B-D), and A-level Maths (graded E-D).

3.10.3.11 Ethical issues

The author asked the ethics committee of the University of Westminster for permission to carry out the study with the students at university. Permission was given to go ahead after it was confirmed that:

- The participants did not belong to a vulnerable group such as children, and adults with physical or mental illnesses.
- The nature of the study was not intrusive in that it would not cause any mental or physical harm.
- The author would obtain the informed consent of the students to use the data for research purposes (verbal consent was sufficient).
- None of the participants would lose out since all participants would use ActiveMath and complete the questionnaires. The study would not be designed to include a control group (which would not use ActiveMath) and a treatment group (which would use ActiveMath), and all students would use the same version of the ILE.

The last point was especially important since the study was conducted in real learning conditions. In the methodology and the description of the study, it was shown how this fact shaped and produced a number of considerations: planning the tasks at an early stage and in such detail that the study could run smoothly and in parallel with the learning process; applying techniques of data collection which are as unobtrusive as possible (e.g. recording web logs); avoiding setting tasks that would disrupt the learning process; and avoiding setting tasks at a time, place and in a way that would be inconvenient for the tutors, the students, and the learning objectives of the course.

Furthermore, the author explained to the students at the beginning of each data collection process (e.g. before administering the tests and when presenting AM), that the data would be used only for research purposes and that the results would be confidential and anonymous. The participants were also asked orally whether they agreed to allow their data to be used for research purposes. All students agreed to allow their data to be used for research purposes. In the case of the administration of ASSIST (which was not part of the normal learning process), students were asked if they would like to participate and it was also made clear to them that they could withdraw their consent to participate at any time. The above questionnaire was also introduced to students according to the guidelines of their creators. With regards to the completion of the ASSIST measurement, 10 students out of the 184 declined to participate and left the class. Feedback on Entwistle's questionnaire was given on demand.

Overall, it was thought that participants would benefit from the experience, and the application was designed in a way to ensure that it would accomplish this. ActiveMath was adjusted, with guidance by the people teaching the course, in order to cater for the needs of the students and the learning outcomes of the maths course. It was also thought that ActiveMath as a computerised and a more interactive way of learning maths would engage students whose subject of study is Computer Science.

In addition, it was very clear to the fellow lecturers that the study would take place for one academic year only.

To preserve the anonymity of the participants, the author used the student ID of the students rather than their surname and name in the data files. Files that would allow matching student IDs to students' names were stored on a password protected machine.

3.11 Strategy for analysis

In the current investigation, the strategy suggested for the statistical analysis has been selected based on the research aims and research questions. Both the research aims and questions aim to examine the relationships between deep and surface approaches towards studying and "interaction" metrics, but also go a step further and examine whether combinations of "interaction" metrics can explain or predict deep and surface approaches towards studying. Hence, correlational analysis and multiple regression analysis have been deemed appropriate statistical methods. In this section, the author examines how these methods are being applied.

3.11.1 Final sample-size involved in the analysis

There were 233 students who actually participated in the study. From this sample, 117 students attended the two tutorial weeks in which AM was used for learning purposes as initially scheduled (i.e. third week and fourth week) and completed the ASSIST measurement. During the process of 'data-cleaning, there was also further exclusion of two more cases due to lack of interaction in AM which resulted in 'NULL' data, resulting in the final sample size of 115[26].

3.11.2 Correlational Analysis

The process of correlational analysis is typically a process which is appropriate for correlational design. Most importantly it serves as a statistical method which indicates which "interaction" metrics are the most relevant to be included in the regression models based on their empirically statistically significant relationships to the ASSIST scales and subscales. Applying correlational analysis prior to multiple regression is for this purpose is quite common (Steele, 2003), but it is especially useful for a "cold-start" type of investigation, such as the current one where there are no really prior empirical findings in a similar context to indicate which metrics are the most relevant for each approach.

As stated in 3.3.1, ASSIST is an ordinal-level measurement, whereas the students' "interaction" metrics are considered interval-level measure. According to De Vaus (2002), a Spearman correlation coefficient (r_s) is the preferred statistic for relations between ordinal variables with a lot of categories (such as the ASSIST scales and subscales), and it is also a suitable method for examining the associations between ordinal-level and interval-level measures. As a result, the Spearman correlation coefficient is calculated to test the hypotheses. The value of the Spearman correlation coefficient is calculated for each hypothesised relationship, with statistical significance set at $p=0.05$.

The analysis was carried out on 115 students of the sample, described in section 3.11.1, who attended and used AM during the third and fourth weeks of the study, and who completed ASSIST.

3.11.3 Multiple Regression – Development of models

3.11.3.1 General strategy and statistical measures

According to the research questions discussed in 3.1, to identify the students' interactions with regards to each studying approach, as measured in ASSIST, it is appropriate to apply multiple regression analysis. It is a typical technique employed in behavioural science when several variables predict a quantitatively measured criterion variable (Meyers and Gamst, 2013). So, in each regression

[26] It was clarified with the DFKI team, that in these two cases there was no data with regards to temporal metrics because they have not visited the AM pages. These two cases are typically called "outliers for data cleaning" (Meyers et. al., 2006).

model the explanatory variables (or independent variables or predictors) are amongst the “interaction” metrics mentioned in 3.4, and the dependent variable (or criterion variable) is each ASSIST deep and surface scale and subscale as discussed in 3.3.2.

The strategy with regards to the development of regression models is formed based on a combination of the accepted statistical practices when developing multiple regression models and specific tactics which have been adopted to aid and enrich the interpretation of the empirical findings in the specific context. The author performed the analysis using the SPSS statistical software. However, it is important to mention that the author does not rely on automated statistical methods (such as “forward” and “backwise” which are typically offered in statistical software such as SPSS) for the development of the models. Field (2009) and Meyers et al. (2006) discuss these methods along with various criticisms and particularly that independent variables with good predictive qualities on their own may be awarded very little weight in the model, in which case the researcher need to exercise some judgement as to which variables to enter. Meyers et al. (2006) recommend a “researcher-controlled” method when developing a model, where the researcher judges which predictors should be included in the model. They also suggest that selection of predictors should be based on a particular theory and/or empirical basis. Field (2009) also supports that the development of models should not be out of the hands of the researcher. The author’s intention is to follow a “research-controlled” method, as it can ensure better that the most relevant predictors are included based on the theory and the empirical findings.

The statistical measures which are used during the development of the models are the ones which are typically suggested by statisticians such as Field (2009), Meyers et al. (2006) and Steele (2003). More specifically, the following statistical measures are used:

- **R^2** which is the variance explained by the model.
- **Adjusted R^2** which takes some account of the inflation error and shows the shrinkage of the explained variance if we were to apply the model in the population. Throughout the process, we should keep an eye on the difference between R^2 and Adjusted R^2, as their values should be as close to each other as possible.
- **Sig.** which is the value for the overall significance of the model.

- **Sig.** which is the value for the significance of each predictor.
- **b value** which is partial regression coefficients (relative contribution of each predictor while controlling for the effects of the other predictors). They reflect raw scores. Because variables are assessed in different metrics, you cannot see based on b weights, which independent variable is the strongest predictor in the model.
- **beta value** which is standardised regression coefficient (weight for standardised predictors).

3.11.3.2 Clarifications on expectation in terms of effect size and variance explained

Before discussion the models representing the surface and deep scale, it is useful to clarify certain statistical issues regarding the thresholds of statistical measures for multiple regression. Cohen (1992) sets thresholds for correlations and multiple regression (see Appendix 3.11.1). Cohen (1992) suggests an "effect size index", which is widely cited when thresholds are needed to interpret statistics in behavioural sciences by statisticians such as Field (2009), but also in learning style research (Law and Meyer, 2009). To understand whether expectations are met in terms of the deep and surface models, it is reasonable, therefore, to use these thresholds in the current research. In terms of multiple regression, there are thresholds for the magnitude of the variance explained -or variance accounted R^2, and for the effect size d^2. As shown in Table 2 in Appendix 3.11.1, a small effect size of 0.02 corresponds to a small R^2 of 1.96% and R of 0.14; a medium effect size of 0.15 corresponds to a medium R^2 of 13.04% and R of 0.36; a large effect of 0.35 corresponds to a large R^2 of 25.92%, and R of 0.51.

It is commonly accepted, of course, that the higher the $R^{2,}$ the better the model fits the data. However, there are varied opinions regarding the magnitude of the R^2 and its importance with regards to the quality of a study (Moksony, 1990). There are also suggestions that in psychology and social science studies a less than large amount of variance is due to the unpredictability of human behaviour (Frost, 2014), and that it can be still valuable (Ramsey and Schafer, 2013).

Furthermore, researchers in the field of statistics suggest that magnitude of R^2 really depends on context or subject area of a study (Chalmer, 1987). So, at this point it is important to point back to our review of learning styles in interactive environments, where it is stated that medium effects (and therefore medium variance explained) are reasonable expectations. This is also enforced by the fact that the study took place in natural settings where unpredictable behaviour can manifest itself. Therefore, according to the aforementioned Cohen's thresholds, it is -at least- expected to be found a medium effect f^2 of 0.15, which corresponds to a medium variance explained of 13.4%.

3.11.3.3 Stages of development of models

The development of models consists of five main stages: 1) Initial selection of predictors, 2) Exclusion of outliers, 3) Exclusion of predictors 4) Decision-making on final recommended model, and 5) Examination of the influence of prior knowledge. As it is discussed in the following sections, these stages are based on the recommendations of statisticians such as Meyers et al. (2006), Field (2009) and Steele (2003), as well on what the author considers relevant and appropriate for investigating the research questions.

1) ***Initial selection of predictors***

During this **first stage of the development**, the following points need to be considered**:**

- Relevance of predictors based on theoretical and empirical connections

The first version of the model is formed based on theoretical connections and empirical connections (i.e. based on the initial correlational analysis between metrics and each scale). The "interaction" metrics with empirical connections to the ASSIST scales will be given priority and be included always in the model, over those metrics whose inclusion is based only on the theoretical connections to ASSIST scales found in the relevant literature related to ASSIST. If, for example, the maximum number of predictors allowed according to sample size (an issue discussed later on) is reached based on empirical connections (i.e. initial correlations between metrics and ASSIST scales), then the first version of the model will simply rely only on them. In the case where there are no correlations between a scale and "interaction" metrics or there are very few of them, then the first version of the model will simply rely more on predictors which occur from the theoretical connections and according to the author's judgement have the most potential to enlighten us in terms of students' interactions according to their studying approach. The idea is to "take advantage" of the maximum number of predictors which are allowed in the first version of the models according to the sample size, in order to obtain an as enriching and as complete a picture as possible of the interactions in AM which may explain a specific approach to studying. Furthermore, in cases where there is no clear indication, according to the theory, as to which metric would be the more "enriching" as a predictor for the model, or when there is a multicollinearity issue (as discussed below) the decision of inclusion will be made based on whether a predictor contributes to variance explained by the model in terms of R^2 and Adjusted R^2, ensuring at the same time that the model remains overall statistically significant. This means that the contribution of some predictors may be examined in terms of R^2 and Adjusted R^2 in a "trial" type version of the models (what the author calls "pre-models"), and then a decision will be made as to their inclusion in the first version of the model. It is also discussed later on that in general the aim is to recommend a model with the highest possible R^2, but also the highest possible Adjusted R^2. A pre-model, for example, where with the inclusion of a specific predictor (or combination of predictors), R^2 is increased while Adjusted R^2 is decreased compared to the rest of pre-models, will not be recommended.

Finally, as indicated by Field (2009) the predictors should have some variation in value, so if a predictor has variation of 0 value (or close to 0), then it will not be considered for inclusion.

- Multicollinearity

When deciding which predictors should be included in the first version of the model, it is important not to include predictors which correlate with each other. A commonly used threshold to judge multicollinearity between predictors is $r > 0.75$ (Meyers et al., 2006). If there is a "multicollinearity" issue, then a decision should be made as to which predictor to include. For example, in the group of temporal metrics, *average time view on content pages* is likely to correlate highly with the *maximum view time on content page*; or *average number of notes clicked per page* very likely correlates to *number of times "notes" link is clicked* (given that they are both based on the number of times the "notes" feature is accessed).

The decision, with regards to which metric is included, can be made according to the empirical connection between these metrics and the ASSIST scale with regards to statistical significance and strength of correlation (i.e. which one has the highest and statistically significant correlation). In the absence of an empirical connection between the metrics and the ASSIST scale, a decision can be made according to the theoretical connection between these metrics and the ASSIST scale (i.e. which one is deemed to have stronger theoretical connections or is deemed to have the most potential to enlighten us in terms of students' interactions according to a specific studying approach). Furthermore, in the absence of an empirical connection, or in cases when there are predictors which are deemed equally important empirically and/or theoretically (or in general when there is no clear indication as to which predictor will be the most "enriching" for the model), both predictors can be tried separately in the model to check which one contributes more. This solution can give a clearer indication as to which one contributes more based on the statistical measures of R^2 and Adjusted R^2. While following this process, there is also a need to keep an eye on the overall significance of the model (Sig. value), making sure that no matter which predictor is chosen, the model still remains overall statistically significant.

- Including appropriate number of predictors

  There are different methods to estimate the appropriate number of predictors included in a regression model. With regards to the first method, Field (2009) and Robson (2002) indicate the simple and quite commonly used rule of thumb of '1 predictor per 15 participants'. With regards to the second method, Field (2009) proposes a method which considers the expected effect size of the relationships between predictors and explanatory variable and the statistical power. With regards to the third method, Cohen (1992) proposes a method which considers expected significance and size-effects with regards to regression models. More specifically, this is how each of these methods calculate the number of predictors based on the current sample-size:

  - First method: The rule of thumb indicates 1 predictor per 15 participants. This means that for 7 predictors there should be 105 participants (7x15=105), or for 8 predictors there should be 120 participants (8x15=120). So, for the current sample of 115 participants there can be between 7 and 8 predictors.
  - Second method: It considers the effect size and indicates that if a medium effect is expected (which is according to the initial expectations as discussed in 2.6.4), then for 6 predictors in a model there should be a sample of at least 100 participants. This is to ensure statistical power of at least 0.8, which is a commonly accepted threshold proposed by Cohen (1988, 1992) (as cited in Field, 2009, p.223).

- Third method: Cohen (1992) indicates that to ensure statistical power of 0.8 (a commonly accepted threshold) and if, with regards to the regression models, the statistical significance is set at a=0.05[27] and the expected effect size is a medium $f^2=0.15$[28], then there should be sample for at least 107 participants corresponding to 8 predictors.

Based on the above three methods, it is reasonable, therefore to conclude that an appropriate number of predictors to include in the first version of each regression model for the sample size of 115 participants is maximum 8 predictors.

It is also worth mentioning that the estimations suggested in the above methods are useful throughout the development of the models. In the following stages, there will be exclusion of outliers as well as exclusion of predictors, meaning that it should be always ensured that there is an appropriate number of predictors according to sample size in the subsequent versions of the models, and especially in the versions suggested by the author (see Chapter 4). In that aspect, the three methods can help with the lowest limits with regards to the number of participants and corresponding number of predictors. Based on the first and third method, it is possible to have between 105 and 107 participants for 7 predictors; and based on the second method there should be at least 100 participants for 6 predictors. Throughout the development of the models, it has to be ensured that the number of participants and number of predictors in the model are well within the aforementioned limits.

2) ***Exclusion of outliers (using Cook's Distance)***

During the **second stage,** after running the model with the initial predictors, there is a need to rerun consecutive models in which one by one the outliers (extreme cases) are excluded. Meyers et al. (2006) calls them "multivariate outliers". The importance of detecting outliers when developing regression models is stressed by Field (2009), as it can cause the model to be biased. Steele (2003) proposes detection of outliers using "Cook's Distance" method. The box plot of Cook's distance shows the outliers through asterisks (the most extreme cases) and circles (less the extreme cases).

---

[27] This is the expected level of significance and all regression models will be developed according to this, which is the default in SPSS.

[28] This is the expected effect size, based on what is discussed in 2.6.4.

As indicated previously, there is a need to ensure that the sample size is according to the number of predictors, so it can be the case that it is not possible to exclude all outliers. A reasonable way to judge which outliers are candidates for exclusion is to focus on the most extreme cases (which are indicated in the boxplot by an asterisk), and then check whether the exclusion of each outlier increases the variance explained by the model based on the measures of R^2 and adjusted R^2. If the exclusion of an outlier decreases the variance, then it should not be excluded. This is a process which is supported in the field of multivariate analysis (Esbensen et al., 2002).

However, as this is a time-consuming process and it is not possible to exclude all outliers (since there is a need to keep an appropriate sample size for the number of predictors), the intention is to mainly focus on the exclusion of the most extreme cases (those indicated in the box-plot by an asterisk and not those indicated with a circle)[29]. This is also supported by Meyers et al. (2006) who suggest that if multivariate outliers are not very extreme, then it is better not to exclude them. This is an example of how the process will work. Suppose the “Cook’s Distance” boxplot indicates more than one extreme outlier (indicated with an asterisk). If there are 3 extreme outliers, cases 11, 22, and 33, then the exclusion starts from the most extreme case (e.g. case 22). Then the variance explained of the model (e.g. called “Model 2a”) is examined. If the exclusion of case 22 increases the variance explained by “Model 2a”, then “Model 2a” is re-run with the exclusion of both cases 22 and the next most extreme case (e.g. case 11). In the new (e.g. called “Model 2b”), if it is found that the exclusion of case 11 decreases the variance explained compared to the variance of Model 2a, then case 22 is not excluded from sample. Next, a new model is re-run (e.g. called “Model 2c”) in which cases 22 and 33 are excluded. If the exclusion of these two cases increase the variance explained by “Model 2c” compared to the one of “Model 2a”, then the exclusion of cases 22 and 33 in “Model 2c” is accepted.

3) *Exclusion of predictors (according to R^2, Adjusted R^2, beta value and Sig. value)*

The **third stage** starts with the rerun of the best version of the model, as determined by the second stage. The model is rerun in a consecutive way by removing the predictors one at a time. In this process the issues that need to be

[29] However, there can be exceptions if there are very few extreme outliers and it is found that less extreme outliers (the ones indicated by circles) can increase the variance of a model.

considered are: what the order of the elimination is and according to which criteria.

More specifically, the decision on which predictors to eliminate and in what order should be based on which predictors are statistically redundant (Field, 2009). The elimination of the predictors is therefore based on: which one is the most insignificant according to the *Sig.* value and the most unimportant according to the *beta* value. More specifically:

- It is sensible to start with the exclusion of predictors from amongst the most insignificant predictors; they can influence the overall significance of the model as well. The question is whether there will be a complete elimination of all predictors which are insignificant, as this is an accepted practice. The intention of the author is to create models which are meaningful and which offer a satisfactory insight with regards to students' interactions and their studying approaches. It is possible that a version of a model which is the "leanest and meanest", where all insignificant factors are eliminated, may not give a meaningful and insightful view of how students interact in AM in relation to the studying approaches. For example, cases where the remaining predictors or predictor cannot really highlight differences between a high deep and a high surface approach towards studying. So, the intention in the current investigation is to check, every time the most insignificant predictor is excluded, whether R^2 and Adjusted R^2 both start decreasing (or only Adjusted R^2 starts decreasing creating in this way an even bigger difference in relation to R^2) while making sure at the same time that the overall significance of model remains always below 0.05. A case for keeping an insignificant predictor in a model (and not obtaining the "leanest and meanest" version of the model), therefore, would be if: the overall significance of model remains below 0.05 and that its exclusion reduces the overall variance explained by the model. Keeping such a predictor in a way makes sense, given that it is included in the model at first place because it helps towards a meaningful and insightful model based on the theory and therefore serves towards the aim and the research questions of the current investigation.
- With regards to the exclusion according to *beta* value (which shows the contribution of a predictor in a standardised and more comparable way), the intention is to check it together with *Sig.* before making a decision about which predictor to eliminate. Usually, those two values go hand in hand in that the predictor with the weakest *beta* is usually also the most insignificant one (Field,

2009). When this is not the case, then the decision for elimination should based on the *Sig.* value of the predictor and the rest of the factors as discussed previously.

4) ***Decision making on suggested (recommended) model***

A decision has to be made as to which version will be the recommended one. As discussed in the third stage, the intention is not to recommend necessarily the "leanest and meanest" version. The recommended version of the model will be one which has overall statistical significance ($p<0.05$), and it is the best in terms of variance explained based on the statistical measures of R^2 and adjusted R^2. This means that the recommended version will have both the highest possible R^2 and Adjusted R^2. Or in absence of such a model version, the recommendation will be for the version of the model with the same or decreased R^2 but an increased Adjusted R^2 compared to the previous version. This means that the recommended model may have predictors which individually are not statistically significant. The reason for this decision is that the "leanest and meanest" version might not give us as enriching and insightful information when interpreting students' interactions in the learning environment for an approach to studying, as the recommended version does. This decision of course can put limitations in terms of generalisation when evaluating and interpreting the model at predictor-level, so the intention is to showcase also the "leanest and meanest" version of the model for each scale and subscale and make comparisons which ultimately may contribute to useful conclusions in the discussion of the empirical findings.

At this stage, it is essential to ensure that the models and especially the recommended version of the model fits well the observed data. In other words, it is possible to generalise to other samples at model-level. As Field (2009) states: "generalisation is a critical additional step, if we find that our model is not generalizable, then we must restrict any conclusions based on the model to the sample used." The generalisation at model-level can be checked according to whether the following assumptions hold for the model (which are commonly recommended by statisticians):

- Assumption regarding normal distribution of residuals, using normal distribution diagram (Steele, 2003).
- Assumption of whether the variance of residuals is constant (i.e. there is homoscedasticity of standardised residuals against predicted ones) (Steele, 2003).

- Assumption regarding normality, with a plot which shows whether points lie on a straight line (Steele, 2003).

5) ***Examination of the influence of prior knowledge***

One of the research questions, indicated in 3.1, examines whether and to what extend different levels of the prior knowledge in mathematics influences the interactions of students with deep and surface approaches towards studying when learning mathematics using the interactive learning environment AM in tutorial sessions in the classroom. To examine the influence of prior knowledge, the intention is to include it as a selector variable in the recommended model, which means splitting the sample into two groups (levels): low prior knowledge and high prior knowledge. SPSS provides the option to split the sample into two groups, as part of the multiple regression process. In this way, it will be possible to examine whether students' interactions are explained better with regards to a specific studying approach in the low or high prior knowledge group. As prior knowledge is a secondary background variable, and the examination of its influence is of complementary and secondary value for this investigation (as indicated in research aims and questions), the intention is to keep the examination of the low and high prior knowledge models for each ASSIST scale at model-level and simply make comparisons with regards to R^2.

According to the aforementioned methodology, in the following chapter, Chapter 4, there will be individual discussion with regards to the development, analysis and interpretation of the multiple regression models for each ASSIST scale.

Chapter 4 - Regression Analysis & Model Interpretation

In this chapter the theoretical assumptions (i.e. hypotheses) in relation to subscales of ASSIST and the "interaction" metrics are discussed in detail. Then the intention is to examine if there is a relationship between ASSIST scales and the way students interact with a mathematical interactive learning environment while practising in their tutorial sessions. The "interaction" metrics are an attempt to measure the way students interact in AM when practising mathematics in their tutorial sessions during weeks 3 and 4 (so the metrics express the combined interactions which took place during these two weeks). By examining the correlations between these measures[30], the intention is to answer the first research question, but also have an indication of what metrics to include in the first version of the model for each scale. Furthermore, this chapter will demonstrate the development of the model for each scale according to the strategy indicated in 3.11. A discussion will follow with regards to the empirical findings for each model.

4.1 The "surface" scale and students' "interaction" metrics

4.1.1 Surface Scale – theoretical assumptions

The "surface" scale measures the extent to which there is: an intention to memorise and treat the content as unrelated bits of knowledge; motivation to avoid failure; an overall negative attitude to studying and an intention to cope minimally with the course requirement; and an intention to follow strictly the instructions and the structure of the learning content, and focus on the minimal requirements of the course (Entwistle, 1997a; Entwistle and Ramsden, 1983). It is based on the research conducted by Marton and Säljö (1976b) who find that some students have an intention to complete the task with very little engagement and with unreflective memorisation.

During the learning process:

- In performance-related metrics, Entwistle and Ramsden (1983) find negative correlations between the surface scale and performance. So, in general educational context, high values on surface scale may result in low values in performance. Their overall negative attitude towards studying, and their intention to cope minimally and with unreflective memorisation, may also

[30] Part of these results were published in a conference paper by Margeti and Mavrikis (2015).

influence their performance when practising their exercises in tutorial sessions in a similar way. It is likely that the higher the score on the surface scale, the lower the number of exercises students are likely to solve on the first attempt, and the higher the number of exercises they solve on third attempt or fail to solve at all. Therefore, it is expected that there will be negative association between:

- *Surface scale and number of exercises solved on the first try*

While, there will be positive associations between:

- *Surface scale and number of exercises solved on the second try*
- *Surface scale and number of exercises solved on the third try*
- *Surface scale and number of exercises finished but not solved.*

- In terms of search-related metrics, because of their lack of engagement and intention to cope minimally (Entwistle and Ramsden, 1983; Entwistle, 1997a), it is also expected that students with a high score on surface scale are not likely to try to elicit more information with regards to mathematical concepts and procedures. This means that students with high values on surface scale may result in low values in metrics related to researching a concept further –an aspect which is represented by number of hyperlinks visited in reading and exercise pages and the metrics related to the search option, as indicated in section 3.4.2. So, it is expected that there will be negative associations between surface scale and:
    - *number of times search option is clicked*
    - *number of submitted queries in search option*
    - *number of search results visited in search option*
    - *number of hyperlinks (concept links) visited in reading and exercise pages*
- In terms of path metrics, for the same reasons mentioned earlier in this section, as students with high surface scores have a tendency to repetitive overlearning of material, they are likely to interact more closely around a certain set of pages. It is expected that there will be positive associations between:
    - *Surface scale and compactness*
- In terms of temporal metrics, students with higher values on surface scale may experience more difficulties when solving exercises, hence spending an

increasing amount of time in an effort to solve them, compared those with lower scores. So, it is expected that students with high scores on surface scale may result in high values on average view time on exercise and reading pages. Another reason for this positive association can be that surface consists of the “fear of failure” subscale. So, students with a high score on the surface scale may work slower, putting more effort into tasks and persisting longer when solving exercises due to their anxiety (Entwistle, 1981). Furthermore, as also indicated in 3.4.4, the metric of maximum view time on exercise and reading pages can indicate some sort of “extreme temporal interaction” due to the aforementioned difficulties. To conclude, positive associations are expected between surface scale:

- *average view time on exercise page*s
- *average view time on a reading pages*
- *maximum view time on a reading page*
- *maximum view time on an exercise page*

4.1.2 Surface Scale – Results on Correlations

Following the proposed methodology in 3.11, we run correlational analysis to identify predictors for the model. There are statistically significant correlations, as expected between:

- *Number of exercises solved on first try with* r_s*=-0.368*
- *Number of exercises solved on third try with* r_s*=0.314*
- *Number of exercises finished but not solved with* r_s*=0.270*
- *Compactness with* r_s*=0.216*
- *Average view time on exercise pages with* r_s*=0.183*

There is also, an unexpected positive (and not negative as initially assumed), statistically significant relationship between the surface scale and *number of hyperlinks (concept links) visited on reading and exercises pages*, with r_s=0.197.

Regarding other metrics, mentioned in section 4.1.1, there are no statistically significant relationships. However, the importance of these metrics for the surface scale is discussed further in the following section with regards to the regression models.

4.1.3 Surface models: Development, Analysis and Discussion

In this section, there is a brief discussion on the development of regression models for the surface scale and analysis and discussion on the results for the selected model version.

4.1.3.1 Surface Scale - Initial selection of predictors

Table 4.1.3.1 shows briefly that the reason behind the initial selection of 6 predictors is their correlations to the scale. As indicated in the methodology, the sample size allows for the inclusion of up to 8 predictors and the intention is to take full advantage of this upper limit. So, the potential of including two more predictors is examined amongst the rest of the predictors mentioned in 4.1.1[31]. The decision is to include two temporal predictors, because of theoretical connections to the scale, and the potential of enriching the discussion (see further information in Table 1 and Table 2 of Appendix 4.1.1).

Selected Predictors	Reason for selection
Number of exercises solved on first try	Statistical (see 4.1.2)
Number of exercises solved on third try	
Number of exercises finished but not solved	
Compactness	
Average view time on exercise pages	
Number of hyperlinks (concept links) visited in reading and exercise pages	
Maximum view time on exercise page	Theoretical connections and enriching the discussion further
Maximum view time on content page	

Table 4.1.3.**1**. Selected predictors for first version of model.

As a result, at this stage, “Model 1b” is the first version for the “surface” scale with all the selected eight aforementioned predictors, which is the maximum number predictors according to this thesis’s strategy, indicated in the methodology.

4.1.3.2 Surface Scale – Development of model

For “Model 1b” the variance explained R^2 is 37.3% and the Adjusted R^2 is currently 32.5%. The process of improving both these measures of variance and the overall significance starts by excluding specific outliers. This results in

[31] Certain of these predictors have not been considered because their variation is close to 0 (i.e. the metrics related to the use of the search option).

“Model 2g” (see Table 4.1.3.**2** below) with R^2 and Adjusted R^2 at 45.6% and 41.3% respectively.

4.1.3.3 Surface Scale – Selection of Model

After excluding the outliers, there is gradual exclusion of predictors, as shown on Table 4.1.3.**2** below. From Model 2g to Model 3, Adjusted R^2 increases while R^2 gradually slightly decreases, making their difference smaller.

After Model 3 the regression is re-run three more times. In model 3 the best candidate for exclusion is *compactness* with the lowest beta (0.078) and high insignificance ($p>0.05$). In the leaner versions, Model 5 and Model 6, there is gradual exclusion of predictors: *number of hyperlinks (concept links) visited on reading and exercises pages*, and *number of exercises finished but not solved*. Model 6 is the leanest and meanest model where all 4 predictors are statistically significant (see Table 9 in Appendix 4.1.3).

	R^2	**Adj. R^2**	**Sig.**
Model 1b (all initially selected predictors)	37.3%	32.5%	0.000
Model 2a (exclusion of case 132)	41%	36.5%	0.000
Model 2b (exclusion of case 132 and 46)	42.3%	37.9%	0.000
Model 2c (exclusion of case 132, and 46 and 76) [Rejected]	41.6%	37%	0.000
Model 2d (exclusion of cases 132, 46 and 58)	42.7%	38.3%	0.000
Model 2e (exclusion of cases 132, 46, 58, and 25)	44.4%	40%	0.000
Model 2f (exclusion of cases 132, 46, 58, 25, and 35) [Rejected]	43.1%	38.6%	0.000
Model 2g (exclusion of cases 132, 46, 58, and 25 and 38)	45.6%	41.3%	0.000

Model 3 (exclusion of cases 132, 46, 58, 25, and 38) (exclusion of predictors: *average view time on exercise pages*)	45.5%	41.8%	0.000
Model 4 (exclusion of cases 132, 46, 58, 25, and 38) (exclusion of predictors: *average view time on exercise pages, and compactness*)	45%	41.8%	0.000
Model 5 (exclusion of cases 132, 46, 58, 25, and 38) (exclusion of predictors: *average view time on exercise pages, compactness, number of hyperlinks (concepts links) visited in reading and exercise pages*)	43.7%	41%	0.000
Model 6-Leanest and Meanest (exclusion of cases 132, 46, 58, 25, and 38) (exclusion of predictors: *average view time on exercise pages, and compactness, number of hyperlinks (concepts links) clicked on exercise and reading pages, and number of exercises finished but not solved*)	42.3%	40.1%	0.000

Table 4.1.3.**2.** Summary of measures of variance and significance for accepted and rejected models

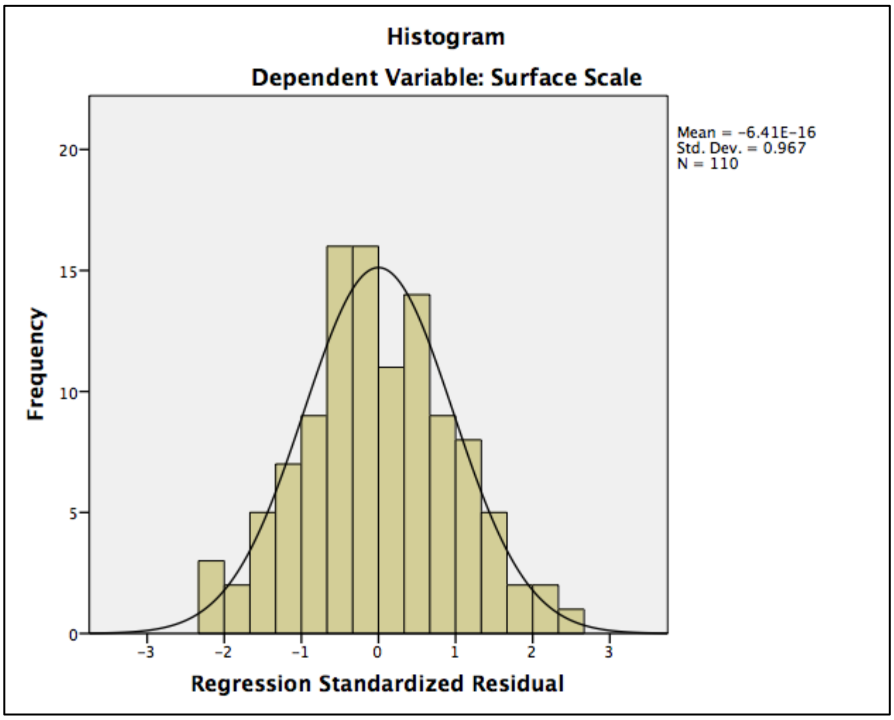
In Table 4.1.3.**2**, it is observed that the overall significance holds for all versions of the model ($p<0.05$). When comparing Model 3 to Model 4, Adjusted R^2 remains the same at 41.8%, whereas R^2 decreases slightly from 45.5% to 45%. In models 5 and 6, both R^2 and Adjusted R^2 decrease gradually to 42.3% and

40.1%, respectively. Therefore, according to our strategy, Model 3 is suggested as the best solution, as it combines simultaneously the highest possible R^2 and Adjusted R^2. Finally, Model 3 has 7 predictors, instead of the initial 8 predictors of Model 1, with a sample size of 110 (after the exclusion of 5 outliers), which is well within the thresholds stated in the strategy, in section 3.11.

4.1.3.4 Surface Scale – Model 3 – Generalisation

The assumptions regarding: normal distribution of residuals, homoscedasticity of standardised residuals against predicted ones, and the normality of residuals hold well.

*Figure 4.1.3.**1**. Histogram of standardised residuals for the final model*



Regarding the assumption about the normal distribution of residuals, as shown in Figure 4.1.3.1, residuals fit quite closely to a normal distribution.

*Figure 4.1.3.**2**. Plot of standardised residuals*

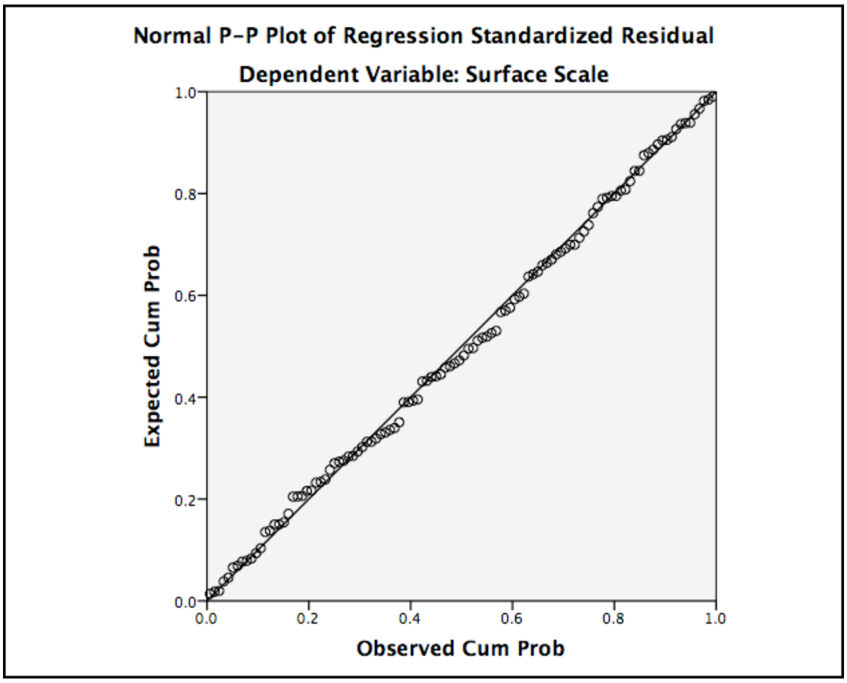
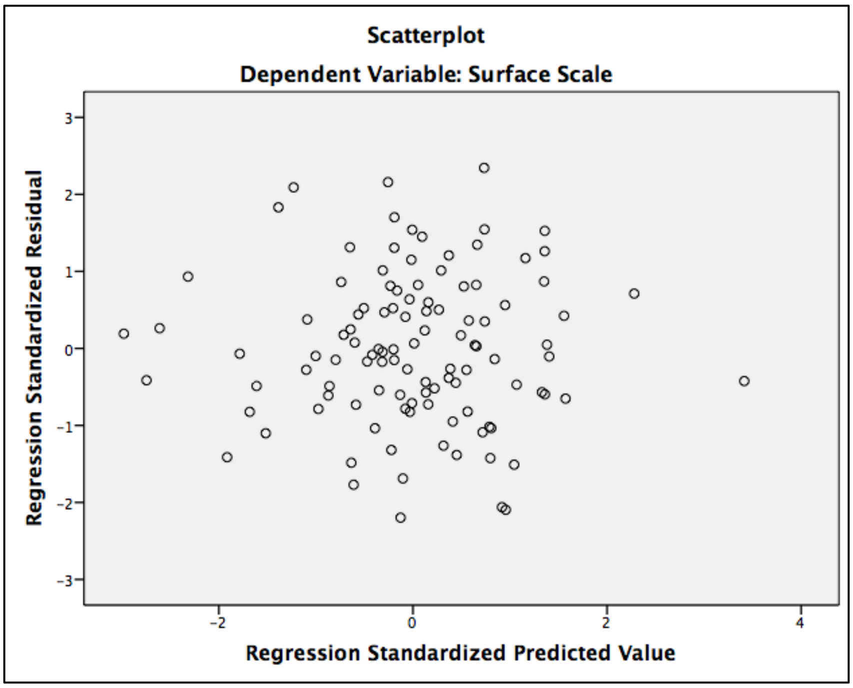


Figure 4.1.3.2 shows that the normality assumption holds since the points lie on the straight line.

*Figure 4.1.3.**3** Plot of the standardised residuals against the predicted ones for the model*



The assumption about the residuals is whether the variance of the residuals is constant, in other words there is homoscedasticity. Figure 4.1.3.3 shows that the scatter plot is reasonably random and the residuals are homoscedastic) and that most residuals are homoscedastic.

To conclude, the assumptions are not violated and therefore we can generalise the findings beyond the sample.

4.1.3.5 Surface Scale – Equation

By selecting Model 3, we can conclude that for both weeks, the surface is expressed through the predictors: number of exercises solved on first try, number of exercises solved on third try, number of exercises finished but not solved, maximum view time on exercise page, maximum view time on reading page, number of hyperlinks (concepts links) visited in reading and exercise pages and compactness. So, the equation, which is formed is as follows:

$SurfaceScale_i$ = b_0 - b_1(Number of exercises solved on first try)$_i$ + b_2(Number of exercises solved on third try)$_i$ + b_3(Number of exercises finished but not solved)$_i$ + b_4(Maximum view time on an exercise page)$_i$ - b_5(Maximum view time on a reading page)$_i$ + b_6(Number of hyperlinks (concepts links) visited in reading and exercises pages)$_i$ + b_7(Compactness)$_i$ (1)

If the b values found in Table 6 in Appendix 4.1.2 are replaced in the above equation, then the equation of the fitted regression model is obtained:

$SurfaceScale_i$ = 46.502 – 0.161(Number of exercises solved on first try)$_i$ + 0.805(Number of exercises solved on third try)$_i$ + 0.340(Number of exercises finished but not solved)$_i$ + 0.002(Maximum view time on an exercise page)$_i$ - 0.004(Maximum view time on a reading page)$_i$ + 0.450(Number of hyperlinks (concepts links) visited in reading and exercises pages)$_i$ +7.580Compactness (2)

4.1.3.6 Surface Scale – Interpretation of model parameters

a) **Direction of relationship between predictors and outcome based on b values**

As the surface scale score increases, there is increment on the following predictors:

- *number of exercises solved on third try*
- *number of exercises finished but not solved*
- *maximum view time on an exercise page*
- *compactness*

- *number of hyperlinks (concepts links) visited in reading and exercises pages*

As the surface scale score increases, there is a decrease on the following predictor:

- *number of exercises solved on first try*
- *maximum view time on a reading page*

It is worth mentioning that the direction of relationships is as expected in the initial theoretical assumptions in 4.1.1. The exceptions are the relationships with regards to the predictor *maximum view time on a reading page,* and *number of hyperlinks (concepts links) visited in reading and exercises pages* (for which an opposite direction was initially assumed). This will be discussed further later on.

The issue at this point is whether it is possible to explain the “surface” approach towards studying based on the combined knowledge of “interaction” metrics, and whether there are distinguishable “interaction” metrics in students’ interactions for this specific approach.

b) **Further discussion of model and predictors**

First of all, the recommended model (model 3) explains approximately 45.5%. This amount of variance is large (see regarding recommended thresholds in Appendix 3.11.1). Furthermore, the leanest and meanest model (model 6) explains also a large amount of variance at 40.1%. Given that a medium amount of variance was expected, this is an indication that the model is generally giving us a good picture of students’ interaction in the AM with regards to the surface approach towards studying.

In addition, despite the fact that the model is overall significant (see Table 4.1.3.**2** above) and holds well with regards to the assumptions, it is observed that not all predictors are significant (see Table 6 in Appendix 4.1.2). The reason for which these predictors were kept is mainly because their inclusion would allow for a richer insight into students’ interaction with regards to the specific scale. So, the question is: does this “allowance” enrich the interpretation of the findings?

Model 3 with its 7 predictors can enrich the interpretation of findings, as it can draw a pretty good picture of (and help to identify) students with high scores on the scale. More specifically, if tutors detect that students –repeatedly- do not manage to solve the exercises on first try and need to try several times (which can be considered a sign of "gaming the system"), then there is cause for intervention. This is especially necessary if the same type of exercises (i.e. which have the same level of difficulty, or require the same type of logic or formula for their solution) continue to be solved on subsequent attempts or not at all. Furthermore, the *maximum view time on an exercise page* increases as the surface score increases. Given the aforementioned performance when solving exercises, this can be interpreted to put it simply as "getting stuck" on a specific exercise or types of exercises, which again can trigger a tutor intervention. Finally, students with high scores on the scale tend to follow a rather compact limited path when going through the learning material AM.

On the other hand, there are some unexpected relationships. With regards to *maximum amount of time on a content (reading) page*, the direction of its relationship to surface scale is negative which is not quite as expected. It seems that, in the context of the current study, students with a high score on surface scale are less likely to spend an increasing amount of time on a specific theoretical page, compared to those with low scores. This is also reinforced by the observations made in class, as a number of students would not go through the theory, even when they were "getting stuck" on specific exercises, unless the tutor advised them to do so. Furthermore, a positive relationship between a surface scale and *number of hyperlinks (concepts links) visited in reading and exercise pages* was also not expected. It can be the case that presenting mathematical concepts through hypertext links in reading and exercise in AM can be used as a mean for unreflective memorization or repetitive overlearning, at least in the context of the specific study.

In comparison to Model 3, Model 6 does not have "contradicting" predictors such as the *number of hyperlinks (concept links) visited in reading and exercise pages*, which can contribute to a more straightforward interpretation, however, it is less enriching for the findings as it does not include the predictors *compactness* and *number of exercises finished but not solved*. So, overall, Model 3 gives a more insightful picture of students' interactions in AM with regards to the surface approach towards studying.

4.2 Surface subscale "unrelated memorising" and students' "interaction" metrics

4.2.1 Unrelated memorising – Theoretical assumptions

This study approach is characterised as rote memorisation and defined by Marton (cited in Entwistle and Ramsden, 1983, p.56) as the process where: "Information can be held in store for longer periods by internal repetition (rehearsal) and if repeated sufficiently often (overlearning) it will become a permanent memory trace in episodic long-term memory (LTM)". Overlearning is related to reviewing chapters and textbooks (Rohrer et al., 2005), so it is reasonable to relate it to revisitation of pages in AM. Based on interviews by Entwistle and Ramsden (1983), science students with an intention for memorisation and unrelatedness are likely to try to remember formulae and procedures and use them to tackle problems – without reference to their mutual relationship.

During the learning process:

- Regarding path-length and visitation metrics, Mimirinis' and Dafoulas' study (2008) finds no significant correlations between the "unrelated memorising" subscale and path-length and number of visits in the system's practical section. However, there could be other associations with other type of "visitation" metrics. It is possible that those with higher "unrelated memorising" scores limit themselves by visiting more of the same rather than a variety of distinct AM pages, because they approach the learning material in a more repetitive, memorising manner (Entwistle and Ramsden, 1983). So, high values on "unrelated memorising" may result in low values in metrics, which indicate some sort of "non-repetitiveness" during practical sessions, like number of exercises accessed and number of distinct pages visited. So, negative associations are expected between "unrelated memorising" subscale and:
    - *number of exercises accessed*
    - *number of distinct pages visited*
- Regarding path metrics (see 3.4.5), students with high "unrelated memorising" scores have a tendency for repetitive overlearning of materials (Entwistle and Ramsden, 1983), so they are likely to interact more closely around a certain set of pages, compared to those with low scores. Hence, a positive association is expected between:
    - *"Unrelated memorising" subscale and compactness*

- Additionally, students with high “unrelated memorising” scores are more likely to follow the given content structure and navigate more passively and linearly, compared to the ones with low score (see 3.4.5). Accordingly, students with high “unrelated memorising” scores are likely to make more use of previous/next buttons, compared to those with low scores (see 3.4.6). Therefore positive associations are expected between the “unrelated memorising” subscale and:
    - *stratum*
    - *number of pages visited using previous/next buttons*
- In terms of revisitation metric, students with high scores on “unrelated memorising” are more likely to revisit parts of the learning material compared to those with low scores (Entwistle and Ramsden, 1983; Rohrer et al., 2005). Thus, it can be assumed that revisitation in AM pages may contribute to rehearsal and overlearning of mathematical concepts, procedures and exercises for those students. Hence, a positive association is expected between:
    - *“Unrelated memorising” subscale and relative amount of revisits*
- For performance-related metrics, in a more general context, there is empirical evidence that associates academic performance and the subscale of “unrelated memorising”: Tait and Entwistle (1996) find a negative statistically significant correlation between “unrelated memorising” subscale and performance (sample of 649 first-year undergraduate students). In the context of mathematics education, Schoenfeld (2006) and Gierl and Bisanz (2003) argue that solving a mathematical problem is a process where memorising procedures and formulas are unlikely to result in a successful outcome. In the current study, it is expected that students with higher “unrelated memorising” scores are unlikely to do well when solving exercises in mathematics, compared to those with lower scores. More specifically it is likely that the higher students score on the “unrelated memorising” subscale, the lower the number of exercises they are likely to solve on the first attempt, and the higher the number of exercises they are likely to solve in subsequent attempts or not at all. Therefore a negative association is expected between:
    - *“Unrelated memorising” subscale and number of exercises solved on the first try*

  Positive associations are expected between:

- *"Unrelated memorising" subscale and number of exercises solved on second try*
- *"Unrelated memorising" subscale and number of exercises solved on third try*
- *"Unrelated memorising" subscale and number of exercises finished but not solved*

- Regarding temporal metrics, students with higher values on "unrelated memorising" may experience more difficulties when solving exercises related to a combination of mathematical concepts, because they tend to treat different concepts as unrelated bits of knowledge (Entwistle, 2001, Entwistle, 1997a). Hence, those students may spend an increasing amount of time on exercises, possibly resulting in high values on *average view time* on exercise and reading pages. Furthermore, the metric *maximum view time* on exercise and reading pages can indicate an extreme interaction due to aforementioned difficulties. Thus, positive associations are expected between "unrelated memorising" subscale:
  - *average view time on exercise pages*
  - *average view time on reading pages*
  - *maximum view time on a reading page*
  - *maximum view time on an exercise page*

4.2.2 Unrelated memorising – Results on Correlations

Following the proposed methodology in 3.11, we run correlational analysis to identify predictors for the model. There are statistically significant correlations, as expected, between "unrelated memorising" subscale and:

- *Number of exercises solved on first try with* r*=-0.377*
- *Number of exercises solved on third try with* r_s*=0.211*
- *Number of exercises finished but not solved with* r_s*=0.198*
- *Compactness with* r_s*=0.230*
- *Average view time on exercise pages with* r_s*=0.224*
- *Relative amount of revisits with* r_s*=0.264*
- *Number of exercises accessed with* r_s*=-0.257*
- *Number of distinct pages visited with* r_s*=-0.218*

For other metrics mentioned in section 4.2.1 there are no statistically significant relationships. However, there was no expectation of a statistically significant correlation between the “unrelated memorising” subscale and:

- *number of hyperlinks (concept links) visited on reading and exercises pages, where there is a statistically significant and positive relationship with r_s=0.209.*
- *average number of times a “notes” link is clicked per page metric, where there is a statistically significant and positive relationship with r_s=0.207.*

4.2.3 Unrelated memorising models: Development, Analysis and Discussion

In this section, there is a brief discussion on the development of regression models for the “unrelated memorising” subscale and analysis and discussion on the results for the selected model version.

4.2.3.1 Unrelated memorising - Initial selection predictors

Table 4.2.3.1 shows that the main reason for initial selection of predictors is their correlations to the subscale. *Number of exercises accessed* and *number of distinct pages* have not been selected due to multicollinearity issues, although there have statistically significant correlations to the subscale (see further information in Table 1 of Appendix 4.2.1).

Selected Predictors	Reason for selection
Number of exercises solved on first try	Statistical (see 4.2.2)
Number of exercises solved on third try	
Number of exercises finished but not solved	
Compactness	
Relative amount of revisits	
Average view time on exercise pages	
Number of hyperlinks (concept links) visited in reading and exercise pages	
Average number of times a “notes” link is clicked per page	

Table 4.2.3.**1** Selected predictors for first version of model.

As a result, “Model 1” is the first version for the “unrelated memorising” subscale with all aforementioned eight predictors, which is the maximum number of predictors according to strategy (see 3.11).

4.2.3.2 Unrelated memorising – Development of model

As previously, the process of improving measures of variance and the overall significance starts by excluding specific outliers. This results in increasing the variance: from Model 1 with R^2 32.9% and Adjusted R^2 27.8% to Model 2c with R^2 40.4% and Adjusted R^2 35.8% (see Table 4 in Appendix 4.2.2).

4.2.3.3 Unrelated memorising – Selection of Model

After excluding the outliers, there is gradual exclusion of predictors, as shown in Table 4.2.3.**2** below.

	R^2	**Adj. R^2**	**Sig.**	**Comments**
Model 3 (exclusion of cases 69, 132, 46, and 127, and *relative amount of revisits*)	40.4%	36.3%	0.000	Note how Adjusted R^2 increases while R^2 gradually slightly decreases, making their difference smaller.
Model 4 (exclusion of cases 69, 132, 46, and 127, and *relative amount of revisits,* and *number of hyperlinks (concepts links) visited on exercise and reading pages*)	40.3%	36.9%	0.000	
Model 5 (exclusion of cases 69, 132, 46, and 127, and *relative amount of revisits, number of hyperlinks (concepts links) visited on exercise and reading pages,* and *average view time on exercise pages*)	40.2%	37.4%	0.000	
Model 6 (exclusion of cases 69, 132, 46, and 127, and *relative amount of revisits, number of hyperlinks (concepts links) visited on exercise and reading pages, average view time on exercise pages,* and *number of exercises solved on third try*)	40%	37.7%	0.000	

Model 7 - Leanest and Meanest (exclusion of cases 69, 132, 46, and 127, and *relative amount of revisits, number of hyperlinks (concepts links) visited on exercise and reading pages, average view time on exercise pages, number of exercises solved on third try,* and *average number of times a "notes" link is clicked per page*)	39.4%	37.7%	0.000	Leanest and meanest model where all three remaining predictors are statistically significant (see Appendix 4.2.4). Note also that all models are overall statistically significant ($p<0.05$).

Table 4.2.3.**2**. Summary of measures of variance and significance

In Table 4.2.3.**2**, Model 6 has simultaneously the highest R^2 and adjusted R^2, and with 4 predictors can offer a more enriching interpretation of the findings (compared to Model 7); hence it is suggested as best solution amongst all versions. Finally, Model 6 has 4 predictors, instead of the initial 8 predictors of Model 1, for a sample size of 111 (after the exclusion of 4 outliers) which is within the thresholds stated in the strategy, in section 3.11.

4.2.3.4 Unrelated memorising – Model 6 – Generalisation

For the selected Model 6 of the subscale, the assumptions are not violated and therefore it is possible to generalise the finding beyond the sample (see the assumptions in Appendix 4.2.3).

4.2.3.5 Unrelated memorising – Equation

Model 6 indicates that the "unrelated memorising" subscale is expressed through the predictors: number of exercises solved on first try, number of exercises finished but not solved, average number of times a notes link is clicked, and compactness. Thus, the following equation is formed:

$$\textbf{\textit{UnrelatedMemorisingSubscale}}_i = b_0 - b_1(\textit{Number of exercises solved on first try})_i + b_2(\textit{Number of exercises finished but not solved})_i + b_3(\textit{Compactness})_i + b_4(\textit{Average number of times a "notes" link is clicked per page})_i \quad (1)$$

If we replace the b values found in Table 7 in Appendix 4.2.2 in the above equation, then we obtain the equation of the fitted regression model:

UnrelatedMemorisingSubscale$_i$ = 10.967 – 0.064(Number of exercises solved on first try)$_i$ + 0.187(Number of exercises finished but not solved)$_i$ + 5.401(Compactness)$_i$ + 1.757(Average number of times a "notes" link is clicked per page)$_i$ (2)

4.2.3.6 Unrelated memorising – Interpretation of model parameters

a) **Direction of relationship between predictors and outcome based on b values**

As the unrelated memorising subscale score increases, there is increment in the following predictors:

- *number of exercises finished but not solved*
- *compactness*
- *average number of time a "notes" link is clicked per page*

As the surface scale score increases, there is a decrease in the following predictors:

- *number of exercises solved on first try*

It is worth mentioning that the direction of the relationships is as expected in the theoretical assumptions in section 4.2.1. The exception is the relationship with regards to *average number of times a "notes" link is clicked per page* for which there was no theoretical assumption. This will be discussed later on.

The issue here is whether it is possible to explain the "unrelated memorising" approach towards studying based on the combined knowledge of "interaction" metrics, and whether there are distinguishable "interaction" metrics in students' interactions for this specific approach.

b) **Further discussion of model and predictors**

First, the recommended model (model 6) explains a large 40% of variance (see regarding recommended thresholds in Appendix 3.11.1). As a medium amount of variance was expected, this indicates that the model generally gives a good picture of students' interaction in AM with regards to the "unrelated memorising" approach.

Additionally, despite the model being overall significant (see Table 4.2.3.**2** above) and holding up well in regards to the assumptions, we observe that one predictor is not significant (see Table 7 in Appendix 4.2.2). This predictor *average number of times a "notes" link clicked per page* is kept chiefly because its inclusion might yield a richer insight into students' interaction.

So the question is: Does the inclusion of *average number of times a "notes" link clicked per page* offer any insight or use for interpretation? In Model 1 it was included purely based on the correlations indicated in 4.2.2, so it is interesting to examine why there might be such statistical indications, based on the observations made in class. More specifically, it was observed that some students would access and use the "notes" feature of AM simply to copy and paste learning material from AM reading pages or simply record answers to exercises. This type of interaction accords with trying to remember formulae and procedures rather than understand – a characteristic of "unrelated memorising". Note-making has been considered in the literature as a sign of deep learning because it shows students' effort to impose their own structure on learning material, however in the current study the "notes" feature in AM was used differently by some students. The current data does not allow for a link between the "unrelated memorising" score and the type of notes students make in AM, however in a similar study this aspect is worth further exploration.

The rest of the predictors in the model can give a useful and insightful interpretation of the findings. For example, tutors may have cause for intervention if they detect that students repeatedly fail to solve exercises on first try or cannot solve them at all and instead allow AM to provide the answer (a behaviour which can be considered as "gaming the system"). Another indication for intervention can be if students tend to follow a highly "compact" path and interact more closely around a limited set of pages than do students with low scores on this subscale.

In comparison to Model 6, the leanest and meanest Model 7 (with the elimination of the predictor *average number of times a "notes" link is clicked per page*) gives a slightly less enriching picture with regards to the "unrelated memorising" subscale, but it certainly gives a clear indication of a surface approach.

Overall, it is the author's opinion that the suggested predictors of Model 6 offer a useful insight into students' interactions with regards to the unrelated memorising approach; however, the inclusion of *average number of times a "notes" link is clicked per page* requires further investigation in future. More specifically, to reinforce the above interpretations, it would be worth examining further in future studies if students with high scores on unrelated memorising subscale tend to use "note-taking" features more to serve "unrelated memorising" tactics. Finally, to enrich further the above interpretations, it would be also worth examining whether those with high scores on the subscale tend to target more identical exercises when following a compact path, compared to those with low scores.

4.3 The surface subscale "fear of failure" and students' "interaction" metrics

4.3.1 Fear of failure - theoretical assumptions

The "fear of failure" subscale measures the extent to which students are motivated to avoid failure (Entwistle, 1997b; Tait et al., 1998; Entwistle, 1981). Students with "fear of failure" have an over-anxious concern about possible failure and this affects the way they tackle their study (Entwistle, 1981).

During the learning process:

- In terms of temporal metrics, as discussed in 3.4.4, it is possible that students with higher score in the "fear of failure" subscale are likely to spend more time on average on an exercise or reading page, than the students with lower scores. Thus, we expect positive associations between "fear of failure" subscale and:
  - *average view time on exercise pages*
  - *average view time on reading pages*
  - *maximum view time on a reading page*
  - *maximum view time on an exercise page*
- In terms of performance-related metrics, based on the empirical findings by Entwistle and Ramsden (1983), it is found in a sample of 865 science first-year undergraduate students that there is a statistically significant, negative correlation between the factor of "fear of failure" and performance. In a more recent study, Tait and Entwistle (1996) find, in a sample of 649 first-year undergraduate students, that there is a statistically significant, negative correlation between "fear of failure" subscale and performance. In the context of the current study, the above findings lead to the general assumption that students with high "fear of failure" scores are not likely to perform well when practising with the AM exercises. Furthermore, as indicated in 2.1.7 and 3.4.7, students' interaction with the exercises can indicate whether students are simply abusing the affordances of the environment to achieve good results but

with questionable learning gains. This "gaming" behaviour, where students try different solutions without a systematic approach and take advantage of the system's answers, has been linked empirically to anxiety about failing. It is expected that there will be negative associations between:

- *"Fear of failure" subscale and number of exercises solved on the first try*

While it is expected that there will be positive associations between:

- *"Fear of failure" subscale and number of exercises solved on the second try*
- *"Fear of failure" subscale and number of exercises solved on the third try*
- *"Fear of failure" subscale and number of exercises finished but not solved*

4.3.2 Fear of failure - Results on Correlations

Following the proposed methodology in 3.11 we run correlational analysis to identify predictors for the model. There are statistically significant correlations, as expected, between "fear of failure" and:

- Number of exercises solved on first try with r_s=-0.263
- Number of exercises solved on third try with r_s=0.262
- Average view time on exercise pages with r_s=0.218
- Maximum view time on an exercise page with r_s= 0.258

For other metrics mentioned in section 4.3.1 there are no statistically significant relationships. However, the importance of these metrics for the "fear of failure" scale is discussed further in the following section with regards to the regression models.

4.3.3 Fear of failure models: Development, Analysis and Discussion

In this section, there is a brief discussion on the development of regression models for the "fear of failure" subscale and analysis and discussion on the results for the selected model version.

4.3.3.1 Fear of failure - Initial selection predictors

Table 4.3.3.1 shows that the main reason for initial selection of predictors is their correlations to the subscale.

Selected Predictors	Reason for selection
Number of exercises solved on first try	Statistical (see 4.3.2)
Number of exercises solved on third try	
Average view time on exercise pages	
Maximum view time on an exercise page	
***Number of exercises finished but not solved*.**	Theoretical connections and enriching further the discussion.
Maximum view time on a reading page	

Table 4.3.3.**1**. Selected predictors for first version of model.

At this stage, these are the 4 predictors which are to be included in the first version of the model. As indicated in the methodology, the sample size allows for the inclusion of up to 8 predictors and the intention is to take full advantage of this upper limit. So, the potential inclusion of the rest of the metrics mentioned in 4.3.1 is examined.

With regards to the metric of *number of exercises finished but not solved*, besides the theoretical connection discussed in 4.3.1, its inclusion allows for useful comparisons to the rest of the “performance-related” metrics which are already included.

With regards to *average view time on reading pages* and *maximum view time on a reading page*, it is also enriching for the discussion of the “fear of failure” subscale to explore whether students with high score on the “fear of failure” subscale tend to persist and spend time, not only on the practical part of the AM learning material, but also on the reading pages of AM, which include the theoretical aspects of the learning material. Because there is a multicollinearity issue, it is not possible to include both. As there is no clear theoretical indication as to which of two metrics can be more enriching for the model, the intention is to choose the one which best contributes in the model. When comparing the variance between pre-models 1a and 1b (see table 1 in Appendix 4.3.1), it is observed that *maximum view time on a reading page* contributes more to the variance in Model 1b (compared to *average view time on reading pages* in Model 1a).

So, the first version of the model, “model 1” for the “fear of failure” subscale will include the 6 predictors indicated in table 4.3.3.**1**. This is according to the strategy indicated in the methodology (see section 3.11).

4.3.3.2 Fear of failure - development of model

As previously, the process of improving measures of variance and the overall significance starts by excluding specific outliers. This results in increasing the variance: from Model 1 with R^2 30.5% and Adjusted R^2 26.6% to Model 2f with R^2 42.1% and Adjusted R^2 38.8% (see Table 4 in Appendix 4.3.2).

4.3.3.3 Fear of failure – Selection of Model

After excluding the outliers, there is gradual exclusion of predictors, as shown in Table 4.3.3.**2** below.

	R^2	Adj. R^2	Sig.	Comments
Model 3 (exclusion of cases 132, 38, 78, and 52 and predictor of *number of exercises finished but not solved*)	42.1%	39.4%	0.000	
Model 4 - Leanest and Meanest (exclusion of cases 132, 38, 78, and 52 and predictors *number of exercises finished but not solved,* and *average view time on exercise pages*)	41.6%	39.4%	0.000	Selected model which is also the leanest and meanest as all predictors are statistically significant

Table 4.3.3.**2**. Summary of measures of variance and significance.

In Table 4.3.3.2, it is observed that the overall significance holds for all versions of the model ($p<0.05$).

In comparison to Model 3, in Model 4 there is a slight decrease for the R^2 from 42,1% to 41.6%, whereas the Adjusted R^2 remains the same at 39.4% (see Table 4.3.3.2). At this stage, we would normally select the more enriching version (that is, Model 3). However, in terms of predictors, the difference between the two models is the *average view time on exercise pages* which is not likely to contribute to a more enriching interpretation, because of contradicting empirical findings. More specifically, the b and beta values in Model 3 indicate a borderline negative relationship, whereas the correlation to the subscale indicates a positive one (see section 4.3.2), and the b value in Model 1 (see Table 3 in Appendix 4.3.2) is 0. Hence, it is suggested that the best solution is Model 4.

Finally, Model 4 has 4 predictors, instead of the initial 6 predictors of Model 1, with a sample size of 111 (after the exclusion of 4 outliers), which is within the thresholds stated in the strategy, in section 3.11.

4.3.3.4 Fear of failure –Model 4 - Generalisation

For the selected Model 4 of the subscale, the assumptions are not violated and therefore it is possible to generalise the finding beyond the sample (see the assumptions in Appendix 4.3.3).

4.3.3.5 Fear of Failure– Equation

By selecting Model 4, we can conclude that for both weeks, "fear of failure" is expressed through the predictors: number of exercises solved on first try, number of exercises solved on third try, maximum view time on an exercise page, and maximum view time on a reading page. So, the equation which is formed is as follows:

FearofFailureSubscale $_i$ ***=*** b_0 ***-*** b_1***(Number of exercises solved on first try)*** $_i$ ***+*** b_2***(Number of exercises solved on third try)*** $_i$ ***+*** b_3***(Maximum view time on an exercise page)*** $_i$ ***-*** b_4***(Maximum view time on a reading page)*** $_i$ ***(1)***

If we replace the b values, found in Table 7 in Appendix 4.3.2, in the above equation, then we obtain the equation of the fitted regression model:

FearofFailureSubscale $_i$ ***= 14.714 – 0.060(Number of exercises solved on first try)*** $_i$ ***+ 0.348(Number of exercises solved on third try)*** $_i$ ***+ 0.001(Maximum view time on an exercise page)*** $_i$ ***-0.001(Maximum view time on a reading page)*** $_i$ ***(2)***

4.3.3.6 Fear of Failure - Interpretation of model parameters

a) **Direction of relationship between predictors and outcome based on b values**

As the "fear of failure" subscale score increases, there is increment on the following predictors:

- *number of exercises solved on third try*
- *maximum view time on an exercise page*

As the "fear of failure" subscale increases, there is a decrease on the following predictors:

- *number of exercises solved on first try*
- *maximum view time on a reading page*

It is worth mentioning that while the direction of relationships is as expected in the initial hypotheses, there is one exception. The relationship between *maximum view time on a reading page* and the “fear of failure” subscale is negative, and not positive as it was initially suggested. This will be discussed further later on.

The issue at this point is whether it is possible to explain the “fear of failure” approach towards studying based on the combined knowledge of “interaction” metrics, and whether there are distinguishable “interaction” metrics in students’ interactions for this specific approach.

b) **Further discussion of model and predictors**

First of all, the recommended model, which is also the leanest and meanest version, (model 4) explains approximately 41.6%. This amount of variance is large (see regarding recommended thresholds in Appendix 3.11.1). Given that a medium amount of variance was expected, this is an indication that the model is generally giving us a good picture of students’ interaction in the AM with regards to the “fear of failure” approach towards studying. In addition, Model 4 is overall significant (see Table 4.3.3.2 above) and holds well with regards to the assumptions and all 4 predictors are also statistically significant (see Table 7 in Appendix 4.3.2).

Model 4 with its 4 predictors can enrich the interpretation of findings, as it can draw a pretty good picture (and help to identify) students with high scores on the subscale. More specifically, students with a high score on the subscale are less likely to solve exercises on first attempt and more likely to solve them on third attempt. Furthermore, the *maximum view time on an exercise page* increases as the “fear of failure” score increases. Combined with the aforementioned performance when solving exercises, this can be interpreted, to put it simply, as “getting stuck” on a specific exercise or types of exercises, which can trigger a tutor intervention.

In addition, with regards to the relationship between “fear of failure” and *maximum amount of time on a reading page*, its direction is negative, as opposed to positive, as initially assumed in section 4.3.1. It seems that in the context of the current study, students with a high score on “fear of failure” subscale are less likely to spend an increasing amount of time on a specific theoretical page, compared to those with low scores. This is also reinforced by the observations made in class, as a number of students would not go through the theory, even when they were experiencing difficulties with specific exercises, unless the tutor advised them to do so. Overall, it is the author's opinion that the predictors of Model 4 can offer a useful insight into students' interactions with regards to the "fear of failure" approach.

4.4 The surface subscale "syllabus boundness" and students' "interaction" metrics

4.4.1 Syllabus boundness - theoretical assumptions

As discussed in 3.3.2, the "syllabus boundness" subscale measures the extent to which there is a preference for clear instructions, clear deadlines and well-defined learning materials with clear structure (Entwistle et al., 1979; Entwistle and Ramsden, 1983; Entwistle, 1997a). Students with a high score on the scale are not likely to be autonomous when studying and they tend to study little beyond what is required to pass (they simply focus on the course's minimum requirements) (Entwistle, 1997a).

During the learning process:

- In terms of path metrics (based on what is discussed in 3.3.2) students with high scores on the "syllabus boundness" scale may go about their activities in a less autonomous and more orderly manner, compared to those with low scores. So, it can be the case that they tend to follow the structure of the learning material as indicated in AM more closely, compared to those with low scores. This means that high scores on "syllabus boundness" may result in high values in a "linearity" metric such as *stratum*. It is also expected that since students with high scores on the "syllabus boundness" subscale are likely to concentrate more on the part of the learning material which is deemed necessary to pass (Entwistle, 1997a), they are likely to interact more closely around a certain set of pages and follow a more "compact" path, compared to those with low scores on the subscale. This means that high scores on the syllabus boundness subscale may result in high values in *compactness*. It is expected that there will be positive associations between the "syllabus boundness" subscale and:
    - *compactness*
    - *stratum*
- In terms of performance-related metrics, it is possible that students with high scores on the "syllabus boundness" subscale, because they tend to study little beyond what is required (Entwistle, 1997a), are not likely to do as well when

solving exercises in mathematics as those with low scores. It is expected that there will be negative association between:

- *"Syllabus boundness" subscale and number of exercises solved on the first try*

While it is expected that there will be positive associations between:

- *"Syllabus boundness" subscale and number of exercises solved on the second try*
- *"Syllabus boundness" subscale and number of exercises solved on the third try*
- *"Syllabus boundness" subscale and number of exercises finished but not solved*

There is also a possible association between "syllabus boundness" and number of exercises cancelled. As discussed in 3.4.7, students with high scores on "syllabus boundness" tend to gear their studying closely just to what seems relevant to their assessment. This means that they may cancel for example, exercises that they consider not to be as relevant to the assessment requirements as other exercises. It is expected that there will be positive associations between:

- *"Syllabus boundness" subscale and number of exercises cancelled*

- In terms of temporal metrics, there can be associations with the "syllabus boundness" subscale, although the theory in studying approaches does not specifically state such an association. It is possible, though, that students with high scores on "syllabus boundness" may spend more time on a specific exercise and reading page, compared to those with low scores, especially if they consider these pages highly relevant to the assessment requirements, and especially if they experience difficulties when solving specific exercises in those pages. So, high scores on the "syllabus boundness" subscale may result in high values on average view time and maximum view times on both reading and exercise pages. It is expected that there will be positive associations between "syllabus boundness" subscale and:
  - *average view time on exercise pages*
  - *average view time spent on reading pages*
  - *maximum view time on a reading page*
  - *maximum view time on an exercise page*

On the other hand, as discussed in 3.4.4 with regards to the minimum view time on an AM page, due to the intention to engage minimally with the learning material and focus only on what they consider relevant to the assessment, it is possible that students with a high score on the “syllabus boundness” subscale may spend a decreasing amount of minimum view time on an AM page which they do not consider relevant to assessment requirements. So, it is possible that there is a negative association between the “syllabus boundness” subscale and:

- *minimum view time on an exercise page*
- *minimum view time on a reading page*

4.4.2 Syllabus boundness - Results on Correlations

Following the proposed methodology in 3.11 we run correlational analysis to identify predictors for the model. There are statistically significant correlations, as expected, between the “syllabus boundness” subscale and:

- Number of exercises solved on first try with r_s=-0.266
- Number of exercises solved on third try with r_s=0.191
- Number of exercises finished but not solved with r_s=0.190
- Number of exercises cancelled with r_s=0.192
- *Compactness* with r_s=0.190
- Average view time on reading pages with r_s=0.189

However, there was no expectation of a statistically significant correlation between the “syllabus boundness” scale and *number of pages visited using the TOC* with rs=0.183.

For other metrics mentioned in section 4.4.1 there are no statistically significant relationships. However, the importance of these metrics for the “syllabus boundness” subscale is discussed further in the following section with regards to the regression models.

4.4.3 Syllabus boundness models: Development, Analysis and Discussion

In this section, there is a brief discussion on the development of regression models for the “syllabus boundness” subscale and analysis and discussion on the results for the selected model version.

4.4.3.1 Syllabus boundness - Initial selection predictors

Table 4.4.3.1 shows that the main reason for initial selection of predictors is their correlations to the subscale. The sample size allows for the inclusion of up to 8 predictors (see section 3.11), so the intention is to consider an 8th predictor amongst those suggested ones in section 4.4.1, as its inclusion has the potential to enrich further the discussion regarding the "syllabus boundness" models. This predictor is *minimum view time on an exercise page,* as indicated in Table 4.4.3.1.

Selected Predictors	Reason for selection
Number of exercises solved on first try	Statistical (see 4.4.2)
Number of exercises solved on third try	
Number of exercises finished but not solved	
Number of exercises cancelled	
Compactness	
Number of pages visited using the TOC	
Average view time on reading pages	
Minimum view time on an exercise page	Its inclusion is enriching for the subscale, as discussed in 4.4.1. Amongst the rest of the temporal metrics, when tried on pre-models, it is the one that contributes best in terms of variance in the model along with the aforementioned predictors (see Model 1b in Table 1 in Appendix 4.4.1).

Table 4.4.3.**1**. Selected predictors for first version of model.

As a result, at this stage, Model 1b is the first version for the "syllabus boundness" subscale with all the selected eight aforementioned predictors, which is the maximum number of predictors, according to the strategy indicated in the methodology (see section 3.11).

4.4.3.2 Syllabus boundness - development of model

As previously, the process of improving measures of variance and the overall significance starts by excluding specific outliers. This results in increasing the variance: from Model 1b with R^2 19.7% and Adjusted R^2 13.6% to Model 2d with R^2 21.7% and Adjusted R^2 15.7% (see Table 4 in Appendix 4.4.2).

4.4.3.3 Syllabus boundness – Selection of Model

After excluding the outliers, there is gradual exclusion of predictors, as shown in Table 4.4.3.2 below.

	R^2	Adj. R^2	Sig.	Comments
Model 3 (exclusion of cases 35 and 31) (exclusion of predictors: *number of pages visited using TOC)*	21.7%	16.4%	0.000	Note how Adjusted R^2 increases while R^2 gradually slightly decreases, making their difference smaller.
Model 4 (exclusion of cases 35 and 31) (exclusion of predictors: *number of pages visited using TOC,* and *number of exercises solved but not finished)*	21.6%	17.1%	0.000	
Model 5 (exclusion of cases 35 and 31) (exclusion of predictors: *number of exercises solved but not finished, number of pages visited using TOC,* and *compactness)*	20.3%	16.5%	0.000	By running the regression model two more times, we obtain the leanest and meanest model. Note how both R^2 and Adjusted R^2 decrease. Model 6 has all 4 predictors statistically significant (see Appendix 4.4.4) Note also that all models are overall statistically significant ($p<0.05$).
Model 6- Leanest and meanest (exclusion of cases 35 and 31) (exclusion of predictors: *number of exercises solved but not finished, number of pages visited from TOC, compactness,* and *number of exercises solved on first try)*	18.7%	15.7%	0.000	

Table 4.4.3.**2**. Summary of measures of variance and significance

In Table 4.4.3.2, Model 4 combines the highest possible R^2 and adjusted R^2; hence it is suggested as the best solution amongst all versions. Finally, Model 4 has 6 predictors, instead of the initial 8 predictors of Model 1, for a sample size of 113 (after the exclusion of 2 outliers), which is within the thresholds stated in the strategy, in section 3.11.

4.4.3.4 Syllabus boundness – Model 4 - Generalisation

For the selected Model 4 of the subscale, the assumptions are not violated and therefore it is possible to generalise the finding beyond the sample (see the assumptions in Appendix 4.4.3).

4.4.3.5 Syllabus boundness – Equation

By selecting Model 4, we can conclude that for both weeks, "syllabus boundness" is expressed through the predictors: number of exercises solved on first try, number of exercises solved on third try, number of exercises cancelled, compactness, average view time on reading pages, and minimum view time on an exercise page. So, the equation which is formed is as follows:

SyllabusBoundnessSubcale$_i$ = b_0 - b_1(Number of exercises solved on first try)$_i$ + b_2(Number of exercises solved on third try)$_i$ + b_3(Number of exercises cancelled)$_i$ + b_4(Compactness)$_i$ + b_5(Average view time on reading pages)$_i$ - b_6(Minimum view time on an exercise page)$_i$ (1)

If the b values, found in table 7 in Appendix 4.4.2, are replaced in the above equation, then the equation of the fitted regression model obtained is as follows:

SyllabusBoundnessSubcale$_i$ = 10.871 – 0.014(Number of exercises solved on first try)$_i$ + 0.190(Number of exercises solved on third try)$_i$ + 0.129(Number of exercises cancelled)$_i$ +3.250(Compactness) + 0.007(Average view time on reading pages) -0.011(Minimum view time on an exercise page) (2)

4.4.3.6 Syllabus boundness - Interpretation of model parameters

a) **Direction of relationship between predictors and outcome based on b values**

As the "syllabus boundness" subscale score increases, there is increment on the following predictors:

- *number of exercises solved on third try*
- *number of exercises cancelled*
- *average view time on reading pages*
- *compactness*

As the "syllabus boundness" scale score increases, there is a decrease on the following predictors:

- *number of exercises solved on first try*

- *minimum view time on an exercise page*

It is worth mentioning that the direction of relationships is as expected in the initial theoretical assumptions in 4.4.1.

The issue at this point is whether it is possible to explain the “syllabus boundness” approach towards studying based on the combined knowledge of “interaction” metrics, and whether there are distinguishable “interaction” metrics in students’ interactions for this specific approach.

b) **Further discussion of model and predictors**

First of all, the recommended model (model 4) explains 21.6% of variance. This amount of variance is medium as expected (see the recommended thresholds in Appendix 3.11.1). However, this is a bit more than 1/5 of the variance of the model explained, and therefore, it is reasonable to say that we do not seem to get the full picture of how students with low and high scores on the “syllabus boundness” scale interact with AM during the tutorial sessions.

In addition, despite the fact that the model is overall significant (see table 4.4.3.2 above) and holds reasonably well all the required assumptions, it is observed that not all predictors are statistically significant (see Table 7 in Appendix 4.4.2). The reason they were kept is mainly because their inclusion will allow for a richer insight into students’ interaction with regards to the specific subscale. So, the question is: does this “allowance” enrich the interpretation of the findings?

The inclusion of all 6 suggested predictors in Model 4 has the potential to offer a useful insight. If tutors, or a system that records student interactions, detect certain tendencies, then there can be cause for intervention. More specifically, these tendencies relate to: not solving the exercises on the first attempt but rather on subsequent attempts; cancelling exercises; and following a rather compact limited path without going through specific exercise pages or spending very little time on them.

In comparison to Model 4, the leanest and meanest Model 6 (with the elimination of the predictors *number of exercises solved on first try* and *compactness*) gives a less enriching picture with regards to the “syllabus boundness” subscale (although Model 6 gives an indication that is a part of a surface approach mainly through the predictor *number of exercises cancelled*).

So, Model 4 is a more “enriching” model and the above interpretations on the findings seem to give a useful insight, however, as discussed earlier, the model still does not give the full picture (the variance explained is not as much as in other surface subscales), so it may be the case that factors such as prior knowledge have influenced the relationship of this subscale with the predictors. Another reason can be that the exercises do not require knowledge beyond what is delivered in class and through AM. Students' interactions might manifest more intensely for high and low scores on the subscale if the exercises required further knowledge (students exploring further concepts or more complicated exercises).

4.5 The surface subscale "lack of purpose" and students' "interaction" metrics

4.5.1 Lack of purpose – Theoretical assumptions

As discussed in 3.3.2, the "lack of purpose" subscale measures the extent to which students intend to cope minimally with the course requirements, because of their lack of interest in the subject (Entwistle et al., 2001; Entwistle and Ramsden, 1983). As a result, those with high scores on the scale are likely to engage less with their studies, compared to those with low scores on the scale.

During the learning process:

- In terms path metrics, because of their lack of interest and intention to cope minimally (Entwistle and Ramsden, 1983; Entwistle et al., 2001), students with high score on the "lack of purpose" subscale are likely to interact more closely around a certain set of pages and follow a more "compact" path, compared to those with low scores on the scale. This means that high scores in the "lack of purpose" subscale may result in high values in c*ompactness*. It is expected that there will be positive association between:
    - *"Lack of purpose" subscale and compactness*
- In terms of performance-related metrics, students with high score on the lack of purpose subscale, due to their lack of interest on subject and intention to cope minimally with their studies (Entwistle and Ramsden, 1983; Entwistle et al., 2001), may be experiencing difficulties when solving exercises which means that they are more likely to solve exercises on the second orthird attempt, or not at all, rather on the first attempt. It is expected that there will be negative association between:
    - *"Lack of purpose" subscale and number of exercises solved on the first try*

While it is expected that there will be positive associations between:

- *“Lack of purpose” subscale and number of exercises solved on the second try*
- *“Lack of purpose” subscale and number of exercises solved on the third try*
- *“Lack of purpose” subscale and number of exercises finished but not solved*

In terms of temporal metrics, in general students with high scores on the “lack of purpose” subscale, due to their lack of interest and their intention to engage minimally with the their studies (Entwistle et al., 2001; Entwistle and Ramsden, 1983), are more likely to spend less time on the reading and exercise pages on AM, compared to those with low scores. On the other hand, it can be also the case that those with high scores on “lack of purpose”, due to the difficulties they experience when solving exercises, tend to spend more time on specific AM pages, compared to those with low scores. So, with regards to temporal metrics and the “lack of purpose” subscale, while there can be an association, it is not possible to indicate its exact direction.

4.5.2 Lack of purpose – Results on Correlations

Following the proposed methodology in 3.11 we run correlational analysis to identify predictors for the model. There are statistically significant correlations, as expected, between “lack of purpose” subscale and:

- Number of exercises solved on first try with r_s= -0.220
- Number of exercises solved on third try with r_s=0.257
- Number of exercises finished but not solved with r_s=0.250
- *Compactness* with r_s=0.207

There was no expectation of statistically significant correlations between the “lack of purpose” subscale and:

- Relative amount of revisits with r_s=0.268
- Number of pages visited from TOC with r_s=0.195
- *Stratum* with r_s= -0.194

Finally, regarding the temporal metrics, there are no any statistically significant relationships, which is not surprising given that there is uncertainty as to the direction of the relationship, as mentioned in section 4.5.1. It seems that sometimes students with high scores on the “lack of purpose” subscale may spend more time on the learning material in AM due to experiencing difficulties compared to those with low scores; and sometimes they may spend less time on the learning material in AM due to their lack of engagement, compared to those with low scores.

4.5.3 Lack of purpose: models: Development, Analysis and Discussion

In this section, there is a brief discussion on the development of regression models for the "lack of purpose" subscale and analysis and discussion on the results for the selected model version.

4.5.3.1 Lack of purpose – Initial selection predictors

Table 4.5.3.1 shows that the reason for initial selection of predictors is their correlations to the subscale.

Selected Predictors	Reason for selection
Number of exercises solved on first try	Statistical (see 4.5.2)
Number of exercises solved on third try	
Number of exercises finished but not solved	
Relative amount of revisits	
Compactness	
Number of pages visited using the TOC	

Table 4.5.3.**1**. Selected predictors for first version of model.

It is worth mentioning that *stratum* is not included, despite its correlation to the subscale, due to a multicollinearity issue with *compactness* (for further information see Appendix 4.5.1).

The sample size allows for the inclusion of up to 8 predictors (see section 3.11). However, the inclusion of the temporal metrics, mentioned in 4.5.1, will not be examined further, as there are no clear theoretical or empirical indications regarding the direction of the relationship between these metrics and the subscale. As a result, at this stage, Model 1 is the first version for the "lack of purpose" subscale with all the selected six aforementioned predictors, which is according to the strategy (see section 3.11).

4.5.3.2 Lack of purpose – development of model

As previously, the process of improving measures of variance and the overall significance starts by excluding specific outliers. This results in increasing the variance: from Model 1 with R^2 16.1% and Adjusted R^2 11.4% to Model 2g with R^2 21% and Adjusted R^2 16.4% (see Table 4 in Appendix 4.5.2).

4.5.3.3 Lack of purpose – selection of model

After excluding the outliers, there is gradual exclusion of predictors, as shown in Table 4.5.3.2 below.

<table>
<tr><th></th><th>R^2</th><th>Adj. R^2</th><th>Sig.</th><th>Comments</th></tr>
<tr><td>Model 3 (exclusion of cases 76, 86, 70, and 24) (exclusion of predictors: number of pages visited from TOC)</td><td>21%</td><td>17.2%</td><td>0.000</td><td rowspan="2">Note how Adjusted R^2 increases, while R^2 remains the same.</td></tr>
<tr><td>Model 4 (exclusion of cases 76, 86, 70, and 24) (exclusion of predictors: number of pages visited from TOC, and compactness)</td><td>21%</td><td>18%</td><td>0.000</td></tr>
<tr><td>Model 5 (exclusion of cases 76, 86, 70, and 24) (exclusion of predictors: number of pages visited from TOC, compactness, and number of exercises solved on third try)</td><td>20.2%</td><td>17.9%</td><td>0.000</td><td rowspan="2">By running the regression model two more times, we obtain the leanest and meanest model. Note how both R^2 and Adjusted R^2 decrease.

Model 6 has all predictors statistically significant (see Appendix 4.5.4).

Note also that all models are overall statistically significant ($p<0.05$).</td></tr>
<tr><td>Model 6 –Leanest and Meanest (exclusion of cases 76, 86, 70, and 24) (exclusion of predictors: number of pages visited from TOC, compactness, number of exercises solved on third try, and number of exercises solved on first try)</td><td>19%</td><td>17.4%</td><td>0.000</td></tr>
</table>

Table 4.5.3.**2**. Summary of measures of variance and significance

In Table 4.5.3.2, Model 4 combines the highest possible R^2 and Adjusted R^2; hence it is suggested as the best solution amongst all versions. Finally, Model 4 has 4 predictors, instead of the initial 6 predictors of Model 1, for a sample size of 111 (after the exclusion of 4 outliers) which is within the thresholds stated in the strategy, in section 3.11.

4.5.3.4 Lack of purpose – Model 4 - Generalisation

For the selected Model 4 of the subscale, the assumptions are not violated and therefore it is possible to generalise the finding beyond the sample (see the assumptions in Appendix 4.5.3).

4.5.3.5 Lack of purpose – Equation

By selecting Model 4, it can be concluded that for both weeks, the "lack of purpose" is expressed through the predictors: *number of exercises solved on first try*, *number of exercises solved on third try*, *number of exercises finished but not solved*, and *relative amount of revisits*. So, the equation which is formed is as follows:

$$\mathit{LackOfPurposeSubcale}_i = b_0 - b_1(\text{Number of exercises solved on first try})_i + b_2(\text{Number of exercises solved on third try})_i + b_3(\text{Number of exercises finished but not solved})_i + b_4(\text{Relative amount of revisits})_i \quad (1)$$

If the b values found in table 7 in Appendix 4.5.2 are replaced in the above equation, then the equation of the fitted regression model is obtained:

$$\mathit{LackOfPurposeSubcale}_i = 6.028 - 0.013(\text{Number of exercises solved on first try})_i + 0.145(\text{Number of exercises solved on third try})_i + 0.124(\text{Number of exercises finished but not solved})_i + 6.129(\text{Relative amount of revisits})_i \quad (2)$$

4.5.3.6 Lack of purpose - Interpretation of model parameters

a) **Direction of relationship between predictors and outcome based on b values**

As the "lack of purpose" subscale score increases, there is increment on the following predictors:

- *number of exercises solved on third try*
- *number of exercises finished but not solved*
- *relative amount of revisits*

As the "lack of purpose" subscale score increases, there is a decrease in the following predictor:

- *number of exercises solved on first try*

It is worth mentioning that the direction of relationships is as expected in the initial theoretical assumptions in 4.5.1, except from the *relative amount of revisits* for which there are no initial theoretical assumptions (an issue which is discussed later on).

The issue at this point is whether it is possible to explain the "lack of purpose" approach towards studying based on the combined knowledge of "interaction" metrics, and whether there are distinguishable "interaction" metrics in students' interactions for this specific approach.

b) **Further discussion of model and predictors**

First of all, the recommended model (model 4) explains only 21% of variance. This amount of variance is medium (see regarding recommended thresholds in Appendix 3.11.1). However, this is only a bit more than 1/5 of the variance of the model explained. Therefore, it is reasonable to say that we do not seem to get the full picture of how students with low and high scores on the "lack of purpose" subscale interact with AM during the tutorial sessions.

In addition, despite the fact that the model is overall significant (see Table 4.5.3.2 above) and holds reasonably well all the required assumptions, it is observed that not all predictors are significant (see Table 7 in Appendix 4.5.2). The reason for which these predictors were kept is mainly because their inclusion would allow for a richer insight into students' interaction with regards to the specific subscale. So, the question is: does this "allowance" enrich at least the interpretation of the results?

Model 4 with its 4 predictors can draw, not quite a complete, but at least a distinguishing and reasonable picture, as it can help to identify students with high and low scores on the subscale. More specifically, students with a high score on the subscale are less likely to solve exercises on first attempt and more likely to solve them on third attempt or not at all, compared to those with a low score. Those with high scores are also more likely to revisit pages, compared to those with low scores. This finding is unexpected specifically with regards to the specific subscale but it fits fairly well with regards to compactness, as mentioned earlier. It is reasonable that students with high scores on the subscale tend to follow a more repetitive compact path when going through the learning material in AM, compared to those with low scores. To conclude: all these are reasonable findings, which can serve towards tutor intervention.

In comparison to Model 4, the leanest and meanest Model 6 with the elimination of the predictors *number of exercises solved on third try,* and *number of exercises solved on first try*, gives a less "enriching" picture with regards to the "lack of purpose" subscale (although there are still indications that it represents part of a surface approach mainly because of the remaining predictor *relative amount of revisits*).

So, Model 4 is a more "enriching" model, and the above interpretations on the findings seem to give a useful insight. However, as discussed earlier, the model still does not give the full picture (also the variance explained is not as much as in other surface subscales). It may be the case, therefore, that factors such as prior knowledge have influenced the relationship of this subscale with the predictors (an issue which is explored further in the next chapter).

4.6 The deep scale and students' "interaction" metrics

4.6.1 Deep Scale – Theoretical assumptions

As discussed in 3.3.2, the "deep" scale measures the extent to which there is (a) an intention to understand for oneself (McCune, 1998; Entwistle, 1997a) (b) an interest in the subject (c) an intention to relate concepts to each other and (d) built-up understanding based on detailed type of information in the learning material (Entwistle, 1997a).

During the learning process:

- For performance-related metrics, as discussed in 3.4.7, there is an overall expectation that students with high scores on the deep scale will do better when practising the exercises during the tutorial sessions, compared to those with low scores. More specifically, the higher students score on the deep scale, the more exercises they are likely to solve on the first attempt and the fewer exercises they are likely to solve on the second and third attempt, or fail to solve at all. Therefore, it is expected that there will be positive association between:
    - *Deep scale and number of exercises solved on the first try*

  While there will be negative associations between:
    - *Deep scale and number of exercises solved on the second try*
    - *Deep scale and number of exercises solved on the third try*
    - *Deep scale and number of exercises finished but not solved*
- Regarding search-related metrics, it is expected that students with high scores on the deep scale are likely to try to elicit more information around concepts, in

order to find deeper meaning and understanding in them, compared to those with low scores. This explorative behaviour is an element of those with a high score in the “interest in ideas” subscale (which is part of the deep scale). Martens et al. (2004) points out a link between intrinsic interest and exploration of concepts. This means that high scores on the deep scale may result in high values in search-related metrics indicated in 3.4.2. There will be positive associations between the deep scale and:

  - *number of times “search” option is clicked*
  - *number of submitted queries in search option*
  - *number of search results visited in search option*
  - *number of hyperlinks (concept links) visited in reading and exercise pages*

- In terms of visiting the AM “notes” feature, as discussed in 3.4.1, students with high scores in deep approach to studying tend to interact more vigorously with the learning material by making notes, compared to those with low scores; so there can be a positive association between the deep approach and the number of visits to the “notes” feature which shows an intention to create notes with regards to the learning material in AM. As there are two metrics related to the visits to the “notes” feature (see sections 3.4.1 and 3.4.8), the potential positive associations are between the deep scale and:
  - *number of times “notes” link is clicked*
  - *average number of times a “notes” link is clicked per page*
- Regarding temporal metrics, as discussed in 3.4.4, it is expected that due to an effort to seek and research further the meaning of a mathematical concept or process it, students with high scores on the deep scale may dedicate more time to their tasks in class compared to those with low scores. So, high scores on deep scale may result in high values in temporal metrics. Thus, there will be positive associations between:
  - *Deep scale and average view time on exercise pages*
  - *Deep scale and average view time on reading pages*
  - *Deep scale and maximum view time on a reading page*
  - *Deep scale and maximum view on an exercise page*

4.6.2 Deep Scale –Results on Correlations

Following the proposed methodology in 3.11, we run correlational analysis to identify predictors for the model. There is one statistically significant correlation between deep scale and *maximum view time on an exercise page* with r_s=0.234, which is expected.

Regarding other metrics mentioned in section 4.6.1, there are no statistically significant relationships. However, the importance of these metrics for the deep scale is discussed further in the following section with regards to the regression models.

4.6.3 Deep models: Development, Analysis and Discussion

In this section, there is a brief discussion on the development of regression models for the deep scale and analysis and discussion on the results for the selected model version.

4.6.3.1 Deep Scale – Initial selection of predictors

Table 4.6.3.1 shows briefly that the reasons behind the initial selection of most predictors are their theoretical connections to the scale, and that they may enrich the discussion by allowing useful comparisons to surface scales and/or by giving a more complete picture as to how students deal with their exercises during their tutorial sessions according to the specific approach to studying (see further information in table 1 of Appendix 4.6.1).

Selected Predictor	Reason for selection
Number of exercises solved on first try	Theoretical connections and enriching further the discussion
Number of exercises solved on second try	
Number of exercises solved on third try	
Number of exercises finished but not solved	
Number of hyperlinks (concepts links) visited on exercise and reading pages	
Average number of times a "notes" link is clicked per page	
Maximum view time on an exercise page	Statistical (see 4.6.2)

Table 4.6.3.**1**. Selected predictors for first version of model.

As indicated in the methodology, the sample size allows for the inclusion of up to 8 predictors and the intention is to take full advantage of this upper limit. So, the potential of including an 8th predictor is examined amongst the rest of the predictors mentioned in 4.6.1. However, certain of these predictors have not been considered because their variation is close to 0 (i.e. the metrics related to the use of the “search” option). In other cases, there are no strong theoretical indications as to which of these predictors will be the most “enriching” for the model, and also when they are tried in pre-models they do not contribute to their variance (see further information in Table 2 of Appendix 4.6.1).

As a result, at this stage, Model 1 is the first version for the deep scale with the seven predictors shown in Table 4.6.3.1.

4.6.3.2 Deep Scale – development of model

For Model 1 the variance explained R^2 is 13.3% and the Adjusted R^2 is 7.7%. The process of improving both these measures of variance and the overall significance starts by excluding specific outliers. This results in Model 2j with R^2 18.5% and Adjusted R^2 12.9% (see Appendix 4.6.2 for detailed discussion).

4.6.3.3 Deep Scale – selection of model

After excluding the outliers, there is gradual exclusion of predictors, as shown on Table 4.6.3.2 below. From Model 2j to Model 4, Adjusted R^2 increases while R^2 gradually decreases slightly, making their difference smaller.

After Model 4, the regression is re-run two more times. In Model 5, there is exclusion of the predictor *number of exercises solved on first try*, since in Model 4 this predictor has the least importance (beta=0.122) and highest insignificance ($p>0.05$). Finally, in model 6 (see Table 6 in Appendix 4.6.2), there is exclusion of predictor *average number of times a “notes” link clicked per page**,*** since in Model 5 this predictor has the least importance (beta=0.143) and highest significance ($p>0.05$). Model 6 is the leanest and meanest model where all three remaining predictors are statistically significant.

	R^2	Adj. R^2	Sig.
Model 1 (all initially selected predictors)	13.3%	7.7%	0.029
Model 2a (exclusion of case 112)	13.4%	7.7%	0.029
Model 2b (exclusion of case 112 and 123) [Rejected]	11.3%	5.3%	0.076
Model 2c (exclusion of case 112 and 116)	14.6%	9%	0.017
Model 2d (exclusion of case 112, 116, and 36) [Rejected]	14.4%	8.6%	0.021
Model 2e (exclusion of case 112, 116, and 81) [Rejected]	14.3%	8.5%	0.022
Model 2f (exclusion of case 112, 116, and 111)	14.9%	9.2%	0.016
Model 2g (exclusion of case 112, 116, 111, and 85)	16.5%	10.9%	0.008
Model 2h (exclusion of case 112, 116, 111, 85, and 76)	17.7%	12.1%	0.005
Model 2i (exclusion of case 112, 116, 111, 85, 76, and 60) [Rejected]	16.2%	10.4%	0.011
Model 2j (exclusion of case 112, 116, 111, 85, 76, and 105)	18.5%	12.9%	0.004
Model 3 (exclusion of case 112, 116, 111, 85, 76, and 105, and *number of exercises solved on third try*)	18.5%	13.7%	0.002
Model 4 (exclusion of case 112, 116, 111, 85, 76, and 105, and *number of exercises solved on third try*, and *number of hyperlinks (concepts links) visited in exercise and content pages)*	18.3%	14.4%	0.001
Model 5 (exclusion of case 112, 116, 111, 85, 76, and 105, and *number of exercises solved on third try*, *number of hyperlinks (concepts links) visited in exercise and content pages,* and *number of exercises solved on first try)*	17.2%	14%	0.001
Model 6-Leanest and Meanest (exclusion of case 112, 116, 111, 85, 76, and 105, and *number of exercises solved on third try*, *number of hyperlinks (concepts links) visited in exercise and content pages, number of exercises solved on first try,* and *average number of times a notes clicked per page)*	15.3%	12.9%	0.001

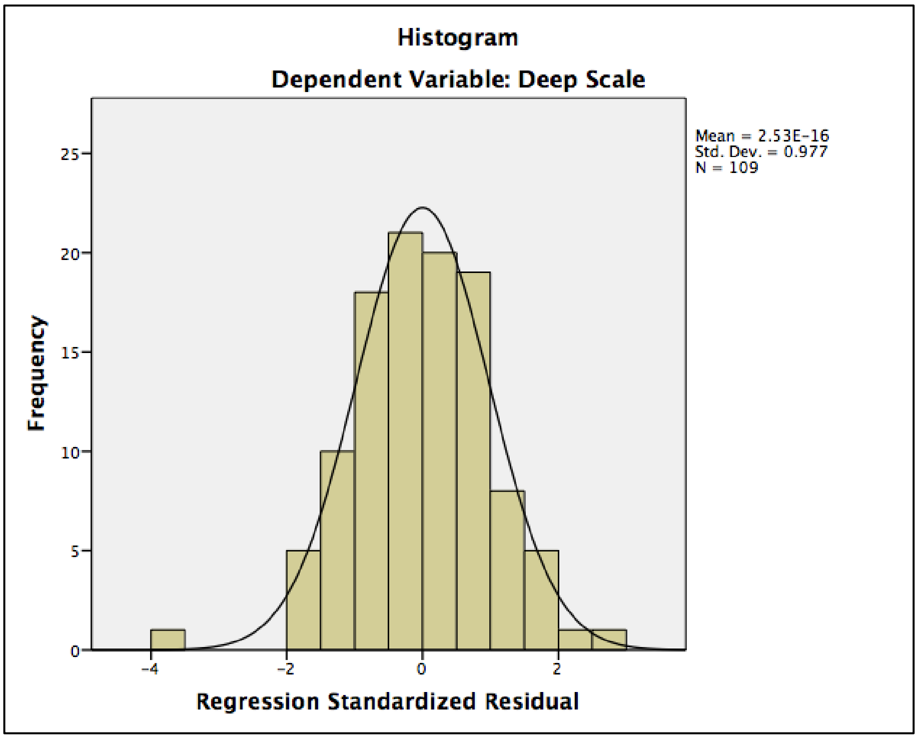
Table 4.6.3.**2**. Summary of measures of variance and significance for accepted and rejected models

In Table 4.6.3.**2**, it is observed that the overall significance holds for all versions of the model ($p<0.05$). When comparing to Model 4, both R^2 and adjusted R^2 gradually decrease in models 5 and 6 from 18.3% to 15.3%, and from 14.4% to 12.9% respectively. Therefore, according to our strategy, Model 4 is suggested as the best solution, as it combines simultaneously the highest possible R^2 and Adjusted R^2. Finally, Model 4 has 5 predictors, instead of the initial 7 predictors of Model 1, with sample size of 109 (after excluding the 6 outliers), which is within the thresholds stated in the strategy, in section 3.11.

4.6.3.4 Deep Scale – Model 4 – Generalisation

The assumptions regarding: normal distribution of residuals, homoscedasticity of standardised residuals against predicted ones, and the normality of residuals hold well.

*Figure 4.6.3.**1**. Histogram of standardised residuals for the final model*



Regarding the assumption about the normal distribution of residuals, as shown in Figure 4.6.3.1, that residuals fit very closely to a normal distribution.

*Figure 4.6.3.**2**. Plot of standardised residuals*

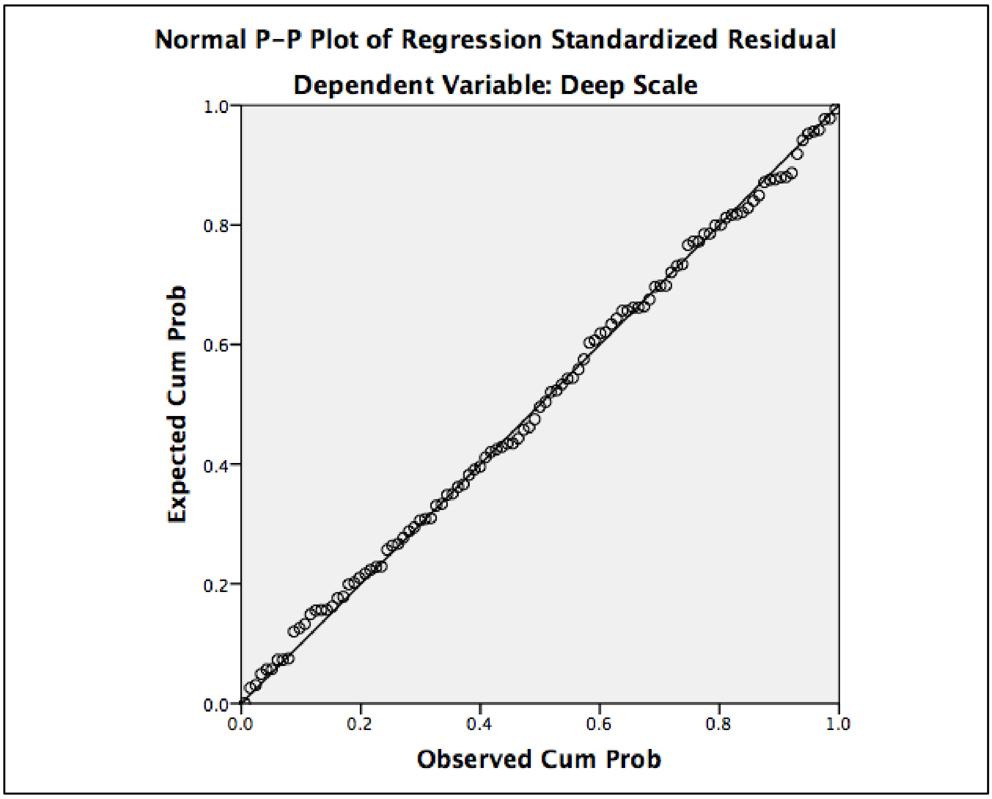
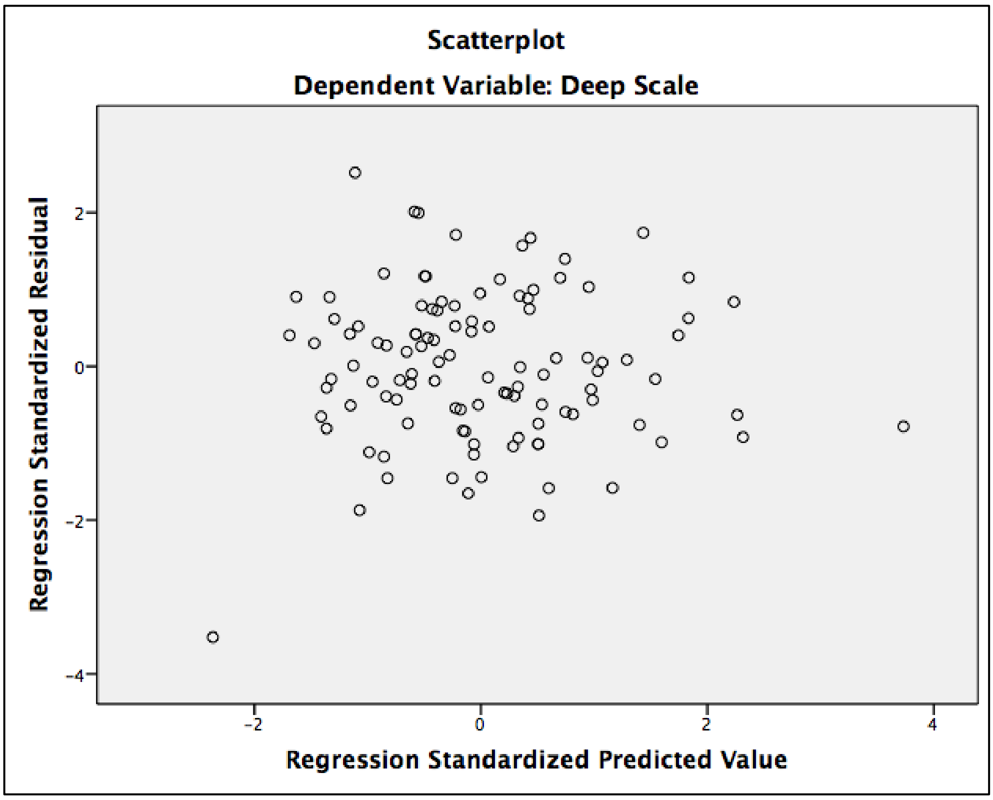


Figure 4.6.3.2 shows that the normality assumption holds since the points lie quite closely on the straight line.

*Figure 4.6.3.**3**. Plot of the standardised residuals against the predicted ones for the model*



The assumption about the residuals is whether the variance of the residuals is constant, in other words whether there is homoscedasticity. Figure 4.6.3.3 shows that the scatter plot is reasonably random and most residuals are homoscedastic (there are only a few cases which are heteroscedastic).

To conclude, the assumptions are not violated and therefore we can generalise the findings beyond the sample.

4.6.3.5 Deep Scale – Equation

By selecting Model 4, it can be concluded that for both weeks, the deep scale is expressed through the predictors: number of exercises solved on first try, number of exercises solved on second try, number of exercises finished but not solved, average number of times a "notes" link clicked per page, and maximum view time on exercise pages. So, the equation which is formed is as follows:

DeepScalei = b_0 + b_1(Number of exercises solved on first try)$_i$ – b_2(Number of exercises solved on second try)$_I$ + b_3(Number of exercises finished but not solved)$_i$ +b_4(Maximum view time on exercise pages)$_i$ +b_5(Average number of times a "notes" link clicked per page)$_i$ (1)

If we replace the b values (see Table 6 in Appendix 4.6.2) in the above equation, then we obtain the equation of the fitted regression model:

DeepScale$_i$ = 56.977 + 0.036(Number of exercises solved on first try)$_i$ – 0.637(Number of exercises solved on second try)$_i$ + 0.488(Number of exercises finished but not solved)i +0.003(Maximum view time on exercise pages)$_i$ +10.619(Average number of times a "notes" link clicked per page)$_i$ (2)

4.6.3.6 Deep Scale – Interpretation of model parameters

a) **Direction of relationship between predictors and outcome based on b values**

As the deep scale score increases, there is increment on the following predictors:

- *number of exercises solved on first try*
- *number of exercises finished but not solved*
- *maximum view time on exercise pages*
- *average number of times a "notes" link clicked per page*

As the deep scale score increases, there is a decrease in the following predictors:

- *number of exercise solved on second try*

It is worth mentioning that the direction of relationships is as expected in the initial assumptions, except from the *number of exercises finished but not solved*.

The issue at this point is whether it is possible to explain the "deep" approach towards studying based on the combined knowledge of "interaction" metrics, and whether there are distinguishable "interaction" metrics in students' interactions for this specific approach.

b) **Further discussion of model and predictors**

First of all, the recommended model (Model 4) explains 18.3% amount of variance. This amount of variance is medium as expected (see recommended thresholds in Appendix 3.11.1). However, this is close to only 1/5 of the variance of the model explained. Furthermore, to make another comparison, the variance explained in the recommended model (Model 4) for the deep scale is less than half of the variance explained for surface scale. Therefore, it is reasonable to say that we do not seem to get the full picture of how students with low and high scores on the deep scale interact with AM during the tutorial sessions.

In addition, despite the fact that the model is overall significant (see Table 4.6.3.2 above) and holds well according to all the required assumptions, it is observed that not all predictors in Model 4 are statistically significant (see Table 6 in Appendix 4.6.2). The reason these predictors were kept is mainly because their inclusion would allow for a richer insight into students' interaction with regards to the specific scale. So, the question is: does this "allowance" enrich the interpretation of the findings?

First of all, it would be difficult to advise - a tutor for example- how to identify a deep approach to studying based on the *number of exercises solved on second try*, which is the most important predictor and the one which "survives" also in the leanest and meanest version Model 6 (see Table 9 in Appendix 4.6.3). This is a predictor which can be interpreted better in combination with other predictors of this type. More specifically, we can say that the more a student is likely to score high on the deep scale, the less likely he or she is to solve an exercise on second try. This would be fine, as it is according to our initial assumption; however, on its own does not say much. It makes more sense if we look at it in combination with the predictor the *number of exercises solved on first try,* and *number of exercises solved but not finished* in Model 4. The positive b and beta values show that students with a high deep approach towards studying are likely to solve exercises on first try. However, the positive b and beta values of *number of exercise solved but not finished* indicate that its relationship to the deep scale is not as initially expected. So, students with a high deep approach towards studying can be distinguished mainly based on the predictor *number of exercises solved on first try*.

In addition, the b and beta values of the remaining predictor of Model 4, which is the *maximum view time spent on exercise page,* indicate the expected positive relationship to the deep scale. It is assumed that students with high scores on the scale who tend to seek meaning for achieving personal understanding are more likely to dedicate time to parts of the learning material in order to achieve personal understanding, compared to those with low scores. However, it is possible that they may experience difficulties with a specific group of exercises, as there is also an aforementioned tendency not to solve exercises at all, (similarly to those with high scores on the surface scale). So, *maximum view time on an exercise page* does not necessarily serve towards distinguishing between a high surface score and a high deep score.

Finally, other complementary manifestations of the deep approach concern the accessing of the AM "notes" feature, which can show an intention for those with high scores on the scale to make more notes, compared to those with low scores on the scale (as discussed in 4.6.1).

In comparison to the suggested Model 4, the leanest and meanest Model 6 (see Table 9 in Appendix 4.6.3) consists of three predictors: *number of exercises solved on second try, number of exercises finished but not solved,* and *maximum exercises spent on exercise page*. As mentioned earlier, these three predictors alone cannot really give a clear and enlightening picture in relation to students' interaction and the deep approach. Therefore, Model 4 is considered more enriching in terms of the interpretation, however, since the model explains only about 1/5 of the variance of the deep scale, the next logical step would be to examine if prior knowledge can explain further the model representing the deep approach (an issue which is explored further in the next chapter).

4.7 The deep subscale "interest in ideas" and students' "interaction" metrics

4.7.1 Interest in ideas – theoretical assumptions

As discussed in 3.4.2, the "interest in ideas" subscale measures the extent to which there is intrinsic interest in the content of a course students are taking. In the general educational context, intrinsic interest in the subject matter has been linked to explorative behaviour, and the freedom to choose beyond the given syllabus.

During the learning process:

- In terms of search-related metrics (use of hyperlinks and search option), there are conceptual links, which can be made as discussed in 3.4.2. Entwistle (1981) describes students with an "interest in ideas" approach as having active interest in the course content; hence those with a high score on the subscale may use more features which can satisfy this active interest in the course content (i.e. by exploring it further) compared to those with low scores. Furthermore, in the context of students' interaction in interactive learning environments, Martens et al. (2004) have found a statistically significant, positive and medium association between the intrinsic interest scores of students and the number of pages with explorative content they visited.
- The search option in AM can allow exploration of a mathematical concept in relation to other concepts, mathematical examples and procedures. The search feature gives also learners the freedom to explore a mathematical concept beyond the AM content by allowing them to find further information on other websites. High "interest in ideas" therefore may lead students to explore mathematical concepts by submitting a high number of query submissions and visit a high number of the search results produced. It is expected that the higher the students score in the "interest in ideas" deep subscale, the more clicks there will be on the search option, the more queries they will submit and the greater number of search results they will click on (in order to explore a concept further). There will be positive associations between the "interest in ideas" subscale and:
    - *number of times "search" option is clicked*
    - *number of submitted queries in search option*
    - *number of search results visited in search option*

- Entwistle (1981) finds that there are certain academic topics which can attract the attention of students with a high score on the "interest in ideas" subscale and prompt them to follow them up. In the reading and exercise pages of AM there are contextual hyperlinks, which allow students to visit new concepts or revisit previous ones. Therefore, it is possible that contextual hyperlinks may entice more students with a high score on the subscale to follow up those pieces of information in the content they find interesting, compared to those with low scores. It is expected that there will be positive association between:
    - *"Interest in ideas" and number of hyperlinks (concept links) visited in reading and exercise pages.*
- In terms of temporal metrics, as discussed in 3.4.4, it is possible that students with a high score in the "interest" in ideas subscale are likely to spend more time on an exercise or reading page, than the students with low scores. Thus, positive associations are expected between "interest in ideas" subscale:
    - *average view time on exercise pages*
    - *average view time on reading pages*
    - *maximum view time on a reading page*
    - *maximum view time on an exercise page*
- In terms of performance-related metrics, in a more general context, there is empirical evidence that associates academic performance and the subscale of "interest in ideas". Based on the empirical findings by Entwistle and Ramsden (1983), it is found in a sample of 865 science first-year undergraduate students that there is a statistically significant, positive and correlation (r=0.24) between the factor of "interest in ideas" and performance. In a more recent study, Tait and Entwistle (1996) find in a sample of 649 first-year undergraduate students that there is a statistically significant positive low correlation (r=0.12) between the "interest in ideas" subscale and performance. On the other hand, in the context of interactive learning environments, empirical evidence in the study of Martens et al. (2004) indicate that the scores of intrinsic motivation do not correlate with students' performance. A possible explanation for these varied results between "interest in ideas" and performance can be found in the wider context of educational psychology. Entwistle and Ramsden (1983) refer to a study conducted by Fransson (1978) where the results reveal that the factor of anxiety can interact with intrinsic motivation. Fransson (1978) finds that the factor of anxiety relates negatively to performance among students with a high

level of intrinsic motivation, but not among those students with a low level of intrinsic motivation.

- There are mixed views therefore with regards to this specific relationship. However, as it is a deep subscale, which expresses intrinsic interest, it is possible that such students may perform well. Therefore, it is expected that there will be positive association between:
    - "Interest in ideas" subscale and number of exercises solved on the first try
- While there will be negative associations between:
    - "Interest in ideas" subscale and number of exercises solved on the second try
    - "Interest in ideas" subscale and number of exercises solved on the third try
    - "Interest in ideas" subscale and number of exercises finished but not solved

4.7.2 Interest in ideas - Results on Correlations

Following the proposed methodology in 3.11 we run correlational analysis to identify predictors for the model. There are statistically significant correlations, as expected, between the "interest in ideas" subscale and:

- Maximum view time on an exercise page with r_s=0.188
- Average view time on exercise pages with r_s=0.185

For other metrics mentioned in section 4.7.1 there are no statistically significant relationships. However, the importance of these metrics for the "interest in ideas" subscale is discussed further in the following section with regards to the regression models.

4.7.3 Interest in ideas models: Development, Analysis and Discussion

In this section, there is a brief discussion on the development of regression models for the "interest in ideas" subscale and analysis and discussion on the results for the selected model version.

4.7.3.1 Interest in ideas –Initial selection of predictors

Table 4.7.3.1 shows briefly that the reasons behind the initial selection of most predictors are their theoretical connections to the subscale, and that they may enrich the discussion by allowing useful comparisons to surface scales and/or by giving a more complete picture of how students deal with their exercises during their tutorial sessions according to the specific approach to studying (see further information in Table 1 of Appendix 4.7.1).

Selected Predictor	Reason for selection
Number of exercises solved on first try	Theoretical connections and enriching further the discussion
Number of exercises solved on second try	
Number of exercises solved on third try	
Number of exercises finished but not solved	
Number of hyperlinks (concepts links) visited on exercise and reading pages	
Maximum view time on an exercise page	Statistical (see 4.7.2)
Average view time on exercise pages	

Table 4.7.3.**1**. Selected predictors for first version of model.

As indicated in the methodology, the sample size allows for the inclusion of up to 8 predictors and the intention is to take full advantage of this upper limit. After examining the potential of including an 8th predictor amongst the rest of the predictors mentioned in 4.7.1, it was decided to include just the seven aforementioned predictors (see further justification in table 2 of Appendix 4.7.1).

4.7.3.2 Interest in ideas – development of model

As previously, the process of improving measures of variance and the overall significance starts by excluding specific outliers. This results in increasing the variance: from Model 1 with R^2 9.9% and Adjusted R^2 4% to Model 2f with R^2 15.8% and Adjusted R^2 10% (see Table 4 in Appendix 4.7.2).

4.7.3.3 Interest in ideas – selection of model

After excluding the outliers, there is gradual exclusion of predictors, as shown in Table 4.7.3.2 below.

	R^2	Adj. R^2	Sig.	Comments
Model 3 **(**exclusion of cases 76, 116, 111, 4, 46, and *Number of exercises solved on first try*)	15.8%	10.9%	0.006	Note how Adjusted R^2 increases while R^2 gradually slightly decreases, making their difference smaller.
Model 4 (exclusion of cases 76, 116, 111, 4, 46, and *Number of exercises solved on first try,* and *number of hyperlinks (concept links) in reading and exercise pages*)	15.7%	11.6%	0.003	
Model 5 (exclusion of cases 76, 116, 111, 4, 46, and *Number of exercises solved on first try, number of hyperlinks (concept links) in reading and exercise pages,* and *average view time on exercise pages*)	14.7%	11.4%	0.002	By running the regression model two more times, we obtain the leanest and meanest model. Note how both R^2 and Adjusted R^2 decrease.
Model 6 –Leanest and Meanest (exclusion of cases 76, 116, 111, 4, 46, and *Number of exercises solved on first try, number of hyperlinks (concept links) in reading and exercise pages, average view time on exercise pages, and number of exercises solved on second try*)	12.2%	9.7%	0.003	Model 6 has all predictors statistically significant (see Appendix 4.7.4) Note also that all models are overall statistically significant ($p<0.05$)

Table 4.7.3.**2**. Summary of measures of variance and significance

In Table 4.7.3.2, Model 4 is suggested as the best solution, as it combines simultaneously the highest possible R^2 and Adjusted R^2. Finally, Model 4 has 5 predictors, instead of the initial 7 predictors of Model 1, with sample size of 109 (after excluding the 6 outliers), which is within the thresholds stated in the strategy, in section 3.11.

4.7.3.4 Interest in ideas – Model 4 – Generalisation

For the selected Model 4 of the subscale, the assumptions are not violated and therefore it is possible to generalise the finding beyond the sample (see the assumptions in Appendix 4.7.3).

4.7.3.5 Interest in ideas – Equation

By selecting Model 4, it can be concluded that for both weeks, the "interest in ideas" is expressed through the predictors: average view time on exercise pages, maximum view time on an exercise page, number of exercises solved on second try, number of exercises solved on third try, and number of exercises finished but not solved. So, the equation, which is formed is as follows:

InterestInIdeasSubcale$_i$ ***=*** b_0 ***-*** b_1***(Number of exercises solved on second try)***$_i$ ***-*** b_2***(Number of exercises solved on third try)***$_i$ ***+*** b_3***(Number of exercises finished but not solved)***$_i$ ***+*** b_4***(Maximum view time on exercise page)***$_{i\,+}$ ***+*** b_5***(Average view time on exercise pages)***$_i$ ***(1)***

If we replace the b values, found in Table 7 in Appendix 4.7.2, in the above equation, then we obtain the equation of the fitted regression model:

InterestInIdeasSubcale$_i$ ***= 13.112 – 0.120(Number of exercises solved on Second Try)***$_i$ ***- 0.262(Number of exercises solved on Third Try)***$_i$ ***+ 0.254(Number of Exercises Finished but not Solved)***$_i$ ***+ 0.001(Maximum View Time on Exercise Page)i + 0.001(Average View Time on Exercise Pages)***$_i$ ***(2)***

4.7.3.6 Interest in ideas – Interpretation of model parameters

a) **Direction of relationship between predictors and outcome based on b values**

As the "interest in ideas" subscale score increases, there is increment on the following predictors:

- *number of exercises finished but not solved*
- *maximum view time on an exercise page*
- *average view time on exercise pages*

As the “interest in ideas” subscale score increases, there is a decrease in the following predictor:

- *number of exercises solved on third try*
- *number of exercises solved on second try*

It is worth mentioning that the direction of relationships with regards to the temporal metrics are as expected in the initial assumptions. However, the direction of the relationship with regards to the number of tries when solving exercises (i.e. number of exercises finished but not solved) is not as expected in the initial assumptions. It seems that having an intrinsic interest in a subject does not necessarily guarantee good performance.

The issue here is whether it is possible to explain the “interest in ideas” approach towards studying based on the combined knowledge of “interaction” metrics, and whether there are distinguishable “interaction” metrics in students’ interactions for this specific approach.

b) **Further discussion of model and predictors**

First of all, the recommended model (Model 4) explains 15.7% amount of variance. This amount of variance explained is medium as expected (see regarding recommended thresholds in Appendix 3.11.1). However, this is close to only 1/6 of the variance of the model explained. Therefore, it is reasonable to say that we do not seem to get the full picture of how students with low and high scores on the interest in ideas subscale interact with AM during the tutorial sessions.

In addition, despite the fact that the model is overall significant (see Table 4.7.3.2 above) and holds reasonably well all the required assumptions, it is observed that not all predictors in Model 4 are statistically significant (see Table 7 in Appendix 4.7.3). The reason for which these predictors were kept is mainly because their inclusion would allow for a richer insight into students’ interaction with regards to the specific subscale. So, the question is: does this “allowance” enrich at least the interpretation of the results?

In the suggested Model 4, it is observed that students with high scores on intrinsic interest are less likely to solve the exercises on the second or third try, but they are more likely to not to solve the exercises at all, compared to those with low scores. These findings, although they do not help in distinguishing a high "interest in ideas" studying approach, are not totally unexpected. As discussed in 4.7.1, although the general expectation was that those with high scores on the subscale would do overall well compared to those with low scores, there were indications in the literature that factors such as anxiety may influence this relationship.

Furthermore, it is possible that the average time spent on exercise pages, as well as the maximum time spent on an exercise page, increases as the intrinsic interest on mathematics increases, because of the difficulties students with high scores seem to be experiencing while solving exercises (i.e. it can be also interpreted as "getting stuck" on specific exercise pages).

With regards to the leanest and meanest Model 6 (where all 3 predictors are statistically significant), students with high intrinsic interest are not more likely to solve exercises on the third try and they are more likely not to solve the exercises, compared to those with low scores. Furthermore, the maximum time spent on an exercise page increases as the intrinsic interest on mathematics increases, which again can be due to the aforementioned performance when solving exercises.

So, while Model 4 gives a slightly more complete image compared to Model 6, they both point to the same issue: they do not manage necessarily to serve towards distinguishing those with high scores on the subscale through their interactions in AM during the tutorial sessions, since their interactions seem to resemble more those with high scores on surface approaches.

Finally, other search-related metrics such as *number of times search option is clicked* and *number of submitted queries in the search option*, which could have contributed in the variance explained by the model, were not included because of lack of variation. It is possible, therefore, that these missing "search-related" interactions can be also responsible for the rest of the variance which is unexplained in the model (also they could give a more "distinguishing" quality to the "interest in ideas" model).

Finally, a factor which may explain the aforementioned unexpected relationships or lack of variance in the model is the students' prior knowledge; an issue which is discussed in detail in the next chapter.

4.8 The deep subscale "seeking meaning" and students' "interaction" metrics

4.8.1 Seeking meaning - theoretical assumptions

As discussed in 3.3.2, the "seeking meaning" subscale shows the extent to which there is an intention for personal understanding, seeking an individual interpretation and imposing their own structure (McCune, 1998; Entwistle, 1997a; Entwistle, 1998). Furthermore, students with an intention to seek meaning are also likely to go through the material fairly slowly (Entwistle and Ramsden, 1998).

During the learning process:

- In terms of search-related metrics, there can be associations with the "seeking meaning" subscale. Students with high scores on the "seeking meaning" subscale may access and use more features in an interactive learning environment which aid towards exploring further concepts and procedures and forming connections between them, compared to those with low scores on the subscale. So, the use of these features may help those with a high score on the subscale towards processing the learning material in a way that aids individual interpretation. This means that there can be positive associations between the subscale and the search-related metrics (as indicated in 3.4.2). More specifically, there can be positive associations between the "seeking meaning" subscale and:
    - *number of hyperlinks (concept links) visited in reading and exercise pages*
    - *number of times "search" option is clicked*
    - *number of submitted queries in search option*
    - *number of search results visited in search option*
- In terms of the path metric *stratum*, students with high scores on "seeking meaning" tend more to impose their own structure on the learning material (Entwistle, 1998), compared to those with low scores. Therefore, those with high scores on the subscale are not likely to go about their activities in an orderly manner compared to those with low scores. This means that high scores on the "seeking meaning" subscale are likely to result in low values on *stratum*. So, there can be a negative association between:
    - *"Seeking meaning" subscale and stratum*

- In terms of temporal metrics, students with high scores on “seeking meaning” tend to go through the material fairly slowly (Entwistle and Ramsden, 1998), so there can be a positive association between the subscale and the temporal metrics of average view time and maximum view time on exercise and content (reading) pages. More specifically, there can be positive associations between “seeking meaning” subscale:
    - *average view time on exercise pages*
    - *average view time on reading pages*
    - *maximum view time on a reading page*
    - *maximum view time on an exercise page*
- In terms of performance-related metrics, as it is a deep subscale which expresses an intention for personal understanding, it can be assumed that students with high scores on the subscale can do well when practising on their exercises during the tutorials. This is also reinforced by literature in mathematics education which supports that students are not likely to do well if they carry out mathematical procedures without really seeking the meaning of the concepts involved (Liston and O'Donoghue, 2009). So, students with a high score on the subscale who tend to seek the meaning of mathematical concepts are more likely to do well, compared to those with low scores. Therefore, a positive association is expected between:
    - “Seeking meaning” subscale and number of exercises solved on the first try
- While there will be negative associations between:
    - “Seeking meaning” subscale and number of exercises solved on the second try
    - “Seeking meaning” subscale and number of exercises solved on the third try
    - “Seeking meaning” subscale and number of exercises finished but not solved

4.8.2 Seeking meaning - Results on Correlations

Following the proposed methodology in 3.11 we run correlational analysis to identify predictors for the model. There is one statistically significant correlation, as expected, between the “seeking meaning” subscale and:

- *Maximum view time on an exercise page with r_s=0.204*

There was no expectation of statistically significant correlations between the "seeking meaning" subscale and:

- *Number of pages visited using TOC with r_s=-0.183*

Regarding other metrics, mentioned in section 4.8.1, there are no statistically significant relationships. However, the importance of these metrics for the "seeking meaning" subscale is discussed further in the following section with regards to the regression models.

4.8.3 Seeking meaning models: Development, Analysis and Discussion

In this section, there is a brief discussion on the development of regression models for the "seeking meaning" subscale and analysis and discussion on the results for the selected model version.

4.8.3.1 Seeking meaning - Initial selection of predictors

Table 4.8.3.1 shows briefly that the main reasons behind the initial selection of most predictors are their theoretical connections to the subscale, and that they may enrich the discussion by allowing useful comparisons to surface scales and/or by giving a more complete picture of how students deal with their exercises during their tutorial sessions according to the specific approach to studying (see further information in table 1 of Appendix 4.8.1).

Selected Predictor	**Reason for selection**
Number of exercises solved on first try	Theoretical connections and enriching further the discussion
Number of exercises solved on second try	
Number of exercises solved on third try	
Number of exercises finished but not solved	
Number of hyperlinks (concepts links) visited on exercise and reading pages	
Stratum	
Maximum view time on an exercise page	Statistical (see 4.8.2)
Number of pages visited using the TOC	

Table 4.8.3.**1**. Selected predictors for first version of model.

As a result, at this stage, Model 1 is the first version for the "seeking meaning" subscale with all the selected 8 aforementioned predictors. According to the strategy indicated in the methodology, this is the maximum number of predictors (see section 3.11).

4.8.3.2 Seeking meaning – development of model

As previously, the process of improving measures of variance and the overall significance starts by excluding specific outliers. This results in increasing the variance: from Model 1 with R^2 13.5% and Adjusted R^2 7% to Model 2g with R^2 20.8% and Adjusted R^2 14.5% (see Table 4 in Appendix 4.8.2).

4.8.3.3 Seeking meaning – selection of model

After excluding the outliers, there is gradual exclusion of predictors, as shown in Table 4.8.3.2 below.

	R^2	Adj. R^2	Sig.	Comments
Model 3 (exclusion of case 36, 123, 61, 85, and 103, and *Stratum*)	20.8%	15.4%	0.001	Note how Adjusted R^2 increases while R^2 gradually slightly decreases, making their difference smaller.
Model 4 (exclusion of case 36, 123, 61, 85, and 103, and *stratum, number of exercises finished but not solved*)	20.1%	15.5%	0.001	
Model 5 (exclusion of case 36, 123, 61, 85, and 103, and *stratum, number of exercises finished but not solved*, *number of exercises solved on second try*)	17.9%	14%	0.001	By running the regression model three more times, we obtain the leanest and meanest model. Note how both R^2 and Adjusted R^2 decrease.
Model 6 (exclusion of case 36, 123, 61, 85, and 103, and *stratum, number of exercises finished but not solved*, *number of exercises solved on second try, number of exercises solved on third try,*)	17%	13.8%	0.001	
Model 7-Leanest and Meanest (exclusion of case 36, 123, 61, 85, and 103, and *stratum, number of exercises finished but not solved*, *number of exercises solved on second try, number of exercises solved on third try, number of hyperlinks (concept links) on reading and exercise pages*)	15.4%	13%	0.000	Model 7 has all predictors statistically significant (see Appendix 4.8.4) Note also that all models are overall statistically significant ($p<0.05$)

Table 4.8.3.**2**. Summary of measures of variance and significance

In Table 4.8.3.2, it is observed that Model 4 is the best solution, as it combines simultaneously the highest possible R^2 and Adjusted R^2. Finally, Model 4 has 5 predictors with sample size of 110 (after excluding the 5 outliers), which is within the thresholds stated in the strategy, in section 3.11.

4.8.3.4 Seeking meaning – Model 4 – Generalisation

For the selected Model 4 of the subscale, the assumptions are not violated and therefore it is possible to generalise the finding beyond the sample (see the assumptions in Appendix 4.8.3).

4.8.3.5 Seeking meaning – Equation

By selecting Model 4, it can be concluded that for both weeks, "seeking meaning" is expressed through the predictors: maximum view time on an exercise page, number of exercises solved on first try, number of exercises solved on second try, number of exercises on third try, number of pages visited using TOC and number of hyperlinks (concept links) visited in reading and exercise pages. So, the equation, which is formed is as follows:

$SeekingMeaningSubcale_i$ = b_0 + b_1(Maximum view time on exercise page)$_i$ + b_2(Number of exercises solved on first try)$_i$ - b_3(Number of exercises solved on second try)$_i$ + b_4(Number of exercises solved on third try)$_i$ + b_5(Number of hyperlinks (concept links) visited in reading and exercise pages)$_i$ – b_6(Number of pages visited using TOC)$_i$ (1)

If we replace the b values, found in Table 7 in Appendix 4.8.2, in the above equation, then we obtain the equation of the fitted regression model:

$SeekingMeaningSubcale_i$ = 14.198 + 0.001(Maximum view time on exercise page)$_i$ + 0.021(Number of exercises solved on first try)$_i$ - 0.111(Number of exercises solved on second try)$_i$ + 0.226(Number of exercises solved on third try)$_i$ - 0.197(Number of hyperlinks (concept links) visited in reading and exercise pages)$_i$ – 0.028(Number of pages visited using TOC)$_i$ (2)

4.8.3.6 Seeking meaning – interpretation of model parameters

a) **Direction of relationship between predictors and outcome based on b values**

As the "seeking meaning" subscale score increases, there is increment on the following predictors:

- *number of exercises solved on first try*

- *number of exercises solved on third try*
- *maximum view time on an exercise page*

As the “seeking meaning” subscale score increases, there is a decrease in the following predictors:

- *number of exercises solved on second try*
- *number of hyperlinks (concept links) visited in reading and exercise pages*
- *number of pages visited using TOC*

It is worth mentioning that the direction of relationships is as expected in the initial assumptions, except from the *number of exercise solved on third try* and *number of hyperlinks (concept links) visited in reading and exercise pages*.

The issue at this point is whether it is possible to explain the “seeking meaning” approach towards studying based on the combined knowledge of “interaction” metrics, and whether there are distinguishable “interaction” metrics in students’ interactions for this specific approach.

b) **Further discussion of model and predictors**

First of all, the recommended model (model 4) explains 20.1% of variance. This amount of variance is medium as expected (see the recommended thresholds in Appendix 3.11.1). However, this is only 1/5 of the variance of the model explained. Therefore, it is reasonable to say that we do not seem to get the full picture of how students with low and high scores on the “seeking meaning” subscale interact with AM during the tutorial sessions.

In addition, despite the fact that the model is overall significant (see Table 4.8.3.2 above) and holds reasonably well all the required assumptions, it is observed that not all predictors in Model 4 are statistically significant (see Table 7 in Appendix 4.8.2). The reason for which these predictors were kept is that their inclusion would give us a richer insight into students’ interaction in the specific subscale. So, the question is: does this “allowance” enrich at least the interpretation of the results?

According to the suggested Model 4, students with a high score on the subscale tend to solve more exercises on the first try, but they tend more to solve exercises on the third try, compared to those with low scores. In terms of the temporal aspect of interactions, the *maximum view time on an exercise page* increases as the score on the subscale increases as well. It is assumed that students with high scores on the subscale who tend to seek meaning for achieving personal understanding are more likely to dedicate time to parts of the learning material in order to achieve personal understanding, compared to those with low scores. However, it is possible that they may experience difficulties with a specific group of exercises in a specific page, as there is also an aforementioned tendency to solve exercises on the third try, like in a "high" surface approach[32].

Other complementary manifestations according to Model 4 of the "seeking meaning" subscale concern the use of AM features. With regards to the use of TOC, it seems that those with high scores on the subscale tend to use it less in order to access the AM pages compared to those with low scores. However, this finding is not very enlightening as there are no relationships with regards to the use of the other AM navigational tools and the subscale to be able to make comparisons.

Furthermore, with regards to the use of hyperlinks (concept links), it is surprising that there is a negative relationship to the subscale, as it was originally thought that it is a feature which would help towards achieving further understanding of mathematical concepts. Possible reasons for this unexpected relationship relate to the influence of prior knowledge and design of the hyperlinks, as discussed further in Chapter 5 and Chapter 6.

In comparison to the suggested Model 4, the leanest and meanest Model 7 (see Table 10 in Appendix 4.8.4) consists of three predictors: *number of exercises on first try*, *maximum view time on an exercise page*, and *number of pages visited using TOC*. One could argue that in Model 7, there are no predictors with a "controversial" relationship to the subscale such as the *number of hyperlinks (concept links) visited in reading and exercise pages*, and *number of exercises solved on the third try,* hence it provides less enriching but clearer interpretation findings. However, with either Model 4 or Model 7 we have the same issue: they explain only about 1/5 and 1/6 of the variance of the "seeking meaning" subscale, respectively.

A possible explanation for the unexplained variance can be that metrics such as

[32] This means students with a high score on the surface scales.

number of times "search" option is clicked and *number of submitted queries in the search option*, which were not included because of lack of variation, could have contributed to the variance explained by the model (see Table 2 in Appendix 4.8.1). It is possible, therefore, that these missing search-related interactions can be also responsible for the small variance explained in the model.

Finally, the reason that neither version of the "seeking meaning" model seems to give us the full picture, may be that factors such as prior knowledge could have influenced the relationship of this subscale with the predictors, and particularly the aforementioned unexpected ones. The influence of prior knowledge on the subscale is discussed further in the next chapter.

4.9 The deep subscale "relating ideas" and students' "interaction" metrics

4.9.1 Relating ideas – theoretical assumptions

As discussed in 3.3.2, the "relating ideas" subscale, which is based on Pask's holist learning style, measures the extent to which there is an intention to form an overview by exploring topics of what may be known, relating one concept to another, and imposing a personal organisation on learning content (Entwistle, 1997b; Entwistle, 1981; Entwistle et al., 1979; Pask, 1976b).

During the learning process:

- In terms of search-related metrics, students with high values on the "relating ideas" subscale may access and use more features in an interactive learning environment to explore further the relationships between the mathematical concepts and procedures, compared to those with low scores. This is also reinforced by research in the field of interactive learning environments which indicates that students who are characterised as holists (based on Pask's theory) tend to have a preference for hypertext links within the content of a subject, exactly because it allows them to find relationships between topics (Chen et al., 2016). Since, there is a conceptual link between the holist learning style and the "relating ideas" approach, it can be assumed that there can be positive relationships between the subscale and the visits to AM features (particularly the feature of hyperlinks) which facilitate further exploration of learning material and creating links among concepts. This means that there can be positive associations between the subscale and the search-related metrics, as indicated in 3.4.2. More specifically, there can be positive associations between the "relating ideas" subscale and:

- *number of hyperlinks (concept links) visited in reading and exercise pages*
- *number of times "search" option is clicked*
- *number of submitted queries in search option*
- *number of search results visited in search option*

- In terms of performance-related metrics, as it is a deep subscale, which expresses an intention to explore and relate concepts to each other, it can be assumed that students with high scores on the subscale will do better when practising on their exercises during the tutorials, compared to those with low scores. This is also reinforced by literature in mathematics education which supports that students are not likely to do well if they focus on each procedure separately rather than trying to find connections between different parts of mathematics (Liston and O'Donoghue, 2009). So, students with a high score in the subscale who tend to build up their understanding in maths by relating concepts are more likely to do well, compared to those with low scores. Therefore, a positive association is expected between:
    - *"relating ideas" subscale and number of exercises solved on the first try*

  While, there will be negative associations between "relating ideas" and:
    - *number of exercises solved on the second try*
    - *number of exercises solved on the third try*
    - *number of exercises finished but not solved*

- In terms of temporal metrics, it is possible that students with high scores on the "relating ideas" subscale, in an effort to explore and draw relationships between concepts and procedures and form their own overview, may dedicate more time on average to the AM reading and exercise page or to a specific reading and exercise page, compared to those with low scores on the subscale. Similarly to what it is discussed in section 4.8.1 with regards to the "seeking meaning" subscale, students with high scores on the "relating ideas" subscale are not likely to skim through the learning material, but go through it fairly slowly in order to build an understanding by drawing relationships between concepts and procedures. In other words, what is suggested here is that trying to build up an understanding of learning concepts and procedures by looking thoroughly into their relationships can be time-consuming. There can be therefore, positive associations between the "relating ideas" subscale and:

- *average view time on exercise pages*
- *average view time on reading pages*
- *maximum view time on a reading page*
- *maximum view time on an exercise page*

- With regards to the metric of *number of times the "notes" link is clicked* there can be an association with the "relating ideas" subscale. According to Einstein et al. (1985), there is evidence that note-taking allows students to engage in processing such as "relating ideas" to one another or integrating the information with one's existing knowledge. So, it is possible that students with a high score on "relating ideas" may access the "notes" link more times with that intention, compared to those with low scores. As there are two metrics related to the visits to the "notes" feature (see sections 3.4.1 and 3.4.8), the potential positive associations are between the "relating ideas" subscale and:
  - *number of times "notes" link is clicked*
  - *average number of times a "notes" link is clicked per page*

4.9.2 Relating ideas – Results on Correlations

Following the proposed methodology in 3.11 we run correlational analysis to identify predictors for the model. There is a statistically significant correlation, as expected, between the "relating ideas" subscale and *average number of times a "notes" link clicked per page* with r_s=0.180.

Regarding other metrics, mentioned in section 4.9.1, there are no statistically significant relationships. However, the importance of these metrics for the "relating ideas" subscale is discussed further in the following section with regards to the regression models.

4.9.3 Relating ideas models: Development, Analysis and Discussion

In this section, there is a brief discussion on the development of regression models for the "relating ideas" subscale and analysis and discussion on the results for the selected model version.

4.9.3.1 Relating ideas - initial selection of predictors

Table 4.9.3.1 shows briefly that the reasons behind the initial selection of most predictors are their theoretical connections to the subscale, that they may enrich the discussion, and that they are the best contributors when they are tried in pre-models (see further justification in Appendix 4.9.1).

Selected Predictor	Reason for selection
Number of exercises solved on first try	Theoretical connections and enriching further the discussion
Number of exercises solved on second try	
Number of exercises solved on third try	
Number of exercises finished but not solved	
Number of hyperlinks (concepts links) visited on exercise and reading pages	
Maximum view time on an exercise page	
Maximum view time on a reading page	
Average number of times "notes" link is clicked per page	Statistical (see 4.9.2)

Table 4.9.3.**1**. Selected predictors for first version of model.

As a result, at this stage, Model 1 is the first version for the "relating ideas" subscale with all the selected eight aforementioned predictors, which is the maximum number of predictors according to the strategy in the methodology.

4.9.3.2 Relating ideas - development of model

As previously, the process of improving measures of variance and the overall significance starts by excluding specific outliers. This results in increasing the variance: from Model 1 with R^2 13.4% and Adjusted R^2 6.8% to Model 2g with R^2 25.8% and Adjusted R^2 19.9% (see Table 4 in Appendix 4.9.2).

4.9.3.3 Relating ideas - selection of model

After excluding the outliers, there is gradual exclusion of predictors, as shown in 4.9.3.2.

	R^2	Adj. R^2	Sig.	Comments
Model 3 (exclusion of cases 112, 85, 42, 98, 105, and 84, and *number of exercises solved on third try*)	25.2%	20%	0.000	In comparison to Model 2g, note how Adjusted R^2 increases while R^2 gradually slightly decreases, making their difference smaller.
Model 4 (exclusion of cases 112, 85, 42, 98, 105, and 84, and *number of exercises solved on third try,* and *maximum view time on exercise page*)	24.1%	19.6%	0.000	By running the regression model two more times, we obtain the leanest and meanest model. Note how both R^2 and Adjusted R^2 decrease. Model 5 has all predictors statistically significant (see Appendix 4.9.4) Note also that all models are overall statistically significant ($p<0.05$).
Model 5 (exclusion of cases 112, 85, 42, 98, 105, and 84, and *number of exercises solved on third try, maximum view time on exercise page,* and *number of hyperlinks (concept links) visited in reading and exercise pages*)	22.5%	18.7%	0.000	

Table 4.9.3.**2**. Summary of measures of variance and significance

In Table 4.9.3.2, we observe that Model 3 is the best solution, as it combines simultaneously the highest possible R^2 and Adjusted R^2. Finally, Model 3 has 7 predictors with sample size of 109 (after excluding the 6 outliers), which is within the thresholds stated in the strategy, in section 3.11.

4.9.3.4 Relating ideas – Model 3 – Generalisation

For the selected Model 3 of the subscale, the assumptions are not violated and therefore it is possible to generalise the finding beyond the sample (see the assumptions in Appendix 4.9.3).

4.9.3.5 Relating ideas – Equation

By selecting Model 3, we can conclude that for both weeks, the "relating ideas" is expressed through the predictors: maximum view time on an exercise page, maximum view time on a reading page, number of exercises solved on first try, number of exercises solved on second try, number of exercises finished but not solved, number of hyperlinks (concept links) visited in reading and exercise pages, and average number of times a "notes" link is clicked per page. So the equation is formed is as follows:

RelatingIdeasSubcale$_i$ = b$_0$ + b$_1$(Number of exercises solved on first try)$_i$ - b$_2$(Number of exercises solved on second try)$_i$ + b$_3$(Number of exercises finished but not solved)$_i$ + b$_4$(Average number of times a "notes" link is clicked per page)$_i$ - b$_5$(Number of hyperlinks (concept links) visited in reading and exercise pages)$_i$ + b$_6$(Maximum view time on an exercise page)$_{i\,+}$ + b$_7$(Maximum view time on a reading page)$_i$ (1)

If we replace the b values, found in Table 7 in Appendix 4.9.2, in the above equation, then we obtain the equation of the fitted regression model:

RelatingIdeasSubcale$_i$ = 13.132 + 0.019(Number of exercises solved on first try)$_i$ - 0.202(Number of exercises solved on second try)$_i$ + 0.203(Number of exercises finished but not solved)$_i$ + 5.789(Average number of time a "notes" link is clicked per page)$_i$ – 0.138(Number of hyperlinks (concept links) visited in reading and exercise pages)$_i$ + 0.000415(Maximum view time on an exercise page)$_i$ + 0.001 (Maximum view time on a reading page)$_i$ (2)

4.9.3.6 Relating ideas - interpretation of model parameters

a) **Direction of relationship between predictors and outcome based on b values**

As the "relating ideas" subscale score increases, there is increment on the following predictors:

- *number of exercises solved on first try*
- *number of exercises finished but not solved*
- *average number of times a "notes" link is clicked per page*

- *maximum view time on an exercise page*
- *maximum view time on a reading page*

As the “relating ideas” scale score increases, there is a decrease in the following predictors:

- *number of exercises solved on second try*
- *number of hyperlinks (concept) links visited in reading and exercise pages*

It is worth mentioning that the direction of relationships is as expected in the initial hypotheses, except from the *number of exercise finished but not solved,* and *number of (hyperlinks) concept links visited in reading and exercise pages*.

The issue at this point is whether it is possible to explain the “relating ideas” approach towards studying based on the combined knowledge of “interaction” metrics, and whether there are distinguishable “interaction” metrics in students’ interactions for this specific approach.

b) **Further discussion of model and predictors**

First of all, the recommended model (Model 3) explains 25.2% of variance. This amount of variance explained is medium as expected (see regarding recommended thresholds in Appendix 1). It explains 1/4 of the variance of the model explained, which amongst the “deep” recommended models, is the highest amount of variance explained.

In addition, despite the fact that the model 3 is overall statistically significant (see Table 4.9.3.2 above) and holds reasonably well all the required assumptions, it can be observed that not all predictors in Model 3 are statistically significant (see Table 7 in Appendix 4.9.2). The reason for which these predictors were kept is mainly because their inclusion would allow for a richer insight into students’ interaction with regards to the specific subscale. So, the question is: does this “allowance” enrich at least the interpretation of the results?

In the suggested Model 3, the inclusion of all 7 predictors can offer a useful insight, if they are examined in a combined way. For example, the factor which seems to contribute the most in the model (*number of exercises solved on second try*) is the kind of predictor which helps towards a more insightful interpretation, but only when combining it with the predictors *number exercises* *solved on first try* and *number of exercises solved but not finished*. More specifically, students with high scores on the subscale tend to solve more exercises on the first try and fewer exercises on the second try, but they also tend to finish more exercises without solving them (in this way allowing AM to give them the answers), compared to those with low scores. Furthermore, the maximum view time on an exercise and a reading page increases as the "relating ideas" score increases. This could be interpreted as follows: those with high scores on the subscale, who tend to relate mathematical concepts to a greater degree compared to those with low scores, may spend more time on a specific exercise or theoretical page in AM. However, given of the aforementioned tendency of not solving exercises at all, this tendency to spend an increasing amount of time on specific pages, may be because students experience difficulties with their exercises and occasionally "get stuck" on an exercise page.

Other complementary manifestations, according to Model 3 of the "relating ideas" subscale, concern the use of AM features: "notes" and hyperlinks (concept links). With regards to the "notes" feature, it seems that those with high scores on the subscale tend to access it more on average per page, compared to those with low scores, which makes sense as those with high scores may use it more as a way to relate mathematical concepts, compared to those with low scores. On the other hand, with regards to the use of hyperlinks (concept links), it is unexpected that there is a negative relationship to the subscale (as stated in 4.9.3.1 it is originally assumed that it is a feature which would help those with high scores on the subscale towards building understanding by relating mathematical concepts).

In comparison to the suggested Model 3, the leanest and meanest Model 5 includes the same performance-related metrics. Furthermore, Model 5 also shows that the *maximum view time* with regards to the theoretical aspects of the learning material has endured the eliminations, compared to m*aximum view time on exercise pages* which has not. With regards to the use of AM features, the predictor showing average access per page of the AM "notes" feature has also endured in the leanest and meanest model. On the other hand, the use of hyperlinks (concept links) has not endured in Model 5; which, considering the aforementioned unexpected negative relationship, makes the findings more compatible to the theoretical assumptions in 4.9.1. Hence Model 5 provides a less enriching, but yet a clearer interpretation in terms of findings, compared to Model 3.

Finally, for both Model 3 and Model 5 there is a need to discuss possible reasons for the unexplained variance of the models. One possible explanation for the unexplained variance can be that metrics such as *number of times search option is clicked* and *number of submitted queries in the search option*, which could have contributed in the variance explained by the model, were not included because of lack of variation. Another reason can be that a factor such as prior knowledge might have influenced the relationship of this subscale with its predictors and can be behind the aforementioned unexpected relationships (an issue which is discussed in detail in the next chapter).

4.10 The deep subscale "use of evidence" and students' "interaction" metrics

4.10.1 Use of evidence - theoretical assumptions

The "use of evidence" subscale measures the extent to which students study by building up meaning from the details, checking evidence to relate it to conclusions, and their preference for a linear sequence in their learning (Entwistle, 1997a; Entwistle et al., 1979). Entwistle based this approach to studying on research conducted by Pask (1976b) on the serialist learning style. Pask (1976) notes the preference of students with such style for step-by-step and tightly structured learning. They also tend to master one topic at a time and adopt a bottom-up approach in which they pay attention to the low level detail, building an overview at a later stage (Entwistle and Hanley, 1977; Riding and Cheema, 1991).

During the learning process:

- In terms of temporal metrics, it is possible that students with high scores on the "use of evidence", in an effort to master one topic at a time and focus on detailed information to build meaning, may dedicate more time on average to the AM reading and exercise pages or to a specific reading and exercise page, compared to those with low scores. Similarly to what it is discussed in section 4.8.1 with regards to the "seeking meaning" subscale, students with high scores on "use of evidence" are not likely to skim through the learning material, but go through it fairly slowly in order to build an understanding by mastering one topic at a time regarding a specific mathematical concept or procedure. In other words, what is suggested here is that trying to build up an understanding of learning concepts and procedures by looking thoroughly into detailed information with regards to mathematical concepts can be time-consuming.

There can be therefore, positive associations between the “use of evidence” subscale and:

- *average view time on exercise pages*
- *average view time on reading pages*
- *maximum view time on a reading page*
- *maximum view time on an exercise page*

* In terms of path metrics, there can be an association with path linearity (i.e. *stratum*). As discussed in 2.4, it can be the case that students with high scores on the “use of evidence” subscale tend to prefer a more linear sequence when studying the learning content, compared to those with low scores. However, in the current investigation this is unlikely to be the case because of the way the learning material is structured in AM, and the tendency for students with a high score on “use of evidence” subscale to start building the meaning of concepts based more on detailed information (Entwistle, 1997a), compared to those with low scores. Students with a “serialist” approach tend to try to thoroughly understand a topic or a sub-section, for example by looking into detailed examples, before moving on to the next topic (Hills, 2003). In the current investigation, this means those with high scores on the subscale may be “forced” to impose their own structure when going through the learning material in AM. This is because the learning material in AM is not structured in a way which specifically favours a bottom up, “serialist” approach[33] (i.e. going from the detailed worked examples to more generic theoretical type of learning material). So, it can be that those with high scores on the “use of evidence” subscale follow a less linear sequence with regards to the learning material in AM, compared to those with low scores; hence it is very likely that there is a negative association between:
  - *“Use of evidence” and stratum*
* In terms of performance-related metrics, theory on studying approaches does not specifically state that the specific subscale is related to good performance. However, as it is a deep subscale which expresses an intention to build up meaning from details and master one topic at a time, it can be assumed that

[33] It is worth mentioning that it does not favour specifically a top down approach either (i.e. going from the generic and theoretical type of learning material to detailed worked examples specific type of information). In other words, the structure of the learning material in AM is designed in a versatile manner so it favours both approaches.

students with high scores on the subscale may do better when practising their exercises during tutorials, compared to those with low scores. More specifically, there can be positive association between:

- *“Use of evidence” subscale and number of exercises solved on the first try*

While there will be negative associations between:

- *“Use of evidence” subscale and number of exercises solved on the second try*
- “Use of evidence” subscale and number of exercises solved on the third try
- “Use of evidence” subscale and number of exercises finished but not solved

4.10.2 Use of evidence - Results on Correlations

There is only one statistically significant correlation between “use of evidence” and *maximum view time on an exercise page* with r=0.180, as expected.

Regarding other metrics, mentioned in section 4.10.1, there are no statistically significant relationships. However, the importance of these metrics for the “use of evidence” subscale is discussed further in the following section with regards to the regression models.

4.10.3 Use of evidence models: Development, Analysis and Discussion

In this section, there is a brief discussion on the development of regression models for the “use of evidence” subscale and analysis and discussion on the results for the selected model version.

4.10.3.1 Use of evidence - initial selection of predictors

Table 4.10.3.1 shows briefly that the reasons behind the initial selection of most predictors are their theoretical connections to the subscale, and that they may enrich the discussion (see further information in Appendix 4.10.1).

Selected Predictor	Reason for selection
number of exercises solved on first try	Theoretical connections and enriching further the discussion.
number of exercises solved on second try	
number of exercises solved on third try	
number of exercises finished but not solved	
Stratum	
maximum view time on a content (reading) page	
maximum view time on an exercise page	Statistical (see 4.10.2)

Table 4.10.3.**1**. Selected predictors for first version of model.

As indicated in the methodology, the sample size allows for the inclusion of up to 8 predictors and the intention is to take full advantage of this upper limit. After examining the potential inclusion of the 9 predictors mentioned in 4.10.1, it was decided to include just the seven aforementioned predictors (see further justification in Appendix 4.10.1).

4.10.3.2 Use of evidence - development of model

As previously, the process of improving measures of variance and the overall significance starts by excluding specific outliers. This results in increasing the variance: from Model 1 with R^2 15.2% and Adjusted R^2 9.7% to Model 2h with R^2 18.3% and Adjusted R^2 13.3% (see Table 4 in Appendix 4.10.2).

4.10.3.3 Use of evidence - selection of model

After excluding the outliers, there is gradual exclusion of predictors, as shown in Table 4.10.3.2 below.

	R^2	Adj. R^2	Sig.	Comments
Model 3 (exclusion of cases 36, 111, 35, and 94, and *number of exercises finished but not solved*)	18.8%	14.1%	0.001	Note how Adjusted R^2 increases while R^2 gradually slightly decreases, making their difference smaller.
Model 4 (exclusion of cases 36, 111, 35 and 94, and *number of exercises finished but not solved*, and *number of exercises on first try*)	18.8%	14.9%	0.000	
Model 5 – (exclusion of cases 36, 111, 35 and 94, and *number of exercises finished but not solved, number of exercises on first try*, and *maximum view time on reading page*)	18.7%	15.6%	0.000	
Model 6 – Leanest and Meanest (exclusion of cases 36, 111, 35 and 94, and *number of exercises finished but not solved, number of exercises on first try, maximum view time on a reading page,* and *number of exercises solved on third try*)	17.1%	14.7%	0.000	By running the regression model one more time, we obtain the leanest and meanest model with all predictors are statistically significant (see Appendix 4.10.4) Note also that all models are overall statistically significant ($p<0.05$).

Table 4.10.3.**2**. Summary of measures of variance and significance

In Table 4.10.3.2, it observed that Model 5 is the best solution as it as it combines simultaneously the highest possible R^2 and Adjusted R^2.

4.10.3.4 Use of evidence – Model 5 – Generalisation

For the selected Model 5 of the subscale, the assumptions are not violated and therefore it is possible to generalise the finding beyond the sample (see the assumptions in Appendix 4.10.3).

4.10.3.5 Use of evidence – Equation

By selecting Model 5, we can conclude that for both weeks, the "use of evidence" is expressed through the predictors: maximum view time on an exercise page, number of exercises solved on second try, number of exercises solved on third try, and stratum. So, the equation, which is formed is as follows:

$UseofEvidenceSubcale_i = b_0 - b_1$(Number of exercises solved on second try)$_i$ + b_2(Number of exercises solved on third try)$_i$ + b_3(Maximum view time on an exercise page)$_i$ - b_4Stratum (1)

If we replace the b values, found in Table 7 in Appendix 4.10.2, in the above equation, then we obtain the equation of the fitted regression model:

$UseofEvidenceSubcale_i$ = 17.314 – 0.248(Number of exercises solved on second try)$_i$ + 0.156(Number of exercises solved on third try)$_i$ + 0.001(Maximum view time on an exercise page)$_i$ - 4.532(Stratum) (2)

4.10.3.6 Use of evidence - interpretation of model parameters

a) **Direction of relationship between predictors and outcome based on b values**

As the "use of evidence" subscale score increases, there is increment on the following predictors:

- *Number of exercises solved on third try*
- *Maximum view time on an exercise page*

As the interest in ideas scale score increases, there is a decrease in the following predictors:

- *Number of exercises solved on second try*
- *Stratum*

It is worth mentioning that the direction of relationships with regards to the temporal metrics are as expected in the initial hypotheses. However, the direction of the relationship with regards to the number of tries when solving exercises is not quite as expected in the initial hypothesis (i.e. we were expecting a negative relationship to the *number of exercises solved on third try*).

The issue at this point is whether it is possible to explain the "use of evidence" approach towards studying based on the combined knowledge of "interaction" metrics, and whether there are distinguishable "interaction" metrics in students' interactions for this specific approach.

b) **Further discussion of model and predictors**

First of all, the recommended model (Model 5) explains 18.7% of variance. This amount of variance is medium as expected (see the recommended thresholds in Appendix 3.11.1). However, this is only a bit less than 1/5 of the variance of the model explained. Therefore, it is reasonable to say that we do not seem to get the full picture of how students with low and high scores on the "use of evidence" subscale interact with AM during the tutorial sessions.

In addition, despite the fact that the model is overall significant (see Table 4.10.3.2 above) and holds reasonably well all the required assumptions, it is observed that not all predictors are significant (see Table 7 in Appendix 4.10.2). The reason for which these predictors were kept is mainly because their inclusion would allow for a richer insight into students' interaction with regards to the specific subscale. So, the question is: does this "allowance" enrich at least the interpretation of the results?

The main issue in the suggested Model 5 is that performance-related predictors like *number of exercises solved on first try*, which could give an indication that is part of a "deep" approach, have not survived the elimination process, as shown previously in Table 7 in Appendix 4.10.2. As to the rest of the performance-related predictors in Model 5, the direction of their relationships to the outcome is not quite as expected, based on the theoretical assumptions in 4.10.1. More specifically, students with high scores on the "use of evidence" subscale are less likely to solve the exercises on the second try, but more likely to solve them on third try, compared to those with low scores.

With regards to the temporal aspect, the *maximum view time on an exercise page* increases as the "use of evidence" score increases. This could be interpreted as trying to building up meaning from details which can be a time-consuming task; however, when considering their aforementioned performance, it could be just that they “get stuck” on a specific exercise page. Hence, it is reasonable to say that the aforementioned findings resemble more the interaction of students with high “surface” scores rather than students with high “deep” scores.

There is, however, a distinguishing aspect of interaction with regards to *stratum* and the specific subscale. It seems that students with a high score on the subscale tend to follow in a less linear way the given structure of the AM material, compared to those with low scores.

At this point it is also worth mentioning that the leanest and meanest Model 6 (where all 3 predictors are statistically significant) with the exception of *stratum*, does not really give any indications that it represents part of a deep approach. This is because the elimination of the predictor *number of exercises solved on third try* does not allow for comparisons to the predictor *number of exercises solved on second try*.

However, the issue of unexplained variance still remains for both models, hence there is a need to explore whether a factor such as prior knowledge can explain further the variance of the model and the aforementioned unexpected relationship.

This chapter indicates which models allow for a richer insight into students’ interaction with regards to the specific subscale, and in which models there is a need to explore further reasons for unexplained variance and unexpected relationships between the scales and the “interaction” metrics. In the next chapter, the intention is to use these insights in order to make comparisons between the deep and surface scales, determine which “interaction” metrics are better predictors for deep and surface approaches towards studying for the current and future studies, and explore further the influence of prior knowledge in the models.

Chapter 5 - General Discussion

In this chapter, following the analysis and discussion of each individual “deep” and “surface” model, the purpose is to summarise the findings and draw conclusions which will facilitate comparisons amongst the models. More specifically, the intention is to examine the variance explained and the contribution of the predictors in both the suggested and the leanest versions of the models. Furthermore, there is a discussion regarding the predictors and their contribution across all models. There will also be an effort to explain the unexplained variance of the models; hence there is further analysis and discussion with regards to the role of prior knowledge and whether it can increase the unexplained variance.

5.1 General discussion on Surface Scales

5.1.1 Surface Scales - size effects and variance explained according to expectations

The model which explains the most variance concerns the main “surface” scale, with R^2 at 45.5% and Adjusted R^2 at 41.8%. In comparison, the models with the least variance are with regards to the “syllabus boundness” subscale with at R^2 21.6% and Adjusted R^2 at 17.1%, and “lack of purpose” with R^2 at 21% and Adjusted R^2 at 18%. The results for variance explained are summarised in a table in Appendix 5.1.

In terms of our expectation for a medium effect size f^2 and explained variance R^2, the expectations are met. The summarised results, in the table of Appendix 5.1, indicate that all the models have an effect size f^2 above the medium threshold of 0.15, and that the variance explained R^2 is above the medium threshold of 13.04%. In fact, the models representing the scales of the “surface” “fear of failure” and “unrelated memorising” approaches are considered, according to the thresholds indicated in section 3.11.3, to have large effect sizes f^2 since their values are above 0.35, and large variance explained R^2, since their values are above 25.92%. The other two surface scales, “lack of purpose” and “syllabus boundness” have the expected medium effect sizes and variance explained.

It is also worth mentioning that the results indicate that the Adjusted R^2 of the models for the “surface” scales (see Appendix 5.1), which shows the shrinkage of the explained variance, also meets the expectations for medium variance explained. The Adjusted R^2 ranges from 17.1% for "syllabus boundness" (which is above the threshold of medium variance explained) to 41.8% for "surface" scale (which is above the threshold of large variance explained). In other words, if we were to apply the model in a different sample, then the variance explained would be expected to range between medium and large. Considering the expectations for medium effect sizes and variance explained, these findings are rather positive in terms of generalisability.

Despite the fact that expectations are met in terms of effect size and variance explained, there is still close to half, or more for some of the "surface" models, unexplained variance. Moksony (1990) supports that independently of the variance explained, a model should be judged in terms of its theoretical reasoning, as a model relies on a theory. In the current investigation, there were initial theoretical assumptions, the majority of which were reflected in the models, as shown during the development and the discussion of the "surface models" in Chapter 4. So, the "surface" models do make sense according to the initial theoretical expectations and they do give useful information, as discussed in Chapter 4. However, as indicated in the individual discussion of the models, especially with regards to the models of "lack of purpose" and "syllabus boundness", there is close to 4/5 of variance unexplained. In Chapter 4, the author mentioned as a possible reason the influence of prior knowledge, an issue which is discussed in greater detail in section 5.5. There can, however, be other reasons for the unexplained variance with regards to the metrics, as discussed in the following section.

5.1.2 Possible reasons for the unexplained variance

First, in terms of the models representing the "surface" approaches to studying, the unexplained variance may be due to the fact that a certain group of temporal metrics which relate to both "reading" (i.e. theoretical) and "exercise" pages of the learning material in AM do not capture certain subtleties in terms of students' interactions. For example, temporal metrics in terms of the "reading" pages can be more indicative in terms of: basic definitions (e.g. definition of function); advanced concepts (e.g. scalar multiplication); theoretical examples; and worked examples. In this way, it is possible to know whether students with high scores on a specific "surface" scale tend to spend more time on a specific type of theoretical page, compared to those with low scores.

Secondly, in order to ensure stronger contributions (i.e. higher beta values), another possibility is to attach more educational meaning to temporal and performance-related metrics by indicating the level of difficulty of the exercises.

In terms of temporal metrics for exercise pages, considering the level of difficulty concerns especially the models representing the surface scales of “syllabus boundness”, “surface” and “fear of failure”, where temporal metrics with regards to the “exercise” pages contribute in both the suggested and the leanest versions of the models, as shown Appendix 5.3. This suggestion can also help enrich the findings with regards to the model of “unrelated memorising”, where the metric of *average view time on exercise pages* is correlated to the subscale initially (see section 4.2.2), but it is excluded from the suggested version of the model, due to its insignificance and weak beta value.

Furthermore, the level of difficulty is an element which can be introduced in all performance-related metrics, as a way to strengthen even more their contribution to the “surface” scales. For example, it could enrich further the findings if it was possible to know whether the exercises, which are not solved on first try, have a basic, medium or advanced level of difficulty. A model that could possibly benefit from this is the one representing the “lack of purpose” studying approach. For example, level of difficulty might have helped to indicate in the leanest version of the model, whether students with a high score on the subscale are more likely not to solve exercises on first try with a medium or advanced level of difficulty compared to those with a low score[34].

However, this suggestion only makes sense if the learning material is quite varied in terms of difficulty. In the current investigation, it is possible that the level of difficulty of the learning material in AM is not varied enough to capture better specific students’ interactions in certain aspects and this might have affected the contribution of certain predictors in specific models. For example, with regards to the model of the “syllabus boundness” studying approach, the low beta values of certain predictors may be due to the fact that the exercises in the specific module do not require knowledge beyond the syllabus. More specifically, the variation in the predictor *compactness* could be increased, if the exercises required further knowledge and had a more varied level of difficulty. In this way, the contrast between those who explore further concepts –and conduct a more extensive path- possibly to solve more advanced exercises, and those would perform a more compact path, would manifest itself more intensely.

[34] As it is, the metric of *number of exercises solved on first try* is eliminated in the leanest and meanest version (see Appendix 5.3).

So, level of difficulty is an element which has the potential to add further educational meaning and increase the contribution of temporal, performance-related and path-related metrics. More specific recommendations with regards to what improvements can be made in terms of data capture are discussed further in 6.2.

5.1.3 A summary on the inclusion of predictors

In this section, the intention is to summarise the conclusions with regards to the inclusion of all predictors across all suggested versions of the "surface" scales, as indicated in sections 4.1-4.5. Table 5.1.3.1 below shows a summary of all the included predictors across all suggested versions of the "surface" models.

Scale	**Unrelated Memorising (based on Model 6)**	**Syllabus Boundness (based on Model 4)**	**Lack of purpose (based on Model 4)**	**Fear of Failure (based on Model 4)**	**Surface (based on Model 3)**
R^2/ Adjusted R^2	**40%/ 37.7%**	**21.6%/ 17.1%**	**21%/ 18%**	**41.6%/ 39.4%**	**45.5% /41.8%**
Number of exercises solved on first try	(-)	(-)	(-)	(-)	(-)
Number of exercises solved on third try		(+)	(+)	(+)	(+)
Number of exercises finished but not solved	(+)		(+)		(+)
Number of exercises cancelled		(+)			
Average number of times a 'notes' link is clicked per page	(+)				
Compactness	(+)	(+)			(+)
Maximum view time on content (reading) page				(-)	(-)
Average view time on content (reading) pages		(+)			
Maximum view time on exercise page				(+)	(+)
Minimum view time on exercise page		(-)			
Relative amount of revisits			(+)		
Number of hyperlinks (concept links) visited in reading and exercise pages					(+)

Table 5.1.3.**1**. A summarised table of all predictors in the suggested versions of the "surface models" (Note: the "(+)" indicates a positive relationship between predictor and scale and the "(-)" indicates a negative relationship between predictor and scale according to the signs of b and beta values).

More specifically, with regards the metrics included in the "surface" models, the following general observations can be made:

a) ***Performance-related predictors***

With regards to performance-related metrics, Table 5.1.3.1 shows that the metric of *number of exercises solved on first try* is included in all “surface” scales. The other performance-related metric, which is included in all models except the one of the “unrelated memorising” subscale, is *number of exercises solved on third try*. It is followed by the metric of *number of exercises finished but not solved*, which is included in 3 out of the 5 “surface” models; and the metric of *number of exercises cancelled* which is included only in the model of the “syllabus boundness” subscale.

The performance-related metrics have the expected direction, according to the initial theoretical assumptions, with regards to the “surface” scales, which also consistent across all the “surface” models. Furthermore, the strength of this type of metrics is when they are interpreted in a combined way, as they can allow for comparisons, which help to give a clear picture about whether a student does well or not with the exercises during the tutorial sessions. For example, students with high “surface” scores are less likely to solve the exercises on first try and more likely to solve exercises on third try or not at all, compared to those with low scores. The key predictor, however, seems to be the metric of *number of exercises solved on first try.* This is because its consistently negative relationship across all “surface” models facilitates interpretation, as it allows meaningful comparisons to the other performance-related predictors; and this can lead to flagging up a high “surface” approach either to a tutor or in an intelligent interactive learning environment.

b) ***Temporal predictors***

Temporal metrics seem to play a role but not across all the “surface” models. It seems that the *maximum view time on exercise page* and *maximum view time on content page* are relatively more important compared to the equivalent *average view times*, since they contribute to two of the scales: the “fear of failure” subscale and the main “surface” scale. With regards to *maximum view time on exercise page*, when comparing to the metric of *average view time on exercise pages,* this can be because it may help to capture more extreme interactions like experiencing problems with a specific group of exercises on a specific page, thus working much slower. This type of interactions particularly relates to anxiety but also to the more general surface approach to studying. Furthermore, with regards to *maximum view time on content page,* its negative relationship to the “surface” scales means that temporal interactions change with regards to the theoretical pages of the learning material. The reason for not having increasingly extreme interactions, as the scales increase, can be that they are simply anxious to get on with the practical part; or that they are helped by the tutor with explanations regarding the theory.

With regards to *minimum view time on exercise page* and *average view time on content pages,* these seem to help only in the interpretation of the "syllabus boundness" model. An increasing score on "syllabus boundness" scale is linked to a decreasing minimum time for a specific exercise page and increasing average on time spent on the "theoretical" pages. This means that those with high scores on the subscale may spend on average more time on theory, but in terms of exercises they tend to spend time only on what they consider relevant to the requirements of the current task (i.e. spending the minimum possible time on a specific exercise page which they consider less relevant).

Finally, the students' interactions with regards to "unrelated memorising" and "lack of purpose" do not seem to be influenced by any temporal metrics in the suggested versions of the models. With regards to the "unrelated memorising" scale, despite the initial theoretical and empirical relevance, *average view on exercise pages* did not survive the exclusion process; whereas for "lack of purpose" there was no theoretical or empirical evidence for potential inclusion.

c) Revisitation predictor

Relative amount of revisits was expected to contribute to the "unrelated memorising" subscale, as there was both theoretical and empirical evidence (see sections 4.2.1 and 4.2.2). However, it survives the exclusion process only for the "lack of purpose" subscale. As discussed further below, this is still quite useful, as it gives a distinguishing aspect of the specific approach.

d) "Compactness" predictor

The characteristic of the path, which concerns how compact it is, also seems to contribute to most "surface" models. *Compactness* contributes, as expected according to the theoretical assumptions, to most "surface" subscales: "unrelated memorising", "syllabus boundness" and the main "surface". Its consistently positive relationship to the scales adds a distinguishing aspect to the models of the surface approaches towards studying.

The question at this point is whether the suggested versions of "surface" models have distinguishing aspects which can help towards identifying them as surface approaches to studying. More specifically:

- **Lack of purpose model,** besides the combined performance-related metrics, is the only model to include the metric representing the revisitation of pages in the AM learning material.
- **Syllabus boundness model,** besides the combined performance-related metrics and *compactness,* is the only one which includes the metric representing the cancellation of exercises.
- **Fear of Failure model,** besides the combined performance-related metrics**,** includes temporal metrics whose relationships to the subscale offer an interesting contrast: the *maximum view time on reading page* has a negative relationship to the subscale; while the *maximum view time on exercise page* has a positive one.
- **Unrelated memorising model:** besides the combined performance-related metrics, another distinguishing aspect is provided by the metrics of *compactness.* Otherwise, it is the only "surface" model to include an initially unexpected metric: accessing the AM "notes" feature. At first glance, this might not make sense theoretically. Note-taking requires a significant degree of cognitive effort (Piolat et al., 2005), which is usually considered part of a deep approach to studying. However, it is possible that during the tutorial sessions, the AM notes feature is used differently, as a text-copying technique (for results or parts of theory) to reproduce knowledge, rather as a note-making which leads to comprehension. Indeed observations in class showed some students used it for copying and pasting text or simply for recording the solutions –a note taking technique which does not lead to deeper understanding (Katayama et al., 2005).

The overall picture we get in section 5.1 is that the metrics which survived the elimination process during the development of the models give a useful insight and a quite distinguishing picture of students' interactions with regards to "surface" approaches towards studying. More specifically, it is not just that there are distinguishing aspects in some models (i.e. predictors such as *relative amount of revisits*, and *number of exercises cancelled*), it is also that we do get an overall picture which allow us to distinguish those with high scores on the subscales through their interactions in AM during the tutorial sessions, since they reflect reasonably well initial expectations.

In Chapter 6, there is further discussion as to the pedagogical implications of these findings and specifically whether a surface approach can be identified and discouraged.

5.2 General discussion on Deep Scales

5.2.1 Deep Scales - size effects and variance explained according to expectations

The model that explains the most variance concerns the "relating ideas" subscale, with R^2 at 25.2% and Adjusted R^2 at 20%. In comparison, the model with the least variance is the one for the "interest in ideas" subscale with R^2 at 15.7% and Adjusted R^2 at 11.6%. The results for variance explained are summarised in table in Appendix 5.1.

In terms of our expectation for a medium effect size f^2 and explained variance R^2, the expectations are met. The summarised results in the table in Appendix 5.1 indicate that all the models have an effect size f^2 above the medium threshold of 0.15, and variance explained R^2 above the medium threshold for of 13.04%. It is also worth mentioning that none of the "deep" models have large effect size or large variance like the "surface" models.

It is also worth mentioning that the results (see Appendix 5.1) indicate that the Adjusted R^2 of the models for the "deep" scales, which shows the shrinkage of the explained variance, also meets the expectations for medium variance explained. The exception is the model for the "interest in ideas" subscale with Adjusted R^2 at 11.6%. The Adjusted R^2 for the other subscales ranges from 14.4% for the "deep" model to 20% for the "relating ideas" subscale (see Appendix 5.1), which according to the thresholds indicated in section 3.11.3 is considered medium variance. This means that, if we were to apply the model in a different sample, then it would be possible to achieve medium a variance explained R^2.

Despite that most expectations in terms of effect size and variance explained are met, there is between 74.8% and 80% of variance which is unexplained in the models. This is an issue which is discussed in the following sections. However, it is also reasonable to discuss whether the findings with regards to the "deep" models make sense theoretically, as independently of the variance explained a model should be also judged in terms of its theoretical reasoning (Moksony,1990). For example, a theoretical issue that needs to be examined in relation to the deep scales is prior knowledge in relation to the deep scales and particularly in relation to the scale of "interest in ideas" (whose model has the lowest R^2 and Adjusted R^2).

5.2.2 Possible reasons for the unexplained variance

Starting with the model with the lowest variance which represents the "interest in ideas" subscale, there could be various reasons for the unexplained variance. Based on the theory, the "interest in ideas" approach is linked to intrinsic interest, syllabus freedom and independent thinking (Entwistle, 2001). According to Brophy (2010), to encourage and maintain the interest in subject, there is a need to "provide opportunities to make choices deciding what to do and to exercise autonomy in doing it". There is certainly an effort to design AM in this way, by giving students the navigational options (i.e. TOC) to choose which topics to engage with and in what order; providing hyperlinks to visit concepts; and providing a "search" feature to explore further mathematical concepts and procedures beyond the given learning material in AM (in Google, Wikipedia and MathWorld). However, all the search-related metrics, such as *number of times "search" option is clicked*, *number of submitted search queries*, and *number of results visited in "search" option,* produced very little data. It seems that the students did not use the search option, hence these metrics could not be used as predictors in the models because of lack of variation, as indicated in Chapter 4. There are possible reasons for the lack of use. Brophy (2010) suggests that there should be emphasis on the curriculum content and on learning activities that connect with students' interests. There was certainly an effort to consider this when adjusting the AM design to suit the purposes of this study[35].

However, it seems that the search-related options are not emphasized visually at least enough in order to attract students' attention, or they are not emphasized when they are most needed (for example, when student spends an enormous amount of time on a specific exercise page with a specific group of exercises). To conclude, the lack of search-related interactions may result in weak manifestation of the specific approach; subsequently, the lack of these search-related predictors in the model can be responsible for part of the unexplained variance.

In addition, the unexplained variance in the model "interest in ideas" might be due to the influence of prior knowledge. Tobias (1994) indicates that the effects of interest in a subject account for less variance than those attributable to prior

[35] The author made changes on the standard AM design to emphasise the search feature, by placing, for example, the search link at the top menu (prior to this it was hidden inside the "dictionary" feature)

knowledge. For example, it can be the case that the elimination of predictor *number of exercises solved on first try* from the suggested version of the model, based on its weak beta value, is due to the influence of prior knowledge (i.e. low prior knowledge may influence in a negative way the number of exercises solved on first try for those with higher scores on the subscale or high prior knowledge may influence in a positive way the specific metric for those with lower scores on the subscale).

With regards to the “relating ideas” and “seeking meaning” models, the initial theoretical assumptions again are that students may use features such as the search option in AM to find and explore further the relationships between the mathematical concepts. As discussed previously, students did not use the search option. So again, the lack of search-related interactions may result in weak manifestation of the specific approaches; subsequently, the lack of these search-related predictors in the models can be responsible for part of the unexplained variance.

Furthermore, lack of a stronger contribution of the *number of exercises solved on first try* can also explain part of the unexplained variance in models of “seeking meaning”, “relating ideas” and “deep” scales. Again, it is possible that prior knowledge might have influenced the contribution of this predictor in these models (i.e. low prior knowledge might influence in a negative way the number of exercises solved on first try for those with the higher scores on the subscale, or high prior knowledge may influence in a positive way the specific metric for those with lower scores on the subscale).

Finally, with regards to the “use of evidence” model, lack of a stronger contribution of the predictor *stratum* can be a reason for part of the unexplained variance. It is worth mentioning that the contribution of the *stratum* depends on the way the learning material in AM is structured. For example, the less the structure suits those with high scores on the “use of evidence” subscale, the more they are forced to follow a non-linear path compared to those with low scores, and the more variation across the “use of evidence” scores there will be in the metric of *stratum*[36]. Furthermore, the absence of the predictor *number of*

[36] Saying this, there is no suggestion that the structure of the material should be changed to suit less those with higher scores on the use of evidence scale (on the contrary, later on in 5.2.1 the author makes a case for designing a versatile ILE). The author simply points out that *stratum* would have more variability and therefore more contribution to the model if the learning material was structured in a way that would force those with high scores to follow a path with even greater non-linearity.

exercises solved on first try can be another reason for the unexplained variance of the model. Again, this could be due to the influence of prior knowledge (i.e. low prior knowledge might influence in a negative way the number of exercises solved on first try for those with the higher scores on the subscale, or high prior knowledge may influence in a positive way the specific metric for those with lower scores on the subscale).

5.2.3 A summary on the inclusion of predictors

In this section, the intention is to summarise the conclusions with regards to the inclusion of all predictors across all suggested versions of the "deep" scales, as indicated in sections 4.6-4.10. Table 5.2.3.1 below shows a summary of all the included predictors across all suggested versions of the "deep" models.

Scale	**Seeking Meaning (based on Model 4)**	**Relating Ideas (based on Model 3)**	**Interest in Ideas (based on Model 4)**	**Use of evidence (based on Model 5)**	**Deep (based on Model 4)**
R^2/ Adjusted R^2	**20.1%/ 15.5%**	**25.2%/ 20%**	**15.7%/ 11.6%**	**18.7%/ 15.6%**	**18.3%/ 14.4%**
Number of exercises solved on first try	(+)	(+)			(+)
Number of exercises solved on third try	(+)		(-)	(+)	
Number of exercises solved on second try	(-)	(-)	(-)	(-)	(-)
Number of exercises finished but not solved		(+)	(+)		(+)
Stratum				(-)	
Maximum view time on exercise page	(+)	(+)	(+)	(+)	(+)
Maximum view time on a content (reading) page		(+)			
Average view time on exercise pages			(+)		
Number of pages visited using the TOC	(-)				
Number of hyperlinks (concept links) visited in reading and exercise pages	(-)	(-)			
Average number of times notes link is clicked per page		(+)			(+)

Table 5.2.3.**1**. A summarised table of all predictors in the suggested versions of the "deep" models (Note: the "(+)" indicates a positive relationship between predictor and scale and the "(-)" indicates a negative relationship between predictor and scale according to the signs of b and beta values).

More specifically, with regards the metrics included in the "deep" models, the following general observations can be made:

a) Performance-related predictors

With regards to the performance-related metrics, Table 5.2.3.1 shows that the metric *number of exercises solved on second try* is included in all “deep” scales, and also its relationship to the deep scales is also consistently negative. However, it is the type of “performance-related” predictor which on its own cannot say much regarding the performance of the students. The predictor *number of exercises solved on first try* can help in this, as it has the expected positive relationship to the “deep” scales, but it contributes only to 3 of the “deep” scales: “seeking meaning”, “relating ideas” and “deep”. Furthermore, the predictor *number of exercises finished but not solved* has a consistent but unexpected positive relationship to the deep scales. In addition, the predictor *number of exercises solved on third try* has the expected negative relationship but only for the “interest in ideas” subscale; it has a negative relationship to the “seeking meaning” and “use of evidence” subscales.

Now, if we examine the performance-related predictors for each deep scale, then in “seeking meaning”, “relating ideas” and “deep” scale, the inclusion of *number of exercises solved on first try* offers some sort of clarity in that it supports the initial general assumption that those with high scores on the deep scales tend to perform better compared to those with low scores. However, in the “use of evidence”, and “interest in ideas” deep scales, those with high scores do not seem to perform better compared to those with low scores. More specifically, those with high scores on the “use of evidence’” scale are more likely to solve exercises on third try and not on second try compared to those with low scores; and those with high scores on the “interest in ideas” scale are more likely not to solve exercises at all and less likely to solve exercises on second and third try, compared to those with low scores.

It is therefore reasonable to conclude that the performance-related metrics do not give a very clear overall pattern for the “deep” scales, or even the expected outcome for two of the subscales in terms of how students do when solving exercises during their tutorial sessions.

A reasonable question at this point is whether these findings contradict the general initial theoretical assumption that the main deep scale indicates a positive relationship to performance. Entwistle and Peterson (2004) support that the deep approach is linked to positive academic performance, but this is more likely to happen when the tasks require a deep level of understanding. In other words, it could be that the practical tasks and in general the level of difficulty of exercises during the tutorials simply do not require a deep level of understanding. This is reinforced by the recorded notes of students, especially with regards to the "functions and graphs" exercises[37]. For example, they commented on the level of difficulty for "real function" exercises with comments such as: *"tooo easy…"*, *"answer is eeeeeeeesy -1", or "i know…so calm down it's very basic".*

Therefore, while there are no overall clear patterns with regards to the performance-related metrics and deep scales, as there are in the surface scales, the findings reflect other observations made in other educational contexts.

b) Temporal predictors

Temporal metrics such as the *maximum view time on exercise page* seem to make a contribution as they are included in all "deep" models and have a consistently positive relationship with all of them. It was initially assumed that students who tend to seek meaning for achieving personal understanding make an effort to relate concepts, try to construct understanding based on details, and have an intrinsic interest, are thus likely to dedicate more time studying specific pages of the learning material, compared to those who do not have these tendencies. However, it can be simply that those with high scores in deep scales may experience difficulties with a specific group of exercises, as across the high scores of the deep scales there are tendencies to not solving exercises at all or solving them on the third try. Similar interpretations can be made for the positive relationship between *average view time on exercise pages* and the "interest in ideas" subscale.

To conclude, maximum view time and average view time on exercise pages can help to explain interactions with regards to a deep approach. However, since their positive relationships to the deep scales may be interpreted as those with high scores experiencing difficulties with the exercises, these type of metrics do not really offer a distinguishing aspect. This is reinforced by the fact that some surface scales have also a positive relationship with the *maximum view time on exercise page*.

With regards to the *maximum view time on a reading page*, there is a positive relationship to "relating ideas" scale, as expected. For those with high scores on the subscale, this can be interpreted as trying to relate more mathematical concepts, compared to those with low scores, or again that they are likely to

[37] Students used the "public" option of the AM "notes" feature to comment on the difficulty of the exercises.

experience more difficulties understanding a specific procedure or concept (given that there is a tendency for not solving exercises at all). However, it is worth mentioning that *maximum view time on a content page* differs with regards to the direction of its relationship between deep and surface scales (i.e. it has a positive relationship to the “relating ideas” scale, but a negative relationship to the “fear of failure” and main “surface” scales).

c) *“Number of hyperlinks (concept links) visited in reading and exercise pages” predictor*

With regards to the *number of concept links visited in reading and exercise pages*, there is an unexpected, according to the initial theoretical assumptions, negative relationship to “seeking meaning” and “relating ideas” scales. It seems that the use of concept links in AM does not help towards encouraging deep approaches such as “seeking meaning” and “relating ideas” towards studying. However, it is worth mentioning that the metric differs with regards to the direction of its relationship between deep and surface scales (i.e. it has a negative relationship to “relating ideas” and “seeking meaning” subscales, but a positive relationship to the main “surface” scale).

d) *“Average number of notes links is clicked per page” predictor*

The metric representing the access in the AM “notes” feature has the expected positive relationship to the “relating ideas” and the main deep scale. However, it is worth mentioning at this point that its relationship has the same direction as the “unrelated memorising” subscale. It seems that note-taking features in an interactive learning environment such as AM can be used as a “copying” strategy for over-rehearsing, as well as note-making to improve understanding. Also, based on observations, some students used it to ask for help, so it is possible that those with “high” scores on the “unrelated memorising” subscale might have used it more to seek help from fellow students (see Appendix 6.1).

e) *“Stratum” predictor*

With regards to the *stratum*, its negative relationship to the “use of evidence” scale reveals the tendency for the students with high scores on the subscale to impose their own structure. In the current investigation, this makes sense as the learning content, especially the theoretical part, in AM is structured in a versatile way (i.e. the concepts and procedures are presented both in an inductive way by starting from the detailed worked examples and leading to more theoretical and generic examples; and in a deductive way by starting from more theoretical generic examples and leading to detailed worked examples or by including summaries of procedures or operations [38] (see structure of theory in AM in Appendix 5.4).

So, this metric can offer a distinguishing aspect in terms of students' interaction with regards to this studying approach, but, at the same time, as discussed in 5.1.3, its variability and relationship to the "use of evidence" scale really depends on the way the learning material is structured in an ILE. For example, if the learning environment suits both those who tend to build up meaning from the details in a quite inductive manner, and those who tend to build up meaning from an overview of what may be known in a quite deductive manner (that is if it is structured in a versatile way[39]), then a tendency to follow a more non-linear path should be expected for those with higher scores on the "use of evidence" subscale compared to those with lower scores.

f) "Number of pages visited using the TOC" predictor

With regards to the *number of pages visited using the TOC,* its negative relationship to the "seeking meaning" subscale is not particularly enlightening, as mentioned in section 4.8.3.6. This is also reinforced by the fact that it is a predictor which is not involved in any suggested versions of deep and surface models, to allow for some potential useful comparisons.

The question at this point is whether the suggested versions of "deep" models have distinguishing aspects which can help towards identifying them as deep approaches to studying through students' interactions. More specifically:

- **The "deep" model** includes the predictor *number of exercises solved on first try*, which can mainly offer a distinguishing aspect with regards to the deep approach towards studying.
- **The "seeking meaning" model** primarily includes the predictor *number of exercises solved on first try*, which can mainly help towards identifying it as part of the deep approach towards studying.

[38] As discussed in 2.3, in other empirical studies an inductive type of structure has been also described as "depth-first" or "bottom-up"; while a deductive type of structure is described as "breadth-first" or "top-down".

[39] It is worth mentioning here that the author makes a case for a versatile ILE, later on in section 5.2.1, based on Pask's (1976a, 1976b) theory of "versatile" learners.

- **The “relating ideas” model** includes the predictors *number of exercises solved on first try* and *maximum view time on a content page*, which can mainly help towards identifying it as part of the deep approach towards studying.
- **The “use of evidence” model** includes primarily one predictor, which may give a distinguishing aspect: *stratum*.
- **The “interest in ideas” model** includes predictors which do not seem to offer distinguishing aspects for a subscale which is considered part of the deep approach towards studying.

The overall picture we get in section 5.2 is that the metrics which survived the selection process during the development of the “deep” models give some insight, but do not quite give an overall distinguishing picture of the deep approaches towards studying. More specifically, there are some isolated distinguishing aspects in most of them, however overall we get a rather mixed picture which does not really serve towards distinguishing those with high scores on the subscales through their interactions in AM during the tutorial sessions, since most of their interactions seem to resemble more those with high scores on the “surface” scales.

In Chapter 6, there is further discussion as to the pedagogical implications of these findings, specifically: whether the deep approach is easy to capture at first place (and if not why not); to what degree the deep approach can be encouraged based on the use and design of an interactive learning environment alone; and whether these findings contradict or reinforce what it is known about deep approaches.

5.3 Comparisons between deep and surface scales

In this section the intention is to make comparisons between the deep and surface scale and with regards to their variance and size effects.

5.3.1 Contribution on deep and surface scales according to beta values

In sections 5.1.3 and 5.2.3, certain indications have been given as to which predictors seem to play an important role in the models. To obtain further understanding with regards to this aspect, it is worth examining the contributions of predictors across surface and deep models, according to their beta value, as shown in table 5.3.1.1.

Scale	Highest Contributing Predictor	Beta value (expressed in absolute values) in suggested versions of models
Fear of failure	*Number of exercises solved on first try*	0.440
Lack of purpose	*Relative amount of revisits*	0.275
Syllabus boundness	*Minimum view time on exercise page*	0.253
Unrelated memorising	*Number of exercises solved on first try*	0.512
Surface	*Number of exercises solved on first try*	0.435
Deep	*Number of exercises solved on second try*	0.326
Interest in ideas	*Number of exercises finished but not solved*	0.385
Relating ideas	*Number of exercises solved on second try*	0.360
Seeking meaning	*Maximum view time on exercise page*	0.297
Use of evidence	*Number of exercises solved on second try*	0.405

Table 5.3.1.**1**. A summary of the predictors with the highest contribution per model, based on the beta values.

Starting with the "surface" models, the highest contributing predictor for "fear of failure", "unrelated memorising" and "surface" models is the *number of exercises solved on first try.* This along with the fact that it is a predictor which contributes to all "surface" models and three of the "deep" models, and offers a distinguishing aspect, makes it an important predictor for future recommendations.

Furthermore, the highest contributing predictor for the "lack of purpose" model is the *relative amount of revisits*. This, in combination with the fact that this predictor gives a distinguishing aspect in this subscale, means that revisitation of learning material is the most important element in this approach towards studying and a predictor which can be recommended for similar future studies.

With regards to the "syllabus boundness" model, the highest contributing predictor is the *minimum view time on exercise page,* and this in combination with the fact that it gives a distinguishing aspect to the subscale, makes it also a good recommendation for future studies.

Amongst the deep scales, the highest contributing predictor for "deep", "relating ideas" and "use of evidence" models is the *number of exercises solved on second try.* This finding gives a useful insight; however, as mentioned earlier, this predictor cannot say much on its own.

Furthermore, the highest contributing predictor for the “interest in ideas” model is the predictor *number of exercises finished but not solved*; however, its positive relationship to the scale means that this predictor cannot really offer a distinguishing aspect for the specific approach.

In a similar way, the highest contributing predictor for the “seeking meaning” model, *maximum view time on exercise page,* does give a useful insight, but due to its positive relationship to both the specific scale and the “surface” scales cannot really offer a distinguishing aspect for the specific approach.

The conclusion reached in sections 5.1.3 and 5.2.3 is reinforced here again. When comparing the models representing the surface and deep approaches towards studying, it is possible to get more useful insights and a clearer and more distinguishing picture of the “surface” models compared to the “deep” models.

5.3.2 Comparing the variance explained by Surface and Deep scales

There is certainly more variance explained for the “surface” models compared to the “deep” models. While in surfaces scales variance ranges between 21% and 45.5%, in the deep scales variance range between 15.7% and 25.2% (see Appendix 5.1). Three of the models representing the surface scales (“unrelated memorising”, “fear of failure” and the main “surface” scale) have almost double the amount of variance compared to the model of the deep scales.

The question at this point is: why there is overall more unexplained variance in the models of the deep scales compared to the models of the surface scales?

Further to what is discussed in section 5.2.3 with regards to this issue, Entwistle (2008), Marton and Säljö (1976), and Entwistle and Peterson (2004) argue that a deep approach towards studying is not quite consistent and it does not manifest itself as strongly as one might have hoped for. This can be because students may have the intention for it, but they may find themselves in a university course with a learning and teaching environment which simply does not encourage it, so they end up behaving in a surface manner. In that sense, it is seems that it is not easy to capture the students’ deep approach towards studying when using an interactive learning environment during tutorial sessions of a specific module (which is only a part of their university course).

It is reasonable to conclude, therefore that the deep scale is a dependent variable with a high degree of complexity; and as Meyers et al. (2006) suggest, in the general field of statistics, unexplained variance can be due to the complexity of a variable.

On the other hand, the surface scale does not seem to have the same degree of complexity and it tends to manifest more strongly. For example, according to Entwistle and Peterson (2004), academic performance has a strong relationship with the surface approach, and specifically indicates a poor academic performance (while deep approach is linked to academic performance only when tasks require a deep level of understanding). It is reasonable to say that while a deep approach manifests itself under certain conditions, for the surface approach this does not seem to be the case.

Furthermore, Entwistle (2008) suggests that the adoption of a deep approach towards studying depends also on students' prior knowledge. More specifically, those students declaring low prior knowledge are more likely to follow a surface approach towards studying (Entwistle and Peterson, 2004). Ramsden (2005) also suggests that prior knowledge is mentioned more in science students as a factor which can lead to lack of comprehension, anxiety, superficial learning and passiveness. So, the unexplained variance in both deep and surface models can be due to the fact that prior knowledge is not considered in the models; hence the decision to examine further empirically and theoretically its influence on the scales (see section 5.5).

5.3.3 "Surface" models and "deep" models – Ensuring Statistical Power

Throughout the development of the models, there was a continuous effort to ensure that the typical threshold for the statistical power of 0.8, as proposed by Cohen (1992), is achieved. Based on the expectations and thresholds indicated in the strategy (see section 3.11.3), initially it was suggested that a reasonable number of maximum 8 predictors should correspond to the study's sample size of 115 participants. Then, throughout the exclusion of outliers and predictors in the models, it was ensured that the limits, indicated in the strategy (i.e. no more than 6 predictors for 100 participants, and no more than 7 predictors for 105-107 participants) were always respected. As shown on Table 1 in Appendix 5.1, all suggested models have an effect size which ranges from medium to large, and for sample sizes between 109-113 participants there are between 4 and 7 predictors (see Appendix 5.1).

5.3.4 Surface scales and deep models– Ensuring Overall Statistical Significance

Throughout the development of the models, there was also an effort to hold the level of significance at the expected a=0.05 (which was set up from the beginning for the whole analysis). The exclusion of outliers and predictors with the highest insignificance assisted toward achieving the expected level of significance at a=0.05, in all the suggested versions of the models, as shown in Appendix 5.1.

5.3.5 Predictors which do not survive in any of the leanest and meanest models

Despite the fact that the suggested versions of the models have overall statistical significance, there is a need to address the fact that certain predictors in the suggested versions models are not significant. These predictors are kept in the suggested versions of the models, because they have the highest possible variance explained both in terms of R^2 and Adjusted R^2 and the same time there is the potential to offer a more enriching interpretation of the models. However, it is also reasonable to examine the leanest and meanest versions of the models for all scales, because it allows for useful comparisons and may also ultimately lead to recommendations for future studies.

At this stage a reasonable question is: which predictors do not survive in any of the leanest and meanest models?

With regards to the "surface" models, as shown in table 2 in Appendix 5.3, the predictor *number of hyperlinks (concept links) visited in reading and exercise pages* does not contribute in any of the leanest and meanest versions of the models. Given that the initial theoretical assumption indicated a negative relationship and not a positive one, its loss from the meanest models can be an indication that the direction of its relationship to the surface scales can alter in a different sample. This can be also the case with regards to the *average number a "notes" link is clicked per page* and the direction of its relationship to the "unrelated memorising" subscale, which also does not survive in the leanest and meanest version of the model.

In the "deep" models, as shown in table 1 of Appendix 5.3, the predictors *average view time on exercise pages,* and *number of hyperlinks (concept links) in reading and exercise pages* do not contribute in any of the leanest and meanest versions of the deep models.

With regards to the *average view time on exercise pages*, its loss does not really take away a lot from the interpretation of the “interest in ideas” scale, as there is another temporal predictor *–maximum view time on exercise page-* which can give a good insight regarding the relationship of temporal metrics to the specific scale. With regards to the *number of hyperlinks (concept links) visited in reading and exercise pages*, there is a negative relationship to the deep scales, while it was initially assumed that it would have a positive one. So, what was stated earlier with regards to the surface scales and the same predictor is reinforced here: its loss from the meanest models means that there may be a relationship between deep scales and the specific predictor, but the direction of its relationship to the deep scales can alter in a different sample.

5.3.6 Surface scales and deep scales – Leanest and Meanest models

In this section the intention is to draw certain conclusions and get a clearer picture as to whether the leanest and meanest versions of the models can still give useful insights (i.e. an enriching picture) and distinguishing aspects with regards to the “deep” and “surface” scales, compared to the suggested version of the models (as shown in Appendix 5.3). Based on the findings in Chapter 4 and the aforementioned observations made in sections 5.1, 5.2, 5.3.5 and 5.3.6, it is possible to summarise the following conclusions:

- **Surface:** As discussed in 4.1.3.6, the eliminated predictors *number of exercises finished but not solved, compactness* and *number of hyperlinks (concepts links) visited in reading and exercise pages* contribute to a more enriching picture of students’ interactions with regards to this approach, albeit a less “contradicting” one. The remaining factors in the leanest version (see Appendix 4.1.3) can still give indications that this model represents the surface approach through: the performance-related metrics; and the interesting contrast of having the *maximum view time on reading page* with a negative relationship to the subscale and the *maximum view time on exercise page* with a positive one, as observed in section 5.1.3. Finally, the amount of variance explained by the leanest model is still large as in the suggested version of the model.
- **Fear of failure:** This is the only scale in which the suggested version of the model is also the leanest and meanest one. As discussed in section 4.3.3.6, overall the model offers useful insights, and a distinguishing aspect can be the contrasting relationships between the subscale and the predictors *maximum view time on reading page* and *maximum view time on exercise page* to the scale, as observed in section 5.1.3. Finally, the amount of variance explained by the leanest model is still large as in the suggested version of the model.

- **Unrelated memorising:** As discussed in sections 4.2.3.6, 5.1.3 and 5.3.5 the only eliminated predictor, compared to the suggested version, is *average number of "notes" links per page*, whose relationship to the subscale can be explained and to certain degree it is enriching, but its role in the model requires further investigation in future studies. The remaining predictors can still give useful insight and distinguishing aspects with predictors such as *compactness*[40] (see Appendix 4.2.4). Finally, the amount of variance explained in the leanest version is still large as in the suggested version of the model.
- **Lack of purpose:** As discussed in section 4.5.3.6, the eliminated predictors *number of exercises solved on third try* and *number of exercises solved on first try* contribute to a more enriching and distinguishing picture of students' interactions with regards to this approach. The remaining two predictors in the leanest version (see Appendix 4.5.4), and especially the *relative amount of revisits* can still give an indication that the model represents a part of a surface approach, as discussed in 5.1.3; however the fact remains that the picture we get from the leanest model does not seem to be complete, and the amount of variance explained, although at medium level, is approximately half the variance explained compared to the aforementioned surface models (see Appendix 5.1).
- **Syllabus boundness:** As discussed in section 4.3.3.6, the eliminated predictors *compactness* and especially *number of exercises solved on first try* contribute to a more enriching picture of students' interactions with regards to this approach. The remaining predictors, and especially the *number of exercises cancelled* can still give an indication that the model represents a part of a surface approach. However, there is still the issue of the amount of variance explained: the syllabus boundness model explains only approximately half of the amount of variance, compared to the variance explained by the aforementioned "surface" models (see Appendix 5.1).
- **Deep:** As discussed in section 4.5.3.6, the eliminated predictor *number of exercises solved on first try* is in essence the main predictor which gives an indication that this model represents a deep approach. The remaining predictors in the leanest version have the same direction in their relationships as with the "surface" scales. The exception is the *number of exercises solved on second try,* which needs, however, to be interpreted and compared with the

[40] It is worth mentioning that *compactness* survives only in the leanest and meanest version of the unrelated memorising model.

other performance-related metrics to make sense, as discussed in section 5.2.3. The variance of leanest version of the model remains at medium level as the one in the suggested version of the model; however, it is less than half compared to the variance explained by the “surface” model and explains only about 1/6 of the model.

- **Interest in ideas:** As discussed in section 4.7.3.6, neither the suggested nor the leanest and meanest version of the model gives an enriching or distinguishing picture that can be considered part of a deep approach. Amongst all the leanest and meanest versions of both “deep” and “surface” models, this the only one where the amount of variance explained is at a small level and not medium as in the suggested version of the model.
- **Relating ideas:** As discussed in section 4.9.3.6, compared to the suggested version the eliminated predictors in the leanest and meanest model are: the *maximum view time on exercise page* (which has the same positive relationship to the surface scales), and the *number of hyperlinks (concept links) visited in reading and exercise pages* (which has an unexpected negative relationship to the subscale). So, the leanest and meanest model may be a less enriching but it gives a clearer indication that it represents part of a deep approach. For example, predictors which contribute to this are: *maximum view time on content page* (which has a unique positive relationship to the “relating ideas” subscale, while it has a negative one to the surface scale); and *number of exercises solved on first try*. The amount of variance explained by the model remains at medium level as in the suggested version of the model; however, it explains a bit less than ¼ of the model.
- **Seeking meaning:** As discussed in section 4.8.3.6, the eliminated predictors are *number of exercises solved on third try*, *number of exercises solved on second try*, and *number of hyperlinks (concept links) visited in reading and exercise pages.* These may give insight, but do not really bring any distinguishing aspects to the model. In the leanest model, the remaining predictor *number of exercises solved on first try* gives an indication that it represents part of a deep approach. The amount of variance explained by the model, which is at medium level as in the suggested version of the model, still explains only about 1/6 of the model.
- **Use of evidence:** As discussed in 4.10.3.6, the predictors of neither the suggested nor the leanest and meanest version of the model give an enriching or distinguishing picture that can be considered part of a deep approach. The only exception is the predictor *stratum* which remains in the leanest version, but

as discussed in sections 5.2.2 and 5.2.3, its relationship to the subscale depends on the structure of the learning material in the ILE. Finally, the amount of variance explained by the model is at a medium level as in the suggested version of the model, but still it explains only about 1/6 of the model.

Based on the above conclusions, and the discussion on predictors in sections 5.1.3 and 5.2.3, when comparing both the suggested and leanest versions of the models at predictor level there is more confidence regarding their distinguishing aspects for the surface models compared to the deep ones. Overall, the leanest version of the "surface" models, with the exception of the "lack of purpose" one, offer a complete image and useful insights compared to the deep ones. Amongst the deep models, the "relating ideas" seems to be the one with the most useful insights and distinguishable aspects. In Chapter 6, the pedagogical implication of these findings is discussed further (i.e. whether it contradicts the findings in different contexts and what it means for future studies in this context).

5.4 Surface and Deep Scales - Generalisation and Limitations

In order to draw conclusions about a population based on regression analysis done on a sample, typically certain assumptions should be checked in regression models (Field, 2009). Therefore, throughout the process of development, as shown in Chapter 4, it is ensured that:

- Predictors with close to zero variance (such as *number of times "search" option is clicked, number of submitted search queries in "search" option, and number of results visited in "search option"*) are not included in the models.
- Multicollinearity is respected and predictors which correlate highly are not included in the models (such as *average view time on content pages* and *maximum view time on content page*).
- For all the suggested versions of the models, the variance of residuals is reasonably homoscedastic and the residuals are reasonably normally distributed.

Furthermore, a statistical measure such as the Adjusted R^2 can indicate the amount of variance explained in regression models, if we were to conduct the study with a different sample. Specific thresholds for the Adjusted R^2 based on empirical evidence in similar context have not been found; so as indicated in the methodology throughout the development of the models, there is an effort to suggest the model with the highest possible Adjusted R^2 (in order to achieve at the same time the smallest possible difference between R^2 and Adjusted R^2 and also the most "enriching" version of the model). With regards to the surface scales, the results in terms of Adjusted R^2 are particularly encouraging, as its values range from medium to large; while with regards to the deep scales, its values are medium for all scales except for one, the "interest in ideas" subscale, where its value is small.

Despite the generally reasonable statistical results in terms of generalisation, there are however limitations. First of all, the sample is a sample of convenience and not a random one. Secondly, the ASSIST instrument poses limitations in terms of generalisation as the students are asked to answer its questions with a specific module in mind, which is part of a specific course and in a specific institution. So, all the data collected is influenced by the context of a specific educational setting. Thirdly, as discussed in section 5.3.2, a student's intention for a deep approach is not consistent as it is influenced by the wider teaching and learning environment in which the ILE is used. Ramsden (2005, p.216) points to research findings which indicate that: "intense effort must be made in course planning, and in the setting of assessment questions, to avoid presenting a learning context which is perceived by students to require, or reward, surface approaches. It is not enough to assume that course materials will encourage students to think deeply about the subject matter, however carefully they have been designed: it is necessary to consider the students' perspective on what is required". More specifically, it is possible that in a low-league university, students with the intention for a deep approach towards studying may interact in a "surface manner", because the wider teaching and learning environment (besides the interactive learning environment) does not encourage a deep approach or students do not perceive that a deep approach is required. Hence the deep approach to studying does not manifest itself strongly or interact with the environment as expected. On the other hand, in a high-league university, where a deep approach towards studying is supposed to be encouraged more, a deep approach towards studying could manifest more strongly and as expected, because the teaching-learning environment encourages it, or because students perceive that a deep approach is required.

The above insights show that it would not be wise to claim generalisation for the current findings, without considering carefully the differences between the educational settings of the current study and any future study. For example, deep approaches towards studying can manifest differently in a high-league university even if it is in a similar course and module, and with a similar interactive learning environment which is used during tutorial sessions. Or, to give another example, if in a future study the way the learning material is designed and structured in an interactive learning environment alters, then the relationships between certain predictors and subscales may also change (e.g. *stratum* and “use of evidence”).

Despite the aforementioned limitations in terms of generalisation, however, the empirical findings and the insights of the current research can give a starting point for forming initial assumptions for a different sample. Even if the sample differs greatly because of the educational setting, adjustments can be made with regards to the initial assumptions which may reflect both these differences and the current findings. Hence it can provide a good starting point in terms of methodological recommendations for future studies.

5.5 Prior knowledge and its relevance to the current study and findings

5.5.1 Prior knowledge as a selection variable in the model of deep and surface approaches

The intention in this section is to address the secondary complementary question with regards to the influence of the students’ prior knowledge. As discussed in 2.5, the level of prior knowledge can influence students’ interaction in a learning environment and it may influence the variance of the models representing the deep and surface scales and subscales. Hence, as indicated in the methodology in section 3.11.3, starting from the suggested versions of the models, prior knowledge is included as a “selection” variable and the sample is split into a “low prior knowledge” group and “high prior knowledge” group. After running the multiple regression for all models, there will be comparisons, at model level, regarding the variance explained by the models of the groups. To facilitate these comparisons, a summary of the variances explained by the models for both groups and the initial sample is shown in Appendix 5.2.

5.5.2 Models in low and high prior knowledge groups

When comparing the variance explained for low and high prior knowledge groups across all scales, it is observed that with regards to the models of the “low prior knowledge” groups, 8 out of the 10 models have increased variance R^2, compared to the suggested models of the whole sample (see the R^2 values highlighted in green in Appendix 5.2). The 2 models which have decreased variance R^2, compared to the suggested models of the whole sample, are those representing the “unrelated memorising” and “fear of failure” subscales. These observations mean that in most models the predictors explain better the variance of the ASSIST scales. This is especially the case for the models of the deep subscales. This is an important finding given that the suggested versions of the models for the whole sample do not explain as much variance as those in the surface subscales. It is an indication that prior knowledge as a selector variable makes a difference in the variance explained.

Furthermore, with regards to the models of the “high prior knowledge” group, it is observed that 8 out of the 10 models have decreased variance R^2, compared to the suggested models of the whole sample (see the R^2 values highlighted in green in Appendix 5.2). The only 2 models with an increased variance R^2, compared to the suggested models of the whole sample, are those representing the “fear of failure” and “use of evidence” subscales[41]. These observations mean that in most models the predictors do not explain better the variance of the ASSIST scales, compared to models of the whole sample.

Therefore, prior knowledge as a selector variable makes a difference in the variance explained, but mainly in the “low prior knowledge” groups, more specifically those representing the deep subscales. This is reinforced by the finding that, in 4 out of the 5 deep models, the variance explained is almost doubled (see Appendix 5.2).

In the following sections, there is further discussion with regards to the possible reasons for which there are changes in terms the variance explained, compared to the suggested versions of the models for the whole sample.

5.5.3 Deep models in low and high prior knowledge groups

In this section, there is an attempt to examine and compare the variance explained by the “deep” models with regards to the low and high prior knowledge groups.

[41] There are also two models which are not significant at 0.05 (see *Sig.* values highlighted in blue in Appendix 5.2). If we were to continue further development of the insignificant model by excluding the most insignificant predictor, then the model would become significant at some point. However, further development of the low/high prior knowledge group models is not intended. The intention here is to simply get some indication at model-level as to which groups are most influenced by prior knowledge.

5.5.3.1 Deep scales in the "low prior knowledge" group

In the "low prior knowledge" group, the variance R^2 increases for all "deep" models, compared to the variance of the "deep" models for the whole sample, and in most models the R^2 value increases as much as double (see Appendix 5.2). More specifically it is observed that there is an overall 17.8% to 21.5% increase: from 18.3% to 36.1% for the "deep" model; from 15.7% to 32.4% for the "interest in ideas" model; from 20.1% to 41.3% for the "seeking meaning" model; and from 25.2% to 43.1% for the "relating ideas" model. The only model in which there is a relatively small increase is the one representing the "use of evidence" subscale with 2.6% (from 18.7% to 21.3%).

In terms of Adjusted R^2, there is an increase for four of the "deep" models, and for as much as approximately double the amount of variance in certain cases. More specifically, it is observed that there is an increase: from 14.4% to 28.1% for the "deep" model; from 11.6% to 23.7% for the "interest in ideas" model; from 15.5% to 32.3% for the "seeking meaning" model; from 20% to 32.6% for the "relating ideas" model. The only model in which there is a slight decrease is the one representing the "use of evidence" subscale with 1.9% (from 15.6% to 13.7%)[42].

These findings indicate that the students' interactions in AM with regards to the deep approaches towards studying are influenced greatly by the "low prior knowledge" group. In the models representing the deep scales, the predictors are able to explain more for students belonging to the "low prior knowledge" group. This finding can have implications as encouraging a deep approach towards studying is especially important in groups with little or no experience in a subject (i.e. it can help tutors in class and also inform the design of an ILE regarding tackling the "low prior knowledge" groups of students which at the same time have a tendency towards a "low" deep approach towards studying).

On the other hand, the specific predictors do not seem to explain as much if a student with a low or high scores on the scale belongs to the "high prior knowledge" group, as discussed below.

---

[42] If we were to continue further with the development of the model by excluding the most insignificant predictor, then the Adjusted R^2 would be very likely increased. However, further development of the low/high prior knowledge group models is not intended. The intention here is simply to get some indication at model-level as to which groups are most influenced by prior knowledge.

5.5.3.2 Deep subscales in the "high prior knowledge" group

In the "high prior knowledge" group, the variance R^2 decreases for 4 of the "deep" models, compared to the variance of the "deep" models for the whole sample (see Appendix 5.2). More specifically, it is observed that there is a decrease: from 18.3% to 13.3% for the "deep" model; from 15.7% to 12.2% for the "interest in ideas" model; from 25.2% to 22.3% for the "relating ideas" model; and from 20.1% to 9% for the "seeking meaning" model. The only model in which there is a relatively slight increase is the one representing the "use of evidence" subscale with 0.6% (from 18.7% to 19.3%).

In terms of Adjusted R^2, there is decrease for all the deep scales[43]. More specifically, it is observed that there is a decrease between 1.7% and 15.5% amongst the "deep" models of the "high prior knowledge" group.

These findings indicate that the students' interactions in AM with regards to the deep approaches towards studying do not seem to be influenced greatly by the "high prior knowledge" group. Whether a student has a low or high score on the deep scales, the predictors are not able to explain as much for the students belonging to the "high prior knowledge" group as they do for the whole sample and especially for the students belonging to the "low prior knowledge" group (with the exception of the "use of evidence" model as indicated above). In the following sections, there is an attempt to find possible reasons behind these findings.

a) Deep Interest in Ideas

According to Tobias (1994) there is empirical evidence supporting that "*the effects of interest account for less variance than those attributable to prior knowledge"*. The empirical evidence comes from studies in different contexts; however, it can be a starting point from which it can be assumed that the involvement of prior knowledge as a selector may increase variance.

---

[43] If we were to continue further development of the model by excluding the most insignificant predictor, then the Adjusted R^2 would be increased. However, further development of the low/high prior knowledge group models is not intended. The intention here is to simply get some indication at model-level as to which groups are most influenced by the prior knowledge.

In the context of the specific study, the variance explained by the "interest in ideas" model is increased substantially, but only in the "low prior knowledge" group. The "interest in ideas" model (see Table 5.2.3.1 in section 5.2.3) seems to represent better those with a low level of prior knowledge compared to those with a high level of prior knowledge. This can be explained by the unexpected finding with regards to the most contributing predictor *number of exercises finished but not solved*; according to which the higher the score on the subscale, the higher the number of unsolved exercises (see section 4.7.3.6). Therefore, it seems that the direction of the relationship in the model between the specific predictor and the "interest of ideas" scale expresses better the interactions of the "low prior knowledge group" rather than those of the "high prior knowledge" group across the specific scale.

b) Deep relating ideas

As indicated earlier, the variance explained by the "relating ideas" model is increased substantially, but only in the "low prior knowledge" group. The "relating ideas" model (see Table 5.2.3.1 in section 5.2.3) seems to represent better the interactions of those with a low level of prior knowledge rather than those with a high level of prior knowledge. This can be explained by the unexpected finding with regards to the second most contributing predictor, *number of exercises finished but not solved*; according to this, the higher the score on the subscale, the higher the number of unsolved exercises (see section 4.9.3.6). Another reason can be the unexpected finding with regards to the predictor *number of hyperlinks (concept links) visited in reading and exercise pages*, which indicates a negative relationship to the subscale instead of a positive one (see section 4.9.3.6). One would expect that a positive relationship might be also more relevant to the interactions of those students in the "high prior knowledge" group, as the increasing use of hyperlinks would show perhaps a tendency to relate old and new concepts. However, this is not the case, which points to a potential issue with regards to the way the hyperlinks are designed in AM; an issue which is discussed further in Chapter 6. Therefore, it seems that the direction of the relationships in the model between the aforementioned predictors and the "relating ideas" subscale seems to express the interactions of the "low prior knowledge" group better than those of the "high prior knowledge" group across the specific scale.

c) Deep seeking meaning

As indicated earlier, the variance explained by the "seeking meaning" model is increased substantially, but only in the "low prior knowledge" group. The "seeking meaning" model (see Table 5.2.3.1 in section 5.2.3) seems to represent better the interactions of those with a low level of prior knowledge rather than a high level of prior knowledge. This can be explained by the unexpected finding with regards to the predictor *number of exercises solved on third try*, according to which the higher the score on the subscale, the higher the number of exercises solved on the third try (see section 4.8.3.6). An indication that this can be the case is the unexpected finding with regards to the predictor *number of hyperlinks (concept links) visited in reading and exercise pages*, which indicates a negative relationship to the subscale instead of a positive one (see section 4.8.3.6). One would expect that a positive relationship might be also more relevant to the interactions of those students in the "high prior knowledge" group, as the increasing use of hyperlinks would show perhaps a tendency to relate old and new concepts. However, this is not the case, which points to a potential issue with regards to the way the hyperlinks are designed in AM; an issue which is discussed further in Chapter 6. Therefore, it seems that the direction of the relationships in the model between certain predictors and the "seeking meaning" subscale seems to express better the interactions of the "low prior knowledge" group, than those of the "high prior knowledge" group across the specific scale.

d) Deep use of evidence

Amongst the models of the three groups the amount of variance explained ranges between 18.7% and 21.3% (see Appendix 5.2), so it is reasonable to say that it does not really differ as greatly as it does for all the other deep models. The slightly more increased variance explained in the model of the "low prior knowledge" group, compared to the one of the "high prior knowledge" group, can be explained by the unexpected finding with regards to the predictor *number of exercises solved on third try*; according to which the higher the score on the subscale the higher the number of exercises solved on the third try (see section 4.10.3.6).

Going back to the variance explained amongst the four models, it seems that the "use of evidence" model is the only deep model in which prior knowledge seem to make relatively little difference in terms of the variance explained by the model. So, with regards to this subscale, the issue is more about why the model overall does not explain more variance, which has been discussed previously in section 5.2.2.

To conclude, certain unexpected relationships found in the deep models can be behind the differences between the variance explained amongst the models of the group. Furthermore, as previously discussed in section 5.1.2, the contribution of certain predictors with distinguishing aspects for the deep scales, such as *number of exercises solved on first try,* could be responsible for the unexplained variance of the model for the whole sample. It is possible that they could be also responsible for this difference in variance between the models of the "low prior knowledge" and "high prior knowledge" groups. It can be the case that the specific predictor, for example, does not have strong enough contribution to represent better the interactions of the "high prior knowledge" group across the scales. Or perhaps it is not possible for this predictor to contribute more to the deep models of the "high prior knowledge" group (i.e. the number of exercises solved on first try by students with a high level of prior knowledge could be simply independent of their high and low scores on the deep scales).

5.5.4 Surface models in low and high prior knowledge groups

In this section, there is an attempt to examine and compare the variance explained by the "surface" models with regards to the low and high prior knowledge groups.

5.5.4.1 Surface scales in the "low prior knowledge" group

In the "low prior knowledge" group, the variance R^2 increases for 3 of the "surface" models, compared to the variance of the "surface" models for the whole sample (see Appendix 5.2). More specifically there is an increase: from 45.5% to 55.8% for the "surface" model, from 21% to 24.4% for the "lack of purpose" model and from 21.6% to 33.3% for the "syllabus boundness" model. On the other hand, there is a decrease: from 41.6% to 33.2% for the "fear of failure" model, and from 40% to 37.5% for the "unrelated memorising" model.

The values of Adjusted R^2 follow the same patterns to those of R^2, with the exception of the "lack of purpose" model[44] where there is a slight decrease from

[44] If we were to continue further development of the model by excluding the most insignificant predictor, then the Adjusted R^2 would be increased. However, further development of the low/high prior knowledge group models is not intended. The intention here is to simply get some indication at model-level as to which groups are most influenced by prior knowledge.

18% to 17% (see Appendix 5.2). More specifically, it is observed that there is an increase between 5.4% and 6.2% for the "surface" model and the "syllabus boundness" model. In addition, there is a decrease between 12.7% and 1% for the "fear of failure" model, the "unrelated memorising" model, and the "lack of purpose" model.

The findings indicate that the students' interactions in AM with regards to some of the surface approaches towards studying (i.e. "surface", "lack of purpose", and "syllabus boundness") seem to be influenced by the level of prior knowledge, specifically in the "low prior knowledge group". In the models representing these surface scales, the predictors are able to explain more for students belonging to the "low prior knowledge" group. This finding can have implications as it can help tutors in class and also inform the design of an ILE regarding tackling the "low prior knowledge" groups of students with a tendency towards a "high" surface approach towards studying.

On the other hand, the predictors of the suggested models for "fear of failure" and "unrelated memorising" do not seem to explain as much, if a student with a low or high scores on these two scales belongs to the "low prior knowledge" group; an issue which is discussed further below.

5.5.4.2 Surface scales in the "high prior knowledge" group

In the "high prior knowledge" group, the variance R^2 decreases for four of the "surface" models, compared to the variance of the "surface" models for the whole sample (see Appendix 5.2). More specifically it is observed that there is a decrease: from 45.5% to 40.7% for the "surface" model, from 40% to 38.9% for the "unrelated memorising" model, from 21% to 13.1% for the "lack of purpose" model, and from 21.6% to 19% for "syllabus boundness". The only model in which there is an increase is the one representing the "fear of failure" subscale with 2.9% (from 41.6% to 44.5%). The values of Adjusted R^2 follow the same patterns to those of R^2.

It is reasonable to say, therefore, that there is not really a clear pattern with regards to the variance explained by the "surface" models amongst the groups (as there is in the "deep" models, where there is more variance explained for the "low prior knowledge" group). In the following sections, there is further discussion regarding each "surface" model, which may shed more light as to why some "surface" models are explained better in the "low prior knowledge" group and some are explained better in the "high prior knowledge" group.

a) Surface fear of failure

The model representing the "fear of failure" subscale, which relates to anxiety during studying, is clearly a model which behaves differently compared to the rest amongst the groups (i.e. the variance explained is increased in the model for the "high prior knowledge" group compared to that of the whole sample and the "low prior knowledge" group). The "fear of failure" model (see table 5.1.3.1 in section 5.1.3) represents better the interactions of those with a high level of prior knowledge across the scores of the subscale. As indicated in 5.5.4.2, the predictors explain more variance in the "fear of failure" model for the "high prior knowledge" group compared to the "low prior knowledge" group. It seems, for example, that the tendency of those with the higher scores on the "fear of failure" subscale to solve more exercises on third try and fewer exercises on first try compared to those with lower scores, is stronger in the "high prior knowledge" group. This is also reinforced by existing literature with regards to the effects of anxiety on prior knowledge. It can be that "it is difficult for the high anxious individual to generate a wide range of response alternatives and to evaluate their appropriateness in relation to his own prior knowledge" (Tobias, 1994). Tobias (1994) also suggests that this can even happen when the response alternatives are presented as multiple-choice problems, (which is the case in AM as most exercises are presented in a multiple-choice format. So, it seems that, even if students have a high level of prior knowledge in the subject of mathematics, the higher the degree of their anxiety while solving exercises during their tutorials, the more likely they are to solve exercises on third, rather on the first try.

b) Surface unrelated memorising

The model representing the "unrelated memorising" subscale (see table 5.1.3.1 in section 5.1.3), is the only one amongst the "surface" models in which the predictors explain better the variance R^2 for the whole sample with 40% (see Appendix 5.2). The findings also indicate that amongst the two groups, the predictors explain a bit better the variance of the "high prior knowledge" group with 38.9% compared to that of the "low prior knowledge" groups with 37.5%. However, overall, it is reasonable to say that there is no really great difference amongst the variances explained, and the predictors seem to explain relatively well[45] the students' interactions with regards to the "unrelated memorising"

[45] It is reasonable at the same time to say that there is always room for improvement, as indicated previously in 5.1.2.

approach independently of students' level of prior knowledge. Furthermore, there is literature to support that high level of prior knowledge does not seem to make a great difference, when there is a high degree of rote learning involved. It is suggested that rote learning does not help to establish proper relationships between new knowledge and prior knowledge for meaningful learning to occur (Terry and Holim, 2008). So, in the context of the current investigation, even if students have a high level of prior knowledge in the subject of mathematics, they are not likely to use it if there is a tendency for rote memorisation. Hence relationships in the model, for example, such as the negative one between the "unrelated memorising" subscale and *number of exercises solved on first try*, can manifest as strong, if not a bit stronger in the model for the "high prior knowledge" group, compared to the "low prior knowledge" group.

c) Surface syllabus boundness

The model representing the "syllabus boundness" subscale (see table 5.1.3.1 in section 5.1.3), explains best the variance R^2 in the "low prior knowledge" group with 33.3%. The variance explained R^2 regarding the model of the "high prior knowledge" group with 19% is relatively close to the one of the whole sample at 21.6%. The "syllabus boundness" model better represents overall the interactions of those with a low level of prior knowledge across the scores of the subscale. It seems, for example, that the more there is a tendency to rely exclusively on the given learning material, the less likely it is to solve exercises on first try; and this manifestation becomes even stronger in the "low prior knowledge" group.

Regarding the relatively low variance explained by the models for the "high prior knowledge" group and the whole sample, it is discussed previously in section 5.1.2 how the level of difficulty of the learning material may be the reason behind this. It is possible that if the level of difficulty of exercises or tasks during the tutorial session required students further to extend their knowledge and perform research beyond the given syllabus, then the differences in interactions in AM across the scores on the subscale would be stronger and the data more varied.

d) Surface scale lack of purpose across all groups

The model representing the "lack of purpose" subscale best explains the variance R^2 in the "low prior knowledge" group with 24.4% (which is 3.4% more than the whole sample) while for the "high prior knowledge" group there is a relatively lower amount of variance explained at 13%. The "lack of purpose" model (see table 5.1.3.1 in section 5.1.3) better represents overall the interactions of those with a low level of prior knowledge across the scores of the subscale. It seems, for example, that the more there is tendency to cope minimally with the requirements of the course and the more there is lack of interest, the less likely it is for a student to solve exercises on first try and more likely to not solve exercises at all; and this manifestation becomes even stronger in the “low prior knowledge” group.

Regarding the relatively low variance explained for the models for the “high prior knowledge” group and the whole sample, it is discussed previously in section 5.1.2 how again the level of difficulty of the learning material may be the reason behind this. It can be that the practical tasks during the tutorial sessions are not challenging enough to distinguish the interactions of those who have lack of interest and tend to cope minimally from those who do not, especially if they belong to a “high prior knowledge” group. It is possible, for example, that the *number of exercises solved on first try* for those students with a high level of prior knowledge is independent of their scores on the subscale[46].

To conclude, throughout section 5.5, there is an effort to examine and explore the influence of prior knowledge across all models of the ASSIST scales, in alignment with investigating the influence of the non-style factor of prior knowledge on the relationship between deep and surface approaches towards studying and “interaction” metrics. Enriching findings did occur, especially with regards to the deep models of the “low prior knowledge” group where there is a quite consistent pattern, but the analysis also indicated some useful insights with regards to the “surface” models and prior knowledge, despite the fact that there was not an overall pattern. These contributions are discussed further in section 6.3.4. Furthermore, possible explanations have been given for these findings; however, it is reasonable to say that the current investigation provides just a good starting point for future studies, where prior knowledge will have a more central role. Further examination is required, with further analysis of the models at predictor-level, where the issue of the influence of prior knowledge will be addressed as a primary question, and will not just have a complementary,

[46] However, this situation can only last for a relatively short time. Tobias (1994) suggests that those with a low interest in the subject and high prior knowledge will eventually end up having low interest in the subject and a low level of knowledge, eventually affecting their performance.

secondary, value as in the current investigation.

Finally, in Chapter 5, comparisons amongst the models and predictors allow us to see that overall the surface models have more distinguishing aspects, while the deep models seem to be explained better when prior knowledge is involved. In the next chapter, it will be shown how the insights of this chapter serve towards helping tutors to identify approaches towards studying during students' interaction in an ILE in order to facilitate intervention and providing recommendations with regards to future improvements in data capture and an ILE's interface design and features.

Chapter 6 - Recommendations and Contributions

This chapter discusses the contribution of the thesis. Section 6.1 looks at methodological contributions both in terms of the statistical analysis and the challenges which can occur when setting up a similar study. Section 6.2 discusses improvements with regards to the ILE features and the predictors of the "deep" and "surface" models in order to achieve a more enriching interpretation of results in future studies. Finally, 6.3 discusses the pedagogical implications of the findings.

6.1 Methodological reflections

This section highlights the expectations regarding the variance of "deep" and "surface" models and the predictors worth considering in future studies. First, there are some practical recommendations regarding the setting up of a similar study.

6.1.1 Dealing with the challenges of setting up a study in real conditions

A key concern underling this thesis's study was its "ecological" validity in order to examine the way students engage with learning material in the same way they would in a real learning situation. As mentioned in 3.7, in order to avoid biases that would challenge the generality of the results to the real world and particularly what Robson (2002) refers to as "demand characteristics", the current study was conducted in real conditions. However, the process of setting up a study which takes place during the tutorial sessions is not without challenges. This is because of what Robson (2002) calls "mundane realism", referring to real life events which are complex and can cause ambiguities. As the author dealt with a number of expected and unexpected complexities due to real conditions, this section can help researchers who run studies in similar contexts to foresee these complexities and plan ahead, or provide awareness at least regarding how they can influence their data.

6.1.1.1 Handling the implications of ethical issues in real teaching-learning conditions

Ethical issues may cause limitations and challenges when a researcher is planning a similar study in real teaching-learning conditions. First of all, it has to be ensured that: none of the students is treated differently (i.e. is given different learning materials); the study is conducted in parallel with the tutorial sessions and is integrated in the whole module schedule in a non-disruptive way; the learning material integrated in the ILE still serves the learning outcomes and aims of module and course; and that the design of the ILE benefits students' learning and does not hinder their learning. So, ultimately it has to be ensured that there is no negative impact on their academic performance due to changes introduced by the study into the learning and teaching process of the module.

In the current investigation, this meant that the author, as the instigator and organiser of the study, had to discuss and agree with the module team on a number of points:

- which chapters from the module's textbook would be integrated (e.g. the module team judged that it would be beneficial for students to include the chapter of Functions-Graphs where graph plotters for functions could be integrated).
- the integration of the study in the module schedule and, specifically, when and how tutors and students would register and familiarise themselves with AM, so the introduction of AM would not disrupt their learning (as discussed in 3.10.3). This is also something which is supported in the general context of educational technology: letting students get familiar with the mathematical system is a reasonable step when going through the process of integrating it in class (Maat, 2010).
- the way the learning material was structured and designed in the ILE. This required, for example, the distribution of detailed storyboards to ensure that there was a common agreement with regards to the structure and the features included in AM. It was particularly important that the structure of topics and subtopics in AM was according to the one given on the textbook which reflected the syllabus and the learning outcomes of the module. In addition, it was important for the module team that there would be features which would potentially promote further active reading, exploration, and engagement with the learning material, such as: hyperlinks (concept links), search and advanced search features, and a graph plotter.
- the way exercises would work. This again required detailed storyboards which indicated: how the exercises included in the textbook would be designed in AM (i.e. fill-in-in the blank, multiple choice, etc.); how many times the students would be allowed to try an exercise until the solution was given by the system; and what type of feedback would be given for each try.

6.1.1.2 Designing an ILE to serve both theoretical assumptions and ethical aspects

There is a fine balance between designing an ILE to serve the theoretical assumptions (i.e. made for the approaches towards studying and the metrics representing the students' interactions), and designing an ILE to serve the above ethical and practical considerations. The design of the ILE should ideally encompass both, and this is certainly possible when both theoretical assumptions made for the study and ethical and practical considerations can coincide. For example, in the current investigation there was certainly an effort to design AM based on theoretical assumptions which linked the deep approaches to use of AM features (e.g. search feature for further exploration of concepts); but this in a way served the ethical aspects as well, as there was an effort to design an ILE which benefits students' deep understanding in concepts. However, it can be the case that respecting the aforementioned ethical aspects means that sometimes it is not possible to make changes that can better serve the theoretical assumptions of the study. In the current investigation, there were such cases:

- Learning material integrated in AM certainly served the learning outcomes of the module, as it did not differ from the learning material of the approved textbook; however, it did not seem to stimulate elements of "high deep" learning such as exploring further concepts, building up understanding based on students' own conclusions, and trying to relate concepts. This was possibly because a lot of the exercises were repetitive, and most exercises were a direct application of the generic formulas and the working examples given, and in very few of them the students had to draw their own conclusions, explore further concepts, or relate different concepts. These conclusions resonate with the suggestions of researchers, such as Sangwin (2004) mentioned in 2.7.3, about designing ILE exercises which are not just direct application of previous worked solutions, but encourage deep learning (e.g. by requiring the generation of examples that satisfy certain criteria, or by trying to provide another answer after having taken tailored feedback into account). Here it is worth noting that the provision of feedback also warrants further investigation as, for example, based on an empirical review by Beaten et al. (2010), it seems that there are no concrete findings as to the relationship between how students use feedback and deep and surface approaches.

- AM eye-tracker which was deemed too disruptive to the learning process was omitted from AM. Initially, its inclusion was considered because it would ensure that view time on AM exercises and reading pages did not include lack of engagement or inactivity and it would give further accuracy in capturing students' interactions.

6.1.1.3 Selecting a big-size module to "afford the potential reductions"

It is commonly acknowledged that the bigger the sample size, the greater the statistical power (Meyers, 2006), so it makes sense to alert researchers who intend to conduct studies in similar context as to how (and to what degree) real teaching-learning conditions can have an impact on the sample-size. More specifically, when setting up a similar study in real conditions, it is possible that the initial sample size, based on the module's size, can be reduced because of students' withdrawals from course, students' lack of attendance, students' lack of interaction with ILE, or not completing ASSIST.

In the current investigation, there were 276 students initially; however, there were reductions as shown in 3.10.3, resulting in a sample size of 115 students

Reduction of the originally estimated sample-size due to events which naturally occur during a university course, as well as during a study in real settings where participants do not necessarily behave as expected, can have an impact on the statistical power of the study's results. In the current investigation, despite the significant reduction (which was 59% of the initial estimated module size), it was still possible further to exclude up to 6 outliers and sustain the typical threshold for statistical power with models (which included up to 8 predictors in their first versions).

Although this reduction in sample size can differ across educational settings, it would be advisable for a researcher ideally to choose a big size module (between 250-300 students) to counteract potential big reductions which can naturally occur when a study in this context takes place in real conditions.

6.1.1.4 Dealing with mundane realism due to unexpected events

In real conditions, students might not use an ILE as expected, or generally behave in the study as expected, causing complexities and ambiguities (i.e. "mundane realism"). More specifically, the following cases were observed:

a) There were students who double-registered in AM (even if there were detailed written and verbal instructions for registration and support during the tutorial sessions by moderator and tutor).
b) There were students who used each other's accounts to interact with AM.
c) There were students who did not use AM at all or at least not throughout during the sessions (e.g. lack of attendance, preference for textbook, leaving tutorial earlier).

These unexpected events may cause complexities during the study and ambiguities in the data, but in most cases, it is possible to deal with them. More specifically:

- With regards to point b), there should be constant monitoring so the moderator of the study can intervene and instruct students to use their own account, and possibly sort out any password and username issues.
- With regards to point c), these cases should be observed and recorded because it can serve the "cleaning of data". This means that, at a later stage, a decision should be made as to whether to exclude cases from the sample. For example, in the current investigation, exclusions took place when a student never visited exercise pages, and there were NULL values in the temporal data for these pages. There was also exclusion of data coming from sessions which were disrupted because of technical reasons. Furthermore, even if the aforementioned cases are not excluded from the sample at the start, it is good to record the reasons for irregular AM use, because it may explain why they are flagged up by the statistical processes as multivariate outliers[47].
- With regards to point a), there is a need for constant monitoring and updating of the registration list with the students' details, as well as ensuring that the sessions with NULL usage, due to double-registration and duplicated accounts, are deleted.

6.1.1.5 Impact of mundane realism as occurred from expected events

There can be expected events, which may add another layer on students' interaction with an ILE and influence their approaches towards studying during the tutorial sessions. More specifically:

[47] We have, for example, the case of a student who left both tutorials sessions quite early for no obvious reason, except perhaps for "lack of interest". This case (case 25) was flagged up as an outlier later on during the development of models possibly because the surface score was not high enough to justify the lack of interaction in AM.

- In big size modules, such as the current one, typically the tutorials are spread out throughout the week after the lecture. So, differences in temporal distance between lecture and tutorial sessions may affect students' interaction with an ILE.
- Tutor intervention can also influence interaction in an ILE. In a big size module, such as the current one, there are usually a number of tutors involved and, unless there are very specific instructions on how to instruct and guide the students during the tutorial sessions, there are different tutoring styles. It is beyond the scope of the current investigation to identify and examine the way tutors instruct and support the students during the tutorial sessions[48]. However, it has to be acknowledged and broadly discussed how differences between the ways tutors instruct and support students in class may have influenced students' interactions in ILE[49]. Here are some examples, based on observations made in class:
    - *Some tutors at the start of the sessions would revise the mathematical concepts (from the lecture) for the whole class by using AM or the board. This revision amongst tutors would also differ in terms of time. So, this type of intervention, for example, can affect students' interaction regarding the time spent on theoretical pages and examples in AM; or even their intention to research a mathematical concept further (they may perceive, for example, that what tutor revises is what they basically need to know for solving the exercises of the specific tutorial session, and they do not need to explore anything further).*
    - *Some tutors would give specific instructions on how to start (i.e. starting from revising a specific mathematical concept, or starting from a specific exercise before moving on to another). So, this intervention, for example, can affect the order in which they view theoretical and exercise pages in AM.*
    - *Some tutors would instruct students to do specific exercises in a given amount of time (so they can progress further to more advanced*

[48] This would be a study on its own right, as there is existing literature dedicated on the way tutors instruct and support students in class (see: De Grave et al. (1999), and Turan et al. (2009)).

[49] The author does not aim here to criticise in any way the different tutoring styles (as all tutors indicated an exceptional degree of professionalism in adapting their intervention according to the needs of the class), but merely to give a few broad examples how differences in intervention may in general influence students' interactions in an ILE in a real learning conditions.

exercises within the time-frame of the tutorial session). This can affect again the time spent on exercise pages (or even push some students to solve exercises on first try or other students to finish exercises without really solving them and getting the answers from the system).

In the current investigation, the impact of these expected events could not really be eliminated in order to prevent the potential ambiguity caused in the statistical results. Regarding the first point, for practical reasons it was not possible to schedule all tutorials of a big size module on the same day. Regarding the second point, for ethical reasons, it was not really possible for the author to stop tutors from intervening or apply pressure on them in order to change their intervention (given that as specialists in the subject they are supposed to know what is best for students' learning). However, the impact of such expected events should be acknowledged as they can be responsible for unexplained variance in the regression models; and in a future study one could investigate further the impact of tutoring styles/profiles on students' interactions in an ILE, as despite the existing literature on tutor profiling (Turan et al., 2009; De Grave et al., 1999), there seems to be a gap in bringing together the areas of tutoring style, approaches to studying and interaction in ILEs.

6.1.2 Researcher as a collaborator and mediator between two teams

The author was the instigator and organiser of the current investigation, so the first challenge, and since there was no funding involved, was to demonstrate how both teams could benefit from the current investigation (i.e. for the module team an opportunity to distribute the learning material in a more engaging way which might benefit students; and for the AM/Action analysis team an opportunity to test AM with a big size "maths" module and get valuable data for action analysis). Once both teams were on board, another challenge for the author was ensuring a smooth collaboration, coordination and mediation between the two teams (which were located in different institutions, different countries, and were never in contact with each other).

These challenges produce valuable experiences on how a researcher can run such an operation with two teams smoothly, ensuring at the same time that they both get what was initially agreed without creating any false expectations.

6.1.2.1 Collaborating with the module team

First of all, to achieve a good collaboration, it is important to find a module team which is willing to incorporate an ILE in their module and work with it. Collis and

Moonen (2001) discuss that when integrating an interactive learning environment in a course it important that there is a certain vision about technology within the institution and that there is a readiness and willingness to change in this direction. Within the specific university, and specifically the computer science department, there was a vision for integrating technology in the teaching and learning processes. Furthermore, although the “maths” module team was positive regarding the integration of AM in the tutorial sessions, it was essential to demonstrate how the results of the current research could benefit the students taking this core “maths” module. At the time of the study, the module team was particularly worried about the discrepancy regarding the level of prior knowledge of the students (some students had A-levels in maths, and some did not) and how they can bridge these differences in class. So, an ILE which had the potential to help connecting basic concepts more effectively and also new, more advanced, concepts (e.g. through hyperlinks and search features) seemed an appealing solution. An integrated search feature, for example, would allow students with high prior knowledge to explore further advanced mathematical concepts. In addition, it was appealing that the integration of chapters such as the one of Functions and Graphs in AM would potentially engage students further, with the use of graph plotters, and also aid their understanding of functions, compared to the textbook.

Furthermore, to ensure a smooth collaboration and communication, it was deemed essential to ensure clarity and agreement on any given decisions and solutions with regards to the AM design; and also to ensure that all tutors involved were updated with all the changes regarding the module schedule and were familiar and comfortable with the use of AM before going into the classroom.

Finally, such a collaboration can create further obligations. There was also an agreement for the students to keep using AM for the following year, only, to ensure consistency in terms of the delivery of the module.

To conclude, for a smooth collaboration with the module team behind the learning material of an ILE, it is important to:

- identify any existing “teaching-learning” concerns in the module and examine how they can be addressed in the design of the ILE.
- try to accommodate in the process of the study and the design of the ILE what the module team thinks can benefit the students.

- distribute detailed storyboards about the design of the ILE prior to the implementation and seek approval (a similar procedure to a corporate “sign-off” in the design process of a digital product). This is also an important step, because it ensures a smoother collaboration with the team responsible for the implementation of the design and provides more clarity and confidence in the whole process for both teams.
- distribute details on the way the study will be conducted, along with all the materials used (e.g. guidelines for registration and the manual of the ILE), and show clearly how it will be incorporated in the module schedule in unobtrusive way.
- ensure that all tutors are familiar and comfortable with the use of the ILE and help them to find ways to incorporate it in their teaching in class.
- ensure that if further demands are made for further use of the ILE beyond the study, they can be met without creating any false expectations.

6.1.2.2 Collaborating with the AM/Action Analysis team

Clear and ongoing communication and negotiation were also crucial elements essential for a smooth collaboration with the AM and Action Analysis team involved in the current study. There were two phases in that collaboration.

The first phase was with regards to adapting AM in order to incorporate the learning material from the module’s textbook. This required an ongoing quality control process, with updates and corrections to ensure that the learning material was presented appropriately according to the requirements of the module team.

The second phase was with regards to the action analysis. At this phase, it was essential to clarify with the team whether metrics suggested by the author were possible to calculate, and whether and how AM can be adapted to ensure that the raw data required for the suggested metrics is collected. For example, with regards to the search feature, its placement in AM was made more prominent to ensure its use, and it was clarified which metrics could be calculated or not. Furthermore, it was also essential to communicate to the team the module schedule, the duration of the tutorial sessions, and unexpected events such as server interruptions and double-registrations to help with data cleaning. Finally, it was essential to clarify the limitations of such a collaboration so there are no false expectations on either side. For example, in terms of the data received, it was clarified that due to ethical reasons the author could not provide the team with data regarding students' final grades in the "maths" module; on the other hand the AM/Action Analysis team could not commit to a second phase where the metrics and AM would be revised (based on the current statistical results) and the study would be repeated.

To conclude, for a smooth collaboration with the team responsible for the implementation of the ILE and the action analysis, it is important to:

- identify with the team from the beginning what is possible to adapt or design and implement in terms of ILE features to serve both requirements set by module team and metrics required for the study
- distribute detailed storyboards about the design of the ILE prior to the implementation and ensure an ongoing quality control process for any potential corrections and updates on the learning material
- identify with the team which of the suggested metrics can be calculated
- communicate detailed schedules for the module and the study
- inform the team of any unexpected events, occurring during the study, which may influence the quality of data, so the team can perform data-cleaning
- be clear about what the expectations and limits of the collaboration are on both sides

6.1.3 Data analysis reflections

6.1.3.1 Decisions on the initial inclusion of predictors in models

The inclusion of predictors in the first version of the models is based on the theory surrounding the approaches towards studying and the correlations that occurred between the "deep" and "surface" scales and the "interaction" metrics.

As shown in Chapter 4, more correlations occurred for the "surface" scales compared to the "deep" scales, so while for the inclusion of predictors in the "surface" models there was more reliance on the empirical results of the correlations, for the "deep" models there was more reliance on the theoretical assumptions. For the "deep" models, the decision was based on which metrics had the strongest indications and the most potential of giving a more enriching[50]

[50] By "enriching", we mean that it offers an additional dimension regarding the students' interaction in an ILE during the tutorials.

and complete picture of students' interactions in AM with regards to a specific approach to studying. For example, this was the case of the performance-related metrics which have been included as a combined group in every model, because: a) they were strongly relevant to the main task of practising the exercises during the tutorials and b) there was the potential for meaningful comparisons if they were included in a combined way[51]. However, there were cases in which a specific aspect in terms of students' interaction was deemed enriching based on the theory, but there were no strong indications as to which specific metric to include (e.g. the inclusion of temporal metrics in the "relate ideas" model as shown in section 4.9.3.1). In such cases, a "trial" type of version of a deep model (i.e. what the author called "pre-model") was used to indicate which predictor contributes best in the model; hence a choice was made based on this finding. Overall, these tactics helped in deciding the best possible choices for the "deep" models; and in the cases of the "deep" scale, the "relating ideas" subscale, and the "seeking meaning" subscale, helped in getting some useful insights and certain distinguishing aspects with regards to the deep approaches to studying.

Based on the strategy and procedures implemented for the initial inclusion of predictors in the models, there are certain conclusions which can be of value in similar future studies. The decision with regards to which predictors to include in a regression model is a straightforward affair, if there are correlations to indicate which metrics are more likely to explain and contribute to the variance of the models. However, if there are no such empirical indications, and there are no prior studies in a similar context to give stronger theoretical indications, the process of inclusion requires different and relatively more complex tactics. First of all, it is suggested to take advantage of the maximum number of predictors which are allowed in the first version of the models according to the sample size, as this will allow comparing the contributions of as many predictors as possible. Secondly, when relying on theoretical assumptions for the inclusion of predictors, it is worth concentrating on those with the strongest theoretical connections to the scale, with the most relevance to the students' practical tasks, and which facilitate in making meaningful comparisons. Thirdly, if certain metrics are considered enriching, but there is no clear indication as to how

[51] This was especially the case for the performance-related in "deep" models because of their unexpected relationships to the "deep" scales. Their combined inclusion extended our understanding on how students perform when practicing tutorial exercises in class with regards to the deep approaches towards studying in the specific educational setting.

exactly they relate to the scale or which ones from a specific group is likely to best contributes to a model (e.g. temporal metrics), then it is worth checking their contribution to the variance by including them in “pre-models”.

6.1.3.2 Thresholds for statistical measures in this context

The thresholds for statistical measures are based on expectations which occur from the literature on learning styles and interaction in ILEs (i.e. there is an expectation for a medium effect size as indicated in section 2.6.4), and they are also based on Cohen’s (1992) suggestions for variance explained R^2 based on the effect size (see Appendix 3.11.1). However, since this study can serve as a starting point for studies of a similar context (where approaches to studying are examined in relation to students’ interaction in ILEs and specifically during tutorial sessions and in real learning conditions), it is reasonable to suggest what thresholds one may expect in terms of the variance explained R^2.

In terms of R^2, there is a difference in terms of expectations between the deep and surface approaches to studying. In most “surface” models, there seems to be approximately double the amount of variance explained compared to the “deep” models (see Appendix 5.1). More specifically, due to reasons discussed in section 5.3.2, it seems that the “deep” models are more sensitive to factors in the teaching and learning environment. Hence, regarding the variance explained R^2, the expectations for a similar study (for a similar study in similar educational settings, ILE, and metrics) are medium between 15.7% and 25.2% (see Appendix 5.1); whereas, for the “surface” model, one can expect in a similar study from medium to large variance explained an R^2 between 21% and 45.5%. Researchers should consider these thresholds, but always in relation to the educational settings in which the study takes place. It is possible, for example, as discussed in section 6.3.1, that in educational settings where a deep approach is encouraged more –or is expected and rewarded according to students’ perceptions- deep approaches may manifest themselves more strongly, hence the amount of variance explained may reach large levels.

6.1.3.3 Importance of the exclusion of multivariate outliers

Some insights gained from the process of the exclusion of outliers as part of development of the models, and which can be useful for future studies, are:

- The exclusion of outliers can have great impact during the development of the models, as it can double the amount of variance explained by the model or increase it in certain cases from small to medium (e.g. the “interest in ideas” model).

- Particular consideration should be given as to how it affects the sample size, and there should be a strategy to ensure that the typical thresholds for statistical power, sample size and number of predictors are always respected. Through the current development process, it is found, for example, that up to 6 “extreme” outliers are a reasonable number to exclude, as this ensures the highest increase in terms of variance, without at the same time compromising the statistical power according to the thresholds set by Cohen (1992) with regards to sample size and number of the predictors included in models.

6.1.4 Recommendations for surviving and non-surviving predictors

6.1.4.1 Surviving predictors

Based on what is discussed in sections 5.1.3 and 5.2.3, there are metrics which serve the models well and can be good recommendations for future studies.

a) Number of exercises solved on first try

This metric has proved to be a particularly important predictor and there are a number of reasons to recommend it in future studies.

First of all, it helps making meaningful comparisons to the rest of the performance-related metrics, and offers a distinguishing aspect between “deep” and “surface” models, which is important if there is a need to flag up a surface approach in an ILE or to inform the tutor about it. More specifically, as discussed in sections 5.1.3 and 5.2.3, the specific metric seems to be particularly valuable for the “deep” models as it seems to be amongst the few predictors to give a distinguishing aspect with regards to the deep approach and which has a consistently negative relationship with the “surface” subscales and the main “surface” scale.

Secondly, as shown in section 5.3.6, it survives in the leanest and meanest models of most “surface” models (see also Appendix 5.3).

Thirdly, as indicated in 5.3.1, it is the predictor which contributes the most to the models of “fear of failure”, “unrelated memorising” and the main “surface” scale.

Overall, these findings reinforce the initial theoretical assumptions made in sections 4.1.1-4.1.10, that the *number of exercises solved on first try* can be a metric which can indicate students’ performance when solving exercises during the tutorials with regards to surface and deep approach towards studying.

Based on these findings, there are also certain recommendations which can be made for future studies in similar context:

- It is reasonable to include it in “unrelated memorising”, “fear of failure”, and main “surface” models. It is worth re-examining its contribution even in the “lack of purpose” and “syllabus boundness’” models, as it can offer a more meaningful comparison to the performance-related metrics which survive in their leanest and meanest versions (see Appendix 5.3).
- It is reasonable to include it in “seeking meaning” and “relating ideas” models.
- It is worth re-examining its contribution to the main “deep” model, since its remaining predictors in the leanest and meanest version cannot really offer a distinguishing aspect with regards to a deep approach (see Appendix 5.3).
- It is worth considering its inclusion in the “interest in ideas” and “use of evidence” models, especially if the researcher judges that the deep approach can manifest more strongly in the educational setting in which the study takes place. This could be because the ILE better supports a deep approach towards studying (see the “design” recommendations in section 6.2.1), or because the whole university course of which the ILE is part has been designed in such a way that a deep approach is encouraged.

b) Compactness

This metric has proved to be particularly important for certain “surface” models.

First, it fits well with the theoretical assumptions (see sections 4.1.1, 4.2.1 and 4.4.1) that the higher the students score on certain “surface” scales, the more compact the path they undertake (that is the more closely they interact around a certain set of pages).

Secondly, it helps enriching the “unrelated memorising”, “syllabus boundness”, and main “surface” models by offering an additional dimension regarding the students’ path characteristics when interacting with a learning environment during tutorials. It has a consistently positive relationship with the aforementioned surface scales (see Appendix 5.3), and along with the rest of the predictors in these models, it can help to flag up a surface approach in an ILE, or to inform the tutor.

Thirdly, as shown in section 5.3.6, it survives in the leanest and meanest model of the “unrelated memorising” subscale. This means that it is a statistically significant predictor, which gives reassurance about its contribution in the specific model for future studies.

Based on these findings, there are certain recommendations which can be made for future studies in similar context:

- It is reasonable to include it in the "unrelated memorising" model.
- It is worth re-examining its contribution to the main "surface" and "syllabus boundness" models, as it is the only path-related metric with regards to these models which has the potential to enrich further and offer a distinguishing aspect with regards to the students' interactions and these surface approaches.
- It is worth considering its inclusion in the model of "lack of purpose" because, as shown in section 4.5.2, there is a positive correlation between compactness and the specific subscale (although the metric does not survive in the suggested version of the model). This is especially the case, if the researcher judges that this type of surface approach can manifest more strongly in the educational setting in which study takes place (e.g. if an issue with students' lack of interest in their studies has been identified as a general problem in a university course).

c) Relative amount of revisits

This metric has the potential to play an important role in certain surface models.

Firstly, it helps enrich the "lack of purpose" model by offering an additional dimension regarding students' page visitation when interacting in an ILE during tutorials; and it also offers a distinguishing aspect to the "lack of purpose" model, as it is only included in the specific model (see section 5.1.3).

Secondly, as it is indicated in 5.3.1, it is the predictor that contributes the most to the "lack of purpose" model.

Thirdly, it survives in the leanest and meanest version of the "lack of purpose" model, which gives reassurance with regards to its contribution in the specific model for future studies.

Based on these findings, there are certain recommendations which can be made for future studies in similar context:

- It is reasonable to include it in the "lack of purpose" model.
- It is worth considering its inclusion in the "unrelated memorising" model. This is because there are quite strong theoretical indications that students with high scores on unrelated memorising are more likely to revisit parts of the learning material compared to those with low scores (Entwistle and Ramsden, 1983, Rohrer et al., 2005). Furthermore, as shown in section 4.2.2, there is a positive correlation (although the metric does not survive in the suggested version of the

model). It is especially worth considering its inclusion if a researcher judges that this type of surface approach can manifest more strongly in the educational setting in which the study takes place (e.g. if an issue with students using rote learning as an approach towards learning has been identified as a general problem in a university course).

d) *Stratum*

The specific metric can help enriching the "use of evidence" model as it adds another aspect to students' interactions in an ILE during the tutorials regarding the linearity of the path they undertake. It can also add a distinguishing aspect to the "use of evidence" model, as it is only included in this specific model (see section 5.2.3). Furthermore, it survives in the leanest and meanest version of the "use of evidence" model.

However, as discussed in sections 5.2.2 and 5.2.3, it seems that its contribution really depends on the way the learning material is structured in the ILE. So, it can be recommended for inclusion in the "use of evidence" model, but, depending on the structure of the learning material, its contribution, and even its relationship to the "use of evidence" scale, can change from negative to positive. For example, this can be the case if the learning material in the ILE is designed in a way that suits more those with higher scores on the subscale compared to those with lower scores.

Overall, this is what one may expect in a similar study, but it is worth clarifying that the above insights do not imply that learning material should be designed in a way that suits those with higher scores on the subscale. Recommendations on how the material in an ILE can be designed are discussed further in section 6.2.1.

e) *Other performance-related metrics with regards to surface approaches*

Number of exercises solved on third try with regards to "surface" models

First, it helps enrich "surface" models by contributing with a positive relationship. For example, its inclusion serves towards making meaningful comparisons with the *number of exercises solved on first try*.

Secondly, it survives in the leanest and meanest version of the "syllabus boundness", "fear of failure" and the main "surface" models (see Appendix 5.3).

However, the specific metric does not have the distinguishing "power" (compared to the predictor *number of exercises solved on first try),* given that it contributes also to positive beta values in the suggested "deep" models for the "seeking meaning" and "use of evidence" subscales.

Based on these findings, there are certain recommendations which can be made for future studies in a similar context:

- It is reasonable to include it in the "syllabus boundess", "fear of failure", and "surface" models. However, it makes sense to include it along with *number of exercises solved on first try* to allow for meaningful enriching comparison in terms of students' performance when solving exercises during tutorials.
- It is worth considering it for inclusion in the "lack of purpose" model because, as shown in section 4.5.2, there is a positive correlation as expected (although the metric does not survive in the suggested model). It is especially worth considering its inclusion if a researcher judges that this type surface approach can manifest more strongly in the educational setting in which the study takes place (e.g. if an issue with students' lack of interest in their studies has been identified as a general problem in a university course). Another condition for its inclusion is that it be included only if *number of exercises solved on first try* is included as well.

Number of exercises finished but not solved with regards to "surface" models

First, it helps enrich the "surface" models. This is because it can point to an aspect of students' interaction called "gaming" behaviour, as discussed in sections 2.1.7 and 3.4.7. As observed in similar systems, students' interaction when solving exercises can indicate whether students are simply abusing the affordances of the environment to achieve good results but with questionable learning gains (Baker et al., 2008; Mavrikis, 2010). This behaviour has been characterized as "gaming", as students try different solutions without a systematic approach and take advantage of the system's feedback (Baker et al., 2008).

Secondly, it survives in the leanest and meanest version of models for the "unrelated memorising" and "lack of purpose" subscales (see Appendix 5.3).

Thirdly, despite the fact that it does not survive in the suggested version of the "fear of failure" model, it is worth considering its inclusion in the specific model. This is because there are quite strong indications in the literature that the behaviour of "gaming the system" is linked empirically to anxiety about failing (Baker et al., 2008); possibly because it allows students to succeed without risking failure (Entwistle et al., 1979). So, it is a predictor that in future similar studies, and especially in educational settings where surface and "fear of failure" approaches can manifest strongly, has still the potential to contribute to the "fear of failure" model.

However, it is worth mentioning that the specific metric does not have the distinguishing power (compared to the predictor *number of exercises solved on first try),* given that it contributes also with positive beta values in the leanest and meanest versions of the deep models for "relating ideas", "interest in ideas" and the main "deep" scales.

Based on these findings, there are certain recommendations which can be made for future studies in a similar context:

- It is reasonable to include it in the "unrelated memorising" and "lack of purpose" models.
- It is worth considering it for inclusion in the "surface" model because, as shown in section 4.1.2, there is a positive correlation and it does survive in the suggested version of the model.
- It is worth re-examining its inclusion in the model "fear of failure" because of the theoretical connections discussed previously.

f) Other performance-related metrics with regards to "deep" models

As discussed in section 5.2.3, the performance-related metrics, except for *number of exercises solved on first try*, give a rather unexpected picture of the students' interactions when solving exercises for most of the "deep" models.

Because of this unexpected picture, a sensible recommendation is to consider those performance-related predictors surviving in the leanest and meanest versions of the "deep" models (see Appendix 5.3). Furthermore, because a deep approach towards studying seems to be very much influenced by the broader educational environment and settings, as discussed in section 5.3.2, it makes sense that one should consider them before making assumptions with regards to the relationships between the performance-related metrics and the deep scales.

If there is a question as to which deep approach the performance-related metrics give the most reasonable, complete and enriching model of with regards to students' interactions when practising tutorial exercises, then after looking at the leanest and meanest versions of the models (see Appendix 5.3), the answer is the "relating ideas" approach. It seems to be the only "deep" model in which the performance-related metrics make the most sense or at least give the most complete picture (that is, students with high scores on the “relating ideas” subscale are more likely to solve exercises on the first attempt or not at all, and not likely to solve exercises on the second attempt, compared to those with low scores). So, it is worth considering for inclusion the predictors *number of exercises solved on second try* and *number of exercises finished but not solved* in the “relating ideas” model. As for the rest of the deep models, as recommended previously, the general educational setting should be considered first before any decisions for their inclusion are made, but also their combined inclusion in models should be considered to allow for more meaningful comparisons.

g) Temporal metrics

Maximum view time on a reading (content) page

As shown in section 5.3.6 and Appendix 5.3, this metric survives in the leanest and meanest versions of the “fear of failure” and the main “surface” models, and the “relating ideas” model. This means that it is a statistically significant predictor, which gives reassurance about its role for the specific models in future studies. Furthermore, it gives a distinguishing aspect to these models, as it has a unique positive relationship to the “relating ideas” subscale, while it has a negative one to the above surface scales. Based on these findings, it is sensible to recommend it for inclusion in the aforementioned models.

Average view time on reading (content) pages

As shown in section 5.3.6, this metric survives only in the leanest and meanest model of the surface model “syllabus boundness”. This means that it is a statistically significant predictor, which gives reassurance about its role for the specific model in future studies. It is worth mentioning that it is not included in any other model (neither in the suggested versions nor in the leanest and meanest versions), so it does give a distinguishing aspect to the specific model. Theoretically speaking, spending on average an increasing amount of time on the theory as the score on the scale is also increasing, also makes sense in the specific approach, as those with high scores on the subscale are more likely to want to ensure that they follow the given syllabus and dedicate more time to it, compared to those with low scores.

Maximum view time on an exercise page

As discussed in sections 5.1.3 and 5.2.3, this is a temporal predictor which plays an important role in both “deep” and “surface” models.

More specifically, it contributes to all the leanest and meanest versions of the “deep” models, with the exception of the “relating in ideas” one where it is included in the suggested version of the model. It also is the highest contributor in the “seeking meaning” model. With regards to the surface scales, it contributes in the leanest and meanest versions of the “fear of failure” and the main “surface” models.

However, it has the same positive relationship to both deep and surface scales. This, as discussed in sections 5.1.3 and 5.2.3, can be for different reasons. Still it is the type of temporal predictor which does not have the distinguishing power other temporal predictors have, such as the *maximum view time on reading (content) page*.

A sensible recommendation is to include it in all the aforementioned models in future studies. However, to aid further the interpretation of its contribution to these models (e.g. whether students with higher scores on the deep scale tend to spend an increasing amount of time on a specific exercise page because of their intention to seek deeper understanding of the logic behind the exercises, or because they experience difficulties with the specific group of exercises) it is worth:

- considering it and interpreting it along with the performance-related metrics (e.g. if students with higher scores on the deep scale tend not to solve the exercises at all, then it could be that they spend an increasing amount of time on a specific group of exercises because they experience difficulties).
- improving the metric as suggested later on in section 6.2.2

To conclude, the recommendations in section 6.1.4.1 can help a researcher in deciding when planning a similar study as to the number of predictors which should be initially included in the models, and whether the targeted sample size is enough with respect to the “desirable” number of predictors and the typical threshold of statistical power, as indicated in 3.11.3.

6.1.4.2 Non-surviving predictors

As shown in section 5.3.6, there are predictors which are not involved in the leanest and meanest versions of the “deep” or “surface” models. However, it may be worth considering them in future studies, because:

- They link to both deep and surface scales both theoretically and empirically (i.e. there are initial correlations or contribution to the suggested models)

- There is a need for more clarity as to their relationship to the scales
- There are design recommendations in relation to them which may improve their contribution if they are applied

  More specifically:

- *The number of hyperlinks (concept links) visited in reading and exercise pages* can be a good candidate for inclusion in the "surface", "relating ideas" and "seeking meaning" models, as there are empirical and theoretical links, and its inclusion offers an opportunity to clarify further the nature of its relationship to these scales.
- The *average number of times a notes link is clicked per page* survives in the leanest and meanest model of the "relating ideas" model only, but there is the potential of further development of the "notes" feature so it can contribute further to other deep models, as discussed in 6.2.1.

6.2 User interface design and data capture

In this section, there is discussion with regards to improvements for the ILE features and the predictors of the "deep" and "surface" models in order to achieve a more enriching interpretation of results in future studies.

6.2.1 User interface design recommendations

6.2.1.1 Search Feature

It was initially expected that there would be possible links between the use of the search feature and the subscales of "relating ideas" and "interest in ideas". Indeed, there are scenarios in which it could have been used. The search option could have been used during the third week of study in AM by those students with an interest in linking mathematical concepts between the chapters on Functions and Graphs. For example, if they typed "function" in the search option, they would get links to the topic of "quadratic graphs". The search option could also support the students' interest to go beyond the syllabus of the current week and make connections between mathematical topics of the third week and mathematical topics of the fourth week. For example, if they typed "simultaneous equations" in the search option, the search would, amongst other results, return links about the topic of "matrix representation of linear equation" (a topic from the chapter of Matrices taught in the fourth week).

However, the search option was not used and there is no data and variability to enable us to examine its relationship to the deep scales. It is possible that the practical nature of the tutorials did not encourage the further exploration of mathematical concepts in graphs and functions. It can also be due to tutor intervention as discussed in section 6.1.1. But the lack of its use most likely relates to the way the search option is presented in AM. The search option in a system like AM can be designed in a way that can gratify students' intrinsic interest and seeking further meaning in mathematics, encouraging in this way a deep approach. Vockell (2006) in his discussion argues that to promote intrinsic motivation, there is a need to stimulate students' sensory curiosity (by making changes that are perceived by the senses) and cognitive curiosity (by making the learner wonder about something) in a learning environment. In addition, Ryan & Deci, (2000) (cited in Martens et al, 2004, p. 371) find that high intrinsic motivation is positively related to curiosity and exploration. They also argue that students with high intrinsic motivation are likely to explore the parts of a programme that gratify their curiosity. It seems that the search option in AM was not seen as a means to satisfy the curiosity of students, not even of those with high scores in the "interest in ideas" subscale.

However, it is possible to improve the AM search option. Fransson (1978) notes that intrinsic interest in a subject is not so much something one creates but rather something one finds. He also notes that if we want to utilise students' intrinsic motivation, we must focus on what they are interested in and link the study material to it. Following this suggestion and taking into account Vockell (2006) who argues that the curiosity of intrinsically motivated students can be triggered by making them notice a change through their senses or by making the learner wonder about something, the search option should not simply be a choice placed statically at the navigation bar, but an option that can emerge at crucial moments and suggest a search at a moment that is likely to be noticed. The idea of encouraging deep understanding in crucial moments in a way is supported in the literature. Entwistle and Patterson (2004) support that presentation of content should be enhanced with the identification of "troublesome knowledge" (i.e. knowledge which students find difficult to understand). Recommended predictors of students' interactions, discussed in section 6.1.4.1 can help with the identification of "troublesome knowledge". For example:

- when students fail to solve on first attempt a specific type of mathematical exercise related to a specific mathematical concept;

- when the time spent on a specific exercise page exceeds a reasonable limit (considering as well the length of the tutorial session);
- when students excessively revisit specific pages of learning material, as revisitation can be linked to the "lack of purpose" scale;
- when students conduct a very compact path and do not expand on material they are supposed to practise on.

  At that point a suggestion could emerge to explore a mathematical concept or a detailed worked example related to the specific type of exercise.

  Tendency not to solve on first try, excessive amount of time spent on an exercise page, and excessive revisitation and compactness can be primary indicators for "troublesome knowledge". But other predictors can be used as well, such as: failing to solve exercises on second or third attempt, or not solving them at all. The module team, which designs the learning material and is aware of the level of difficulty, can be involved in deciding the predictors involved and the thresholds. To conclude, the main recommendations for the design of search are as follows:

- identify "troublesome knowledge" based on predictors which are linked to surface approaches towards studying (for more discussion on such indicative predictors see section 6.3.2)
- design a search which is responsive to "troublesome knowledge" and emerges at these crucial moments

6.2.1.2 Notes Feature

It is discussed in section 6.1.4.2 that there are indications that students with high scores on the "unrelated memorising", "relating ideas" and "deep" scales are more likely to access the "notes" feature.

So, use of this feature is not necessarily linked to the deep approaches, as initially assumed. This is also reinforced also by the observations made in class, and the recorded notes (see Appendix 6.1). They indicate that there are students who used notes in a rather "surface" manner (e.g. to copy and paste learning material in them or simply record the answers of the exercises), and other students who use them in a "deeper" manner (e.g. to record the logic behind the solution of an exercise).

It is possible that if the use of the "notes" feature could help the deep approach to manifest itself more strongly, then further contribution to the "deep" models could be attributed to the predictor representing it. This can be a pedagogical issue, but it is also a design issue. The question is whether there is a way to improve the design of the “notes” feature to encourage further a deep approach towards learning. One way to do this is further to develop the social aspect of the “notes” feature. This is also supported in the literature. Entwistle and Peterson (2004) support the provision of opportunities for group discussion of both content and learning processes, and Baeten et al. (2010) point to empirical evidence which indicates that feedback received online from peer-groups encourages a deep approach.

Currently, in the “notes” feature, there is the option “*Allow others to see this note”*. This public aspect of the “notes” feature can be further enhanced by allowing students to create a study group of their choice (e.g. specific study group, tutorial group or whole module). The reason behind this suggestion is that it can serve two major “intrinsic motivation” elements: autonomy and relatedness (Ryan and Deci, 2000). In the field of user interface design, researchers such as Lockton (2012) emphasises the importance of designing for intrinsic motivation by finding ways to make users feel that they interact with choice (i.e. autonomy) and that they are part of a community (i.e. relatedness). By allowing students to share information and give support to each other within a group of their choice, can help them to feel more autonomous and offer a feeling of belonging to a specific learning community[52].

Another recommendation is to allow students and tutors to rate highly the public notes that show deep understanding such as explaining the logic behind an exercise. Based on the recorded notes in the current study, it occurs that students record in notes the logic of an exercise and are happy to share it (see Appendix 6.1). If the design of the “notes” feature allows users to reward this type of students’ actions by favouring them, then it can be a way to express support and encouragement by peers and tutors, which in return can motivate towards a deep approach towards studying (Baeten et al., 2010). In addition, it can again increase the aspect of intrinsic motivation which has to do with relatedness, as it is another way to make the students feel that they belong in the learning community in which their peers and tutors show appreciation for

[52] There are already indications, based on the recorded notes, that these design suggestions can be quite well-received, as students used the “notes” feature to ask for help or share amongst themselves advice on the logic and solutions of the exercises during the tutorial sessions (see Appendix 6.1).

their efforts.

To conclude, the recommendation is: to develop further the current “public” option of the “notes” feature, so that it allows students the autonomy to share information with a group of their choice, and allow peers and tutor to favour “public notes”, increasing in this way support, encouragement, relatedness and ultimately the students’ intrinsic motivation.

6.2.1.3 Hyperlinks (Concept Links) Feature - creating links between concepts

As discussed in sections 6.1.4.2, the number of hypertext concept links has an unexpected relationship in terms of direction to the deep and surface scales. It seems that those with higher scores on the “relating ideas” and “seeking meaning” subscales are less likely to use them, compared to those with lower scores. In contrast, those students with higher scores on the “surface” scale are more likely to use them, compared to those with lower scores. So, there are indications through the suggested versions of these models, that the use of hyperlinks is not necessarily linked to the deep approaches, as initially assumed, and there is even an indication that they are used by those with higher scores on the “surface” scale.

With regards to the “deep” models, a possible reason can be the influence of students’ prior knowledge. As discussed in 5.5.3.2, it seems that the direction of the relationships in these models seem to express better the interactions of the “low prior knowledge” group, than those of the “high prior knowledge” group across the specific scales. In a way, this is also supported in the literature, which indicates that those with low prior knowledge can have issues with cognitive overload and not being able to integrate new pieces of information to the whole of what they learn when they study in an ILE (Mampadi and Mokotedi, 2012). It is possible, therefore, that even those with higher scores in the deep scales tend to not use the hyperlinks because perhaps because they felt cognitively overloaded by all the available choices in AM, or because the hyperlink’s current design does not help them enough to understand the connections between the concepts. In the field of educational technology, there have been various efforts and suggestions for designing links in ways that better suit students’ individual characteristics (e.g. link hiding, link annotation, direct guidance) (De Bra and Calvi, 1998; Brusilovsky, 2001; Mampadi and Mokodedi, 2012).

Based on these suggestions, AM could be made more responsive with regards to the hyperlinks by highlighting the hyperlinks similarly to the search feature, at crucial moments when they are likely to be noticed. Recommended predictors of students' interactions, discussed in section 6.1.4.1 (and further discussed in section 6.3.2) can help with the identification of "troublesome knowledge" when content needs to be enhanced (Entwistle and Patterson, 2004). For example:

- when students fail to solve on first attempt a specific type of mathematical exercise related to a specific mathematical concept;
- when the time spent on a specific exercise page exceeds a reasonable limit (considering as well the length of the tutorial session);
- when students excessively revisit specific pages of learning material, as revisitation can be linked to the "lack of purpose" scale;
- when students conduct a very compact path and do not expand on material they are supposed to practise on.

Furthermore, the reason behind the tendency to use hyperlinks for those with the higher scores on the "surface" scale, could be because the content of the hyperlinks in AM simply repeats information about concepts and procedures found in AM pages, possibly contributing to repetitive overlearning. So, what can be recommended here is providing further information through the hyperlinks, for example extra examples with regards to a concept, especially if students have already visited pages with relevant content (in other words the content of the hyperlink for a specific page can be changed according to what has been previously seen). So, instead of repeating information, which might have been previously seen, the content of hyperlinks will aim to extend further what is currently known.

As the function of facilitating links between mathematical concepts is important for encouraging a deep approach, it is worth trying to find another way to introduce it in an interactive learning environment. Another feature which can be considered is a digital concept map. There are currently stand-alone applications and websites for creating digital concept maps[53]; however, the idea is to incorporate such a tool in AM. It can work in to two ways:

- encourage students to find links themselves between current concepts and apply this visually by creating nodes; and

[53] They are also called mind maps. Some examples are: Cmap, Giffy Online, and Bubb.us.

- activate learners' prior knowledge of by highlighting and differentiating the old concepts, and allow students to connect them to new concepts.

  The idea of concept maps has support in the context of mathematics at university level. Beng and Yunus (2013) support that a concept map can help apply a holistic approach (which relates strongly to the "relating ideas" approach) and most importantly strengthen and activate prior knowledge; hence they apply it in the mathematical concept of calculus.

  This recommended feature, an embedded digital concept map in an ILE, besides enhancing understanding regarding the relationship between old concepts and new concepts, can serve also as way for the tutor to check on students' prior knowledge and any misconceptions regarding the connections and structure of students' concept maps. By relating teaching in this way to prior knowledge and facilitating students to reflect on the relationship between old and new concepts, one can create a more student-oriented learning environment that encourages a deep approach (Entwistle and Peterson, 2004).

6.2.1.4 Distinguishing detailed worked examples from the rest of the theoretical generic material

Sections 5.1.3, and 5.2.3 discuss that it is possible that the way the material is structured and the AM interactive learning environment is designed (although flexible and versatile) do not really encourage the "use of evidence" approach to emerge. So, it is worth looking at possible solutions which can encourage this specific approach. Based on Hills (2003) "design" recommendations for an ILE, AM can display suggestions to check detailed worked examples, but in a timely manner, especially when there are signs of "troublesome knowledge", as discussed previously. According to Hill (2003), this type of help is likely to benefit those who pay attention to the low-level detail, and master one topic at a time. These are tendencies for those with higher scores on the "use of evidence" subscale, (a scale which is based on Pask's (1976) serialist approach towards studying, as discussed in section 4.10.1). Students with such tendencies are also likely to ask for help before moving to the next topic (Chen et al., 2016), which in a way reinforces the ideas of giving them "timely suggestions".

Based on the above insights, it is possible to make the following recommendations:

- First, make a clear distinction between the way the mathematical concepts and operations are presented and structured in the interactive learning environment by distinguishing detailed worked examples (see Appendix 3.4.11) from more theoretical and generic learning material (see Appendix 3.4.1)[54].

- Secondly, make "timely suggestions" for students to check on a relevant detailed worked example when there are signs that they are dealing with "troublesome knowledge".

  In a similar way, distinctions in the learning material and "timely suggestions" can be made for those who tend to form an overview by exploring topics of what may be known, and relating a concept to another, as discussed in 4.9.1. These are tendencies for those with the higher scores on the "relating ideas" subscale, (a scale which is based on Pask's (1976) holist approach towards studying, as discussed in section 4.9.1). Based on these insights, a possible recommendation is to create more summarised generic material (see an example of such a page in Appendix 3.4.12), which brings together related mathematical concepts and examples, and which can be suggested in a timely manner, when for example there are signs of "troublesome knowledge".

6.2.1.5 Need for a versatile interactive learning environment

Finally, recommendations regarding the design of an ILE should aim for a versatile learning environment. The current structure of the learning material in AM does support a versatile structure by incorporating both an inductive and a deductive manner of delivering the learning material, as discussed in section 5.2.3. However, more could be done to achieve a more versatile ILE.

The current recommendations are based on Pask's (1976b, 1976a) insights with regards to the concept of "versatile learners". These are individuals who tend to study both the low-level detail following a "serialist" approach, and the overview and relationship of concepts to build understanding following a "holist" approach. Entwistle and Ramsden (2015) and Entwistle (2001) indicate that both approaches (the "serialist" one which points to the "use of evidence" approach and the "holist" one which points to the "relating ideas" approach) are needed in order to obtain a deep level of understanding. So, the

[54] When the given learning material was incorporated in AM, there was an effort to make such a distinction, but this was not always possible (see Appendix 3.4.10) because of the way the learning material was written and produced in a textbook by the lecturer in charge of mathematics module. The way the learning material was structured and presented in AM could not be altered significantly compared to the textbook, as it required the lecturer's approval. There were also ethical reasons: the approved module syllabus represented in the textbook serves specific learning outcomes, hence it cannot be changed because it can affect those learning outcomes and ultimately students' performance.

recommendation here is for an ILE which incorporates both approaches in its design. The aim is to facilitate these two approaches to studying, but move at the same time towards combining both to encourage versatility in studying. Indeed, as discussed previously, the incorporation of both a digital concept map and "timely suggestions" for checking on summarised theoretical generic material, which can encourage a "relating ideas" approach, and the "timely suggestions" to check detailed worked examples which can encourage a "use of evidence" approach, may serve towards creating a more 'versatile' ILE.

Finally, as indicated previously, both the "timely suggestions" for detailed worked examples or summarised theoretical generic learning material can occur when there are signs of "troublesome knowledge". To support students' autonomy[55] in terms of choices and versatility in terms of learning, however, it is worth considering giving them choice on whether they prefer visiting content with relevant detailed worked examples or are content with more generic theoretical material, independently of their scores on the "relating ideas" and "use of evidence" scales.

6.2.1.6 Identifying a more "concrete" picture of an approach in an interactive learning environment

It is possible to create a feature in an "intelligent" ILE which identifies students who interact, for example, in a "surface" manner. This can assist tutors to identify, quickly and effectively, students especially with surface approaches.

To give a more concrete picture of how a student with a very high surface score, for example, would interact in the specific system and in real learning conditions, one idea is to use "values of interest" with regards to the regression model representing a scale. The idea derives from an example given by Field (2009) regarding how to use a regression model. More specifically, the extreme values of interest, which are the minimum and maximum values of each predictor in the model, are used according to the direction of the relationship to the outcome.

For example, with regards to Model 4 representing the main surface approach, the minimum values of the predictors are used where there is a negative relationship with the outcome, and the maximum values of the predictors are used where there is a positive relationship with the outcome (see Appendix 6.2).

---

[55] This can also contribute to students' intrinsic motivation, as discussed in section 5.2.1.

It is possible to say that a student with the highest score on the surface scale could interact in the specific system as follows:

- Given that the specific system allows a student to try up to 3 times to find the correct answer and at the end it reveals the answer, the student is likely to solve four exercises on first try, 15 exercises on third try, and 32 exercises will be finished but not solved (which means on the third try the answer was incorrect) over two tutorial sessions. To give a more proportional picture, we can say that out of all the exercises solved on first and third try or not at all, a student with the highest score on the surface scale is likely to solve just 7.8% of the exercises the first time, 29.4% of the exercises the third time, and approximately 62.7% of the exercises will be solved but not really finished.
- For the maximum amount of view time on an exercise page, a student may well spend around 3750 seconds (62.5 minutes) (which is basically half the two-hour tutorial session on one specific exercise page); while for the maximum amount of view time on a content (reading) page, a student may spend just around 108.9 seconds (1.8 minutes) on the theory. To give a more proportional picture, the maximum number of minutes spent on an exercise page can be approximately 35 times greater than the maximum number of minutes spent on a page with theory.
- They are also likely to follow a highly compact path, and visit up to 11 hyperlinks (concepts links) on reading and exercise pages.

These “extreme interaction profiles” can be presented for training tutors who can in this way develop an appreciation of the kinds of interaction they should expect, or not, from the students. In addition, an intelligent system could use the previous data and models to identify the students who tend to interact in a “surface” manner.

6.2.2 Improvements in data capture

In Chapter 5, there is discussion regarding the reasons behind the unexplained variance in the models (i.e. not strong enough contributions of existing predictor and absence of other predictors). The contribution of current predictors can be strengthened by improving the predictors (both those which are involved and those which are not involved in the suggested models).

Furthermore, in both Chapters 4 and 5, there is an attempt to interpret students’ interactions based on the models representing each approach towards studying.

In deep models particularly, there is an attempt to explain unexpected findings; whereas in the surface models, where results are mostly the expected ones, there is an attempt to interpret students' interactions and even distinguish one subscale from another. Although the current interpretations are reasonable and give a good indication of certain type of interactions, improvements to ensure a more enriching interpretation in similar studies can be made with regards to the predictors.

6.2.2.1 Recommendations for improvements on path metrics

Path-length related metrics such as path length and number of reading pages and number of exercise pages have not been found to correlate with any of the scales, and have not contributed to the models of the deep and surface scales. This can be due to the fact that, despite the distinction between reading and exercise pages, they still seem to be generic and are probably missing semantic elements (i.e. elements which can reveal more about the learning content of a page) which would make them more sensitive to capturing changes in the variability of students' page visitation across the scores of the scales.

Something that can be recommended to address this issue is distinguishing in reading pages between: those which include definitions, generic theoretical examples, summarized pages which show how concepts related to each other (see examples in Appendices 3.4.1, 3.4.13, and 3.4.12); and those which include detailed working examples (see Appendix 3.4.11). By making this distinction, it is expected that predictors representing visitation of these two different types of reading pages can relate or contribute better to the "relating ideas" and "use of evidence" scales respectively. More specifically, and based on the definitions given by Entwistle et al. (1979), Entwistle (1981) and Pask (1976b) for these two different approaches, it is expected that those with a high score on the "relating ideas" scales are likely to visit more pages which have a broader view of the theoretical content (e.g. pages with theoretical examples explaining the general logic of operations (see an example in Appendix 3.4.14), or pages with an overview/summary of the concepts and their relationships (see an example in Appendix 3.4.12); while those with high scores on the "use of evidence" scale are likely to visit more pages which have more specific detailed worked examples (see Appendix 3.4.11). The inclusion of the suggested predictors in the "relating ideas" and "use of evidence" in future studies may reinforce the previous design recommendation for creating a versatile interactive learning environment where both the "relating ideas" and the "use of evidence" approaches may manifest much more strongly.

6.2.2.2 Recommendations for improvements on temporal and performance-related metrics

In a similar way, more semantic elements can be attached to temporal and performance-related metrics to ensure stronger relationships and contributions for these predictors in the models and decrease the amount of unexplained variance.

For example, the element of "level of difficulty" can be incorporated in temporal and performance-related metrics. However, the educational research on the level of difficulty and how it affects students' learning is vast and it requires thorough investigation in order for a researcher to make decisions on how it can be incorporated in the data. There is a field of educational research which examines, for example, how awareness of the difficulty of learning tasks or the perceived level of difficulty by students can affect performance, motivation and anxiety (Kukla, 1974; Martin and Manning, 1995). So, a researcher, for example, has to make decisions with regards to whether the level of difficulty can be defined by the lecturer/tutor (based on his or her expertise on the specific subject and teaching experience in a specific module and course), or whether it can be defined as the perceived level of difficulty by students. Another decision could be whether the assigned levels of difficulty will be known to the students or not.

In the following sections, there are various examples of how the aspect of level of difficulty could be involved in temporal, performance-related and path metrics (without, though, going into further detail, as further investigation on the level of difficulty is beyond the scope of this thesis). More specifically, with regards to:

a) Temporal metric - Average time view time on exercises pages

As shown in Chapter 4, *average view time on exercise pages* has produced statistically significant relationships to the "unrelated memorising", "fear of failure", and the main "surface" scales. However, its contribution is not enough to be included it in their suggested models. So, for example, there can be a distinction into: *average view time on exercise pages of basic level difficulty*, *average view time on exercise pages of medium level of difficulty*, and *average view time on exercise pages of high level of difficulty*; which may offer more variability for the metric across the scores of the aforementioned scales. For example, it could be the case that students with a higher degree of anxiety (i.e.

those with the higher scores on the “fear of failure” scale) tend to spend on average more time on exercise pages with a high level of difficulty, compared to those with a lower degree of anxiety.

b) ***Temporal metric - Maximum view time on exercise page***

As shown in section 5.3.6, *maximum view time one exercise page* survives in the leanest and meanest models of the “fear of failure” and the main “surface” scales. Capturing the *maximum view time on an exercise page* with regards to the level of difficulty can be more enriching in terms of interpretation. In section 5.1.3, it is discussed that the specific metric can help to reveal more extreme interactions like a student experiencing problems with a specific group of exercises on a specific page and thus working much more slowly. This interpretation can be reinforced if it is shown, for example, that students with high “surface” scores tend to spend maximum view time on a group of exercises with a high level of difficulty. A distinction, therefore can be made with regards to the specific metric, such as: *maximum view time on exercise pages of basic level difficulty*, *maximum view time on exercise pages of medium level of difficulty*, and *maximum view time on exercise pages of high level of difficulty.*

Furthermore, in section 5.2.3, it is discussed that there are two possible interpretations for the positive contribution of *maximum view time on an exercise page* to the “deep” models (either that it is a time-consuming effort to build understanding of concepts and procedures, or that students may experience difficulties with a specific group of exercises with a high level of difficulty). The previous distinction may aid further with the interpretation of the deep scales as well.

c) ***Temporal metric - Maximum view time on reading (content) page***

Although the specific metric has more distinguishing power compared to *maximum view time on exercise page* as the direction of its relationship between deep and surface scale differs (see discussion in section 5.2.3), it is worth considering similar improvements. For example, the element “level of difficulty” may aid the interpretation of the positive association between the metric and the “relating ideas” scale (i.e. whether increasing maximum view time means an increasing intention and effort to relate mathematical concepts, or whether it means that students occasionally “get stuck” on a specific page with exercises of a high level of difficulty).

d) Performance-related metrics

If the difficulty of a task is defined as an informational component which conveys the probability of success in accomplishing it, then it can certainly affect performance (Martin and Manning, 1995). It is an element, therefore, which is worth considering in performance-related metrics, as it has the potential to enrich further the interpretation and increase the contribution of these metrics on both “deep” and “surface” models.

The importance of performance-related predictors in “surface” models is discussed in section 5.1.3 and 5.3.1. For example, predictors such as *number of exercises solved on first try* contribute to all scales, and in three of the “surface” models, including the main “surface” one, is the highest contributor. Adding another aspect to these predictors has the potential to provide even more enriching comparisons. It is possible to compare, for example, if those students with high surface scores tend to not solve on first try exercises with a high level of difficulty; whereas they tend to solve exercises on third try (or not at all) with a medium or low level of difficulty.

Adding the aspect of level of difficulty can serve towards increasing the variance of the models; especially those with the least variance, as indicated in section 5.1.3. For example, the variance of the model representing the “lack of purpose” scale can benefit from a predictor which indicates the level of difficulty for the number of exercises solved on first try (or third try or not solved at all). It can show whether students with an intention to cope minimally with the course requirements engage with all levels of difficulty when solving exercises (it can be the case for example, that they engage mainly with exercises of a low level of difficulty).

Finally, level of difficulty is an aspect which can aid in increasing the contribution and aid the interpretation of *number of exercises cancelled* in the “syllabus boundness” model. This is also reinforced by literature which indicates that quitting a learning task is linked to students’ evaluation that a task is difficult (a type of self-regulation that is deemed to be counterproductive) (Winne and Hadwin, 2009). It is possible that students with a high score on the “syllabus boundness” scale tend to quit exercises with a high level of difficulty because they tend to do the wrong type of self-regulation.

e) New metrics occurring from design recommendations

In section 6.2.1, there are design recommendations for further development of existing AM features. The implication of these design recommendations is that there should be new metrics to capture the use of these features.

Metrics for concept map

With regards to student's use of the concept map, the proposed metrics can capture, for example: access of the concept map (i.e. number of clicks on the relevant link); interactions regarding the creation of the concept map (i.e. number of clicks on "create", "edit" and "save" buttons); tutor's rating on students' concept maps.

Furthermore, as discussed in section 6.2.1, the use of the concept map as a way of facilitating students to create links between mathematical concepts can encourage a deep approach. Therefore, there are expectations of positive relationships between the deep scales and the proposed metrics. The specific metrics are particularly expected to contribute to the "relating ideas" model, as it represents a studying approach which is about the intention to relate one concept to another (Entwistle, 1997a; Pask, 1976b).

Metrics for notes feature

With regards to the use of the "notes" feature, the proposed metrics can capture, for example: an indication that a note is created (i.e. number of clicks on "save" buttons); choice in terms of creating a public or private note (i.e. number of private and public notes); creation of peer-group when creating a public note; access to notes feature (i.e. number of clicks on relevant link); and indication that a note has been favoured or rated by peers or tutor based on its quality or usefulness.

Furthermore, as discussed in 6.2.1, the students' choice to create a peer group or share notes with the peer-group can increase motivation and encourage interest in the subject, so metrics indicating sharing notes with the peer-group may have positive relationships with the "interest in ideas" scale. In addition, metrics indicating the notes' rating based on its quality can offer a distinguishing aspect for both "deep" and "surface" models (i.e. positive relationships for "deep" scales and negative for "surface" scales). For example, a note with a high rating, which explains the rationale of a mathematical solution, can be an indication of students' intention to understand for themselves: an intention associated with the deep approach towards studying (Entwistle, 1997a).

To conclude, the proposed improvements on AM features and interaction metrics aim to encourage further the deep approaches, while offering at the same time the potential for more distinguishing and enriching aspects with regards to the interaction metrics.

6.3 Pedagogical Insights

In this section, there is discussion as to: how the pedagogical insights occurred in the current investigation can help towards supporting tutors in class; serve researchers who intend to conduct similar studies; and help towards addressing criticisms in the field of learning styles. Finally, there is discussion as to whether and how prior knowledge influences students' interactions with regards to the deep and surface approaches towards studying.

6.3.1 Deep approach in specific context and in other contexts

6.3.1.1 Does the deep approach have distinguishable aspects?

What mainly comes out of the discussion of the empirical findings in section 5.2 is that there is unexplained variance in the deep models, and that the metrics which survived the selection process during the development of the models give an indication of students' interactions with regards to deep approaches towards studying, but not a clear picture. These findings can be explained. According to Entwistle (2008) a deep approach during a degree course is neither as consistent nor as strong as one might hope. An indication that this is the case is the inconsistent picture which occurs from the performance-related metrics, with the exception of *number of exercises solved on first try* for three of the "deep" scales. According to Entwistle (2008) and Marton and Säljö (1976a), it appears also that it is not easy to move students towards a deep approach. For example, the hyperlinks (concept links) in reading and exercise pages are less likely to be accessed by those students with high scores on the deep scales, and access to the AM "notes" feature is also likely to be done by those with high scores on the "unrelated memorising" scale.

Also, the difficulty in capturing the deep approach towards studying can be due to "dissonant orchestrations" – a concept introduced by Entwistle and Peterson (2004). This means that the intention to understand can be sometimes associated with surface processes, so while students have the intention to understand, and declare it in the questionnaire, it may be the case that the way the course is designed (e.g. design of examination and coursework requirements) prevents that approach. In the context of the current study, they may declare an intention for a deep approach, but, in reality, they interact in AM in a "surface" manner. This is reinforced by Ramsden (2005), who points to research findings which show that a deep approach can also be influenced by the general course planning and the setting of the assessment questions. The module material might not be enough to encourage students to think deeply about the subject matter. If students have the perception that the general learning environment (besides the interactive one provided during the tutorials) rewards a surface approach, they may simply behave in this way, despite their deep intentions. This is also reinforced by Entwistle and Peterson (2004) who argue that a deep approach is linked with high academic performance only when tasks require a deep level of knowledge.

Similar insights to the aforementioned research findings and arguments have been also found with regards to teaching mathematics in higher education. Sangwin (2004) argues that simply changing the assessment and introducing a sophisticated ILE (i.e. one that gives tailored feedback to students' answers) is not necessarily sufficient to orient students towards a deep approach towards studying. More specifically, he supports the argument, citing Crawford et al. (1998), that in order to orient towards a deep approach, students' whole learning experience in a course has to change (e.g. students' conceptions about the nature and purpose of the subject, students' perceptions about tutors' attitudes etc.).

This could be mainly why students with high scores on the “deep” scales tend not to perform well when solving exercises and not to use “concept links” as expected, and also the reason why the AM “notes” feature seems to be used in both a “deep” and a “surface” manner. So, when using an interactive learning environment in tutorial lessons and for a specific module, it is advisable not to forget that it is part of a wider teaching and learning ecosystem, which includes among other things students' perceptions regarding the course design and other assessment requirements and learning materials. It seems that the whole teaching and learning environment should encourage a deep approach and not just the interactive learning environment. Otherwise, performance and use of features may not be as expected.

Furthermore, the direction of some of the performance-related and temporal predictors might have been influenced by the element of students' prior knowledge –an issue which is discussed further in section 6.3.5-.

6.3.1.2 Possible recommendations for tutors

As a result of the discussion in 6.3.1.1 and as occurred from the study's findings, there are recommendations which can be made for tutor, however they do come with certain cautions.

First of all, tutors should be aware that to encourage a deep approach, hence help it manifest in a stronger way, it is not just a matter of improvements in the context of a module or an ILE, but also in the wider context of teaching and learning of the course in which the module and ILE are part of. For example, when they start a course, managing students' expectations and perceptions about the course, and even challenging or correcting their conceptions about what "learning" means. This can be done by, for example, informing and mentoring them about how they are expected to approach their studying, make clear what their responsibilities are throughout the course, raise awareness about the different approaches to studying. In relation to these suggestions, Entwistle and Peterson (2004) give specific guidelines for learning that support a deep approach: encourage reflection and metacognitive alertness, and self-regulation in studying, and provide opportunities for group discussions of learning process. They also provide other guidelines with regards to the connections and alignment of aims, assessment, learning material, and student support (see Entwistle and Peterson (2004), p.424).

Secondly, despite the aforementioned cautions, it is possible for a tutor to detect signs of a deep approach during tutorial sessions, based on a combination of certain characteristics as expressed by the predictors of each deep model (see sections 5.2.3 and 5.3.6, and Appendix 5.3). More specifically, such indications can be:

- A tendency to solve exercises on the first try, especially with regards to models of the "seeking meaning" and "relating ideas" approaches. This can be also the case with regards to the main "deep" approach, but possibly not across various educational settings. This is because the metric *number of exercises solved on first try* does not survive on the leanest and meanest model of this subscale. However, with regards to the rest of performance-related, predictors their combined inclusion in the "deep" models do not provide a meaningful picture, so it is not advisable to make any further recommendations. There is also another reason. Based on the findings according to which the influence of the prior knowledge explains better the "deep" models in the "low prior knowledge" group, it is reasonable to recommend that tutors should be aware of students' prior knowledge in a systematic way (e.g. via a diagnostic test at the start of course, similarly to the process followed by University of California, as discussed in 2.1.1.3 *3)*). In this way, it will be more clear whether their performance during the tutorial activities is due lack of prior knowledge or lack of a deep approach towards studying.

- A tendency to make notes, especially with regards to especially with regards to model of the “relating ideas” approach, possibly in an effort to relate mathematical concepts and procedures. This can be also the case with regards to the main “deep” approach, but possibly not across various educational settings. This is because the metric of *average number a ‘notes’ link is clicked per page,* does not survive on the leanest and meanest model of this subscale.
- A tendency to follow a less linear path (i.e. stratum) with regards to model of the “use of evidence” approach; however, there is a caution. This is especially the case if the structure of the learning material in the ILE does not support building the meaning of concepts starting from the more detailed information.
- A tendency for spending an increasing amount of time on the theoretical aspects, especially with regards to the “relating ideas” model, possibly in an effort to create relationships between concepts. Although this is a distinguishable aspect for this “deep” model, it needs to be interpreted by the tutor in combination with the performance in the tutorial exercises, as it is also possible that students may experience difficulties understanding a specific procedure or concept (e.g. if it coexists with the tendency for not solving exercises at all).

6.3.1.3 Deep approach – Conclusion on pedagogical insights

To conclude, researchers who conduct similar studies should consider that the deep approach towards studying is sensitive to the broader learning and teaching environment in which the interactive one is used, and that this, together with the perceptions and prior knowledge of students, can influence it. Therefore, it is also not that easy to capture it and encourage it. Whether the deep approach manifests strongly or not depends on the educational setting and conditions in which the study takes place. Hence, the initial theoretical assumptions made for predictors and deep subscales should consider this before deciding which predictors are the most important or can be considered for inclusion in the deep models, as discussed in 6.1.4.1. Finally, if we want the deep approach to manifest strongly, then there should be improvements not only in the interactive learning environment and the predictors, as discussed in 6.2, but also in the broader learning and teaching environment (i.e. the design and delivery of the whole degree course)[56].

6.3.2 Surface approach in specific context and other contexts

6.3.2.1 Does the surface approach have distinguishable aspects?

Based on the discussion of the findings in sections 5.1 and 5.3, when examining both the suggested and leanest and meanest models of the surface approaches –both at model level and at predictor level- there is more confidence in them compared to the “deep” models. First of all, there is more variance explained in most “surface” models. At predictor level, there are predictors which help them to have distinguishable aspects as “surface” models[57]. So, the surface approach does manifest in the specific educational setting more strongly (in particular through the performance-related metrics).

Hence, it seems that surface approach is more easily captured, flagged-up and potentially discouraged. These conclusions are supported by relevant literature. Entwistle (2008) supports the argument that it is easier to move away from a surface approach rather than move towards a deep approach. Entwistle and Peterson (2004) indicates that a surface approach relates strongly with poor academic performance, hence it is easier to identify.

So, identifying and flagging up the surface approach is possible, and it is more easily discouraged, but, still, improvements in the interactive environment cannot do this alone; there should be improvements in the broader learning and teaching environment (i.e. the degree course) as discussed previously.

6.3.2.2 Possible recommendations for tutors

The implication of this is that the findings can help a tutor to decide whether to intervene; or to intervene at an early stage of the tutorial sessions when there

---

[56] It is beyond the scope of this thesis to make suggestions for improvements in the broader teaching and learning environment of a degree course; however Entwistle (2008), Entwistle and Peterson (2004), Fusaro (2008), and Baeten et al. (2010) make suggestions as to how a deep approach can be encouraged in degree courses.

[57] Although the author cannot go as far as claiming that they can be distinguished from each other, especially when considering the leanest and meanest models. For example, the “leanest and meanest” models of both the “fear of failure” and the “surface” approaches consist of the same predictors (see Appendix 5.3). But in every surface model there are “distinguishable” enough predictors to allow us to characterise them as surface approaches (even in the “lack of purpose” model with only two predictors).

are indications of a surface approach, based on a combination of certain characteristics as expressed by the predictors of each surface model (see sections 5.1.3 and 5.3.6, and Appendix 5.3). More specifically, such indications can be:

- A tendency not to solve exercises on the first try but in subsequent ones, especially with regards to models of the main “surface” and “fear of failure” approaches. This can be also the case with regards to the “syllabus boundness” and “lack of purpose” approaches, but possibly not across various educational settings. This is because the metric *number of exercises solved on first try* does not survive on either of the leanest and meanest models of these two subscales, while *number of exercises solved on third try* does not survive in the leanest and meanest model of the “lack of purpose” subscale).
- A tendency not to solve exercises on the first attempt, but also not solving exercises at all and getting the answers from the ILE (pointing to a possible “gaming” behaviour), especially with regards to the model of “unrelated memorising” and “lack of purpose” approaches. This can be also the case with regards to the main surface approach, but possibly not across various educational settings (as the metric *number of exercises finished but not solved* does not survive in its leanest and meanest model).
- A tendency to conduct a compact path and not possibly expand on the learning material which is supposed to be covered during the tutorials. This is especially the case for the model of the “unrelated memorising” approach. This can be also the case with regards to the main “surface” and “syllabus boundness” approaches, but possibly not across various educational settings (as the metric of *compactness* does not survive in their leanest and meanest models).
- A tendency to revisit the same pages, especially with regards to the model of the “lack of purpose” approach.
- A tendency to cancel exercises, which can be the case for the model of the “syllabus boundness” approach.
- A tendency for an increasing amount of time to be spent on a page with a group of exercises, especially with regards to the “fear of failure” and the main “surface” models. However, because the metric has also a positive relationship with all the deep approaches, there is a need to be considered with the rest of the predictors of these two approaches and especially the performance-related metrics.

The above summary of tendencies can give tutors reasonable indications of a surface approach and help in terms of deciding on intervention when students interact in an ILE during tutorial sessions. However, based on the author's experience as a tutor, it is not realistic to expect a tutor to keep track of all the different combinations corresponding to each approach while being in a class with 20 students during the tutorial sessions[58]. At the same time, in certain cases, such as the *maximum view time on an exercise page*, which has also a positive relationship with the deep scales, it is not advisable to consider a predictor in isolation.

So, if one wants the intervention to be based on greater detail and include the combined predictors of a model, then it is reasonable to do this with the help of a feature in the ILE which does exactly that. For example, in the educational settings of the current study, an AM feature could inform after the first tutorial session of a combination of predictors which characterises the surface approach of a student, and flag it up based on the following combined interactions (or tendencies):

a) tendency not to solve exercises on the first but on a subsequent third attempt;
b) tendency to not solve exercises but get answers from the ILE;
c) tendency to spend an increasing amount of maximum time on a page with a group of exercises;
d) tendency to spend a decreasing amount of a maximum time on a page with theoretical learning content;
e) tendency to use a compact path and not expand on the learning material designated for the tutorial session;
f) tendency to use "concept links".

Furthermore, if one wants to generalise the "surface" tendencies for any future systems, then one would rely on the first four shown above.

However, it is worth mentioning at this point that the study does not necessarily provide evidence that, based on the current "surface" models, it is possible to distinguish one surface approach from another (e.g. "fear of failure" from the rest of the surface approaches). It seems that, in each surface model, there are enough "distinguishable" predictors to characterise it as a surface approach. To support this, it is worth drawing our attention to the leanest and meanest model

[58] This might be also the case for a module leader who is managing a big-size module like in current study.

of the “lack of purpose” approach. In the specific model, there are only two predictors left; however, the predictors *number of exercises finished but not solved* which can point to low performance and possibly a “gaming” behaviour, and *relative amount of revisits* which can point to repetitive overlearning, are good indicators of a surface approach (as was argued in the initial theoretical assumptions).

6.3.2.3 Surface approach – Conclusion on pedagogical insights

To conclude, according to the findings, it is reasonable and possible to:

- give indications of surface approaches to a tutor in a short but easily-comprehensible manner so it can be used in a direct and practical way in class, as follows: a tendency not to solve exercises on first attempt but on subsequent ones or not at all (pointing to a “gaming” behaviour); a tendency to an increasing amount of time spent on a page with a group of exercises (combined with not being able to solve them on first try but on subsequent ones or not at all); a quite compact path (i.e. students do not expand on the learning material they are supposed to examine); a tendency for revisiting the same pages; and a tendency to cancel exercises.
- have a system through which a tutor can be informed in a much more detailed manner about how combinations of characteristics (i.e. tendencies) can point to a surface approach, as suggested previously.

6.3.3 How findings address criticisms in the field of learning styles

As discussed in Chapter 2, there are pedagogical criticisms regarding learning styles. When starting the current investigation, there was a question as to whether a theory which has produced so many theoretical frameworks, constructs and measurements, and inspired a huge amount of empirical work in a variety of contexts in education, still has anything useful to offer. The literature review discussed how the qualitative and quantitative research methods with which Entwistle and his team produced the Approaches and Study Skills Inventory for Students (ASSIST) seemed a way to address these pedagogical criticisms.

The first criticism is with regards to matching or mismatching learning style to instruction. As discussed in Chapter 2, the main aim of most studies which use learning style in a very similar context is eventually to propose ILEs which can be adapted to suit students' style (Papanikolaou et al., 2003; Bajraktarevic et al., 2003; Stash et al., 2004; Paredes and Rodriguez, 2004). However, the current investigation shows that the frequently used, but much criticized pedagogical approach of matching learning style to instruction (Coffield et al., 2004; Curry, 1990; Moran, 1991), is not the only way to use the learning styles in the context of student interaction in learning environments. Entwistle's ASSIST can offer a different pedagogical approach in this context. More specifically, the current empirical findings show that this pedagogical approach can be insightful and enriching with regards to students' interaction in ILEs, and has the potential to improve learning and teaching by helping systems and tutors identify a surface approach towards studying (and ultimately discourage it and improve students' performance).

More specifically, the findings (based on the statistical significance of predictors and their survival in their leanest and meanest models, as shown in Appendix 5.3) show that "surface" approaches consist of certain distinguishable characteristics, which may also manifest in similar studies and similar ILEs. Such characteristics are: the tendency not to solve the exercises on first attempt but on subsequent ones; the tendency for an increasing maximum view time on a page with a specific group of exercises, and a decreasing view time on a page with theory; the tendency to conduct a compact path; the tendency to get the answers to the exercises from the system (pointing to a "gaming" behaviour); the tendency to cancel exercises; and the tendency to revisit of learning material.

The second criticism is whether learning styles can be measured or capture subtleties and complexities of individual human behaviour in real educational settings. In order to address this criticism, this study is set in real educational setting, when students are practising their tutorial exercises during the tutorial sessions and as part of a specific module in a specific course. AM is also populated with the learning material of the module syllabus. Students' intention towards studying is declared in that context, and although this does pose limitations in terms of generalisation, at the same time it addresses the severe aforementioned criticism. The findings show that these subtleties and complexities of individual human behaviour can be captured, and there is a realistic view of students' interaction in a learning environment.

More specifically:

- the "unclear" image we get of the deep approach is not contradictory, as explained in 6.1.4.1. On the contrary, it shows a realistic complexity: students may declare the intention for deep studying but they do not always follow it through, hence their performance can suffer (i.e. solving exercises on third attempt or not at all). This can be due to the broader educational setting, or that the students are not challenged enough, or because of their lack of prior knowledge eventually results in a surface approach.

- the direction of the relationship between the deep approaches and the use of certain tools in AM is not as expected. Concept links tend to be used by those with high scores on surface scales instead of those with high scores on deep scales, which simply shows that a deep approach is not easily encouraged and there is a need for improvements in the interactive learning environment (see section 6.2.2), as well as the broader learning environment (see section 6.3.1). In addition, the AM "notes" feature tends to be accessed both by those with high scores on the "unrelated memorising" scale and by those on the "relating ideas" scale; so the way it is designed again it does not seem to discourage a surface approach, hence the design recommendations in section 6.2.2.
- the surface approach breaks down to approaches which are explained by different combinations of predictors, and in which the highest contributed predictor can be also different (see sections 5.1.3 and 5.3.1). Based on the author's teaching experience, this resonates with the complex but realistic situation a tutor typically faces in a class, where a student can behave in a "surface" manner for different reasons (e.g. anxiety, tendency for rote memorisation, tendency to cope minimally with course requirements etc.). It is reasonable therefore that each one of the surface approaches have similarities but can also dictate different characteristics in the way it interacts with the learning environment, as there are different reasons for triggering each surface approach. This complex but realistic situation is reflected in the current empirical findings. For example, those with high scores on the "fear of failure" scale (which is heavily influenced by the element of anxiety) and on "unrelated memorising" scale (which is influenced by the tendency of rote memorisation) tend not to solve exercises on the first attempt. However, there are differences between the two surface approaches: while those with high scores on the "fear of failure" scale tend to solve exercises on subsequent attempts, it seems that those with high scores on the "unrelated memorising" scale are likely to be given the answers of the exercises by the system (pointing to a "gaming" behaviour). So, the reason behind a surface approach (e.g. anxiety or rote

memorisation) can be responsible for differences in students' interaction in an ILE[59].

To conclude, while a deep approach cannot simply manifest itself strongly, and cannot be captured and encouraged easily, it seems that the surface approach can be captured, and flagged up. It even seems possible potentially to distinguish one surface approach from another, based on differences in interactions, so that a tutor or a system would have evidence whether a "surface" interaction is due to anxiety or to rote memorization. Therefore, the current empirical findings show that the use of ASSIST in real educational settings helps in revealing realistic complexities with regards to students' interaction in learning environments and, in this way, it manages to address the criticism of the field of learning styles.

6.3.4 The influence of prior knowledge

Prior knowledge can influence the surface and deep approach to studying. Ramsden's (2005) insights remind us how important its influence is for the studying approaches. He argues that attempts to understand the material may be disrupted by inadequate "background knowledge" of the relevant field. He emphasises that this is especially the case when the learning task demands that the student has grasped a fundamental concept in scientific subjects. For example, in the context of the current investigation, students during the tutorials had to have grasped the concept of "transpose of a matrix" in order to solve the relevant exercises. Hence, prior knowledge is more frequently related to the approach a student takes towards a task in science disciplines, than in arts and in social science disciplines. It can also influence students' performance, and the way students interact with an ILE by influencing path-related metrics, temporal metrics, and metrics related to use of navigational options (Chen and Ford, 2000; Chen and Paul, 2003; Chen et al., 2016).

Therefore, based on the literature only, there are strong arguments for a researcher to investigate further its influence in the current context. However, what it is not known is how exactly it influences the models representing the deep and surfaces approaches towards studying, models which indicate how

[59] By saying this the author does not claim that they can be completely distinguished based on the findings of the current study, but simply that there are strong indications that there are different elements which constitute the interaction of each surface approach, and that in future studies after applying the recommended improvements a better distinction is possible.

students interact in ILEs during their tutorials sessions in mathematics with regards to their studying approaches.

The discoveries made in the current investigation reinforce what is known in the literature regarding prior knowledge in relation to approaches to studying and students' interaction in ILEs. However, they also contribute further as presented below.

6.3.4.1 Discovery number #1 – Prior knowledge split is increasing the variance of all the "deep" models in the "low prior knowledge" group

The first discovery is that the split into "prior knowledge" groups seems to increase the variance explained by some of the models. More specifically, the variance of all the deep models is increased in the "low prior knowledge" group compared to the "high prior knowledge" group and the whole sample (see Appendix 5.2). To give an indication of how important this increase is for some "deep" models, it is worth re-examining the variance explained when placing them in the "low prior knowledge" group. The variance explained by models representing the main "deep" approach, and by the "seeking meaning" and the "interest in ideas" approaches in the "low prior knowledge" group almost doubles compared to the ones for the whole sample. "Placing" also the "relating ideas" model in the "low prior knowledge" group is responsible for 41.5% of its variance[60]. So, with the exception of the "use of evidence" model where there is a relatively small increase, it seems that prior knowledge has impact on the "deep" models.

6.3.4.2 Discovery number #2 – Prior knowledge split does not really have the same impact in terms of increasing the variance in the surface models

On the other hand, prior knowledge does not have a similar consistent and dramatic impact in terms of increasing the variance in the surface models (either in low or high prior knowledge groups). More specifically, as shown in Appendix 5.2, the models of the main "surface", "fear of failure", and "unrelated memorising" explain relatively large amounts of variance in both high and low prior knowledge groups (and the whole sample). Also, there is no consistency as to where they explain more variance. For example, in main "surface" model and in the "lack of purpose" and "syllabus boundness" models, more variance is

[60] Author estimated the proportional difference versus $R^2_{LowPKgroup}$ in % as follows: $(R^2_{LowPKgroup} - R^2_{WholeSample}) / R^2_{LowPKgroup}$

explained in the “low prior knowledge” group, while in the “fear of failure” model, more variance is explained in the “high prior knowledge” group.

However, as indicated in 5.5.4.2, there are also some useful insights with regards to the “surface” models and prior knowledge, despite the fact that there is not an overall pattern. One of the insights relates to the “fear of failure” subscale, where it seems that even for those students with a high level of prior knowledge, the higher the degree of their anxiety, the more negatively it will affect their performance when solving exercises[61]. Furthermore, with regards to the “unrelated memorising” subscale, independently of the level of prior knowledge, the degree of rote memorisation seems to be the one which mainly affects students’ interactions[62].

The findings described in sections 6.3.4.1 and 6.3.4.2 reinforce that students with low prior knowledge in a subject are more likely to follow a surface approach towards studying (Entwistle and Peterson, 2004); and prior knowledge can influence the decision to follow a deep approach towards studying (Entwistle, 2008). In addition, the current investigation indicates that those students who declared a deep approach towards studying, might just have behaved in a manner with “surface” characteristics or tendencies while practising their exercises in class, because their low prior knowledge is a hindrance. This can explain why there is such a big difference in terms of variance explained between the “low prior knowledge” group of most of the “deep” models, and the “high prior knowledge” group. This confirms that deep approaches to studying –except from the “use of evidence” one – seem to be more sensitive to this specific factor. It also resonates with the conclusions in section 6.3.1 which also show that the deep approaches are generally sensitive to the broader learning environment in which a study takes place.

[61] As shown in Appendix 5.2, the variance explained R^2, although large in both ‘fear of failure’ models of low and high prior knowledge groups, is larger in the “high prior knowledge” group, compared to the “low prior knowledge” group by 11.3%. So, it seems that the performance-related metrics seem to explain better the model in the “high prior knowledge” group.

[62] As shown in Appendix 5.2, the variance explained R^2 between the ‘unrelated memorising’ models of low and high prior knowledge groups differs by only 1.4%.

6.3.4.3 Discovery number #3 – Prior knowledge split seems to explain better the "paradoxical" findings at least in "low prior knowledge" groups of deep models

It seems to explain the paradoxical mixed "deep" and "surface" tendencies found in the deep models. More specifically, it makes sense that:

- In the "low prior knowledge" group, highly intrinsically motivated students tend not to solve exercises on the second or even third try. On the contrary, students seem not solve the exercises at all and get the answers from AM (see the suggested model in section 5.2.3). It seems that the students' prior knowledge greatly influences performance when solving tutorial exercises with regards to the "interest in ideas" approach.
- In the "low prior knowledge" group, while students with high scores on the "seeking meaning" scale tend to solve exercises on the first try, there is at the same time a tendency to not to solve exercises on the second, but on the third try.
- In the "low prior knowledge" group, while students with high scores on "relating ideas" scale tend to solve exercises on the first try, there is also a tendency not to solve exercises on the second try and not solve exercises at all but get the answers from AM. Furthermore, those with high scores on the subscale have also the tendency not to use hyperlinks in order to relate concepts. This can be a design issue, as discussed in 6.2.1, but it can be also an issue of low prior knowledge as there are not as many old concepts to make connections to.
- In the "low prior knowledge" group, while students with high scores on the main "deep" scale tend to solve exercises on the first try, there is also a tendency not to solve exercises on the second try or solve exercises at all but get the answers from AM.

6.3.4.4 Discovery number #4 –Prior knowledge seems to be particularly influential for the "interest in ideas" deep scale

The model "*interest in ideas"* has the highest increase (compared to any other model) in terms of variance R^2 in the "low prior knowledge" group. It increases by 16.4% from 15.7% for the whole sample to 32.4% for the "low prior knowledge" group (as shown in section 5.5.3.2 and Appendix 5.2). This is not surprising, as literature shows interest in a subject and prior knowledge go hand in hand, and one should be seen in light of the other. Ramsden (2005) argues that it is expected to find that interest and background knowledge are related to each other in the natural setting of student learning. More specifically, it is argued that comprehending entails the use of prior knowledge, and interest in a subject empowers students to direct this knowledge in order to synthesise it with new pieces of knowledge (Brophy, 2010). So, high interest in the subject in combination with a deep approach towards studying may activate prior knowledge in order to make connections with new concepts. It is reasonable, therefore, to say that the lack of prior knowledge does not give the opportunity to students to use their declared interest in the subject in this way, hence it does not manifest strongly. The current findings for the "interest in ideas" approach and prior knowledge show in a stronger way (compared to any other deep scales) that students in the "low prior knowledge" group have "high surface" tendencies in the way they interact, because declaring a high interest or low interest is not as influential as their lack of knowledge. This is further supported in the literature, where it is argued that prior knowledge has a more direct effect in a learning environment compared to interest in the subject (Tobias, 1994). In the current investigation, this is confirmed, but the findings go a step further as they indicate that this direct effect is stronger for the "low prior knowledge" group.

6.3.4.5 Discovery number #5 – Prior knowledge is not particularly influential for the "use of evidence" deep scale

It seems that being in a low or high prior knowledge group does not really explain the variance of the "use of evidence" model more. As shown in section 5.5.3.2 and Appendix 5.2, the variance R^2 remains at medium range and increases very little for both high and low prior knowledge groups, with 18.7% for the whole sample, and 19.3% for the "high prior knowledge" group and 21.3% for the "low prior knowledge" groups. It is possible that this is because the design of an ILE may influence the interactions of students with high scores on the "use of evidence" subscale (those with tendencies to a "serialist" approach as discussed in 6.2.1) more, compared to prior knowledge. First of all, the influence of the design of an ILE for those with tendencies to a "serialist" approach is indicated in the current background research, which shows that since the late 1990s there is a body of empirical work with regards to design of interactive environments and their preferences. Secondly, there is more recent empirical evidence which shows that this connection is still relevant. More specifically, Chen et al. (2016) point to empirical evidence which shows that if an ILE does not have suitable design choices for those who have tendencies to a “serialist” approach, then it can have negative effects for them[63]. This supports what is claimed previously that the design environment can have more influence on the “use of evidence” approach (i.e. it can be one of the reasons it does not emerge strongly). It also resonates with what it is discussed in 6.2.1, where there are suggestions to offer in terms of design more suitable features for those with high scores on the “use of evidence” subscale.

So, overall, these discoveries indicate that mostly the online interactions of deep approaches towards studying (except for the “use of evidence” one) seem to be explained better in the “low prior knowledge” group. It also shows that students’ intention to interact in a “deep” manner is not enough and those students may interact in a manner that manifests “surface” tendencies, when there is lack of prior knowledge of a subject. This resonates with what has been observed specifically with regards to prior knowledge in mathematics. Ramsden (2005) observes that prior knowledge is mentioned more in science students as a factor which can lead to lack of comprehension, anxiety, superficial learning and passiveness. Ramsden (2005, p.201) gives the example of a science student who “describes how his previous knowledge of a type of problem helps him to take a deep approach, while his weakness in a basic mathematical concept makes his approach to another part of the same question anxious, passive and superficial*”.*

These conclusions can be useful and quite enriching information for tutors and for the design of ILEs. For example, it may be useful for a tutor to know that the reason a student interacts with an ILE in a manner which indicates “surface” tendencies is lack of prior knowledge in a concept. Practically speaking, though, in a class with 20 students during a tutorial session it is not always easy to identify lack of prior knowledge of a concept and tackle it. However, an ILE and particularly an intelligent one as we described in 6.2.1 can help do that in a more systematic way, evidenced from the interaction with the ILE. In addition, in 6.2.1, there is already a design suggestion of a digital concept map incorporated in ILE (where students will have the opportunity to highlight old concepts and link them to new concepts) and which will be checked and given feedback by the tutor. In this way, the tutor will be able more easily to identify

[63] In the study of Chen et al. (2016), the holist/serialist construct is not measured through ASSIST; but through SPQ measurement which is based on Pask’s theoretical framework, similarly to the “use of evidence” and “relating ideas” approaches.

and tackle gaps in students’ prior knowledge. Furthermore, the “notes” feature, as suggested in 6.2.1, can also help a tutor understand if there are gaps in a student’s knowledge, as students can publish questions which can reveal lack of prior knowledge in a concept (current records on students’ notes in this study indicate that this can occur).

Besides helping the tutor to identify effectively and quickly students’ gaps in prior knowledge, the “notes” and “concept map” features offer a direct solution which facilitates students’ prior knowledge. As discussed in 6.2.1, a feature like a concept map can strengthen and activate students’ prior knowledge, so it can serve both those students who need to identify gaps in their prior knowledge, and those who have a good degree of prior knowledge but need to start activating and reflecting on it.

The above design recommendations point to an ILE which encourages a deep approach: by relating teaching to prior knowledge and facilitating students to reflect on the relationship between old and new concepts, one can create a more student-oriented learning environment that supports a deep approach (Entwistle and Peterson, 2004).

6.4 Latest developments in the field of ILEs since data collection took place

Since the time the data collection took place, there are a number developments in the field of ILEs. The following sections discuss the ones, which according to the author’s opinion seem to be the most commonly discussed and important with regards to future work. They also discuss how the current investigation can be still relevant to these developments, which are inevitable as the technology in ILEs evolves rapidly.

6.4.1 An evolution toward more flexible and most cost-effective ways of updating and implementing materials on ILEs serving different learning purposes.

As discussed in 2.1.1.2 there are practical considerations which may influence the choice of an ILE both for use in class and research. Since the time during which the data collection took place, there has been an evolution in terms of the development of mathematical ILEs. This positive evolution, has allowed practitioners to find more effective ways time-wise and cost-wise to deal with the creation and updating of learning materials as well as incorporating those materials in ILEs. Older open source ILEs such as Tall’s graphic calculus approach[64], discussed in the review by (Crowe and Zand, 2001; Crowe and Zand, 2000b), although useful at the time, could not allow much variety in terms of features and materials, and flexibility in terms of adapting and incorporating learning materials for the learning objectives of a specific course or module because of aforementioned practical reasons. At the time the current research was taking place, there were of course open source learning management systems such as Moodle which could offer relatively easy incorporation of learning materials. However this type of ILEs has a more generic learning purpose as it serves a number of different subjects.

Based on the example of the evolution of a mathematical ILE such as GeoGebra (see Hohenwarter, 2015), we can now say that there are more effective solutions for practitioners to deal with the aforementioned practical considerations by offering a flexible authoring environment for creating and uploading materials or offering to download applications such as graphic calculators [65]. This type of development helps an ILE to serve as a better “means to an end”, as discussed later on in 6.4.5.

6.4.2 Massive Open Online Courses (MOOCs)

MOOCs is not a newfound type of ILE [66], however, the importance and popularity has increased in recent years in higher education (Kaplan and Haenlein, 2016; Yuan and Powell, 2013; Artigue, 2013). Its two key features are: open access (i.e. anyone can participate in an online course for free); and scalability (i.e. courses are designed to support an indefinite number of participants) (Yuan and Powell, 2013). Yan and Powell (2013) also indicate common features which encourage collaboration, apply visualisation (videos), and offer personalised experience. At the same time, they raise concerns with regards to the quality assurance (related to the lack of structure as there is no instructor in central role), and the assessment and credit.

A good example of a non-profit MOOC is Khan’s Academy, which along with other subjects for different educational levels, it also offers algebra topics for

---

[64] See: http://www.graphiccalculus.co.uk/

[65] See: https://www.geogebra.org/m/W7dAdgqc

[66] Yuan and Powell (2013) place the first early examples in 2008. Although, the author observes that the foundations for one of the most pioneering non profit MOOC, MIT OPEN Courseware, started back in 2001 (see: https://ocw.mit.edu/about/).

higher education. It is an example which serves well the criteria, discussed in 2.1.1.1, with a variety of features (see Appendix 6.3). It is also a well-balanced example of effective scaffolding and self-directed learning. While it gives the freedom for individual exploration, at the same time it indicates a clear structure regarding the order of topics and gives guidance through self-diagnosed activities and tests to start from an appropriate topic and level of knowledge. It is an example which satisfies experts who support that ILEs should consider individual preferences and allow students more control (Joshi, 2017), but it also satisfies experts who are concerned about ILEs relying solely on discovery, and who support a more guided mode of learning and more structured ways of collaboration and solving problems (Engelbrecht and Harding, 2005a).

6.4.3 Use of programming languages have gained popularity

A recent development seems to be that programming languages such as MATLAB[67] seems to win over CAS in terms of the use of ILEs in classrooms in higher education (and more specifically in the context of mathematical courses which offer also computer programming modules in UK universities) (Sangwin and O'Toole, 2017).

However, it seems that MATLAB was not always the most popular choice for courses with pure mathematics. In older reviews such as Lavicza (2010), CAS seemed to be a relatively more popular choice. Sangwin and O' Toole (2017) indicate that in mathematical undergraduate courses MATLAB is more frequently used in first-year courses compared to CAS, such as Maple. Furthermore, according to their findings Mathematica and GeoGebra are not used by single institutions, while open source ILEs such as Derive, Octave, Scilab, Maxima are not used anymore. They suggest that the use of an ILE depends on its tight practical integration, referring also to its availability outside the labs and guidance for its use. However, based on the opinions of teaching staff inquired in the same study, there seem to be also other reasons for the involvement of programming languages in mathematical courses (e.g. belief that programming should be compulsory and/or that there is a need for employability reasons).

[67] It has been developed over the years from numerical analysis in linear algebra classes to a programming language package with features for advanced visualisation, large scale modelling, data analysis, etc. (Moler, 2004; Moler, 2006; MathWorks, 2017).

6.4.4 Gravitation towards M learning tools

Since the review of Crowe and Zand (2000), where web-based ILEs seemed to be one of the mathematical ILE choices, now not only they are a necessity but there is also a gravitation to be accessed in mobile devices. (Artigue, 2013) points out to this evolution. Reviews such as (Taleb et al., 2015) support that the use of m learning tools can: facilitate distance learning, be more cost-effective, promote collaborative learning, and engage more students. They also support that m learning tools which provide programming aspects for mathematics and/or multimodality (through writing, voice, and graphics) can increase comprehension of mathematical concepts and students' motivation and self-confidence. However, it is reasonable to suggest that incorporating ILEs on mobile devices, would also require aspects and features, which Taleb et al. (2015) characterise as 'trust-built' factors, such as collaborative features, and activities which offer support (e.g. feedback, rewards, external auditing, and personalisation).

6.4.5 Latest developments and current investigation

The above developments give food for thought as to how the present and future of ILEs is shaping and pose a challenge for every researcher who conducts an investigation in the field. As Hoyles and Noss (2003) point out there is always a danger that the rapid changes in technology often result in research being outdated by the time it is completed. In this sense, there is always going to be a question as to whether any latest developments in ILEs are still relevant to an investigation in the specific field by the time it is completed.

A way to deal with such an issue is to use any ILE involved in an investigation as a "means to an end". More specifically, the use of AM as the ILE of choice for this research, although is justified as discussed in 2.1.1.4, it serves towards establishing associations between students' interaction and their approaches towards studying. Although there can be constraints as to the generalisation of findings, the methodological reflections and pedagogical insights suggested in this section can serve as good starting point for future investigations involving the recent developments discussed in 6.4.2, 6.4.3, and 6.4.4. For example, understanding which predictors can be the most contributing in future models representing students' interaction in a MOOC, in a MATLAB software, or in a mobile application, can be a good starting point for a future investigation. Or understanding that a deep approach does not necessarily manifest strongly based solely on the use of an ILE, but it depends also on the wider educational setting[68], can be a useful insight which can help towards having reasonable expectations from findings. In this sense the current investigation offers insights that can be relevant and transferable to future investigations regardless of the exact ILE; an issue which is also further discussed in sections 7.2 and 7.3.

Finally, with regards to the development discussed in 6.4.1, for an ILE to serve as a “means to end” in the future investigations it is essential to include a cost-effective and time-effective authoring aspect, as it has become increasingly important to find ways which allow researchers to catch up with the rapid technological developments and changes in trends in the field of mathematical ILEs.

[68] It is worth mentioning at this point that if the educational setting consists solely of a MOOC environment, which offers the possibility of a complete online course and not just a module, then this will certainly give more control to researchers doing similar investigations to ensure the stronger manifestation of the deep approach.

Chapter 7 - Conclusion

The current research starts with the aim of investigating how first-year undergraduate students with a deep approach and those with a surface approach to studying interact when using an ILE such as ActiveMath for mathematics in tutorial sessions, in the classroom, in real learning conditions. This investigation can ultimately help towards finding ways to support first-year undergraduate students in the subject of mathematics by capitalising on the potential of interactive learning environments, and looking into how they can be more effective in classroom by taking into account students' individual characteristics, and specifically their studying approaches. The focus of the current investigation into supporting first-year students in their practical sessions in mathematics is motivated by the challenges tutors face in the classroom. These challenges have long been identified and discussed in the literature of higher education, but they also resonate with the author's own experience in teaching first-year undergraduates.

7.1 The challenges

As stated in the literature, first-year undergraduates can demonstrate different approaches towards their studying which have their origins in their schooling experiences (Entwistle and Peterson, 2004). They are likely to have already formed preferred or habitual ways of approaching studying, which can influence the way they approach their undergraduate studies in either a negative or a positive way. A deep approach towards studying, for example, which is about seeking the meaning of the concepts and procedures and trying to find meaningful relationships amongst them, may lead to good academic performance (Entwistle, 1997a; Entwistle and Ramsden, 1983). On the other hand, a surface approach towards studying, which is about treating the learning material as unrelated bits of knowledge and coping minimally with the course requirements, can eventually have a negative influence on academic performance (Entwistle, 1997a; Entwistle and Ramsden, 1983). Changing habitual surface approaches towards studying can be challenging for tutors in classroom, as it requires first identifying and keeping track of students' studying approaches at an early stage (which becomes even more challenging when tutors have to manage and teach large undergraduate classes).

Furthermore, challenges with regards to the way students approach their studying have been also identified in mathematics education. For example, the challenge of discouraging rote memorisation of formulas and rules and encouraging a deep understanding of mathematical procedures is quite prominent (Ladson-Billings, 1997; Crowe and Zand, 2000b; Sangwin, 2004). Similar concerns have been raised by Liston and O'Donoghue (2009) and Saha et al. (2015) who point to research which indicates that students often carry out mathematical procedures without really understanding the concepts involved, and that they focus on each procedure separately rather than trying to find connections between different parts of mathematics, or between methods and concepts.

The thesis demonstrated that "interaction" metrics derived from the use of an interactive learning environment when students are practising their exercises in tutorial sessions can help towards identifying students' surface approaches to studying. But it is important to go a step further and design learning environments which stimulate students to develop productive strategies (Entwistle and Peterson, 2004). Especially with regards to digital environments, an empirical investigation on their use can help us understand how it is possible to use and design them in a way that encourages a deep approach towards studying and discourages a surface one.

7.2 Implications and limitations of findings

During the current investigation, it has become apparent that Entwistle's instrument reflects an educational philosophy that is based on a rich theoretical background and decades of research that recognises subtleties and offers a realistic view of a student's intentions and actions in a learning environment.

However, there was initially a question as to whether the rich theory behind the ASSIST inventory could allow for theoretical assumptions in the specific context. More specifically, it was questionable whether the inventory would eventually help to identify surface and deep approaches towards studying, given there was no similar research to uncover relationships between the students' approaches towards studying and students' interaction with ILEs during tutorial sessions in class.

The findings indicate that it is possible to identify tendencies in students' interactions in an ILE such as AM, which are linked to surface approaches (e.g. the performance-related interaction metrics and particularly *number of exercises solved on first try,* which consistently contributes to all the "recommended" surface models, and has the highest contribution and survives in 3 out of 5 of the leanest and meanest versions of the surface models). This is a positive outcome of the current investigation, as it can serve towards creating interactive environments which assist tutors during practical sessions with the challenge of identifying and keeping track of students' approaches towards studying quickly and effectively, especially in large undergraduate classes.

The findings also indicate that it is not easy to identify deep approaches, as there are tendencies in the deep models which theoretically and empirically are linked to surface approaches (i.e. unexpected relationships between "deep" scales and "interaction" metrics which better suit the assumptions made for the surface scales). Furthermore, the findings show that in the "deep" models there is simply not as much variance explained as there is in the "surface" models. However, valuable conclusions can be also drawn from the findings with regards to the deep approaches. In the context of the current investigation, deep approaches seem to be sensitive to the broader learning and teaching environment in which the interactive environment is used. In addition, it is shown empirically that prior knowledge of students in mathematics can influence deep approaches. More specifically, the variance of most "deep" models increases, and in certain models it is twice as much, in the "low prior knowledge" group compared to the "deep" models for the whole sample. As discussed in sections 5.5.3 and 6.3.4, a possible reason behind this increase is that there are certain "surface" tendencies in the "deep" models, which manifest more strongly in the group with low prior knowledge in mathematics, compared to the "high prior knowledge" group.

Despite the fact that the findings with regards to the "deep" models do not give as many distinguishable aspects in the students' interaction with the ILE as those for the "surface" models, overall the findings do show that the approaches towards studying as measured by ASSIST can be linked to the realistic complexities of the students' interaction with an ILE during tutorial sessions. For example, the "deep" model consists of predictors with both an unexpected relationship to the "deep" scale (such as *number of exercises solved but not finished*) and with an expected relationship, which also gives a distinguishable aspect, to the "deep" scale (such as *number of exercises solved on first try)*. This finding expresses the realistic complexity, that a student with a high score on the "deep" scale tends to solve exercises on first try, but tends also not to solve exercises at all (possibly because of his low prior knowledge, considering that the contributions of the predictors in the "deep" model are explained better in the "low prior knowledge" group).

To conclude, certain unexpected relationships found in the "deep" models can be behind the differences between the variance explained amongst the models of the group. Furthermore, as previously discussed in section 5.1.2, the contribution of certain predictors with distinguishing aspects for the deep scales such as *number of exercises solved on first try* could be responsible for the unexplained variance of the model for the whole sample. It is possible that they could also be responsible for this difference in variance between the models of the “low prior knowledge” and “high prior knowledge” groups. It could be the case that the specific predictor, for example, does not have a strong enough contribution to represent better the interactions of the “high prior knowledge” group across the deep scales. Or perhaps it is not possible for this predictor to contribute more to the deep models of the “high prior knowledge” group (i.e. the number of exercises solved on first try by students with a high level of prior knowledge could be simply independent of their high and low scores on the deep scales).

Capturing these realistic complexities of students’ interactions with regards to the approaches towards studying can however be at the expense of the generalization of findings. The students’ intention towards studying is declared in the specific context of a module, which belongs to a specific course and in a specific university. The advantage is that such usage of learning styles provides a realistic and complex view with regards to students’ interaction in ILEs. However, this at the same time comes at an expense, as it poses limits to the generalisation of the findings, and ultimately recommendations for future studies. So, making recommendations based on findings which cannot be fully generalised can be challenging, but it does not mean that that it cannot be done and that it cannot give a good starting point for future studies in different educational settings. To add clarity to the whole process of predictor selection and give more concrete recommendations, a differentiation can be made between the predictors which survive the selection process and make it into the suggested (or recommended) models (the ones with the highest R^2 and Adjusted R^2) and those which survive and make it in the leanest and meanest models (where all predictors are statistically significant). More specifically, a distinction is possible according to:

- the predictors which should be considered, because they survive in the leanest and meanest versions of the models and therefore can be relied on more for future studies;

- the predictors which can be considered but with a certain degree of caution because they enrich the suggested (or recommended) versions of the model (by adding further aspects with regards to students' interactions), but they can be sensitive across different educational settings and students' prior knowledge. This is especially the case for the models of the "deep" approach which, as discussed previously, is sensitive to change (e.g. there are predictors such as the number of tries on exercises whose relationship to the deep scale may change if the broader educational setting encourages more a deep approach towards studying).

Furthermore, tendencies found in both "deep" and "surface" models with regards to the use of the AM "notes" and "concept links" features pointed towards possible improvements in these AM features (see section 6.2.1), which may help towards encouraging a deep approach to studying. However, in this case as well there is a need for further investigation which will indicate whether the recommended changes in those features can indeed strengthen the links between their use and the deep approaches towards studying.

7.3 Future work

With regards to future work, the current investigation has opened up possibilities for a number of studies in the field.

First of all, a natural continuation would be a study with a similar sample and in similar education settings (e.g. a university of a similar league, a similar course and module, and similar variation in terms of prior knowledge). It would require, however, the recommended changes in terms of predictors, discussed in chapter 6, to take place in an effort to capture better the deep approaches towards studying. It could also take into account the recommendations for design modifications and features that could further encourage a deep approach. Furthermore, a similar study would give an opportunity to re-examine the predictors of the "surface" models and re-affirm their distinguishing aspect but also go a step further and distinguish one surface approach from another. For example, it is worth verifying further tendencies which can distinguish the "fear of failure" approach from the "unrelated memorising" approach, as it would be useful for a tutor to know that a student behaves in a "surface" manner because of anxiety or rote memorisation.

Secondly, as deep approaches are sensitive to the educational setting, it would be worth investigating them in a teaching and learning environment where they might manifest more strongly. For example, in a high league university but in a similar course and module, students' intention to follow a deep approach might manifest more strongly because as Ramsden (2005) indicated students might perceive that the teaching and learning environment required it from them.

Thirdly, another possible line of future investigation could be with regards to the influence of prior knowledge on the deep and surface scales in a study where it would be the prime concern. It would be worth verifying, for example, whether its influence is still stronger on the deep approaches compared to surface approaches. However, it would also be worth going a step further and examining the models at predictor-level, as this would give further insights into what specific interactions in an ILE represent students' deep and surface approaches at a specific level of prior knowledge.

The current investigation opens up, therefore, various lines of investigation whose combined knowledge has the potential to lead to ILEs with more intelligent features which can help tutors deal with challenging situations which typically manifest in classrooms when teaching first-year undergraduates. It can, for example, help tutors identify at an early stage students' habitual "surface" approaches towards studying, as well as helping tutors support students in adopting more productive approaches.

References

ABDULWAHED, M., JAWORSKI, B. & CRAWFORD, A. 2012. Innovative approaches to teaching mathematics in higher education: a review and critique. *Nordic Studies in Mathematics Education,* 17**,** 49-68.

AL-AZAWEI, A. & BADII, A. 2014. State od the Art of Learning Styles-Based Adaptive Educational Hypermedia Systems (LS-BAEHSs). *International Journal of Computer Science and Information Technology,* 6.

ALESSI, S. & TROLLIP, S. 2001. *Multimedia for Learning: Methods and Development,* Boston, Allyn & Bacon.

ALI, R., BAKAR, Z. A. & AKHTAR, N. 2014. The demarcation of cognitive and learning style: Myth or reality as an impediment in educational research. *Journal of Psychological and Educational Research,* 22**,** 76-101.

ALLINSON, C. & HAYES, J. 1997. *RE: Cognitive Style Index (CSI).*

ALLINSON, L. 1992. Learning Styles and Computer-Based Learning Environments. *In:* TOMEK, I. (ed.) *Lecture Notes in Computer Science.* Wolfville: Springer.

ARMSTRONG, S., PETERSON, E. & STEPHEN, R. 2011. Understanding and defining cognitive style and learning style: a Delphi study in the context of educational psychology. *Educational Studies,* 38.

BAETEN, M., KYNDT, E., STRUYVEN, K. & DOCHY, F. 2010. Using student-centred learning environments to stimulate deep approaches to learning: Factors encouraging or discouraging their effectiveness. *Educational Research Review,* 5**,** 243-260.

BAJRAKTAREVIC, N., HALL, W. & FULLICK, P. 2003. ILASH: Incorporating Learning Strategies in Hypermedia. *14th ACM Conference on Hypertext and Hypermedia.* Nottinham.

BAKER, R., WALONOSKI, J., HEFFERNAN, N., ROLL, I., CORBETT, A. & KOEDINGER, K. 2008. Why Students in "Gaming the System" Behavior in Interactive Learning Environments. *Journal of Interactive Learning Research,* 19**,** 185-224.

BARAB, S., BOWDISH, B. & LAWLESS, K. 1997. Hypermedia Navigation: Profiles of Hypermedia Users. *Educational Technology Research and Development,* 45**,** 23-41.

BEATY, L., GIBBS, G. & MORGAN, A. 1997. Learning Orientations and Study Contracts. *In:* MARTON, F., HOUNSELL, D. & ENTWISTLE, N. (eds.) *The Experience of Learning - Implication for Teaching and Studying in Higher Education.* Scottish Academic Press.

BENG, Y. H. & YUNUS, A. S. B. M. 2013. A holistic approach to activate and enhance prior knowledge a tertiary learners in the upcoming lectures of calculus. *World Applied Scienes Journal,* 21**,** 156-161.

BERENDT, B. & BRENSTEIN, E. 2001. Visualizing individual differences in Web navigation: STRATDYN, a tool for analyzing navigation patterns. *Behavior Research Methods, Instruments, & Computers,* 33**,** 243-257.

BERRY, J., GRAHAM, E. & SMITH, A. 2006. Observing student working style when using graphic calculators to solve mathematical problems. *International Journal of Mathematical Education,* 37**,** 291-308.

BORBA, M. 2009. Potential scenarios for Internet use in the mathematics classroom. *ZDM The International Journal on Mathematics Education,* 41,**,** 453-465.

BORBA, M., CLARKSON, P. & GADANIDIS, G. 2013. Learning with the use of the Internet. *In:* CLEMENTS, M. A., BISHOP, A. J., KEITEL, C., KILPATRICK, J. & LEUNG, F. K. S. (eds.) *Third International Hanbook of Mathematics Education.*

BOTAFOGO, R., RIVLIN, E. & SHNEIDERMAN, B. 1992. Structural Analysis of Hypertexts: Identifying Hierarchies and Useful Metrics. *ACM Transactions on Information Systems,* 10**,** 142-180.

BROPHY, J. 2010. *Motivating students to learn,* New York, Routledge.

BRUSILOVSKY, P. 2001. Adaptive Hypermedia. *User Modeling and User-Adapted Interaction,* 11**,** 87-110.

BUTEAU, C. & MULLER, E. Evolving technologies integrated into undergraduate mathematics education. *In:* SON, L. H., SINCLAIR, J. B., LAGRANGE & HOYLES, C., eds. Proceedings of the ICMI Study 17 Conference: Digital technologies in mathematics education-Rethinking the terrain, 2006 Hanoi, Vietnam.

CAZES, C., GUEUDET, G., HERSANT, M. & VANDEBROUCK, F. 2006. Using e-exercises bases in mathematics: case studies at university. *International Journal of Computers for Mathematical Learning,* 11,**,** 327-350.

CHALMER, B. 1987. *Understanding Statistics,* New York, Marcel Denker Inc.

CHEN, S. & FORD, N. 2000. Individual differences, hypermedia navigation and Learning: An empirical study. *Journal of Educational Multimedia and Hypermedia,* 9**,** 281-311.

CHEN, S., HUANG, P.-R., SHIH, Y.-C. & CHANG, L.-P. 2016. Investigation of multiple human factors in personalised learning. *Interactive Learning Environments,* 24**,** 119-141.

CHEN, S. & MACREDIE, R. 2002. Cognitive Styles and Hypermedia Navigation: Development of a Learning Model. *Journal of the American Society for Information Science and Technology,* 53**,** 3-15.

CHEN, S. & PAUL, R. 2003. Editorial: Individual differences in web-based instruction - an overview. *British Journal of Education Technology,* 34**,** 385-392.

COCKBURN, A. & MCKENIE, B. 2001. What do users do? An empirical analysis of the web use. *International Journal of Human-Computer Studies,* 54**,** 903-922.

COFFIELD, F., MOSELEY, D., HALL, E. & ECCLESTONE, K. 2004. Learning styles and pedagogy in post-16 learning. A systematic and critical review. London: Learning and Skills Research Centre.

COHEN, J. 1988. *Statistical Power analysis for behavioral science,* Hillsdale, NJ, Eribaum.

COHEN, J. 1992. Quantitative Methods in Psychology - A power primer. *Psychological Bulletin,* 112**,** 155-159.

COHEN, L., MANION, L. & MORRISON, K. 2000. *Research Methods in Education,* London, RoutledgeFalmer.

COLLIS, B. & MOONEN, J. 2001. *Flexible Learning in digital world*, Cogan Page Limited.

CROWE, D. & ZAND, H. 2000a. Computers and undergraduate mathematics 3: Internet resources. *Computer & Education,* 35,**,** 123-147.

CROWE, D. & ZAND, H. 2000b. Computers and undergraduate mathematics I: setting the scene. *Computer & Education,* 35**,** 95-121.

CROWE, D. & ZAND, H. 2001. Computers and undergraduate mathematics 2: on the desktop. *Computer & Education,* 37,**,** 317-344.

CURRY, L. An organisation of Learning Styles Theory and Constructs. Proceedings of the Annual Meeting of the American Educational Research Association, 1983 Montreal. U.S. Depertment of Education.

CURRY, L. 1987. Integrating Concepts of Cognitive or Learning Style: A review with attention to psychometric standards. Jamaica, N.Y: Centre for the Study of Learning and Teaching Styles, St John's University.

CURRY, L. 1990. A Critique of the Research on Learning Styles. *Educational Leadership,* 48**,** 50-56.

CURRY, L., PHD 1999. Cognitive and Learning Styles in Medical Education. *Pedagogue,* 9,**,** 1-6.

DE BELLO, T. C. 1990. Comparison of eleven major learning models: variables, appropriate populations, validity of instrumentations and the research behind them. *Reading, Writing, and Learning Disabilities,* 6,**,** 203-222.

DE BRA, P. & CALVI, L. 1998. AHA: a Generic Adaptive Hypermedia System.

Proceedings of the Second Workshop on AdaptiveHypertext and Hypermedia. Pittsburgh.

DE GRAVE, W., DOLMANS, D. & VAN DER VLEUTEN, C. 1999. Profiles of effective tutors in problem-based learning: scaffolding student learning. *Medical Education,* 33**,** 901-906.

DE VAUS, D. 2002. *Surveys in Social Research,* London, Routledge.

DEBORAH, L. J., BASKARAN, R. & KANNAN, A. 2014. Learning styles assessment and theoretical origin in an E-learning scenario: a survey. *Artificial Intelligence Review,* 42**,** 801-819.

DIX, A., FINLAY, J., ABOWD, G. & BEALE, R. 1998. *Human-computer interaction,* Harlow, Pearson Education Limited.

DRIJVERS, P. 2016. EVIDENCE FOR BENEFIT? REVIEWING EMPIRICAL RESEARCH ON THE USE OF DIGITAL TOOLS IN MATHEMATICS EDUCATION. *13th International Congress on Mathematical Education.* Hamburg.

DRISSI, S. & ABDELKIRIM, A. 2012. An adaptive educational hypermedia intergrating learning styles: Model and experiment. *International Conference on Education and e-Learning Innovations (ICEELI).* IEEE.

DUNN, R. 2000. Capitalising on College Students' Learning Styles: Theory, Practice and Research. *In:* DUNN, R. & GRIGGS, S. (eds.) *Practical Approaches to Using Learning Styles in Higher Education.* London: Greenwood Publishing Group.

DUNSER, A. & JIRASKO, M. 2005. Interaction of Hypertext forms and global versus sequential learning styles. *J.Educational Computing Research,* 32**,** 79-91.

EHRMAN, M. 1990. Psychological Factors and Distance Education. *The American Journal of Distance Education,* 4**,** 10-24.

EINSTEIN, G., MORRIS, J. & SMITH, S. 1985. Note-Taking, Individual Differences, and Memory for Lecture information. *Journal of Educational Psychology,* 77**,** 522-532.

ENGELBRECHT, J. & HARDING, A. 2005. Teaching Undergraduate Mathematics on the Internet - Part 1: Technologies and Taxonomy. *Educational Studies in Mathematics,* 58**,** 235-252.

ENTWISTLE, N. 1981. *Styles of Learning and Teaching- An integrated outline of Educational Psychology,* Chichester, John Wiley and Sons.

ENTWISTLE, N. 1997a. *Approaches and Study Skills Inventory for Students (ASSIST)* [Online]. ETL Project. Available: http://www.etl.tla.ed.ac.uk/questionnaires/ASSIST.pdf [Accessed 07 / 1 / 13].

ENTWISTLE, N. 1997b. Contrasting Perspectives on Learning. *In:* MARTON, F., HOUNSELL, D. & ENTWISTLE, N. (eds.) *The Experience of Learning - Implication for Teaching and Studying in Higher Education.* Scottish Academic Press.

ENTWISTLE, N. 1998. Supporting students' framework for conceptual understanding: knowledge objects and their implications. *In:* RUST, C. (ed.) *Improving student learning: Improving Students as Learners.* Oxford: Oxford Brookes University, Oxford Centre for Staff and Learning Development.

ENTWISTLE, N. 2001. Styles of learning and approaches to studying in higher education. *Kybernetes,* 30**,** 593-603.

ENTWISTLE, N. Taking stock: teaching and learning research in higher education. Teaching and Learning Research in Higher Education, 2008 Guelph, Ontario.

ENTWISTLE, N. & HANLEY, M. 1977. Personality, Cognitive Style, and Students' Learning Strategies. *Higher Education Bulletin,* 6**,** 23-43.

ENTWISTLE, N., HANLEY, M. & HOUNSELL, D. 1979. Identifying distinctive approaches to studying. *Higher Education,* 8,**,** 365-389.

ENTWISTLE, N., MCCUNE, V. & WALKER, P. 2001. Conceptions, styles and approaches within higher education: analytic abstractions and everyday experience. *In:* STERNBERG, R. J. & ZHANG, L.-F. (eds.) *Perspectives on thinking, learning and cognitive styles.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

ENTWISTLE, N. & PETERSON, E. 2004. Conceptions of learning and knolwege in higher education: Relationships with study behaviour and influences of learning environments. *International Journal of Educational Research,* 41**,** 407-428.

ENTWISTLE, N. & RAMSDEN, P. 1983. *Understanding Students Learning,* New York, Nichols.

ENTWISTLE, N. & RAMSDEN, P. 2015. *Understanding Student Learning*, Routledge Revivals.

ESBENSEN, K. H., GUYOT, D., WESTAD, F. & HOUMOLLER, L. P. 2002. *Multivariate Data Analysis: In Practice: An Introduction to Multivariate Data Analysis and Experiemental Design*, Camo Process AS.

EYSENCK, M. W. 1994. *The Blackwell Dictionary of Cognitive Psychology,* Oxford, Blackwell.

FIELD, A. 2009. *Discovering Statistics Using SPSS*, SAGE Publications Ltd.

FIORINA, L., ANTONIETTI, A., COLOMBO, B. & BARTOLOMEO, A. 2007. Thinking style, browsing orimes and hypermedia navigation. *Computers and Education,* 49**,** 919-941.

FORD, N. & CHEN, S. 2000. Individual Differences, Hypermedia Navigation, and Learning: An empirical Study. *Journal of Educational Multimedia and Hypermedia,* 9**,** 281-311.

FORD, N. & CHEN, S. 2001. Matching/mismatching revisited: an empirical study of learning and teaching. *British Journal of Education Technology,* 32**,** 5-22.

FRANSSON, A. 1978. *Test anxiety and motivation to learn.*, Acta Univertsitatis Gothoburgensis.

FREEMAN, S., EDDY, S., MCDONOUGH, M., SMITH, M., OKOROAFOR, N., JORDT, H. & WENDEROTH, M. P. 2014. Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences of the United States of America,* 111.

FROST, J. 2014. How to Interpret a Regression Model with Low R-squared and Low P values. *The Minitab Blog.* Pennsylvania: Minitab Inc.

FUSARO, M. 2008. *What is Teaching for Understanding? A new framework sets the stage for more effective practice* [Online]. Available: https://www.gse.harvard.edu/news/uk/08/05/what-teaching-understanding [Accessed 10/6/16 2016].

GALBRAITH, P. 2006. Students, mathematics, and technology: assessing the present – challenging the future. *International Journal of Mathematical Education in Science and Technology,* 37**,** 277-290.

GALBRAITH, P. & HAINES, C. 1998. Disentangling the nexus: Attitudes to mathematics and technology in a computer learning environment. *Educational Studies in Mathematics,* 36,**,** 275-290.

GIERL, M. & BISANZ, E. 2003. Identifying content and cognitive skills that produce gender differences in mathematics: A demonstration of the DIF analysis framework. *Journal of Educational Measurement,* 40**,** 281-306.

GÓMEZ-CHACÓN, I. & HAINES, C. STUDENTS' ATTITUDES TO MATHEMATICS AND TECHNOLOGY. COMPARATIVE STUDY BETWEEN THE UNITED KINGDOM AND SPAIN. Proceedings of 11th International Congress on Mathematical Education, 2008 Monterrey, Mexico.

GOULDING, M. & KYRIACOU, C. A systematic review of the role of ICTs in learning algebra. *In:* KUCHEMANN, D., ed. Proceedings of the British Society for Research into Learning Mathematics, 2007 London. British Society for Research into Learning Mathematics.

GRAFF, M. 2005. Indivudual differences in hypertext browsing strategies. *Behaviour*

and Information Technology, 23**,** 93-99.
HAGERTY, G. & SMITH, S. 2005. Using the web-based interactive software ALEKS to enhance college algebra. *Mathematics and Computer Education,* 39**,** 183-194.
HANDAL, H. & HERRINGTON, A. 2003. Re-Examining Categories of Computer-Based Learning in Mathematics Education. *Contemporary Issues in Technology and Teacher Education,* 3**,** 275-287.
HEID, K. & TODD, E. 2001. Computer Algebra Systems: Revolution or Retrofit for Today's Mathematics Classrooms? *Theory Into Practice,* 40**,** 128-136.
HENRY, M. J. 1995. Remedial Math Students' Navigation Patterns Through Hypermedia Software. *Computers in Human Behavior,* 11**,** 481-493.
HERDER, E. Revisitation Patterns and Disorientation. Proceedings of German Workshop on Adaptivity and User Modeling in Interactive Systems ABIS, 2003 Karlsruhe. 291-294.
HERDER, E. & JUVINA, I. Discovery of Individual User Navigation Styles. Adaptive Hypermedia AH2004 Workshop on Individual Differences, 2004 Eindhoven. 316-325.
HILLS, H. 2003. *Individual Preferences in e-learning,* Hants, Gower Publishing Company.
HOHENWARTER, J. 2015. *GeoGebra Quickstart-Evolution of interactive tutorials* [Online]. [Accessed 01/10/2017].
HONG, Y. Y. & THOMAS, O. J. M. 2015. Graphical construction of a local perspective on differentiation and integration. *Mathematics Education Research Journal,* 27**,** 183-200.
HOYLES, C. & NOSS, R. 2003. What can digital technologies take from and bring to research in mathematics education. *In:* BISHOP, A. (ed.) *Second International Handbook of Research in Mathematics Education.*
HUSCH, L. 2001a. *Visual Calculus - Drill -Domains of Functions* [Online]. Available: http://archives.math.utk.edu/visual.calculus/0/domain.1/index.html [Accessed 12/10/2017].
HUSCH, L. 2001b. *Visual Calculus - Pre-Calculus* [Online]. Available: http://archives.math.utk.edu/visual.calculus/0/index.html [Accessed 03/10/2017].
HUSCH, L. 2001c. *Visual Calculus - Symmetry of functions* [Online]. Available: http://archives.math.utk.edu/visual.calculus/0/symmetry.5/index.html [Accessed 01/10/2017].
ISSA, T. & ISAIAS, P. 2015. *Sustainable Design,* London, Springer-Verlag.
JAMES, W. & BLANK, W. 1993. Review and Critique of Available Learning-Style Instruments for Adults. *New Directions for Adult and Continuing Education,* 1993**,** 47-57.
JENKS, M. & SPRINGER, J. 2002. A view of the Research on the efficacy of CAI. *Electronic Journal for the Integration of Technology in Education,* 1**,** 43-58.
JONASSEN, D. 1994. *Technology as Cognitive Tools: Learners as Designers* [Online]. ITForum. Available: http://itech1.coe.uga.edu/itforum/paper1/paper1.html [Accessed 10/12/09].
JOSHI, D. R. 2017. Influence of ICT in Mathematics Teaching. *International Journal for Innovative Research in Multidisciplinary Field,* 3.
JUAN, A., HUERTAS, A., STEEGMANN, C., CORCOLES, C. & SERRAT, C. 2008. Mathematical e-learning: state of the art and experiences at the Open University of Catalonia. *International Journal of Mathematical Education in Science and Technology,* 39**,** 455-471.
JUVINA, I. & VAN OOSTENDORP, H. Predicting user preferences: from semantic to pragmatic metrics of Web navigation behaviour. Proceedings of the ACM International Conference Series, 2004 Amsterdam, Holland.
JUVINA, I. & VAN OOSTENDORP, H. 2006. Individual differences and behavioral metrics involved in modeling web navigation. *Universal Access in the Information Society,* 4**,** 258-269.
KAPLAN, A. & HAENLEIN, M. 2016. Higher education and the digital revolution: About

MOOCs, SPOCs, social media and the Cookie Monster. *Business Horizons,* 59**,** 441-450.

KATAYAMA, A., SHAMBAUGH, N. & DOCTOR, T. 2005. Promoting knowledge transfer with electronic note taking. *Computer in Teaching,* 32.

KEADY, G., FITZ-GERALD, G., GAMBLE, G. & SANGWIN, C. 2006. Computer-aided assessment in mathematical sciences. *UniServe Science Assessment Symposium Proceedings.*

KHOJU, M., JACIW, A. & MILLER, G. 2005. Effectiveness of Graphing Calculators in K-12 Mathematics Achievement. Palo Alto, CA: Emperical Education.

KLEITMAN, D. 2010. *MIT Open Courseware - Chapter 32: Some Linear Algebra* [Online].Available:http://ocw.mit.edu/ans7870/18/18.013a/textbook/MathML/chapter32/contents.xhtml [Accessed 03/10/2017].

KLEITMAN, D. 2010b. *MIT Open Courseware - Java Applets for Calculus* [Online]. Available:http://ocw.mit.edu/OcwWeb/Mathematics/18-013ASpring-2005/Tools/index.htm [Accessed 04/09/2017].

KOLB, D. 1984. *Experiential learning: experience as the source of learning and development.,* New Jersey, Prentice Hall.

KOLB, D. A. 1985. *Learning Style Inventory, Revised Edition.,* Boston, MA, Hay Group.

KOLB, D. A. 1999. *Learning Style Inventory, Version 3,* Boston, MA, Hay Group.

KUKLA, A. 1974. Performance as a function of resultant achievement motivation (perceived ability) and perceived difficulty. *Journal of Research in Personality,* 7**,** 374-383.

KULIK, J. & FLETCHER, J. D. 2016. Effectiveness of Intelligent Tutoring Systems: A Meta-Analytic Review. *Review of Educational Research,* 86**,** 42-78.

LADSON-BILLINGS, G. 1997. It doesn't add up: African American Students' Mathematics Achievement. *Journal for Research in Mathematics Education,* 28**,** 697-708.

LAGRANGE, J.-B., ARTIGUE, M., LABORDE, C. & TROUCHE, L. 2003. Technology and Mathematics Education: A Multidimensional Study of the Evolution of Research and Innovation. *In:* BISHOP, A. J., CLEMENTS, M. A., KEITEL, C., KILPATRICK, J. & LEUNG, F. K. S. (eds.) *Second International Handbook of Mathematics Education.* Dordrecht: Springer International Handbooks of Education.

LAVICZA, Z. 2010. Integrating technology into mathematics teaching at the university level. *Mathematics Education,* 42**,** 105-119.

LAW, D. & MEYER, H. F. 2009. Relationships between students' demographic background, subject areas and learning patterns in post-secondary education of Hong-Kong. *ICHL.* Springer-Verlag Berlin Heidelberg.

LAURILLARD, D. 2002. *Rethinking University Teaching - A framework for the effective use of learning technologies,* Oxon, Routledge Falmer.

LEMIRE, D. 2000. Research Report - A Comparison of Learning Styles Scores: A Question of Concurrent Validity. *Journal of College Reading and Learning Association,* 31**,** 109-116.

LI, Q. & EDMONDS, K. A. 2005. Mathematics and At-Risk Adult Learners: Would Technology Help? *Journal of Research on Technology in Education,* 38**,** 143-166.

LIDWELL, W., HOLDEN, K. & BUTLER, J. 2010. *Universal Principles of Design*, Rockport Publishers.

LIN, C.-H. & DAVIDSON-SHIVERS, G. V. 1996. Effects of Linking Structure and Cognitive Style on Students' Performance and Attitude in a Computer-based Hypertext Environment. *Journal of Educational Computing Research,* 15**,** 317-329.

LISTON, M. & O'DONOGHUE, J. 2009. Factors influencing the transition to university service mathematics: part I a quantitative study. *Teaching Mathematics and Its Applications,* 28**,** 77-78.

LIU, Y. & GINTHER, D. 1999. Cognitive Styles and Distance Education. *Online Journal of Distance Learning Administration,* 2,.

LOCKTON, D. 2012. Attitudes, meaning, emotion and motivation in design for behaviour change. Available: http://ssrn.com/abstract=2123495.

LUNTS, E. 1997. What does the Literature Say about the Effectiveness of Learner Control in Computer-Assisted Instruction. *Electronic Journal for the Integration of Technology in Education,* 1**,** 59-75.

MAAT, S. M. B. 2010. Using a Computer Algebra System in Teaching and Learning of Ordindary Differential Equations among Engineering Technology Students. *The 3rd Regional Conference on Engineering Education and Research in Higher Education.* Sarawak.

MAMPADI, F. & MOKODEDI, P. A. 2012. Towards Effective Combination of Prior Knowledge and Cognitive Styles in Adaptive Educational Hypermedia Systems. *International Journal of Emerging Technologies in Learning (iJET),* 7.

MARTENS, R. L., GULIKERSW, J. & BASTIAENSW, T. 2004. The impact of intrinsic motivation on e-learning in authentic computer tasks. *Journal of Computer Assisted learning,* 20**,** 368–376.

MARGETI, M. & MAVRIKIS, M. 2015. Students' Deep and Surface Approach: Links to Interaction in Learning Environments. *10th European Conference on Technology Enhanced Learning*. Toledo (Spain): Springer.

MARTIN, B. A. & MANNING, D. 1995. Combined effects of normative information and task difficutlty of the commitment-performance relations. *Journal of Management,* 21.

MARTON, F. & SÄLJÖ, R. 1976a. On qualitative differences in learning - II Outcome as a function of the learner's conception of the task. *British Journal of Educational Psychology,* 46,**,** 115-127.

MARTON, F. & SÄLJÖ, R. 1976b. On qualitative differences in learning: I-Outcome and process. *British Journal of Educational Psychology,* 46,**,** 4-11.

MAVRIKIS, M. 2010. Modelling Student Interactions in Intelligent Learning Environments: Constructing Bayesian Networks from Data. . *International Journal on Artificial Intelligence Tools,* 19**,** 733-753.

MAVRIKIS, M., MACIOCIA, A. & LEE, J. Targeting the affective state of students studying mathematics in a web-based ILE. Proceedings of 11th International Conference on Artificial Intelligence in Education, 2003 Sydney.

MCCUNE, V. 1998. Academic development during the first year at university. *In:* RUST, C. (ed.) *Improving student learning: Improving Students as Learners.* Oxford: Oxford Brookes University, Oxford Centre for Staff and Learning Development.

MCDONALD, C. 2016. STEM Education: A review of the contribution of the disciplines of science, technology, engineering and mathematics. *Science Education International,* 27**,** 530-569.

MCENEANEY, J. 2001. Graphic and numerical methods to assess navigation in hypertext. *International Journal of Human-Computer Studies,* 55**,** 761-786.

MELIS, E., GOGUADZE, G., HOMIK, M., LIBBRECHT, P., ULRICH, C. & WINTERSTEIN, S. 2006. Semantic-aware components and services of ActiveMath. *British Journal of Educational Technology,* 37**,** 405-423.

MEYERS, L. & GAMST, G. 2013. *Applied Multivariate Research Design and Interprentation,* London, Sage.

MEYERS, L., GAMST, G. & GUARINO, A. J. 2006. *Applied Multivariate Research - Design and Interpretation,* London, Sage.

MILLER, A. 1991. Personality Types, Learning Styles and Educational Goals. *Educational Psychology,* 11**,** 217-234.

MIMIRINIS, M. & DAFOULAS, G. 2008. Patterns of use of Virtual Learning Environments and Students' Approaches to Learning:

a Case Study of Undergraduate Students. *In:* LUCA, J. & WEIPPL, E., eds. Proceedings of World Conference on Educational Multimedia Hypermedia and Telecommunications, 2008 Vienna. Vienna, 6349-6356.

MOKSONY, F. 1990. Small is beautiful. The use and intepretation of R2 in social research. *Review of Sociology (Szoctologiai Szemle)***,** 130-138.

MATHWORKS. 2017. *Rediscover MATLAB* [Online]. Available: https://uk.mathworks.com/products/matlab.html [Accessed 5/10/2017.

MOLER, C. 2004. *The Origins of MATLAB* [Online]. Available: https://uk.mathworks.com/company/newsletters/articles/the-origins-of-matlab.html [Accessed 01/10/2017].

MOLER, C. 2006. *The growth of MATALAB and The MathWorks over two decades* [Online]. [Accessed 3/10/2017].

MORAN, A. 1991. What can Learning Styles Research Learn from Cognitive Psychology? *Educational Psychology,* 11**,** 239-245.

MORENO, R. Who learns best with multiple representations? Cognitive Theory implications for individual differences in multimedia learning. Proceedings of 14th ED-MEDIA World Conference on Educational Multimedia, Hypermedia & Telecommunications, 2002 Colorado, Denver.

MYERS BRIGGS, I. & MCCAULLEY, M. 1986. *Manual: A guide to the development and use of the Myers-Briggs Type Indicator,* Palo Alto, Consulting Psychologists Press.

NARCISS, S. 2007. Feedback Strategies for Interactive Learning Tasks. *In:* SPECTOR, M., MERRILL, D., VAN MERRIENBOER, J. & DRISCOLL, M. (eds.) *Handbook of Research Educational Communications and Technology.* London: Lawrence Erlbaum Associates.

NIEDERHAUSER, D. 2007. Educational Hypertext. *In:* SPECTOR, M., MERRILL, D. & VAN MERRIENBOER, J. (eds.) *Hanbook of Research on Educational Communications and Technology.* London: Lawrence Erlbaum Associates.

OBENDORF, H., WEINREICH, H., HERDER, E. & MAYER, M. Web Page Revisitation Revisited: Implications of a Long-term Click-stream Study of Browser Usage. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2007 San Jose, USA. 597-606.

OSMON, P. 2009. Post-16 maths and university courses: numbers and subject interpretation. *In:* JOUBERT, M. (ed.) *Proceedings of the British Society for Research into Learning Mathematics.* Loughborough.

PAPANIKOLAOU, K., GRIGORIADOU, M., KORNILAKIS, H. & MAGOULIAS, G. 2003. Personalising the Interaction in a Web-based Educational Hypermedia System: the case of INSPIRE. *User Modeling and User-Adapted Interaction,* 13**,** 213-267.

PAREDES, P. & RODRIGUEZ, P. 2004. A Mixed Approach to Modelling Learning Styles in Adaptive Educational Hypermedia. *Third IASTED International Conference on Web-Based Education* Innsbruck.

PASK, G. 1976a. Conversation techniques in the study and practice of education. *British Journal of Education Psychology,* 46,**,** 12-25.

PASK, G. 1976b. Styles and Strategies of Learning. *British Journal of Educational Psychology,* 46,**,** 128-148.

PASK, G. 1988. Learning strategies and teaching strategies and conceptual or learning style. *In:* SCHMECK, R. (ed.) *Learning strategies and learning styles.* New York: Plenum.

PERKIN, G., CROFT, T. & LAWSON, D. 2013. The extent of mathematics learning support in UK higher education-the 2012 survey. *Teaching Mathematics and its Applications,* 32**,** 165-172.

PIOLAT, A., OLIVE, T. & KELLOG, R. 2005. Cognitive Effort during Note Taking.

Applied Cognitive Psychology, 19**,** 291-312.
PROTOPSALTIS, A. 2006. *Reading in web-based hypertexts: cognitive processes strategies and reading goals.* University of Westminster.
RAMSDEN, P. 2005. The context of learning in academic departments. *The experience of learning: Implications for teaching and studying in higher education.* Edinburgh: University of Edinburgh, Centre for Teaching, Learning and Assessment.
RAMSEY, F. & SCHAFER, D. 2013. *The Statistical Sleuth - A course of Methods of Data Analysis,* Boston, Brooks/Cole Cengage Learning.
RASMUSSEN, K. L. & DAVIDSON-SHIVERS, G. V. 1998. Hypermedia and Learning Styles: Can Preformance Be Influenced? *Journal of Educational Multimedia and Hypermedia,* 7**,** 291-308.
RAYNER, S. & RIDING, R. J. 1997. Towards a Categorisation of Cognitive Styles and Learning Styles. *Educational Psychology,* 17**,** 5-27.
RIDING, R. 2001. Cognitive Style Analysis. Birmingham: Learning and Training Technology.
RIDING, R. & CHEEMA, I. 1991. Cognitive Styles - an overview and integration. *Educational Psychology,* 11**,** 193-214.
RIDING, R. J. 1997. On the Nature of Cognitive Style. *Educational Psychology,* 17**,** 29-48.
ROBSON, C. 2002. *Real World Research,* Oxford, Blackwell.
ROHRER, D., TAYLOR, K., PASHLER, H., WILTED, J. & CEPEDA, N. 2005. The effect of Overlearning on Long-Term Retention. *Applied Cognitive Psychology,* 19**,** 361-374.
ROSEN, Y. & SALOMAN, G. 2007. The differential learning achievements of constructivist technology-intensive learning environments as compared with traditional ones: A meta-analysis. *Journal of Educational Computing Research,* 36**,** 1-14.
RUBLE, T. & STOUT, D. 1994. A critical assessment of Kolb's Learning Style Inventory. U.S. Department of Education.
RYAN, R. & DECI, E. 2000. Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being. *American Psychologist,* 55**,** 68-78.
SAAB, E. 1999. *Welcome to the online math tests* [Online]. Available: http://mathonline.missouri.edu/ [Accessed 03/10/2017].
SADLER-SMITH, E. 1997. 'Learning Style': frameworks and instruments. *Educational Psychology,* 17**,** 51-63.
SAHA, S., AGARWAL, B. & MEHTA, P. Teaching Advanced Algebra to Engineering Majors: Dealing with the classroom challenges. WCI' 15: Proceedings of the Third International Symposium on Women in Computing and Informatics, 2015. ACM Digital Library.
SANGWIN, C. & O'TOOLE, C. 2017. Computer programming in the UK undergraduate mathematics curriculum. *International Journal of Mathematical Education in Science and Technology,* 48**,** 1133-1152.
SANGWIN, C., CAZES, C., LEE, A. & WONG, K. L. 2010. Micro-level Automatic Assessment Supported by Digital Technologies. *In:* HOYLES, C. & LANGANGE, J.-B. (eds.) *Mathematics Education and Technology-Rethinking the Terrain - The 17th ICMI Study.* London: Springer.
SANGWIN, C. 2004. Section A - Assessing mathematics automatically using computer algebra and the internet. *Teaching Mathematics and its Applications,* 23.
SCANDURA, J. M. 2012. The role of automation in instruction: Recent advances in authorIT and TutorIT solve fundamental problems in developing intelligent tutoring systems. *Technology, instruction, cognition and learning,* 9**,** 3-8.
SCHENKER, J. 2007. *The Effectiveness of Technology Use in Statistics Instruction in*

Higher Education: A Meta-analysis Using Hierarchical Linear Modeling. Kent State University.

SCHERLY, D., ROUX, L. & DILLENBOURG, P. 2000. Evaluation of hypertext in an activity learning environment. *Journal of Computer Assisted Learning,* 16**,** 125-136.

SCHMID, R., BERNARD, R., BOROKHOVSKI, E., TAMIM, R., ABRAMI, P., WADE, A., SURKES, M. A. & LOWERISON, G. 2009. Technology's effect on achievement in higher education: a Stage I meta-analysis of classroom applications. *Journal of Computing in Higher Education,* 21,**,** 95-109.

SCHOENFELD, A. 2006. Learning to think mathematically: problem solving, metacognition, sense making in mathematics. *In:* GROUWS, D. (ed.) *Handbook of Research on Mathematics Teaching and Learning.* Reston: Information Age Publishing.

SHAHABI, C., ZARKESH, A., ADIBI, J. & SHAH, V. 1997. *Knowledge discovery from users web-page navigation* [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.45.2824 [Accessed 1 / 10/2015].

STASH, N., CRISTEA, A. & DE BRA, P. 2004. Authoring of learning styles in adaptive hypermedia: problems and solutions. 13th International Conference on World Wide Web, 2004 New York, USA. New York, 114-123.

STEELE, F. 2003. Quantitative Statistics II-Multivariate Analysis. London: Institute of Education.

STERNBERG, R. J. & GRIGORENKO, E. L. 1997. Are Cognitive Styles Still in Style? *American Psychologist,* 52**,** 700-712.

STERNBERG, R. J. & GRIGORENKO, E. L. 2001. A capsule History of Theory and Research. *In:* STERNBERG, R. J. & ZHANG, L.-F. (eds.) *Pespectives on thinking, learning, and cognitive styles.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

STROPPEL, M., WOHLMUTH, B., BAUMANN, G., NEBE, G. & REIF, U. 2007. *Mathematics online test: Vector Calculus Test 3* [Online]. Available: http://www.mathematics-online.org/tests/test97/ [Accessed 05/10/2017].

SUN, Z., XIE, K. & ANDERMAN, L. 2018. The role of self-regulated learning in students' success in flipped undergraduate math courses. *The Internet and Higher Education,* 36**,** 41-53.

SUROWIEC, R. L. 2004. *The Higher Maths Education Stats & Academy Reviews: Software* [Online]. The Higher Maths Education Stats & Academy. Available: http://ltsn.mathstore.ac.uk/reviews/software.shtml - simvis [Accessed 12/09/2009 2009].

TAIT, H. & ENTWISTLE, N. 1996. Identifying students at risk through ineffective study strategies. *Higher Education,* 31**,** 97-116.

TAIT, H., ENTWISTLE, N. & MCCUNE, V. 1998. ASSIST: a reconceptualisation of the Approaches to Studying Inventory. *In:* RUST, C. (ed.) *Impoving student learning: Improving Students as Learners.* Oxford: Oxford Brookes University, Oxford Centre for Staff and Learning Development.

TALEB, Z., AHMADI, A. & MUSAVI, M. 2015. The effect of m-learning on mathematics learning. *Social and Behavioral Science,* 171**,** 83-89.

TAUSCHER, L. & GREENBERG, S. 1997. How people revisit web pages: emperical findings and implications for the design of history systems. *International Journal of Human-Computer Studies,* 47**,** 97-137.

TERRY, K. & HOLIM, S. 2008. *Handbook of Research on Instructional Systems and Technology,* New York, Hershey.

THOMAS, M. & HOLTON, D. 2003. Technology as a tool for teaching undergraduate mathematics. *In:* BISHOP, A. J., CLEMENTS, M. A., KEITEL, C., KILPATRICK, J. & LEUNG, K. S. (eds.) *Technology as a tool for teaching undergraduate mathematics.* Dordrecht, The Nederlands: Kluwer Academic Publishers.

THURSTON, W. 1990. MATHEMATICAL EDUCATION. *Notices of the american mathematical society,* 37**,** 844-850.

TOBIAS, S. 1994. Interest, Prior Knowledge, and Learning. *Review of Educational Research,* 64**,** 37-48.

TOKPAH, C. 2008. *The effects of computer algebra systems on students' achievement in mathematics.* Kent State University.

TRENHOLM, S., ALCOCK, L. & ROBINSON, C. 2015. An investigation of assessment and feedback practices in fully asynchronous online undergraduate mathematics courses. *International Journal of Mathematical Education in Science and Technology,* 45.

TURAN, S., ELCIN, M., ODABASI, O., WARD, K. & SAYEK, I. 2009. Evaluating the role of tutors in problem-based learning sessions. *Procedia Social and Behavioral Sciences,* 1**,** 5-8.

VANLEHN, K. 2011. The relative effectiveness of human tutoring, intelligent systems and other tutoring systems. *Educationak Psychologist,* 46.

VOCKELL, E. 2006. *Educational Psychology: A Practical Approach* [Online]. Purdue University, School of Education. Available: http://education.purduecal.edu/Vockell/EdPsyBook/ [Accessed 10/10/2014].

WANER, S. 2007. *Tutorials for Finite Math and Calculus - 1.3 Linear Functions and Models* [Online]. Available: http://www.zweigmedia.com/RealWorld/tutorialsf0/frames1_3.html [Accessed 02/10/2017].

WANER, S. 2010. *Finite mathematics and applied calculus: Applied calculus utility: function evaluator and grapher* [Online]. Available: http://www.zweigmedia.com/RealWorld/functions/func.html [Accessed 20/08/2017].

WEIGAND, H.-G. & WELLER, H. 2001. Changes of working styles in a computer algebra environment - The case of functions. *International Journal of Computers for Mathematical Learning,* 6,**,** 87-111.

WINNE, P. & HADWIN, A. 2009. Studying of self-regulated learning. *In:* HACKER, D. J., DUNLOSKY, J. & GRAESSER, A. (eds.) *Metacognition in Educational Theory and Practice.* Routledge.

WITKIN, H. A., OLTMAN, P. K., RASKIN, E. & KARP, S. A. 1971. *Embedded Figure Test: Group Embedded Figures Test (GEFT),* Palo Alto, California, Consulting Psychologists Press.

WITKIN, H. A., OTTMAN, P. K., RASKIN, E. & KARP, S. A. 1977. Field dependent and field independent cognitive styles and their educational implication. *Review of Educational Research,* 47**,** 1-64.

WONG, N.-Y. 2003. The Influence of Technology on the Mathematics Curriculum. *In:* BISHOP, A., CLEMENTS, M. A., KEITEL, C., KILPATRICK, J. & LEUNG, F. K. S. (eds.) *Second International Handbook of Research in Mathematics Education.* Dordrecht: Springer International Handbooks of Education.

YUAN, L. & POWELL, S. 2013. MOOCs and Open Education: Implications for Higher Education. *In:* STANDARDS, J. C.-C. F. E. T. I. (ed.). JISC.

Appendices

Appendices – Chapter 3 – Methodology

.

Appendix 3.3 – The ASSIST instrument

A S S I S T
Approaches and Study Skills Inventory for Students

This questionnaire has been designed to allow you to describe, in a systematic way, how you go about learning and studying. The technique involves asking you a substantial number of questions which overlap to some extent to provide good overall coverage of different ways of studying. Most of the items are based on comments made by other students. Please respond truthfully, so that your answers will **accurately**
describe your **actual** ways of studying, and work your way through the questionnaire quite **quickly.**

Background information

Name or Identifier .... **Age .... years** **Sex** M / F

University or College .... **Faculty or School ....**

Course .... **Year of study ....**

A. What is learning?

When you think about the term 'LEARNING ', what does it mean to you?

Consider each of these statements carefully, and rate them in terms of how close they are to **your own** *way of thinking about it.*

		Very close	*Quite close*	*Not so close*	*Rather different*	*Very different*
a.	Making sure you remember things well.	5	4	3	2	1
b.	Developing as a person.	5	4	3	2	1
c.	Building up knowledge by acquiring facts and information.	5	4	3	2	1
d	Being able to use the information you've acquired.	5	4	3	2	1
e.	Understanding new material for yourself.	5	4	3	2	1
f.	Seeing things in a different and more meaningful way.	5	4	3	2	1

Please turn over

B. Approaches to studying

The next part of this questionnaire asks you to indicate your relative agreement or disagreement with com- ments about studying again made by other students. Please work through the comments, giving your **immediate** response. In deciding your answers, think in terms of **this particular lecture course.** It is also very important that you answer **all** the questions: check you have.

5 *means agree (   )* ***4*** *= agree somewhat (* ***?*** *)* ***2*** *= disagree somewhat (* ***x?*** *)* ***1*** *= disagree (* ***x*** *).*

Try <u>not</u> to use ***3*** *= unsure (* ***??*** *), unless you really have to, or if it cannot apply to you or your course.*

			?	**??**	**x?**	**x**
1.	I manage to find conditions for studying which allow me to get on with my work easily.	5	4	3	2	1
2.	When working on an assignment, I'm keeping in mind how best to impress the marker.	5	4	3	2	1
3.	Often I find myself wondering whether the work I am doing here is really worthwhile.	5	4	3	2	1
4.	I usually set out to understand for myself the meaning of what we have to learn.	5	4	3	2	1
5.	I organise my study time carefully to make the best use of it.	5	4	3	2	1
6.	I find I have to concentrate on just memorising a good deal of what I have to learn.	5	4	3	2	1
7.	I go over the work I've done carefully to check the reasoning and that it makes sense.	5	4	3	2	1
8.	Often I feel I'm drowning in the sheer amount of material we're having to cope with.	5	4	3	2	1
9.	I look at the evidence carefully and try to reach my own conclusion about what I'm studying.	5	4	3	2	1
10.	It's important for me to feel that I'm doing as well as I really can on the courses here.	5	4	3	2	1
11.	I try to relate ideas I come across to those in other topics or other courses whenever possible.	5	4	3	2	1
12.	I tend to read very little beyond what is actually required to pass.	5	4	3	2	1
13.	Regularly I find myself thinking about ideas from lectures when I'm doing other things.	5	4	3	2	1
14.	I think I'm quite systematic and organised when it comes to revising for exams.	5	4	3	2	1
15.	I look carefully at tutors' comments on course work to see how to get higher marks next time.	5	4	3	2	1
16.	There's not much of the work here that I find interesting or relevant.	5	4	3	2	1
17.	When I read an article or book, I try to find out for myself exactly what the author means.	5	4	3	2	1

18.	I'm pretty good at getting down to work whenever I need to.	5	4	3	2	1
19.	Much of what I'm studying makes little sense: it's like unrelated bits and pieces.	5	4	3	2	1
20.	I think about what I want to get out of this course to keep my studying well focused.	5	4	3	2	1
21.	When I'm working on a new topic, I try to see in my own mind how all the ideas fit together.	5	4	3	2	1
22	I often worry about whether I'll ever be able to cope with the work properly.	5	4	3	2	1
23.	Often I find myself questioning things I hear in lectures or read in books.	5	4	3	2	1
24.	I feel that I'm getting on well, and this helps me put more effort into the work.	5	4	3	2	1
25.	I concentrate on learning just those bits of information I have to know to pass.	5	4	3	2	1
26.	I find that studying academic topics can be quite exciting at times.	5	4	3	2	1
27.	I'm good at following up some of the reading suggested by lecturers or tutors.	5	4	3	2	1
28.	I keep in mind who is going to mark an assignment and what they're likely to be looking for.	5	4	3	2	1
29.	When I look back, I sometimes wonder why I ever decided to come here.	5	4	3	2	1
30.	When I am reading, I stop from time to time to reflect on what I am trying to learn from it.	5	4	3	2	1

		?	??	x?	x
31. I work steadily through the term or semester, rather than leave it all until the last minute.	5	4	3	2	1
32. I'm not really sure what's important in lectures so I try to get down all I can.	5	4	3	2	1
33. Ideas in course books or articles often set me off on long chains of thought of my own.	5	4	3	2	1
34. Before starting work on an assignment or exam question, I think first how best to tackle it.	5	4	3	2	1
35. I often seem to panic if I get behind with my work.	5	4	3	2	1
36. When I read, I examine the details carefully to see how they fit in with what's being said.	5	4	3	2	1
37. I put a lot of effort into studying because I'm determined to do well.	5	4	3	2	1
38. I gear my studying closely to just what seems to be required for assignments and exams.	5	4	3	2	1
39. Some of the ideas I come across on the course I find really gripping.	5	4	3	2	1
40. I usually plan out my week's work in advance, either on paper or in my head.	5	4	3	2	1
41. I keep an eye open for what lecturers seem to think is important and concentrate on that.	5	4	3	2	1
42. I'm not really interested in this course, but I have to take it for other reasons.	5	4	3	2	1
43. Before tackling a problem or assignment, I first try to work out what lies behind it.	5	4	3	2	1
44. I generally make good use of my time during the day.	5	4	3	2	1
45. I often have trouble in making sense of the things I have to remember.	5	4	3	2	1
46. I like to play around with ideas of my own even if they don't get me very far.	5	4	3	2	1
47. When I finish a piece of work, I check it through to see if it really meets the requirements.	5	4	3	2	1
48 Often I lie awake worrying about work I think I won't be able to do.	5	4	3	2	1
49 It's important for me to be able to follow the argument, or to see the reason behind things.	5	4	3	2	1
50. I don't find it at all difficult to motivate myself.	5	4	3	2	1
51. I like to be told precisely what to do in essays or other assignments.	5	4	3	2	1
52. I sometimes get 'hooked' on academic topics and feel I would like to keep on studying them.	5	4	3	2	1

C. Preferences for different types of course and teaching

5 *means definitely like ( )* ***4*** *= like to some extent (* ***?*** *)* ***2*** *= dislike to some extent (* ***x?*** *)* ***1*** *= definitely dislike (* ***x*** *).*
Try *<u>not</u>* *to use* ***3*** *= unsure (* ***??*** *), unless you really have to, or if it cannot apply to you or your course.*

			?	??	x?	x
a.	lecturers who tell us exactly what to put down in our notes.	5	4	3	2	1
b.	lecturers who encourage us to think for ourselves and show us how they themselves think	5	4	3	2	1
c.	exams which allow me to show that I've thought about the course material for myself.	5	4	3	2	1
d.	exams or tests which need only the material provided in our lecture notes.	5	4	3	2	1
e.	courses in which it's made very clear just which books we have to read.	5	4	3	2	1
f.	courses where we're encouraged to read around the subject a lot for ourselves.	5	4	3	2	1
g.	books which challenge you and provide explanations which go beyond the lectures.	5	4	3	2	1
h.	books which give you definite facts and information which can easily be learned.	5	4	3	2	1

Finally, how well do you think you have been doing in your assessed work overall, so far?

Please rate yourself ***objectively***, *based on the grades you have been obtaining*

Very well		*Quite Well*		*About average*		*Not so well*		*Rather badly*
9	8	7	6	5	4	3	2	1

Thank you very much for spending time completing this questionnaire: it is much appreciated.

Note from the author: The author acknowledges that she has no copyright of the ASSIST instrument presented on pages 316-321 and that it has been taken from the following source: ENTWISTLE, N. 1997a. Approaches and Study Skills Inventory for Students (ASSIST) [Online]. ETL Project. Available: http://www.etl.tla.ed.ac.uk/questionnaires/ASSIST.pdf [Accessed 07 / 1 / 13].

Appendix 3.4 .1 - Typical ‘reading’ page and features in AM

[AM screenshot was removed because of potential copyright restriction]

Appendix 3.4.2 – Typical ‘exercise’ page and ‘exercise’ window’ in AM

[AM screenshot was removed because of potential copyright restriction]

Appendix 3.4.3 - Search option

[AM screenshot was removed because of potential copyright restriction]

Appendix 3.4.4 - Exercise that is completed but not solved

[AM screenshot was removed because of potential copyright restriction]

Appendix 3.4.5 - Exercise solved on first try

[AM screenshot was removed because of potential copyright restriction]

Appendix 3.4.6 - Exercise solved on second try

[AM screenshot was removed because of potential copyright restriction]

Appendix 3.4.7 - Exercise solved on third try

[AM screenshot was removed because of potential copyright restriction]

Appendix 3.4.8.a – ‘Notes’ option –Create

[AM screenshot was removed because of potential copyright restriction]

Appendix 3.4.8.b – ‘Notes’ option – Making a note

[AM screenshot was removed because of potential copyright restriction]

Appendix 3.4.8.c – ‘Notes’ option – Editing and Deleting a note

[AM screenshot was removed because of potential copyright restriction]

Appendix 3.4.9 – Home Page

[AM screenshot was removed because of potential copyright restriction]

Appendix 3.4.10 – Worked Example

[AM screenshot was removed because of potential copyright restriction]

Appendix 3.4.11 – Worked Example and Hyperlinks

[AM screenshot was removed because of potential copyright restriction]

Appendix 3.4.12 – Worked Example - Summary

[AM screenshot was removed because of potential copyright restriction]

Appendix 3.4.13 – Theoretical Example

[AM screenshot was removed because of potential copyright restriction]

Appendix 3.4.14 – Worked Example

[AM screenshot was removed because of potential copyright restriction]

Appendix 3.10.1 - ActiveMath manual

ActiveMath – Interactive Notes

for

Module: 2COS404 "Information Fundamentals"

What is ActiveMath

ActiveMath has been created in order to provide you with interactive exercises and examples which are essential in this module. Interactive features such as the Graph Plotter and detailed coloured graphs will help you to better understand the mathematical concepts involved in this module.

ActiveMath includes the complete subject material for the following chapters:

- Chapter 4 – Functions
- Chapter 5 – Graphs
- Chapter 8 – Matrices

Here is a short guide that will help you understand how to use ActiveMath.

A short guide

First open the browser (Internet Explorer or Netscape) and type the following URL:

http://iclass.activemath.org/ActiveMath2/

1. Log-in page

You can access ActiveMath, **only** if you have registered and created your own personal account. For each session, you will need to use **your own unique username and password.**

2. Home page

At the home page, you can select among the following chapters:

- Chapters 4 – Functions
- Chapter 5 – Graphs
- Chapter 8 – Matrices

3. Main Menu

The menu (at the top right corner) has the following options:

- **Home page**: This always leads to the home page in case you want to select another chapter.
- **Contact:** In case you are experiencing log-in problems **in ActiveMath** (such as forgetting your password etc), please contact Maria Margeti in the following e-mail address: **M.M.Margeti@wmin.ac.uk**
- **Help:** This provides brief but useful information about ActiveMath.
- **Logout:** Please remember to logout as soon as you finish your session.
- **Graph Plotter**: The graph plotter will help you to comprehend certain mathematical concepts and solve the exercises. You will use it mainly in the examples and exercises of Chapter 5-Graphs.

There are two different types of graph plotters that you will use:

- The first graph plotter consists of plotting a parabola and a line.
- The second graph plotter consists of plotting two lines.

Each graph plotter serves the needs and purposes of different examples and exercises. However, there are exercises in which you can use either of the two (e.g. the exercises that you need to plot only a line). Each exercise and example usually indicates which type of graph plotter you should use.

4. How to move around in ActiveMath

- **Search**
  This consists of three options:
  - The **Simple** option helps you to jump quickly to a specific piece of information in the content.
  - The **Advanced** option helps you to locate more accurately information using options such as "contain phrase". You can also search many keywords at the same time.
  - The **History** option stores automatically your searches for each session.

Each time you conduct a simple or advanced search about a mathematical concept, the results will be displayed in a pop-up window. At the bottom of the pop-up window, there are also three external links:

- **Google**
- **Wikipedia**
- **MathWorld**

The above three links will permit you to do further research about a mathematical concept.

- **Table of contents**
  The table of contents is always placed on the left-hand side of the screen and allows you to move around the subchapters of the chapter. By clicking on the main subchapters more subchapters will be revealed in certain cases. The information of each subchapter is presented to the right part of the screen.

- **Previous and Next Buttons**
  At the bottom of each page you can find the previous and the next buttons that can help you to move from one page to another of each chapter in a linear way (like in a book).

- **Explanation of mathematical terms**
  Sometimes you may forget certain mathematical terms and concepts. In order to help you remember these terms, we have highlighted them **in bold,** and by clicking on them a pop-up window comes up with an explanation of the term.

5. Other features of ActiveMath

- **Theory – Examples –Exercises**

The titles of the subchapters indicate whether they are theory, examples, or exercises. For example:

  - **Theory:** They are presented in an orange/pink background.
  - **Examples:** They are presented with two vertical orange lines.
  - **Exercises:** They are presented in a blue background.

- **Exercises**

  To start each exercise, you should click on “Start exercise” which opens up a pop-up window. An exercise can be single choice, multiple choice, or fill-in the blank.

  Select or type your answer and then click on “Evaluate” to get the feedback. ActiveMath will allow you three attempts before it gives you the correct answer.

- **Notes**

ActiveMath gives you the option of keeping notes for each theory, example, or exercise. For example, you can keep notes about the correct answer of an exercise by clicking on .

You will find this icon at the top right corner of a theory, example, or exercise.

The Notes, you make, are kept permanently and you can view them whenever you access your account. You can also change them (“Edit”), or delete them (“Delete”).

- **Print**

You can print the right part of the screen in PDF by clicking on “Print this page as PDF” (at the bottom of each page). You can also print the whole chapter by clicking on “print the whole book” (at the table of contents).

Appendix 3.10.2 - ActiveMath registration guidelines

ActiveMath - Interactive Notes

for

Module: 2COS404 "Information Fundamentals"

Accessing and using ActiveMath is easy. **You just need a personal username and password that is unique for each student.** In order to create a personal username and password you need to create an account.

Please follow the following guidelines to create an account:

- Type in your browser (Internet Explorer or Netscape) the URL: **http://iclass.activemath.org/ActiveMath2/**

  After typing the URL, the following screen will be displayed:

- Click on **"register?"** to open the registration form.
- In the registration form you should complete **accurately all your personal details.**
- **Students ID:** It should be typed as: w12345678
- **Date of Birth**
  - It should be typed as: dd/mm/yyyy

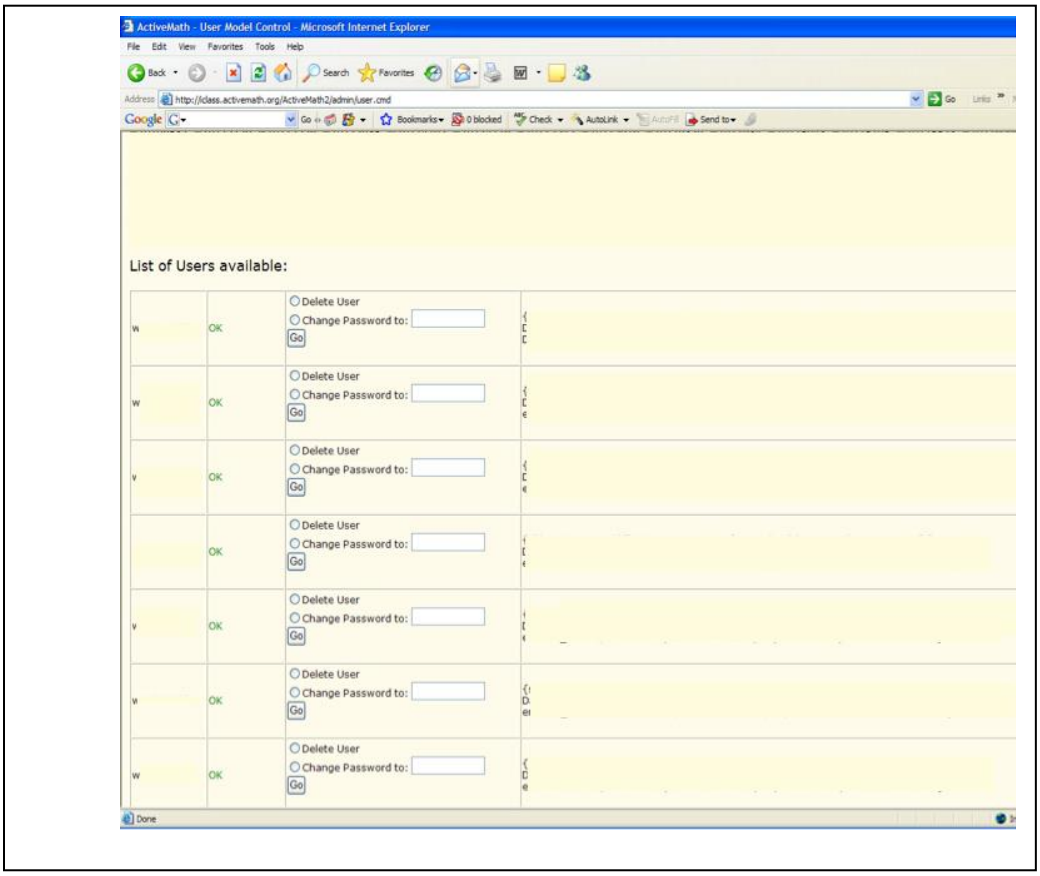
Appendix 3.10.2 – ActiveMath registration guidelines (continue)

- **Username**
  - The username should be your **student ID** (for example, w12345678). The username should be **typed in lower-case**.
- **Password**
  - The password should be your **date of birth**. It should be typed as: ddmmyyyy. The password is case sensitive.

Note: Do not submit the registration form, before your tutor checks it with you.

Should you have any problems logging in, please contact Maria Margeti.

Appendix 3.10.5 – ActiveMath administration account



Note from author: the details of users have been obscured to protect the identity of participants.

Appendix 3.11.1 - Thresholds for effect size and variance

r correlation coefficient (effect size)	0.1 small	0.3 medium	0.5 large
r^2 shared variance explained (or variance accounted for)	1%	9%	25%

Table **1**. Thresholds - Correlation coefficients -effect sizes r – shared variance explained r^2 based on Cohen, (1992).

f^2 (effect size) $R^2 / 1- R^2$	Small 0.02	Medium 0.15	Large 0.35
R^2 (variance explained)	0.0196 (1.96%)	0.1304 (13.04%)	0.2592 (25.92%)
R	0.14	0.36	0.51

Table **2**. Thresholds – Multiple Regression – effect size f^2 – variance explained R^2 – Correlation coefficient R, based on Cohen (1992).

Appendices – Chapter 4 - Regression Analysis and Model Interpretation

Appendices 4.1 – Surface Models

Appendix 4.1.1 –Further Justification for inclusion of predictors

Selected Predictors	Reason for selection
Number of exercises solved on first try	Statistical (see 4.1.2)
Number of exercises solved on third try	
Number of exercises finished but not solved	
Compactness	
Average view time on exercise pages	
Number of hyperlinks (concept links) visited in reading and exercise pages	
Maximum view time on exercise page	It is enriching for the discussion to explore whether students with high score on surface scale are spending an, increasingly, excessive amount of time on a specific exercise page, compared to those with low scores
Maximum view time on content page	-It is enriching for the discussion of the surface scale to explore whether students with high score on surface scale tend to persist and spend time, not only on the practical part of the AM learning material, but also on the content (reading) pages of AM. -It is preferred compared to *Average view time on content (reading) pages* (with which there is a multicollinearity issue). When tried in premodels 1a and 1b, *maximum view time on content page* contributes more in the variance (see table 2 below)

Table 1. Reasons for selection

Pre-models	R^2	Adj. R^2	Sig.
Model 1b (***Maximum view time on content (reading) page***, *Average View Time on Exercise Pages, Compactness, Maximum View Time on Exercise Page, Number of exercises solved on third try, Number of concept links visited in reading, exercise pages, Number of exercises solved on first try, Number of exercises finished but not solved)*	37.3%	32.5%	0.000
Model 1a (***Average view time on content (reading) pages***, *Average View Time on Exercise Pages, Compactness, Maximum View Time on Exercise Page, Number of exercises solved on third try, Number of concept links visited in reading, exercise pages, Number of exercises solved on first try, Number of exercises finished but not solved)*	36.6%	31.8%	0.000

Table 2. Pre-models

Appendix 4.1.2 – Detailed discussion on development of model

Model 1b – Surface Scale - All predictors based on initial selection

Model Summary[b]				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.610[a]	.373	.325	8.287

a. Predictors: (Constant), Maximum View Time on Content Page, Average View Time on Exercise Pages, Compactness, Maximum View Time on Exercise Page, Number of Exercise Solved on Third Try, Number of hyperlinks (concept links) visited in reading and exercise pages, Number of Exercise Solved on First Try, Number of Exercise Finished But not Solved

b. Dependent Variable: Surface Scale

Table 1 – Model Summary

ANOVA[a]						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4323.297	8	540.412	7.868	.000[b]
	Residual	7280.146	106	68.681		
	Total	11603.443	114			

a. Dependent Variable: Surface Scale

b. Predictors: (Constant), Maximum View Time on Content Page, Average View Time on Exercise Pages, Compactness, Maximum View Time on Exercise Page, Number of Exercises Solved on Third Try, Number of hyperlinks (concept links) visited in reading and exercise pages, Number of Exercises Solved on First Try, Number of Exercises Finished But not Solved

Table 2 – Overall Significance of model

Coefficients[a]						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	41.340	5.745		7.196	.000
	Number of hyperlinks (concept links) visited in reading, exercise pages	.044	.371	.010	.120	.905
	Number of Exercise Finished But not Solved	.348	.246	.156	1.419	.159
	Number of Exercise Solved on First Try	-.153	.034	-.416	-4.560	.000
	Number of Exercise Solved on Third Try	.661	.414	.173	1.597	.113
	Compactness	16.955	7.822	.178	2.168	.032
	Maximum View Time on Exercise Page	.002	.001	.142	1.760	.081
	Average View Time on Exercise Pages	-.001	.002	-.031	-.340	.734
	Maximum View Time on Content Page	-.002	.002	-.115	-1.415	.160

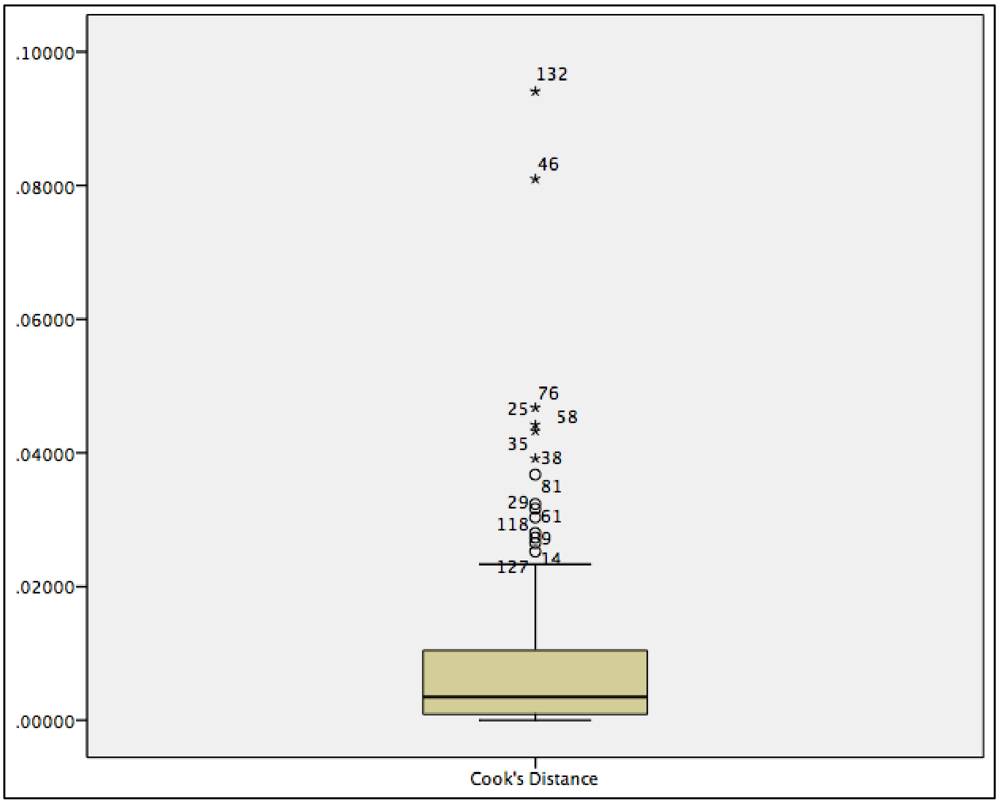
a. Dependent Variable: Surface Scale

Table 3 – B, Beta, and Sig. values for predictors

Surface Scale – Exclusion of outliers

To exclude the outliers from Model 1 the intention is to use the Cook's Distance method. The box plot of Cook's distance below shows a number of outliers. However, the intention is to exclude amongst the most extreme ones (those indicated with an asterisk), those which improve the measures of variance R^2 and adjusted R^2.

Figure 1. Cook's Distance Box Plot



In Fig.1, it is observed that the most extreme outliers are cases: 132, 46, 76, 58, 25, 35 and 38. These cases are excluded gradually on models Model 2a, 2b, 2c, 2d, 2e, and 2f, and 2g. It is observed that the exclusion of cases 132, 46, 58, 25, and 38 has increased both R^2 and adjusted R^2 (from 37.3% in Model 1 to 45.6% in Model 2g for R^2, and from 32.5% in Model 1 to 41.3% in Model 2g for adjusted R^2). However, outliers 76 and 35 decrease the variance, thus they are not excluded from the sample. So, the five cases which are excluded are: 132, 46, 58, 25 and 38.

Model 3 – Surface Scale- Recommended version

Model Summary[b]				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.675[a]	.455	.418	7.542

a. Predictors: (Constant), Maximum View Time on Content Time, Number of Exercises Solved on First Try, Compactness, Number of Exercises Solved on Third Try, Maximum View Time on Exercise Page, Number of concept links visited in reading, exercise pages, Number of Exercises Finished But not Solved

b. Dependent Variable: Surface Scale

Table 4 – Model Summary

ANOVA[a]						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4848.324	7	692.618	12.175	.000[b]
	Residual	5802.594	102	56.888		
	Total	10650.918	109			

a. Dependent Variable: Surface Scale

a. Predictors: (Constant), Maximum View Time on Content Time, Number of Exercises Solved on First Try, Compactness, Number of Exercises Solved on Third Try, Maximum View Time on Exercise Page, Number of concept links visited in reading, exercise pages, Number of Exercises Finished But not Solved

Table 5 – Overall significance of model

Coefficients[a]						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	46.502	5.187		8.965	.000
	Number of hyperlinks (concept links) visited in reading, exercise pages	.450	.357	.101	1.262	.210
	Number of Exercises Finished But not Solved	.340	.225	.157	1.512	.134
	Number of Exercises Solved on First Try	-.161	.028	-.435	-5.737	.000
	Number of Exercises Solved on Third Try	.805	.386	.213	2.085	.040
	Compactness	7.580	7.498	.078	1.011	.314
	Maximum View Time on Exercise Page	.002	.001	.167	2.168	.032
	Maximum View Time on Content Time	-.004	.002	-.188	-2.412	.018

a. Dependent Variable: Surface Scale

Table 6 – B, Beta, and Sig. values for predictors

Appendix 4.1.3 – Surface Scale – Model 6 – Leanest and Meanest

Exclusion of outliers 132 and 46 and 58 and 25 and 38 and predictors *Average View Time on Exercise Pages and Compactness and Number of concepts clicked on exercise and reading pages and Number of Exercises Finished but not Solved*

Model Summary[b]				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.650[a]	.423	.401	7.651

a. Predictors: (Constant), Maximum View Time on Content Time, Number of Exercises Solved on First Try, Number of Exercises Solved on Third Try, Maximum View Time on Exercise Page

b. Dependent Variable: Surface Scale

Table 7 – Model Summary

ANOVA[a]						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4504.207	4	1126.052	19.236	.000[b]
	Residual	6146.712	105	58.540		
	Total	10650.918	109			

a. Dependent Variable: Surface Scale

b. Predictors: (Constant), Maximum View Time on Content Time, Number of Exercises Solved on First Try, Number of Exercises Solved on Third Try, Maximum View Time on Exercise Page

Table 8 – Overall Significance of model

Coefficients[a]						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	52.995	2.545		20.822	.000
	Number of Exercises Solved on First Try	-.170	.028	-.458	-6.032	.000
	Number of Exercises Solved on Third Try	1.312	.283	.347	4.639	.000
	Maximum View Time on Exercise Page	.002	.001	.146	1.903	.060
	Maximum View Time on Content Time	-.003	.001	-.174	-2.308	.023

a. Dependent Variable: Surface Scale

Table 9 – B, Beta, and Sig. values for predictors

Appendices 4.2 –Unrelated Memorising Models

Appendix 4.2.1 –Further Justification for inclusion of predictors

Non-Selected Predictors	Reason for non-selection
Number of exercises accessed	-Multicollinearity issue with *Number of Exercises solved on First Try* *-Number of Exercises solved on First Try* was selected instead because of the: stronger correlation to the subscale; the potential of enriching for the discussion of the model, as it allows for useful comparisons with regards to the rest of the performance-related metrics.
Number of distinct pages	-Multicollinearity issue with *Average view time on exercise pages* - *Average view time on exercise pages* was selected instead because of the: stronger correlation to the subscale; the potential of enriching for the discussion of the model, as it represents the temporal aspect of students' interactions in AM - *Number of distinct pages* can indicate a repetitive tendency when students are going through the learning material, but this is also represented well by *relative amount of revisits* (which has been already included in the model)

Table1. Reasons for not selecting predictors

Appendix 4.2.2 – Further information on development of model

Model 1 – Unrelated Memorising Scale - All predictors based on initial selection

Model Summary[b]				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.574[a]	.329	.278	2.879

a. Predictors: (Constant), Average View Time on Exercise Pages. Compactness, Number of Exercises Solved on Third Try, Avg number a notes link is clicked per page, Number of hyperlinks (concept links) visited in reading, exercise pages, Number of Exercises Solved on First Try, Relative amount of revisits, Number of Exercises Finished But not Solved

b. Dependent Variable: Surface subscale Unrelated Memorising

Table 1 – Model Summary

ANOVA[a]						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	430.658	8	53.832	6.496	.000[b]
	Residual	878.438	106	8.287		
	Total	1309.096	114			

a. Dependent Variable: Surface subscale Unrelated Memorising

b. b. Predictors: (Constant), Average View Time on Exercise Pages. Compactness, Number of Exercises Solved on Third Try, Avg number a notes link is clicked per page, Number of hyperlinks (concept links) visited in reading, exercise pages, Number of Exercises Solved on First Try, Relative amount of revisits, Number of Exercises Finished But not Solved

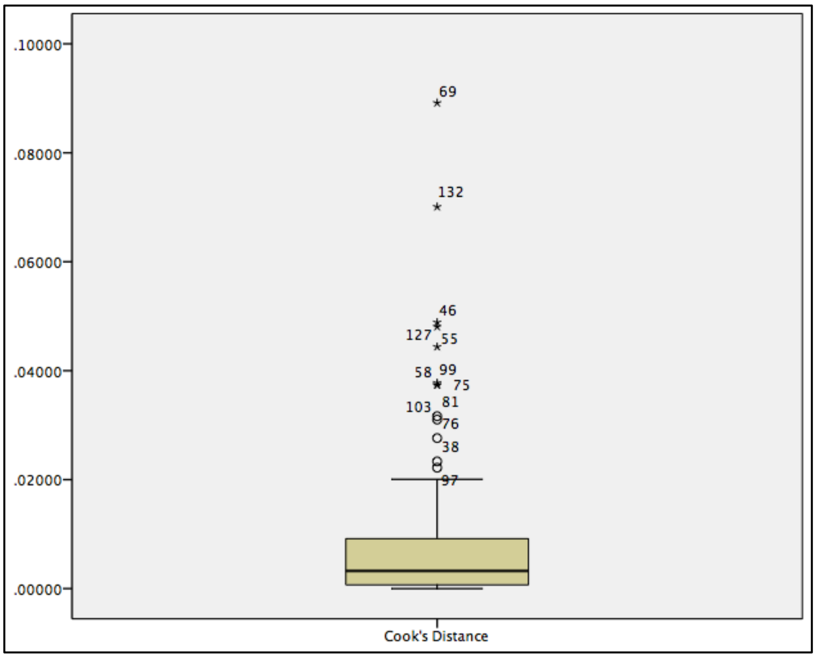
Table 2 – Overall significance of model

Coefficients[a]						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	11.135	1.915		5.815	.000
	Number of hyperlinks (concept links) visited in reading, exercise pages	.034	.131	.024	.260	.795
	Relative amount of revisits	.879	2.445	.038	.360	.720
	Number of Exercises Solved on First Try	-.056	.012	-.452	-4.816	.000
	Number of Exercises Solved on Third Try	.099	.143	.077	.694	.489
	Number of Exercises Finished But not Solved	.131	.084	.175	1.567	.120
	Compactness	4.280	3.224	.134	1.328	.187
	Avg number an notes link is clicked per page	1.297	1.882	.059	.689	.492
	Average View Time on Exercise Pages	.000	.001	-.027	-.286	.775

a. Dependent Variable: Surface subscale Unrelated Memorising

Table 3 – B, Beta, and Sig. values for predictors

Figure 1. Cook's Distance Box Plot



	R^2	Adj. R^2	Sig.
Model 1 (all initially selected predictors)	32.9%	27.8%	0.000
Model 2 (exclusion of cases 69)	34.2%	29.2%	0.000
Model 2a (exclusion of cases 69, 132)	37%	32.2%	0.000
Model 2b (exclusion of cases 69, 132, and 46)	37.9%	33.1%	0.000
Model 2c (exclusion of cases 69, 132, 46, and 127)	40.4%	35.8%	0.000
Model 2d (exclusion of cases 69, 132, 46, 127, 55) [Rejected]	38.2%	33.4%	0.000
Model 2e (exclusion of cases 69, 132, 46, 127, and 58) [Rejected]	40.2%	35.4%	0.000
Model 2f(exclusion of cases 69, 132, 46, 127, and 99) [Rejected]	39.9%	35.1%	0.000
Model 2g(exclusion of cases 69, 132, 46, 127, and 75) [Rejected]	38.3%	33.3%	0.000

Table 4. Summary of measures of variance and significance for accepted and rejected models for exclusion of outliers

Model 6 – Unrelated Memorising – Recommended model

Model Summary[b]				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.633[a]	.400	.377	2.657
a. Predictors: (Constant), Avg number an notes link is clicked per page, Number of Exercises Finished But not Solved, Number of Exercises Solved on First Try, Compactness				
b. Dependent Variable: Surface subscale Unrelated Memorising				

Table 5 – Model Summary

ANOVA[a]						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	499.178	4	124.794	17.675	.000[b]
	Residual	748.408	106	7.060		
	Total	1247.586	110			
a. Dependent Variable: Surface subscale Unrelated Memorising						
b. Predictors: (Constant), Avg number an notes link is clicked per page, Number of Exercises Finished But not Solved, Number of Exercises Solved on First Try, Compactness						

Table 6 – Overall significance of model

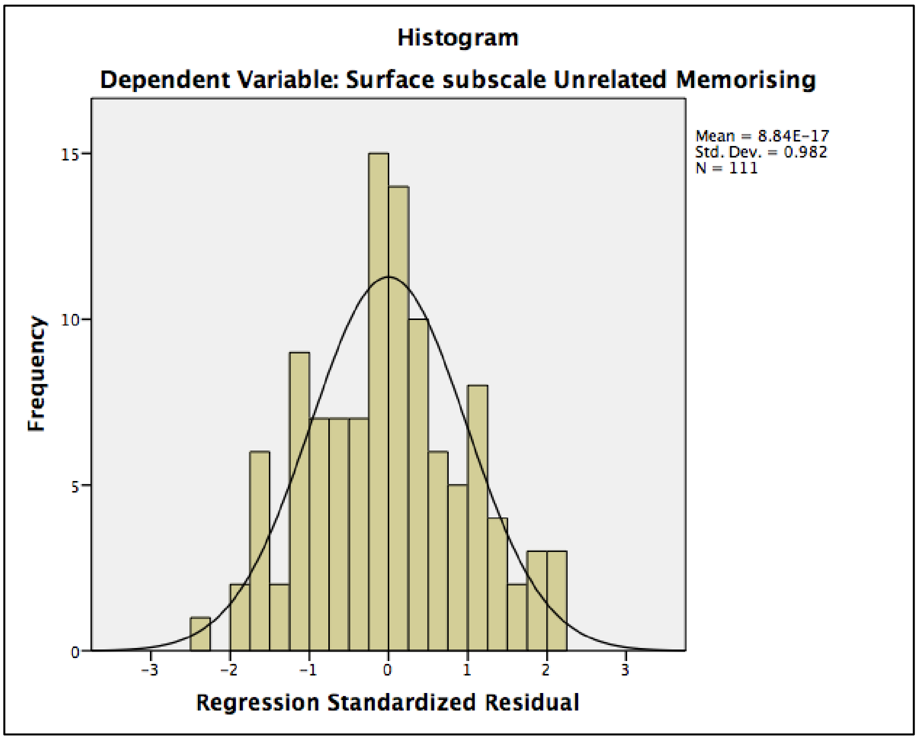
Coefficients[a]						
		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Std. Error	Beta	t	Sig.
1	(Constant)	10.967	1.685		6.507	.000
	Number of Exercises Solved on First Try,	-.064	.009	-.512	-6.745	.000
	Number of Exercises Finished But not Solved	.187	.056	.253	3.309	.001
	Compactness	5.401	2.554	.166	2.115	.037
	Avg number an notes link is clicked per page	1.757	1.715	.080	1.025	.308
a. Dependent Variable: Surface subscale Unrelated Memorising						

Table 7 – B, Beta, and Sig. values for predictors

Appendix 4.2.3 – Unrelated memorising – Model 6 – Generalisation - Assumptions

The assumptions regarding: normal distribution of residuals, homoscedasticity of standardised residuals against predicted ones, and the normality of residuals hold well.

Figure 2. Histogram of standardised residuals for the final model



Regarding the assumption about the normal distribution of residuals, as shown in Figure 2, that residuals fit quite closely to a normal distribution.

Figure 3. Plot of standardised residuals

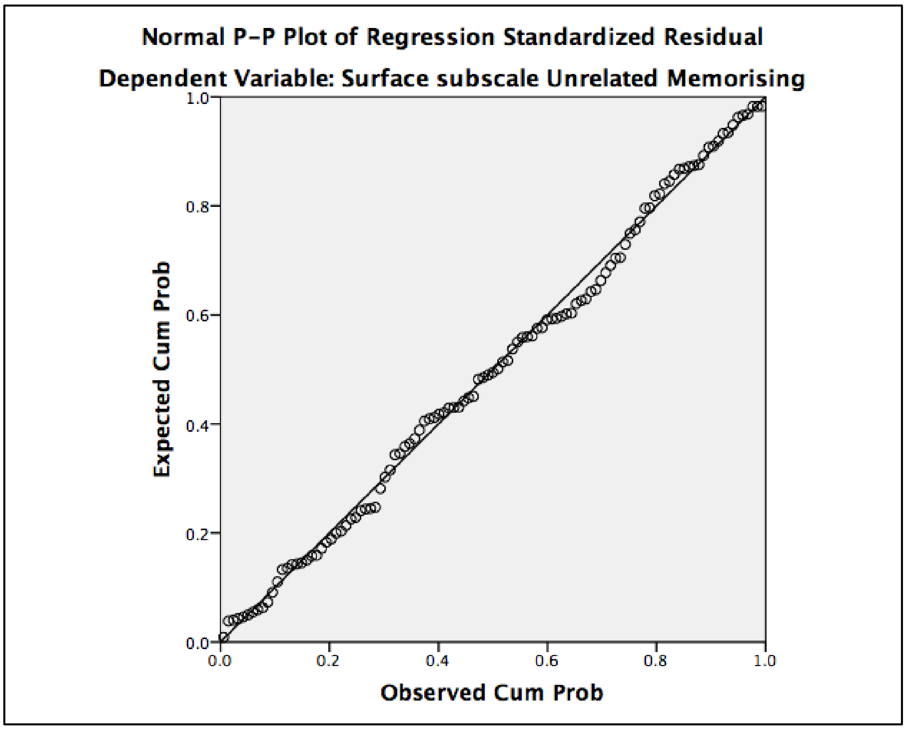
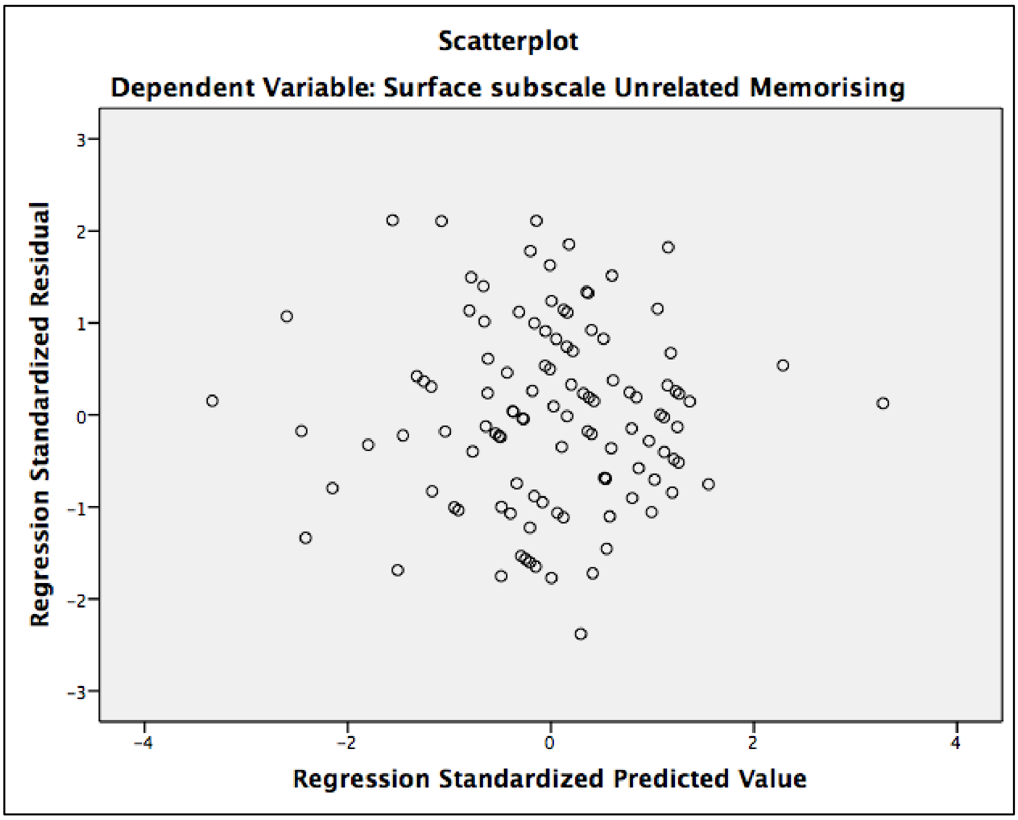


Figure 3 shows that the normality assumption holds since the points lie on the straight line.

Figure 4. Plot of the standardised residuals against the predicted ones for the model



The assumption about the residuals is whether the variance of the residuals is constant, in other words there is homoscedasticity. Figure 4 shows that the scatter plot is reasonably random and the residuals are homoscedastic with a few exceptions.

Appendix 4.2.4 – Unrelated Memorising subscale –Model 7 – Leanest and Meanest

Excluding cases 69 and 132 and 46 and 127 and *Relative amount of revisits, Number of hyperlinks (concept links) visited in reading and exercise pages, Avg View Time Ex.Pages, Number of exercises solved on third try, and Avg number a notes links is clicked per page.*

Model Summary[b]				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.628[a]	.394	.377	2.658

a. Predictors: (Constant), Compactness, Number of Exercises Solved on First Try, Number of Exercises Finished But not Solved

b. Dependent Variable: Surface subscale Unrelated Memorising

Table 8 – Model Summary

ANOVA[a]						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	491.766	3	163.922	23.206	.000[b]
	Residual	755.820	107	7.064		
	Total	1247.586	110			

a. Dependent Variable: Surface subscale Unrelated Memorising

b. Predictors: (Constant), Compactness, Number of Exercises Solved on First Try, Number of Exercises Finished But not Solved

Table 9 – Overall significance of model

Coefficients[a]						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	10.781	1.676		6.432	.000
	Number of Exercises Solved on First Try	-.065	.009	-.520	-6.895	.000
	Number of Exercises Finished But not Solved	.182	.056	.246	3.233	.002
	Compactness	6.036	2.478	.186	2.436	.017

a. Dependent Variable: Surface subscale Unrelated Memorising

Table 10 – B, Beta, and Sig. values for predictors

Appendices 4.3 –Fear of Failure Models

Appendix 4.3.1 –Further Justification for inclusion of predictors

Pre-models	**R^2**	**Adj. R^2**	**Sig.**
Model 1a *(Average View Time on Content Pages, Maximum View Time on Exercise Page, Number of Exercises Solved on Third Try, Average View Time on Exercise Pages, Number of Exercises Solved on First Try, Number of Exercises Finished But Not Solved)*	29.4%	25.5%	0.000
Model 1b *(Maximum View Time on Content Pages, Maximum View Time on Exercise Page, Number of Exercises Solved on Third Try, Average View Time on Exercise Pages, Number of Exercises Solved on First Try, Number of Exercises Finished But Not Solved)*	30.5%	26.6%	0.000

Table1. Pre-models

Appendix 4.3.2 – Further information on development of model

Model 1 – Fear of Failure Scale - All predictors based on initial selection

Model Summary[b]				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.552[a]	.305	.266	3.232

a. Predictors: (Constant), Maximum View Time on Content Page, Average View Time on Exercise Pages, Number of Exercises Solved on Third Try, Maximum View Time on Exercise Page, Number of Exercises Solved on First Try, Number of Exercises Finished But Not Solved

b. Dependent Variable: Surface subscale Fear for Failure

Table 1 – Model Summary

ANOVA[a]						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	494.294	6	82.382	7.884	.000[b]
	Residual	1128.489	108	10.449		
	Total	1622.783	114			

a. Dependent Variable: Surface subscale Fear for Failure

b. Predictors: (Constant), Maximum View Time on Content Page, Average View Time on Exercise Pages, Number of Exercises Solved on Third Try, Maximum View Time on Exercise Page, Number of Exercises Solved on First Try, Number of Exercises Finished But Not Solved

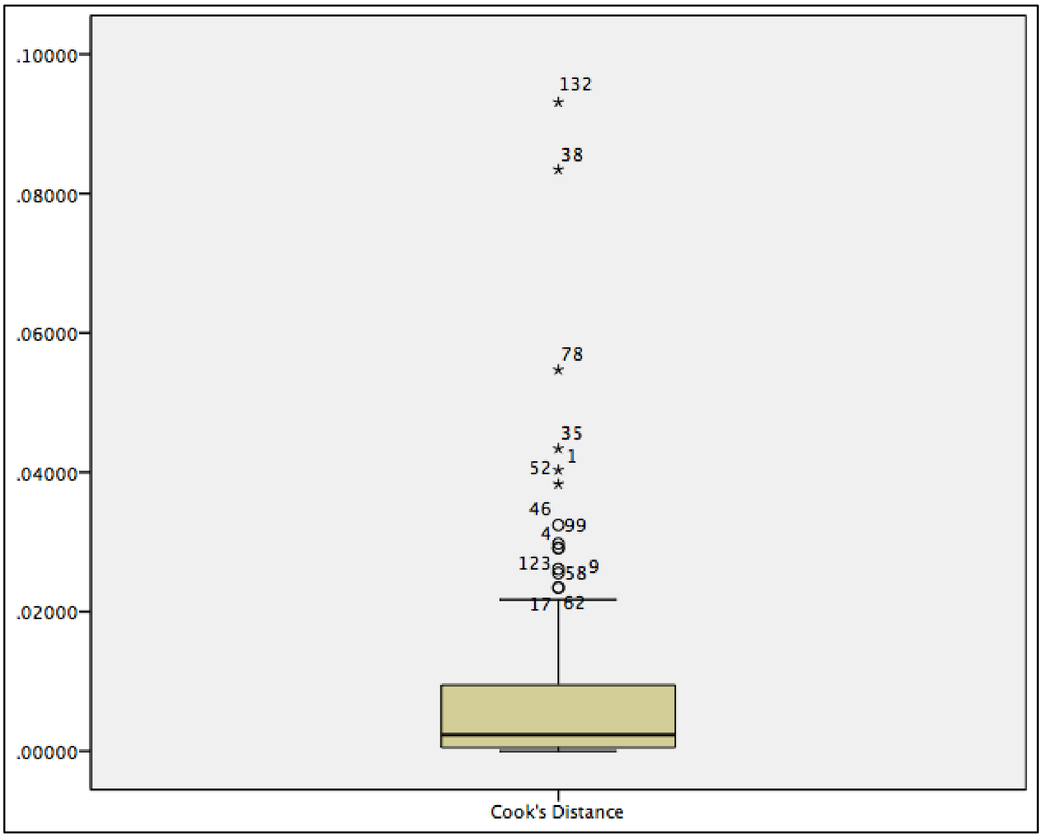
Table 2 – Overall significance of model

Coefficients[a]						
		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Std. Error	Beta	t	Sig.
1	(Constant)	14.161	1.347		10.513	.000
	Number of Exercises Solved on First Try	-.048	.013	-.349	-3.747	.000
	Number of Exercises Solved on Third Try	.332	.157	.232	2.117	.037
	Number of Exercises Finished But Not Solved	.043	.095	.051	.450	.654
	Maximum View Time on Exercise Page	.001	.000	.253	3.012	.003
	Average View Time on Exercise Pages	.000	.001	-.028	-.302	.763
	Maximum View Time on Content Page	-.001	.001	-.160	-1.925	.057

a. Dependent Variable: Surface subscale Fear for Failure

Table 3 – B, Beta, and Sig. values for predictors

Figure 1. Cook's Distance Box Plot



	R^2	Adj. R^2	Sig.
Model 1 (all initially selected predictors)	30.5%	26.6%	0.000
Model 2a (exclusion of case 132)	35.1%	31.5%	0.000
Model 2b (exclusion of cases 132 and 38)	38.4%	34.9%	0.000
Model 2c (exclusion of cases 132, 38, and 78)	40.2%	36.8%	0.000
Model 2d (exclusion of cases 132, 38, 78, and 35) **[Rejected]**	39.3%	35.1%	0.000
Model 2e (exclusion of cases 132, 38, 78, and 1) **[Rejected]**	36.9%	33.3%	0.000
Model 2f (exclusion of cases 132, 38, 78 and 52)	42.1%	38.8%	0.000

Table 4. Summary of measures of variance and significance for accepted and rejected models for exclusion of outliers

Model 4 – Fear of Failure – Recommended version (and Leanest and Meanest)

Model Summary[b]				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.645[a]	.416	.394	2.848
a. Predictors: (Constant), Maximum View Time on Content Page, Number of Exercises Solved on First Try, Number of Exercises Solved on Third Try, Maximum View Time on Exercise Page				
b. Dependent Variable: Surface subscale Fear for Failure				

Table 5 – Model Summary

ANOVA[a]						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	611.943	4	152.986	18.856	.000[b]
	Residual	860.021	106	8.113		
	Total	1471.964	110			
a. Dependent Variable: Surface subscale Fear for Failure						
b. Predictors: (Constant), Maximum View Time on Content Page, Number of Exercises Solved on First Try, Number of Exercises Solved on Third Try, Maximum View Time on Exercise Page						

Table 6 – Overall significance of model

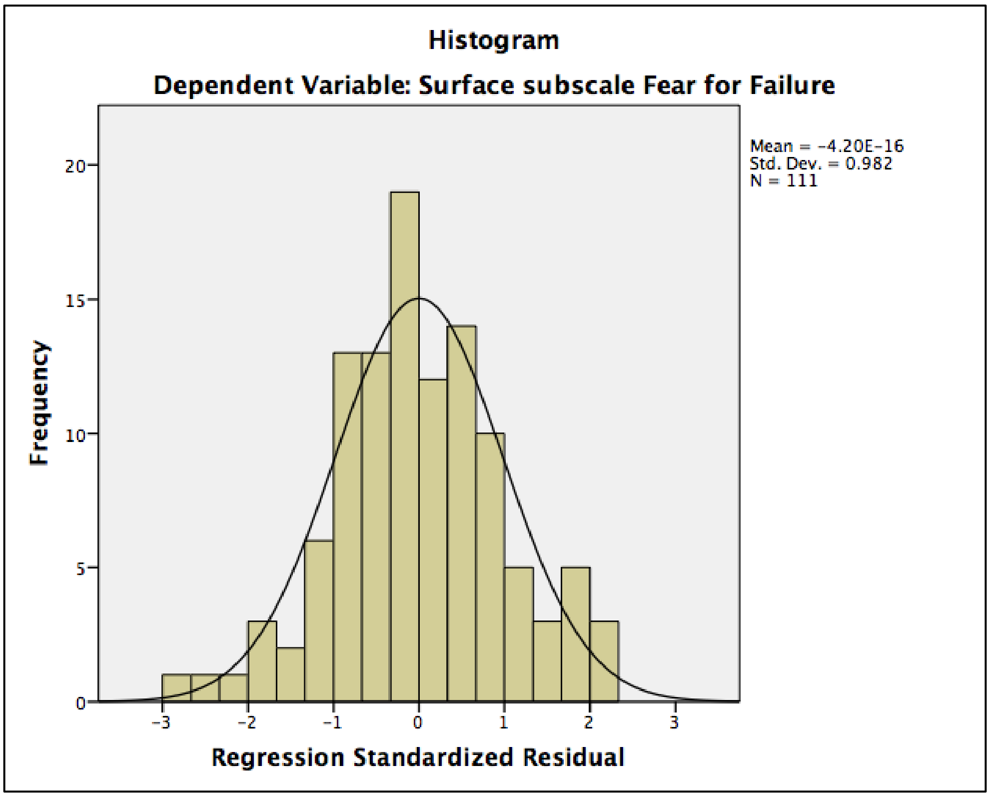
Coefficients[a]						
		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Std. Error	Beta	t	Sig.
1	(Constant)	14.714	.936		15.717	.000
	Number of Exercises Solved on First Try	-.060	.010	-.440	-5.785	.000
	Number of Exercises Solved on Third Try	.348	.104	.251	3.351	.001
	Maximum View Time on Exercise Page	.001	.000	.281	3.657	.000
	Maximum View Time on Content Page	-.001	.001	-.208	-2.733	.007
a. Dependent Variable: Surface subscale Fear for Failure						

Table 7 – B, Beta, and Sig. values for predictors

Appendix 4.3.3 – Fear of Failure – Model 4 – Generalisation - Assumptions

The assumptions regarding: normal distribution of residuals, homoscedasticity of standardised residuals against predicted ones, and the normality of residuals hold well.

Figure 2. Histogram of standardised residuals for the final model



Regarding the assumption about the normal distribution of residuals, as shown in Figure 2, that residuals fit quite closely to a normal distribution.

Figure 3. Plot of standardised residuals

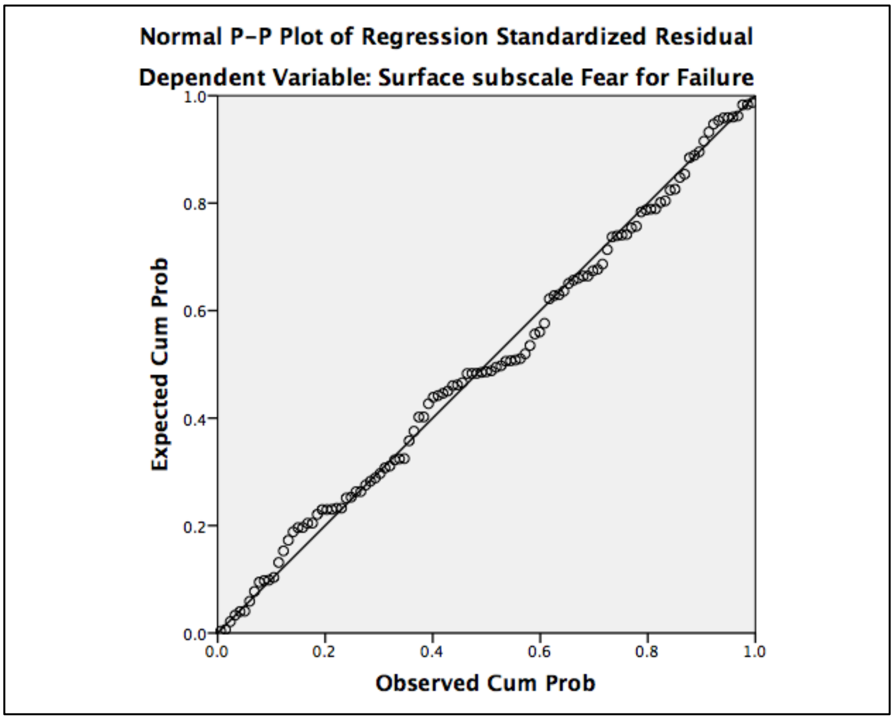
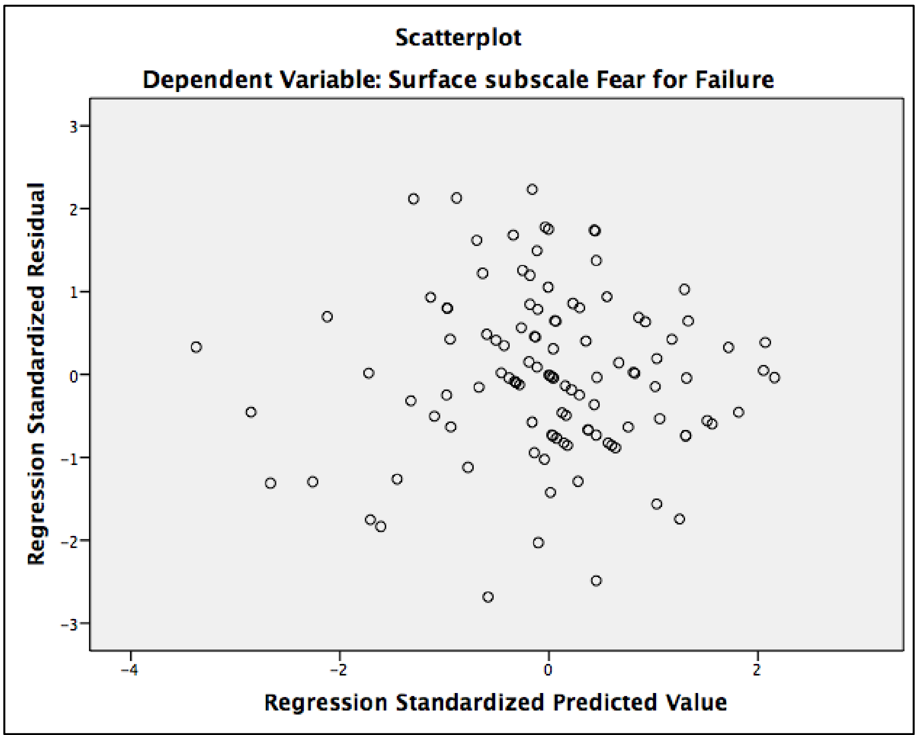


Figure 3 shows that the normality assumption holds since the points lie on the straight line.

Figure 4. Plot of the standardised residuals against the predicted ones for the model



The assumption about the residuals is whether the variance of the residuals is constant, in other words there is homoscedasticity. Figure 4 shows that the scatter plot is reasonably random and most residuals are homoscedastic.

Appendices 4.4 –Syllabus Boundness Models

Appendix 4.4.1 –Further Justification for inclusion of predictors

Pre-models	R^2	Adj. R^2	Sig.
Model 1a *(Average View Time on Content Pages, Compactness, Number of Exercises Cancelled, Number of Exercises Solved on Third Try, Number of Exercises Solved on First Try, Number of pages visited from TOC, Number of Exercises Finished But Not Solved)* **[With only 7 predictors]**	14.4%	8.8%	0.017
Model 1b ***Minimum View Time on Exercise Page****, Number of Exercises Finished But Not Solved, Number of Exercises Solved on First Try, Compactness, Number of Exercises Cancelled, Average View Time on Content Pages, Number of pages visited from TOC, Number of Exercises Solved on Third Try* **[Recommended model]**	19.7%	13.3%	0.002
Model 1c ***Minimum View Time on Content Page****, Number of Exercises Solved on Third Try, Compactness, Number of Exercises Cancelled, Number of Exercises Solved on First Try, Average View Time on Content Pages, Number of pages visited from TOC, Number of Exercises Finished But Not Solved*	15.9	9.5	0.016

Model 1d ***Average View Time on Exercise Pages,*** *Compactness, Average View Time on Content Pages, Number of Exercises Cancelled, Number of Exercises Solved on Third Try, Number of pages visited from TOC, Number of Exercises Solved on First Try, Number of Exercises Finished But Not Solved*	14.5%	8%	0.029
Model 1e ***Maximum View Time on Exercise Page****, Number of Exercises Cancelled, Compactness, Average View Time on Content Pages, Number of Exercises Solved on Third Try, Number of Exercises Solved on First Try, Number of pages visited from TOC, Number of Exercises Finished But Not Solved*	14.6%	8.1%	0.029

Table 1. Pre-models

Note: *maximum view time on content pages* is not tried in a pre-model because of the multicollinearity issue with *average view time on content pages*

Appendix 4.4.2 – Further information on development of model

Model 1b – Syllabus Boundness subscale - All predictors based on initial selection

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.444[a]	.197	.136	2.673

a. Predictors: (Constant), Minimum View Time on Exercise Page, Number of Exercises Finished But Not Solved, Number of Exercises Solved on First Try, Compactness, Number of Exercises Cancelled, Average View Time on Content Pages, Number of pages visited from TOC, Number of Exercises Solved on Third Try

b. Dependent Variable: Surface subscale Syllabus-boundness

Table 1 – Model Summary

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	185.604	8	23.200	3.246	.002[b]
	Residual	757.526	106	7.146		
	Total	943.130	114			

a. Dependent Variable: Surface subscale Syllabus-boundness

b. Predictors: (Constant), Minimum View Time on Exercise Page, Number of Exercises Finished But Not Solved, Number of Exercises Solved on First Try, Compactness, Number of Exercises Cancelled, Average View Time on Content Pages, Number of pages visited from TOC, Number of Exercises Solved on Third Try

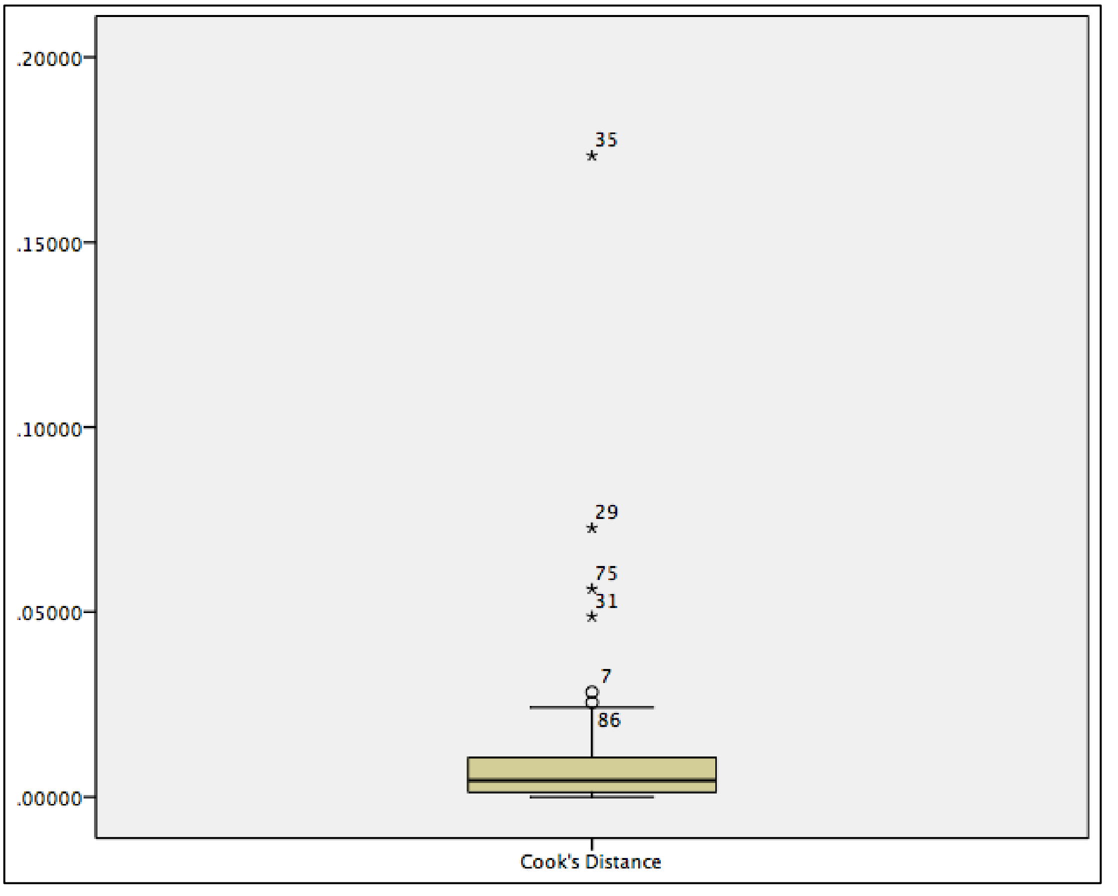
Table 2 – Overall significance of model

Coefficients[a]						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	11.820	1.839		6.428	.000
	Number of Exercises Solved on First Try	-.020	.010	-.192	-1.987	.050
	Number of Exercises Solved on Third Try	.141	.133	.129	1.055	.294
	Number of Exercises Finished But Not Solved	.041	.079	.065	.519	.605
	Number of Exercises Cancelled	.109	.076	.141	1.428	.156
	Number of pages visited from TOC	.000	.017	.002	.022	.982
	Compactness	2.737	2.614	.101	1.047	.297
	Average View Time on Content Pages	.004	.002	.162	1.584	.116
	Minimum View Time on Exercise Page	-.011	.004	-.251	-2.634	.010

a. Dependent Variable: Surface subscale Syllabus-boundness

Table 3 – B, Beta, and Sig. values for predictors

Figure 1. Cook's Distance Box Plot



	R^2	Adj. R^2	Sig.
Model 1b (all initially selected predictors)	19.7%	13.6%	0.002
Model 2a (exclusion of case 35)	20.9%	14.9%	0.001
Model 2b (exclusion of case 35 and 29) **[Rejected]**	18.1%	11.8%	0.006
Model 2c (exclusion of case 35 and 75) **[Rejected]**	20%	13.9%	0.002
Model 2d (exclusion of cases 35 and 31)	21.7%	15.7%	0.001

Table 4. Summary of measures of variance and significance for accepted and rejected models for exclusion of outliers

Model 4 – Syllabus Boundness – Recommended model

Model Summary[b]				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.465[a]	.216	.171	2.591
a. Predictors: (Constant), Minimum View Time on Exercise Page, Number of Exercises Solved on Third Try, Number of Exercises Solved on First Try, Compactness, Number of Exercises Cancelled, Average View Time on Content Pages				
b. Dependent Variable: Surface subscale Syllabus-boundness				

Table 5 – Model Summary

ANOVA[a]						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	195.764	6	32.627	4.862	.000[b]
	Residual	711.369	106	6.711		
	Total	907.133	112			
a. Dependent Variable: Surface subscale Syllabus-boundness						
b. Predictors: (Constant), Minimum View Time on Exercise Page, Number of Exercises Solved on Third Try, Number of Exercises Solved on First Try, Compactness, Number of Exercises Cancelled, Average View Time on Content Pages						

Table 6 – Overall significance of model

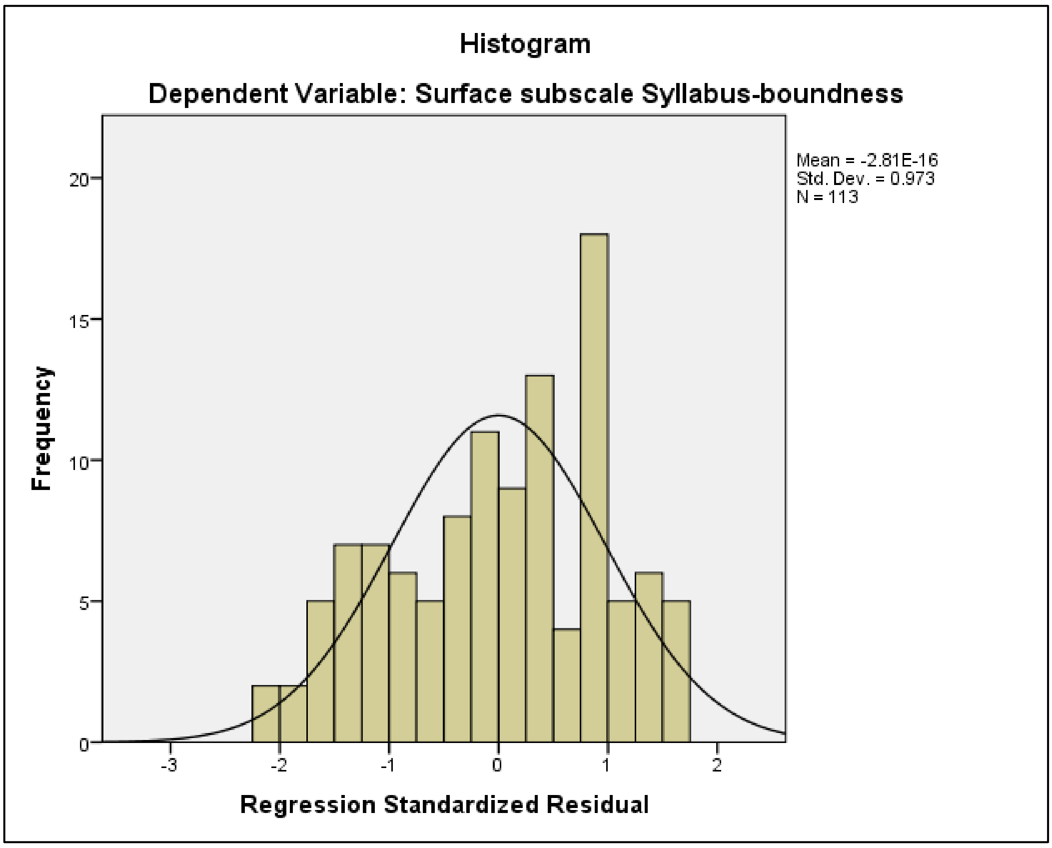
Coefficients[a]						
		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Std. Error	Beta	t	Sig.
1	(Constant)	10.871	1.758		6.182	.000
	Number of Exercises Solved on First Try	-.014	.010	-.138	-1.428	.156
	Number of Exercises Solved on Third Try	.190	.099	.177	1.931	.056
	Number of Exercises Cancelled	.129	.069	.170	1.861	.065
	Compactness	3.250	2.446	.119	1.329	.187
	Average View Time on Content Pages	.007	.003	.233	2.295	.024
	Minimum View Time on Exercise Page	-.011	.004	-.253	-2.753	.007
a. Dependent Variable: Surface subscale Syllabus-boundness						

Table 7 – B, Beta, and Sig. values for predictors

Appendix 4.4.3 – Syllabus Boundness – Model 4 – Generalisation - Assumptions

The assumptions regarding: normal distribution of residuals, homoscedasticity of standardised residuals against predicted ones, and the normality of residuals hold well.

Figure 2. Histogram of standardised residuals for the final model



Regarding the assumption about the normal distribution of residuals, as shown in Figure 2, that residuals fit quite closely to a normal distribution.

Figure 3. Plot of standardised residuals

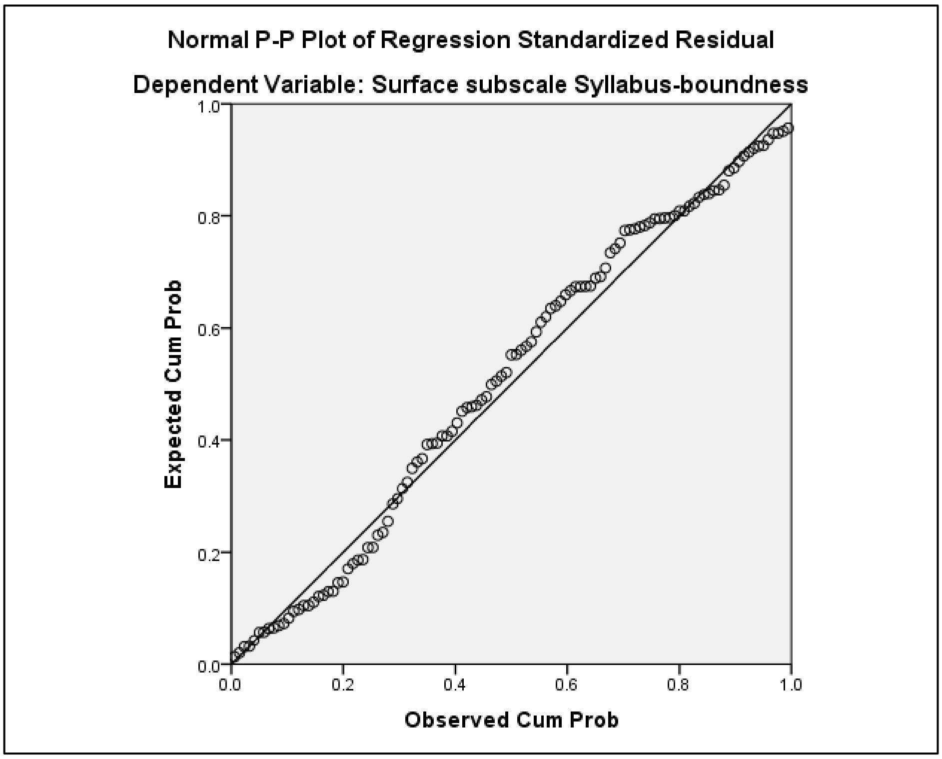
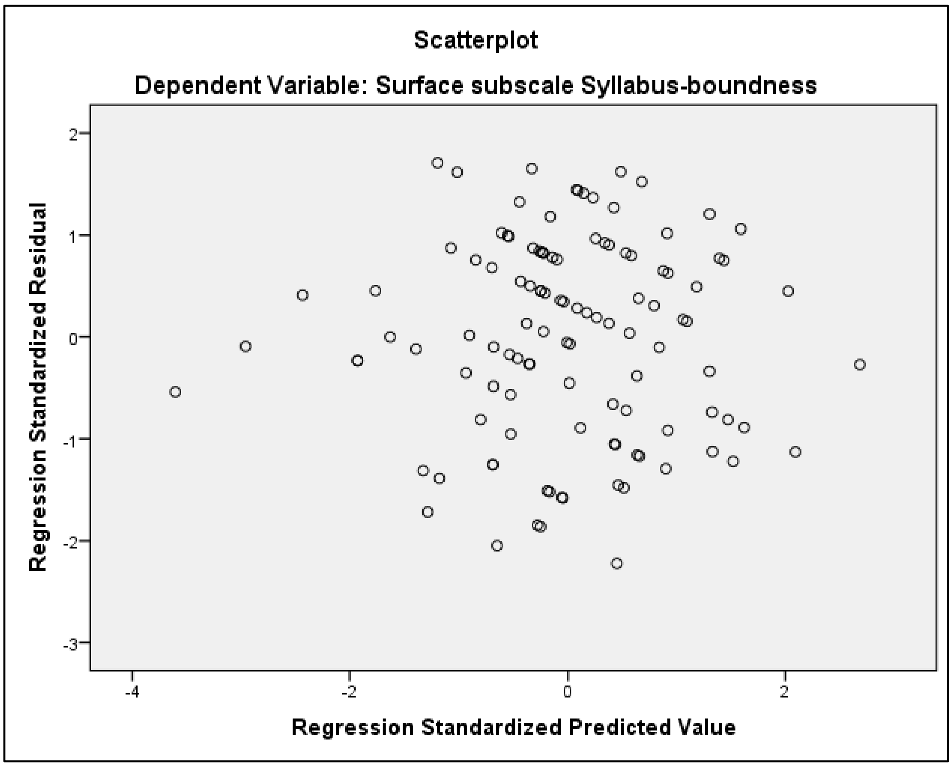


Figure 3 shows that the normality assumption holds since the points lie on the straight line.

Figure 4. Plot of the standardised residuals against the predicted ones for the model



The assumption about the residuals is whether the variance of the residuals is constant, in other words there is homoscedasticity. Figure 4 shows that the scatter plot is reasonably random and that most residuals are homoskedastic.

Appendix 4.4.4 – Syllabus Boundness subscale –Model 6 – Leanest and Meanest

Exclusion of outliers 35 and 31 and Number of pages visited from TOC and Number of Exercises Finished but not Solved and Compactness and Number of exercises solved on first try

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.433[a]	.187	.157	2.613

a. Predictors: (Constant), Minimum View Time on Exercise Page, Number of Exercises Solved on Third Try, Number of Exercises Cancelled, Average View Time on Content Pages

b. Dependent Variable: Surface subscale Syllabus-boundness

Table 8 – Model Summary

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	169.756	4	42.439	6.216	.000[b]
	Residual	737.377	108	6.828		
	Total	907.133	112			

a. Dependent Variable: Surface subscale Syllabus-boundness

b. Predictors: (Constant), Minimum View Time on Exercise Page, Number of Exercises Solved on Third Try, Number of Exercises Cancelled, Average View Time on Content Pages

Table 9 – Overall significance of model

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	11.883	.612		19.415	.000
	Number of Exercises Solved on Third Try	.227	.097	.211	2.347	.021
	Number of Exercises Cancelled	.152	.068	.201	2.245	.027
	Average View Time on Content Pages	.009	.003	.312	3.372	.001
	Minimum View Time on Exercise Page	-.012	.004	-.274	-3.001	.003

a. Dependent Variable: Surface subscale Syllabus-boundness

Table 10 – B, Beta, and Sig. values for predictors

Appendices 4.5 – Lack of Purpose Models

Appendix 4.5.1 –Further Justification for inclusion of predictors

Non-Selected Predictors	**Reason for non-selection**
Temporal metrics	There are no clear theoretical or empirical indications regarding the direction of the relationship between these metrics and the subscale.
Stratum	*Stratum* correlates highly with predictor *Compactness* (r=-0.790). Therefore, as explained in the strategy, only one of them will be included in the first version of the model. As they both correlate to the subscale, the inclusion will be based on which one has the strongest correlation. As shown in 4.5.2, amongst the two metrics the one with the strongest correlation is *Compactness*.

Table 1. Reasons for not selecting predictors

Appendix 4.5.2 – Further information on development of model

Model 1 – Lack of Purpose- All predictors based on initial selection

Model Summary[b]				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.401[a]	.161	.114	3.208

a. Predictors: (Constant), Number of pages visited from TOC, Number of Exercises Solved on First Try, Number of Exercises Solved on Third Try, Compactness, Relative amount of revisits, Number of Exercises Finished But Not Solved

b. Dependent Variable: Surface subscale Lack of purpose

Table 1 – Model Summary

ANOVA[a]						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	212.835	6	35.473	3.448	.004[b]
	Residual	1111.113	108	10.288		
	Total	1323.948	114			

a. Dependent Variable: Surface subscale Lack of purpose

b. Predictors: (Constant), Number of pages visited from TOC, Number of Exercises Solved on First Try, Number of Exercises Solved on Third Try, Compactness, Relative amount of revisits, Number of Exercises Finished But Not Solved

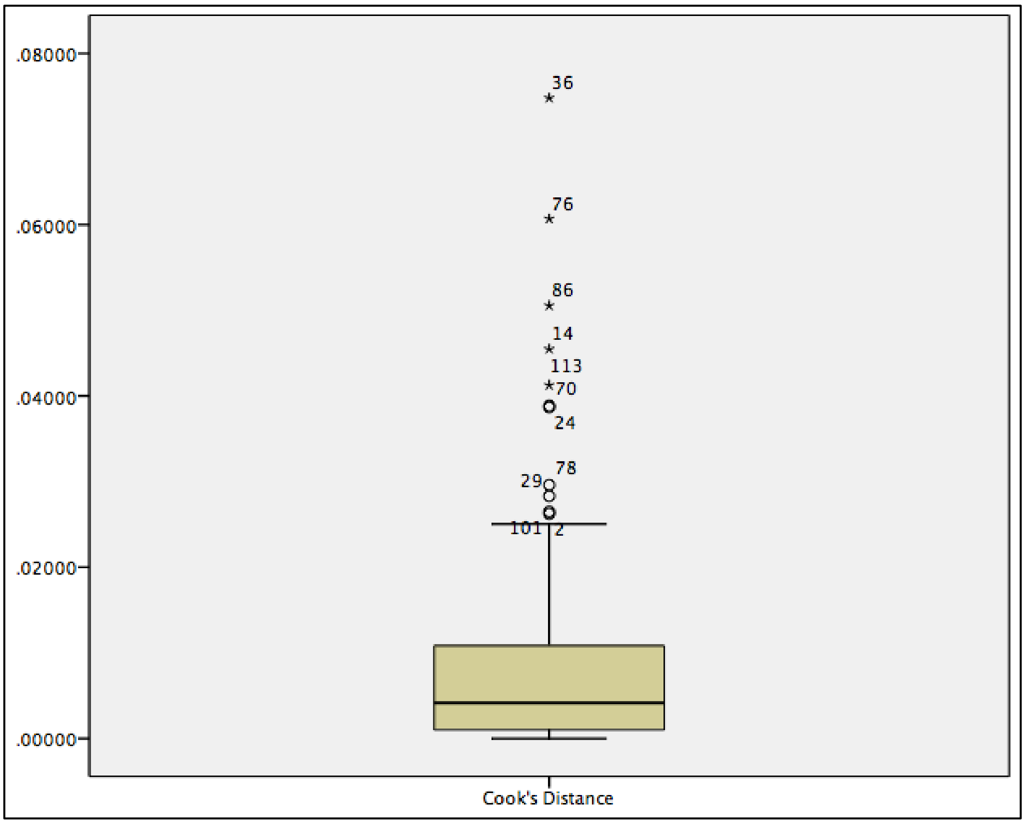
Table 2 – Overall significance of model

Coefficients[a]						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	5.386	1.985		2.713	.008
	Number of Exercises Solved on First Try	-.022	.012	-.175	-1.892	.061
	Number of Exercises Solved on Third Try	.172	.157	.133	1.097	.275
	Number of Exercises Finished But Not Solved	.111	.094	.147	1.174	.243
	Compactness	4.714	3.542	.147	1.331	.186
	Relative amount of revisits	1.194	2.843	.051	.420	.675
	Number of pages visited from TOC	.003	.020	.016	.145	.885

a. Dependent Variable: Surface subscale Lack of purpose

Table 3 – B, Beta, and Sig. values for predictors

Figure 1. Cook's Distance Box Plot



	R^2	Adj. R^2	Sig.
Model 1 (all initially selected predictors)	16.1%	11.4%	0.004
Model 2a (exclusion of case 36) [Rejected]	15.7%	10.9%	0.005
Model 2b (exclusion of case 76)	16.6%	12%	0.003
Model 2c (exclusion of cases 76, and 86)	18.9%	14.3%	0.001
Model 2d (exclusion of cases 76, 86, and 14) [Rejected]	18.8%	14.1%	0.001
Model 2e (exclusion of cases 76, 86, and 113) [Rejected]	18.5%	13.8%	0.001
Model 2f (exclusion of cases 76, 86, and 70)	19.6%	15%	0.001
Model 2g (exclusion of cases 76, 86, 70, and 24)	21%	16.4%	0.000

Table 4. Summary of measures of variance and significance for accepted and rejected models for exclusion of outliers

Model 4 – Lack of Purpose – Recommended model

Model Summary[b]				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.458[a]	.210	.180	2.871
a. Predictors: (Constant), Relative amount of revisits, Number of Exercises Finished But Not Solved, Number of Exercises Solved on First Try, Number of Exercises Solved on Third Try				
b. Dependent Variable: Surface subscale Lack of purpose				

Table 5 – Model Summary

ANOVA[a]						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	231.761	4	57.940	7.031	.000[b]
	Residual	873.536	106	8.241		
	Total	1105.297	110			
a. Dependent Variable: Surface subscale Lack of purpose						
b. Predictors: (Constant), Relative amount of revisits, Number of Exercises Finished But Not Solved, Number of Exercises Solved on First Try, Number of Exercises Solved on Third Try						

Table 6 – Overall significance of model

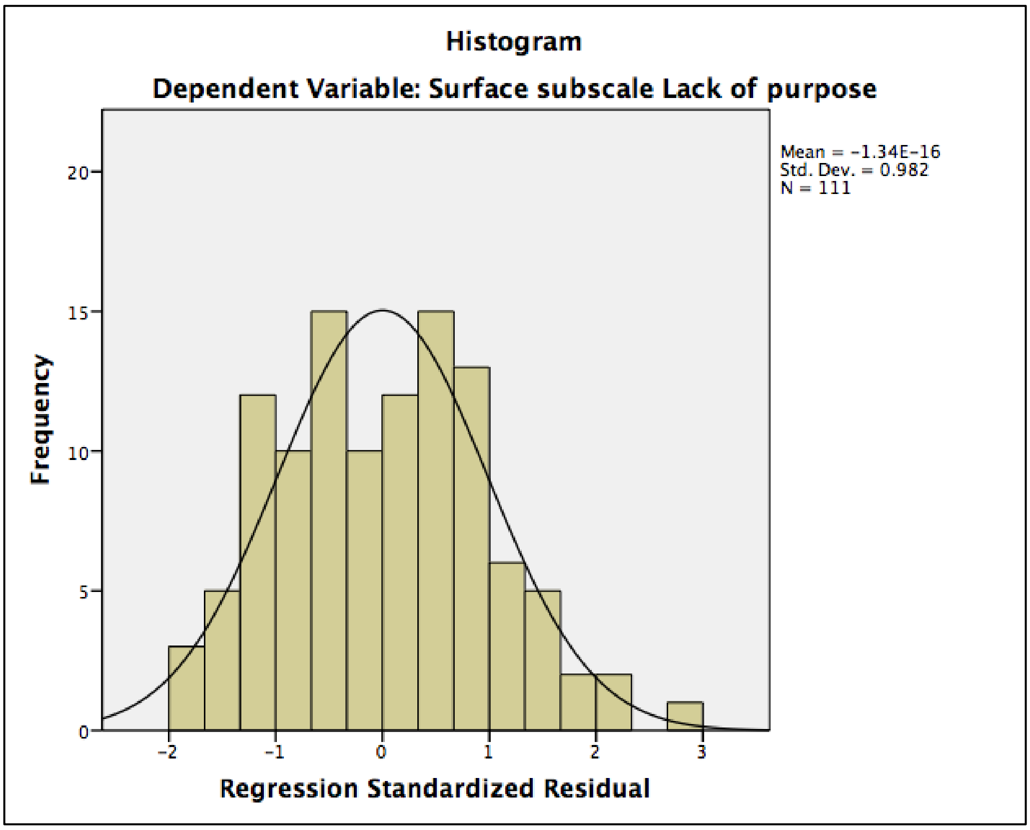
Coefficients[a]						
		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Std. Error	Beta	t	Sig.
1	(Constant)	6.028	1.007		5.984	.000
	Number of Exercises Solved on First Try	-.013	.010	-.114	-1.278	.204
	Number of Exercises Solved on Third Try	.145	.141	.122	1.035	.303
	Number of Exercises Finished But Not Solved	.124	.081	.178	1.524	.131
	Relative amount of revisits	6.129	2.025	.275	3.027	.003
a. Dependent Variable: Surface subscale Lack of purpose						

Table 7 – B, Beta, and Sig. values for predictors

Appendix 4.5.3 – Lack of Purpose – Model 4 – Generalisation - Assumptions

The assumptions regarding: normal distribution of residuals, homoscedasticity of standardised residuals against predicted ones, and the normality of residuals hold well.

Figure 2. Histogram of standardised residuals for the final model



Regarding the assumption about the normal distribution of residuals, as shown in Figure 2, that residuals fit quite closely to a normal distribution.

Figure 3. Plot of standardised residuals

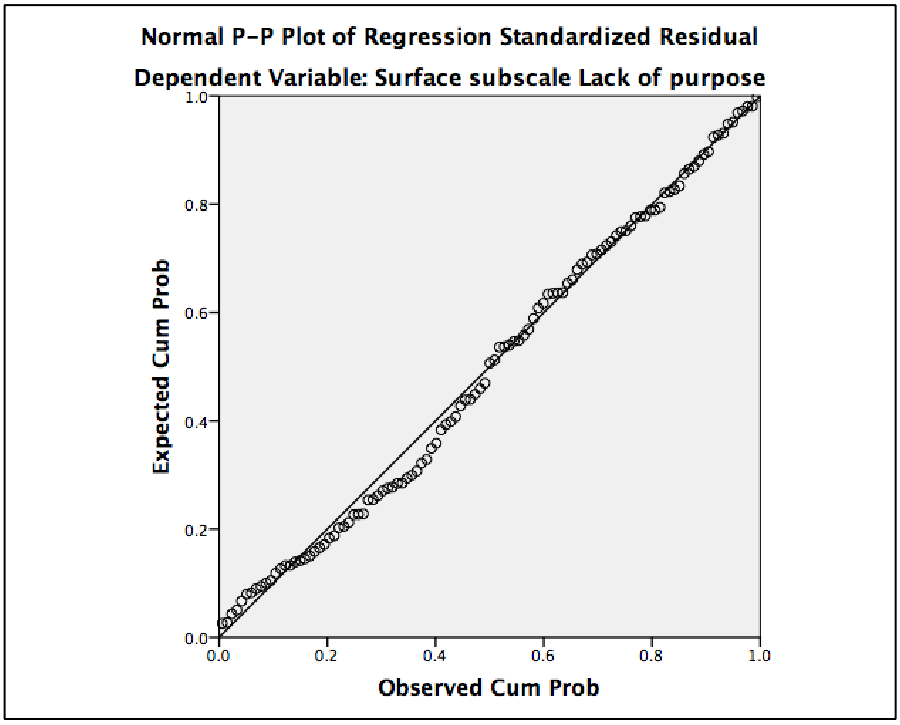
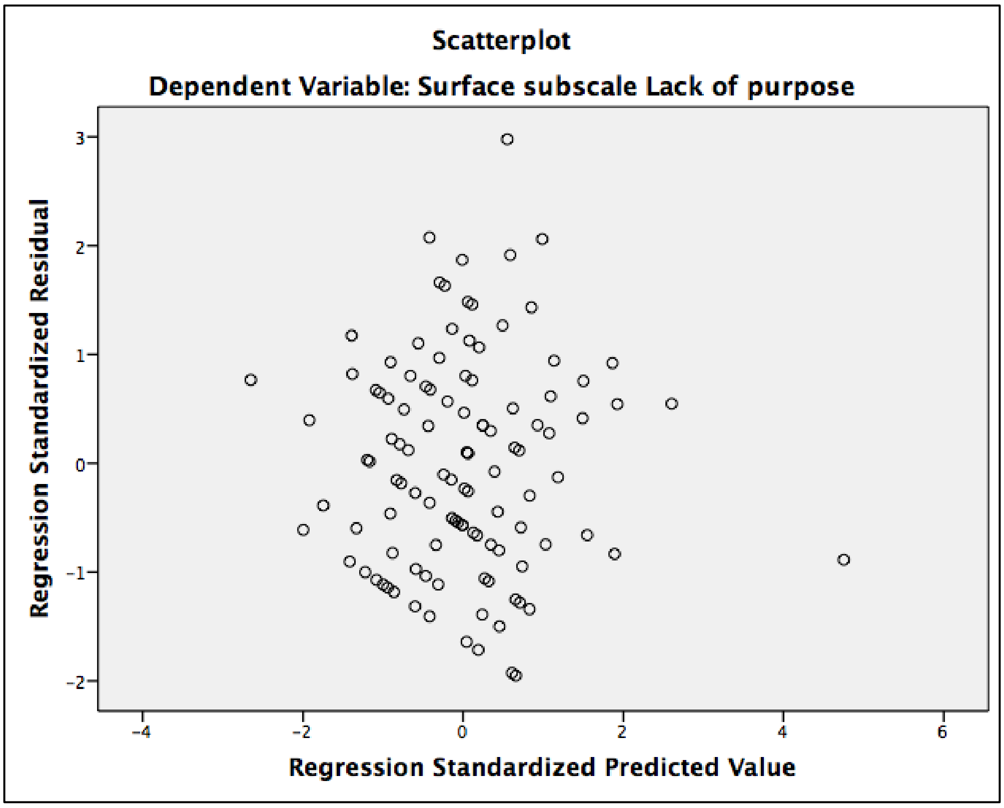


Figure 3 shows that the normality assumption holds since the points lie on the straight line.

Figure 4. Plot of the standardised residuals against the predicted ones for the model



The assumption about the residuals is whether the variance of the residuals is constant, in other words there is homoscedasticity. Figure 4 shows that the scatter plot is reasonably random and the residuals are homoscedastic with a few exceptions.

Appendix 4.5.4 – Lack of Purpose subscale –Model 6 – Leanest and Meanest

Exclusion of outliers 76 and 86 and 70 and *number of pages visited from TOC*, *compactness, number exercises solved on third try*, and *number of exercises solved on first try*

Model Summary[b]				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.435[a]	.190	.174	2.880
a. Predictors: (Constant), Relative amount of revisits, Number of Exercises Finished But Not Solved				
b. Dependent Variable: Surface subscale Lack of purpose				

Table 8 – Model Summary

ANOVA[a]						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	209.457	2	104.729	12.626	.000[b]
	Residual	895.840	108	8.295		
	Total	1105.297	110			
a. Dependent Variable: Surface subscale Lack of purpose						
b. Predictors: (Constant), Relative amount of revisits, Number of Exercises Finished But Not Solved						

Table 9 – Overall significance of model

Coefficients[a]						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	5.293	.796		6.653	.000
	Number of Exercises Finished But Not Solved	.178	.061	.256	2.926	.004
	Relative amount of revisits	7.039	1.951	.316	3.608	.000
a. Dependent Variable: Surface subscale Lack of purpose						

Table 10 – B, Beta, and Sig. values for predictors

Appendices 4.6 – Deep Models

Appendix 4.6.1 –Further Justification for inclusion of predictors

Selected Predictor	Reason for selection
Number of exercises solved on first try ***Number of exercises solved on second try*** ***Number of exercises solved on third try*** ***Number of exercises finished but not solved***	-Theoretical connections discussed in section 4.6.1 -Enriching further the discussion by allowing useful comparisons to the surface scales, as shown in sections 4.1 - 4.5 - Their combination gives a more complete picture as to how students deal with their exercises during their tutorial sessions according to the specific approach to studying
Number of hyperlinks (concepts links) visited on exercise and reading pages ***Average number a 'notes' link is clicked per page***	-Theoretical connections discussed in section 4.6.1 -Enriching further the discussion by allowing useful comparison to the surface scales (in which they have been found to have statistical connections). Note: *Average number a 'notes' link is clicked per page* is statistically more relevant to the deep scales (i.e. it correlated to 'relating ideas' subscale), so it is selected instead of *number of times 'notes' link is clicked.*
Maximum view time on an exercise page	Statistical (see 4.6.2)

Table 1. Reasons for selection

Non-Selected Predictors	Reason for non-selection
Average view time on content (reading) pages ***Maximum view time on a content (reading) page*** ***Average view time on exercise pages***	-No strong indications, based on the theory, as to which of these predictors would be the most 'enriching' for the model. -When tried in pre-models 1a, 1b and 1c neither of these metrics contribute to the 'deep' model with both a higher R^2 and a higher Adjusted R^2 compared to Model 1 (see table 3 below).

Table 2. Reasons for not selecting predictors

Pre-models	R^2	Adj. R^2	Sig.
Model 1 *(Avg number an notes link is clicked per page, Number of Exercises Solved on Second Try, Maximum View Time on Exercise Page, Number of concept links visited in reading and exercise pages, Number of Exercises Solved on First Try, Number of Exercises Finished But Not Solved, Number of Exercises Solved on Third Try)*	13.3%	7.7%	0.029
Model 1a *(Average view time on exercise pages, Avg number an notes link is clicked per page, Number of Exercises Solved on Second Try, Maximum View Time on Exercise Page, Number of concept links visited in reading and exercise pages, Number of Exercises Solved on First Try, Number of Exercises Finished But Not Solved, Number of Exercises Solved on Third Try)*	13.8%	7.3%	0.40
Model 1b *(Maximum view time on a reading page, Avg number an notes link is clicked per page, Number of Exercises Solved on Second Try, Maximum View Time on Exercise Page, Number of concept links visited in reading and exercise pages, Number of Exercises Solved on First Try, Number of Exercises Finished But Not Solved, Number of Exercises Solved on Third Try)*	14%	7.6%	0.036

Model 1c *(Average view time on reading (content) pages, Avg number an notes link is clicked per page, Number of Exercises Solved on Second Try, Maximum View Time on Exercise Page, Number of concept links visited in reading and exercise pages, Number of Exercises Solved on First Try, Number of Exercises Finished But Not Solved, Number of Exercises Solved on Third Try)*	13.7%	7.2%	0.042

Table 3. Pre-models

Appendix 4.6.2 – Detailed discussion on development of model

Model 1 – Deep Scale - All predictors based on initial selection

Model Summary[b]				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.365[a]	.133	.077	8.348

a. Predictors: (Constant), Avg number an notes link is clicked per page, Number of Exercises Solved on Second Try, Maximum View Time on Exercise Page, Number of concept links visited in reading and exercise pages, Number of Exercises Solved on First Try, Number of Exercises Finished But Not Solved, Number of Exercises Solved on Third Try

b. Dependent Variable: Deep Scale

Table 1 – Model Summary

ANOVA[a]						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1145.816	7	163.688	2.349	.029[b]
	Residual	7455.957	107	69.682		
	Total	8601.774	114			

a. Dependent Variable: Deep Scale

b. Predictors: (Constant), Avg number an notes link is clicked per page, Number of Exercises Solved on Second Try, Maximum View Time on Exercise Page, Number of concept links visited in reading and exercise pages, Number of Exercises Solved on First Try, Number of Exercises Finished But Not Solved, Number of Exercises Solved on Third Try

Table 2 – Overall Significance of model

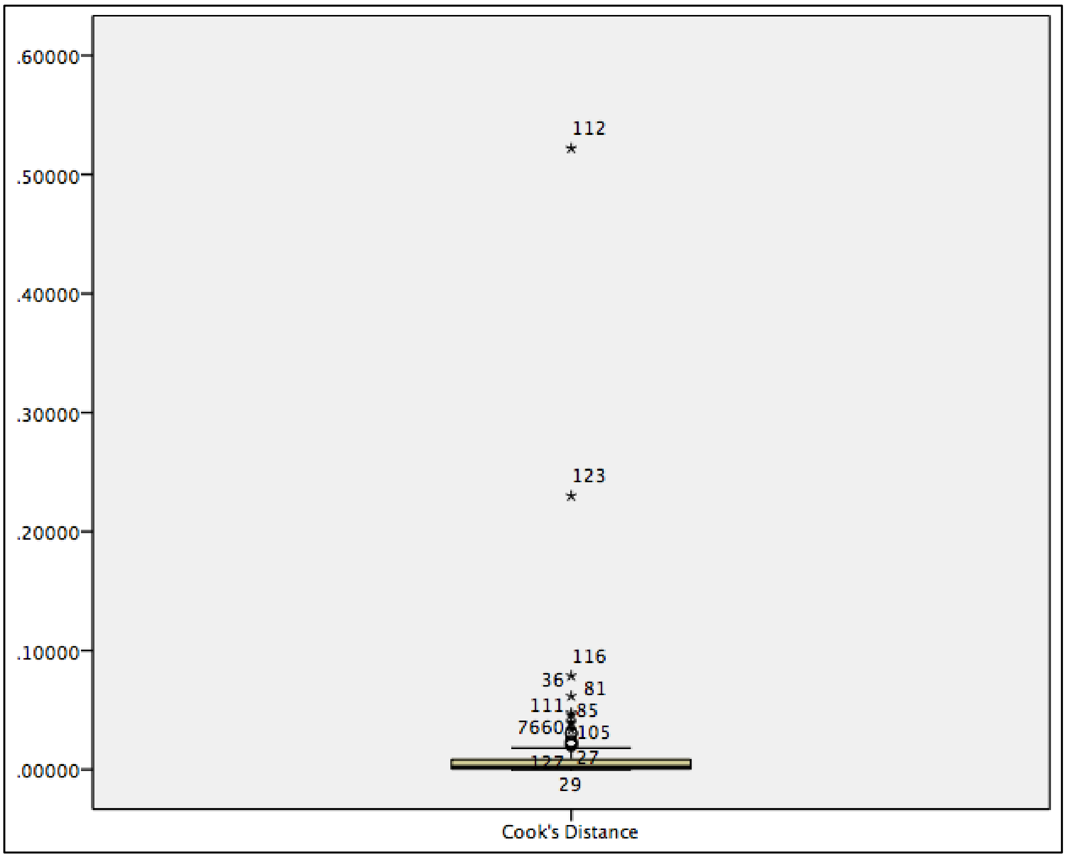
Coefficients[a]						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	57.933	2.719		21.310	.000
	Number of Exercises Solved on First Try	.030	.033	.095	.926	.357
	Number of Exercises Solved on Second Try	-.644	.234	-.347	-2.757	.007
	Number of Exercises Solved on Third Try	.229	.444	.070	.517	.606
	Number of Exercises Finished But Not Solved	.418	.250	.218	1.671	.098
	Maximum View Time on Exercise Page	.003	.001	.246	2.603	.011
	Number of concept links visited in reading and exercise pages	-.274	.357	-.074	-.767	.445
	Avg number an notes link is clicked per page	-1.746	5.410	-.031	-.323	.748
a. Dependent Variable: Deep Scale						

Table 3 – B, Beta, and Sig. values for predictors

Deep Scale – Exclusion of outliers

To exclude the outliers from Model 1 the intention is to use the Cook's Distance method. The box plot of Cook's distance below shows a number of outliers. However, the intention is to exclude amongst the most extreme ones (those indicated with an asterisk), those which improve the measures of variance R^2 and adjusted R^2.

Figure 1. – Cook's Distance Box Plot



In Fig.1, it is observed that the most extreme outliers are cases: 112, 123, 116, 36, 81, 111, 85, 76, 60, and 105. These cases are excluded gradually on models Model 2a, 2b, 2c, 2d, 2e, 2f, 2g, 2h, 2i, 2j. It is observed that the exclusion of cases 112, 116, 111, 85, 76, and 105 has increased both R^2 and adjusted R^2 (from 13.3% in Model 1 to 18.2% in Model 2j for R^2, and from 7.7% in Model 1 to 12.9% in Model 2j for adjusted R^2). The only exception is the exclusion of outliers 123, 36, 81, and 60 in Models 2b, 2d, 2e, and 2i; since these models do not increase the variance, they are rejected.

Model 4 – Deep Scale- Recommended version

Model Summary[b]				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.428[a]	.183	.144	7.551
a. Predictors: (Constant), Avg number an notes link is clicked per page, Number of Exercises Finished But Not Solved, Number of Exercises Solved on First Try, Maximum View Time on Exercise Page, Number of Exercises Solved on Second Try				
b. Dependent Variable: Deep Scale				

Table 4 – Model Summary

ANOVA[a]						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1319.549	5	263.910	4.628	.001[b]
	Residual	5873.386	103	57.023		
	Total	7192.936	108			
a. Dependent Variable: Deep Scale						
b. Predictors: (Constant), Avg number an notes link is clicked per page, Number of Exercises Finished But Not Solved, Number of Exercises Solved on First Try, Maximum View Time on Exercise Page, Number of Exercises Solved on Second Try						

Table 5 – Overall significance of model

Coefficients[a]						
		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Std. Error	Beta	t	Sig.
1	(Constant)	56.977	2.524		22.574	.000
	Number of Exercises Solved on First Try	.036	.029	.122	1.225	.223
	Number of Exercises Solved on Second Try	-.637	.216	-.326	-2.945	.004
	Number of Exercises Finished But Not Solved	.488	.195	.267	2.506	.014
	Maximum View Time on Exercise Page	.003	.001	.251	2.614	.010
	Avg number an notes link is clicked per page	10.619	6.460	.154	1.644	.103
a. Dependent Variable: Deep Scale						

Table 6 – B, Beta, and Sig. values for predictors

Appendix 4.6.3 – Deep Scale – Model 6 – Leanest and Meanest

Excluding outliers 112, 116, 111, 85, 76, and 105 – *Number of exercises solved on third try, Number of concept links visited in reading and exercise pages, Number of exercises solved on first try and Avg number a notes link is clicked per page*

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.391[a]	.153	.129	7.617

a. Predictors: (Constant), Maximum View Time on Exercise Page, Number of Exercises Finished But Not Solved, Number of Exercises Solved on Second Try

b. Dependent Variable: Deep Scale

Table 7 – Model Summary

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1100.391	3	366.797	6.321	.001[b]
	Residual	6092.545	105	58.024		
	Total	7192.936	108			

a. Dependent Variable: Deep Scale

b. Predictors: (Constant), Maximum View Time on Exercise Page, Number of Exercises Finished But Not Solved, Number of Exercises Solved on Second Try

Table 8 – Overall Significance of model

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	58.788	2.228		26.382	.000
	Number of Exercises Solved on Second Try	-.564	.203	-.289	-2.781	.006
	Number of Exercises Finished But Not Solved	.423	.189	.231	2.236	.027
	Maximum View Time on Exercise Page	.003	.001	.269	2.931	.004

a. Dependent Variable: Deep Scale

Table 9 – B, Beta, and Sig. values for predictors

Appendices 4.7 – Interest in Ideas Models

Appendix 4.7.1 –Further Justification for inclusion of predictors

Selected Predictor	Reason for selection
Number of exercises solved on first try	-Theoretical connections discussed in section 4.7.1 -Enriching further the discussion by allowing useful comparisons to the surface scales, as shown in sections 4.1-4.5 - Their combination gives a more complete picture as to how students deal with their exercises during their tutorial sessions according to the specific approach to studying.
Number of exercises solved on second try	
Number of exercises solved on third try	
Number of exercises finished but not solved	
Maximum view time on an exercise page	Statistical (see 4.7.2)
Average view time on exercise pages	
Number of hyperlinks (concepts links) visited on exercise and reading pages	-It is enriching to know whether students with high score on intrinsic interest tend to use more such a common ILE feature in order to follow up mathematical concepts, compared to those with low scores. -It allows for comparisons with the surface scales, since it has been found to correlate unexpectedly in a positive direction with some of them (see sections 4.1 and 4.2).

Table 1. Reasons for selection

Non-Selected Predictors	Reason for non-selection
Average view time on content (reading) pages	-No strong indications, based on the theory, as to which of these predictors would be the most 'enriching' for the model. -When tried in pre-models 1a, 1b neither of these metrics contribute to the 'interest in ideas' model with both a higher R^2 and a higher Adjusted R^2 compared to Model 1 (see table 3 below).
Maximum view time on a content (reading) page	
Metrics related to use of search option mentioned in 3.7.1	Their variation is close to 0, so they cannot be included as predictors.

Table 2. Reasons for not selecting predictors

Pre-models	**R^2**	**Adj. R^2**	**Sig.**
Model 1 *(Number of concept links visited in reading and exercise pages, Maximum View Time on Exercise Page, Number of Exercises Solved on Second Try, Average View Time on Exercise Pages, Number of Exercises Finished But Not Solved, Number of Exercises Solved on First Try, Number of Exercises Solved on Third Try)*	9.9%	4%	0.121
Model 1a *(**average view time on reading pages**, Maximum View Time on Exercise Page, Number of concept links visited in reading and exercise pages, Number of Exercises Finished But Not Solved, Average View Time on Exercise Pages, Number of Exercises Solved on First Try, Number of Exercises Solved on Second Try, Number of Exercises Solved on Third Try)*	10 %	3.2%	0.175
Model 1b *(**maximum view time on a reading page**, Maximum View Time on Exercise Page, Number of concept links visited in reading and exercise pages, Number of Exercises Finished But Not Solved, Average View Time on Exercise Pages, Number of Exercises Solved on First Try, Number of Exercises Solved on Second Try, Number of Exercises Solved on Third Try)*	10%	3.2%	0.174

Table 3. Pre-models

Appendix 4.7.2 – Further information on development of model

Model 1 –Interest in Ideas - All predictors based on initial selection

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.315[a]	.099	.040	2.938

a. Predictors: (Constant), Number of concept links visited in reading and exercise pages, Maximum View Time on Exercise Page, Number of Exercises Solved on Second Try, Average View Time on Exercise Pages, Number of Exercises Finished But Not Solved, Number of Exercises Solved on First Try, Number of Exercises Solved on Third Try

b. Dependent Variable: Deep subscale Interest in Ideas

Table 1 – Model Summary

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	101.607	7	14.515	1.682	.121[b]
	Residual	923.488	107	8.631		
	Total	1025.096	114			

a. Dependent Variable: Deep subscale Interest in Ideas

b. Predictors: (Constant), Number of concept links visited in reading and exercise pages, Maximum View Time on Exercise Page, Number of Exercises Solved on Second Try, Average View Time on Exercise Pages, Number of Exercises Finished But Not Solved, Number of Exercises Solved on First Try, Number of Exercises Solved on Third Try

Table 2 – Overall significance of model

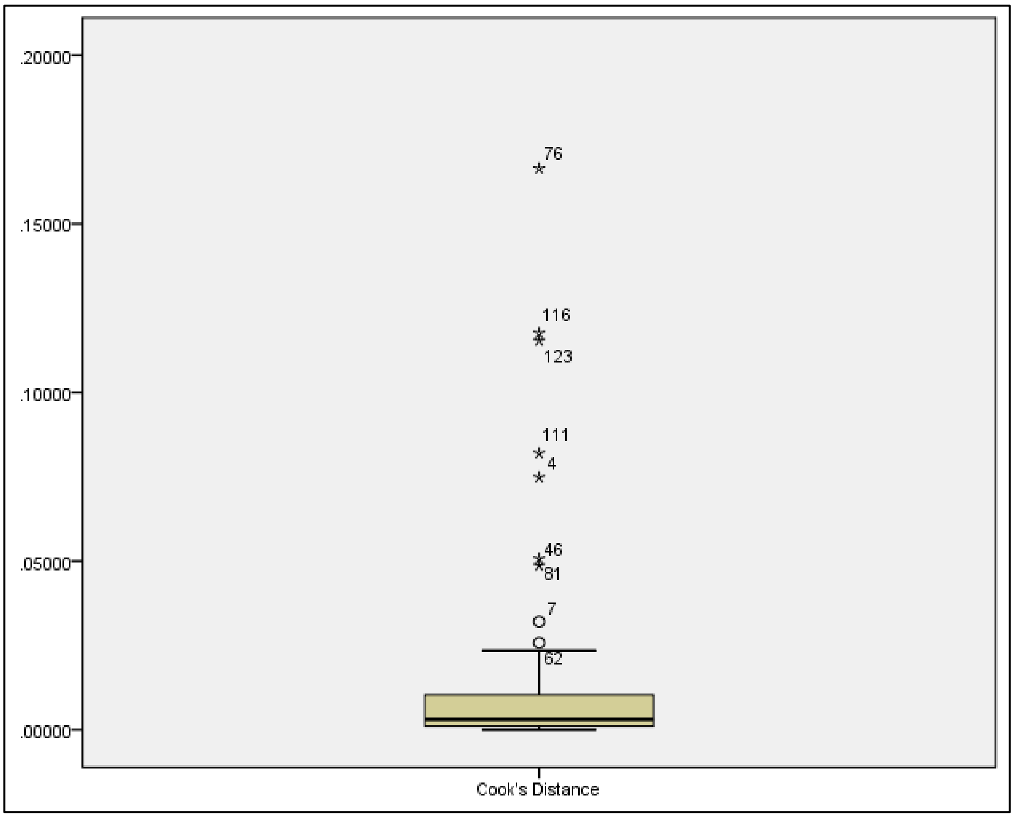
Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	13.210	1.232		10.724	.000
	Number of Exercises Solved on First Try	-.003	.013	-.024	-.209	.835
	Number of Exercises Solved on Second Try	-.044	.082	-.069	-.539	.591
	Number of Exercises Solved on Third Try	-.313	.156	-.275	-1.999	.048
	Number of Exercises Finished But Not Solved	.231	.089	.348	2.580	.011
	Average View Time on Exercise Pages	.001	.001	.085	.780	.437
	Maximum View Time on Exercise Page	.001	.000	.173	1.814	.072
	Number of concept links visited in reading and exercise pages	.052	.124	.041	.422	.674

a. Dependent Variable: Deep subscale Interest in Ideas

Table 3 – B, Beta, and Sig. values for predictors

Figure 1. Cook's Distance Box Plot



	R^2	Adj. R^2	Sig.
Model 1 (all initially selected predictors)	9.9%	4%	0.121
Model 2a (exclusion of case 76)	13.4%	7.7%	0.029
Model 2b (exclusion of case 76, 116)	14.4%	8.7%	0.019
Model 2c (exclusion of cases 76, 116, and 123) **[Rejected]**	12.3%	6.4%	0.051
Model 2d (exclusion of cases 76, 116, and 111)	14.7%	8.9%	0.018
Model 2e exclusion of cases 76, 116, 111, and 4)	15.3%	9.5%	0.015
Model 2f (exclusion of cases 76, 116, 111, 4, 46)	15.8%	10%	0.012
Model 2g (exclusion of cases 76, 116, 111, 4, 46 and 81) **[Rejected]**	15.6%	9.8%	0.014

Table 4. Summary of measures of variance and significance for accepted and rejected models for exclusion of outliers

Model 4 – Interest in Ideas – Recommended version

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.396[a]	.157	.116	2.758

a. Predictors: (Constant), Maximum View Time on Exercise Page, Number of Exercises Solved on Third Try, Average View Time on Exercise Pages, Number of Exercises Solved on Second Try, Number of Exercises Finished But Not Solved

b. Dependent Variable: Deep subscale Interest in Ideas

Table 5 – Model Summary

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	147.232	5	29.446	3.872	.003[b]
	Residual	790.959	104	7.605		
	Total	938.191	109			

a. Dependent Variable: Deep subscale Interest in Ideas

b. Predictors: (Constant), Maximum View Time on Exercise Page, Number of Exercises Solved on Third Try, Average View Time on Exercise Pages, Number of Exercises Solved on Second Try, Number of Exercises Finished But Not Solved

Table 6 – Overall significance of model

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	13.112	.939		13.960	.000
	Number of Exercises Solved on Second Try	-.120	.081	-.170	-1.485	.140
	Number of Exercises Solved on Third Try	-.262	.151	-.234	-1.741	.085
	Number of Exercises Finished But Not Solved	.254	.085	.385	2.988	.004
	Average View Time on Exercise Pages	.001	.001	.108	1.134	.259
	Maximum View Time on Exercise Page	.001	.000	.226	2.400	.018

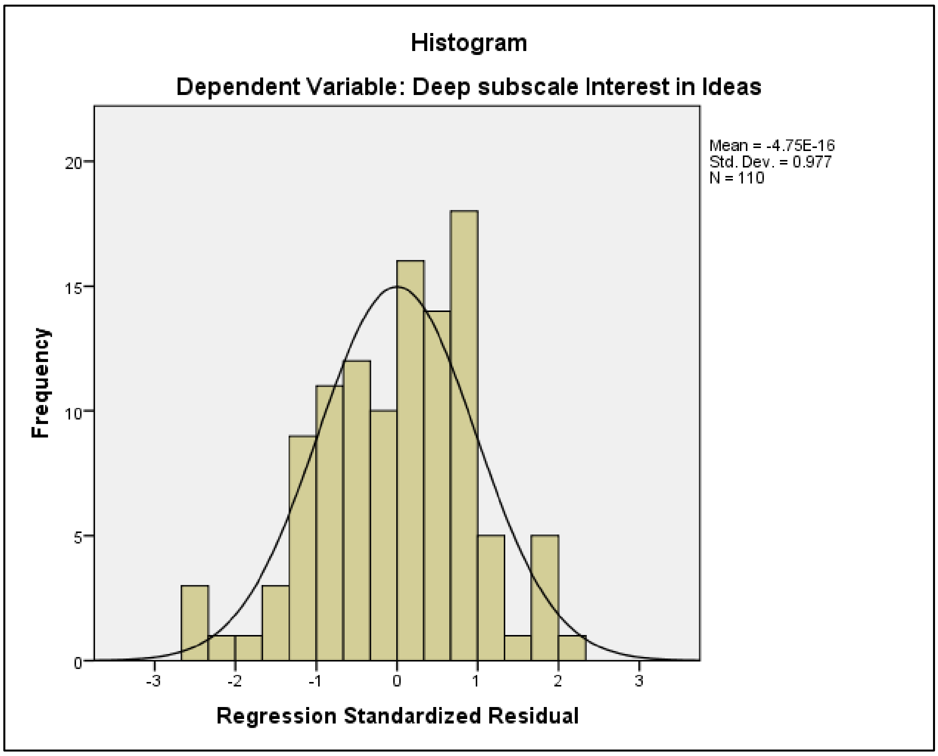
a. Dependent Variable: Deep subscale Interest in Ideas

Table 7 – B, Beta, and Sig. values for predictors

Appendix 4.7.3 – Interest in Ideas – Model 4 – Generalisation - Assumptions

The assumptions regarding: normal distribution of residuals, homoscedasticity of standardised residuals against predicted ones, and the normality of residuals hold well.

Figure 2. Histogram of standardised residuals for the final model



Regarding the assumption about the normal distribution of residuals, as shown in Figure 2, that residuals fit quite closely to a normal distribution.

Figure 3. Plot of standardised residuals

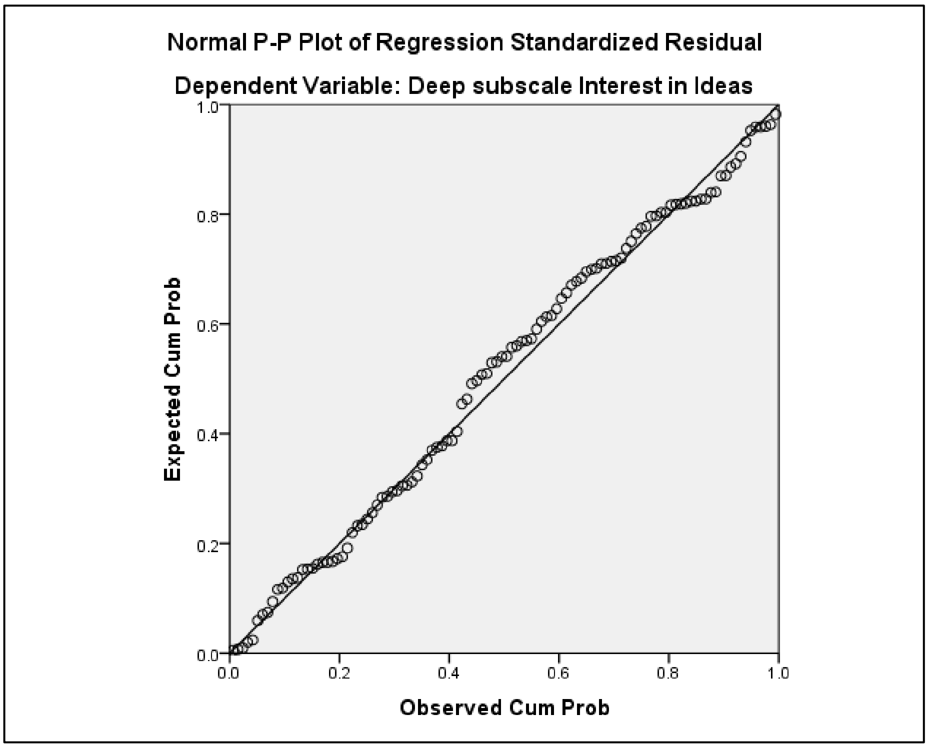
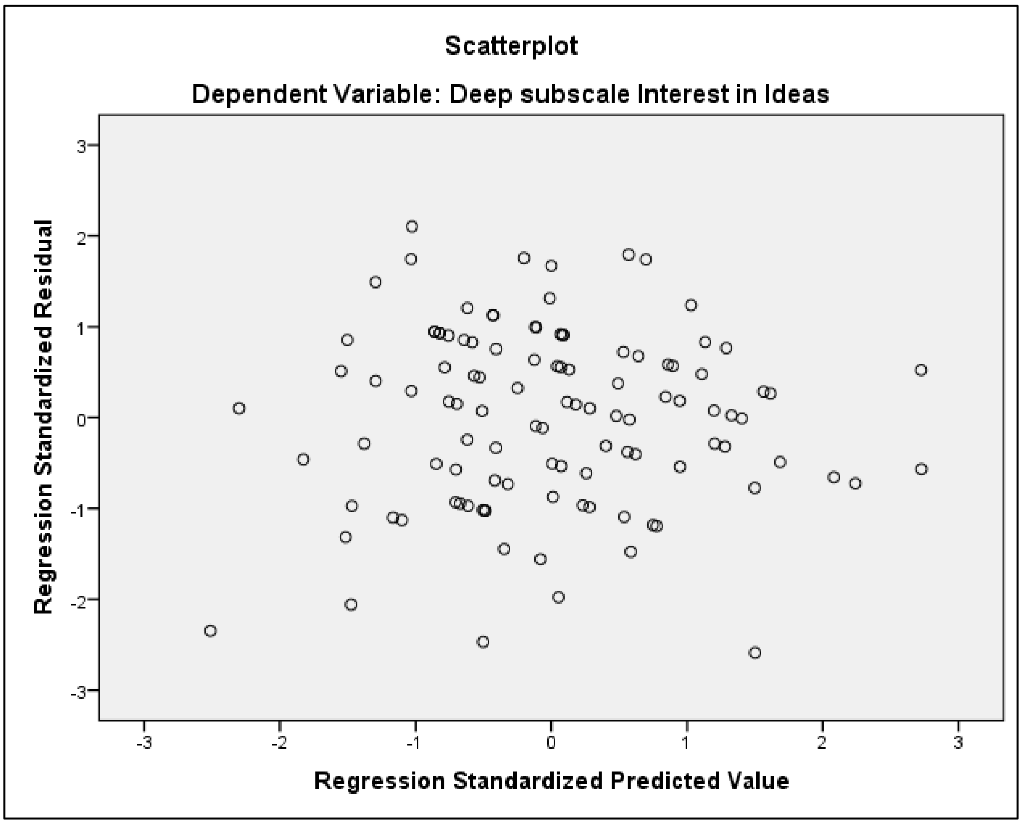


Figure 3 shows that the normality assumption holds since all points are quite close to the straight line.

Figure 4. Plot of the standardised residuals against the predicted ones for the model



The assumption about the residuals is whether the variance of the residuals is constant, in other words there is homoscedasticity. Figure 4 shows that the scatter plot is reasonably random and the residuals are homoscedastic with a few exceptions.

Appendix 4.7.4 – Interest in Ideas subscale –Model 6 – Leanest and Meanest

Exclusion outliers 76, 116, 111, 4, and 46 and *Number of Exercises solved on first try* and *Number of concept links in reading and exercise pages* and *Average view time on exercise pages* and *Number of exercises solved on second try*

Model Summary[b]				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.349[a]	.122	.097	2.788
a. Predictors: (Constant), Maximum View Time on Exercise Page, Number of Exercises Solved on Third Try, Number of Exercises Finished But Not Solved				
b. Dependent Variable: Deep subscale Interest in Ideas				

Table 8 – Model Summary

ANOVA[a]						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	114.320	3	38.107	4.903	.003[b]
	Residual	823.870	106	7.772		
	Total	938.191	109			
a. Dependent Variable: Deep subscale Interest in Ideas						
b. Predictors: (Constant), Maximum View Time on Exercise Page, Number of Exercises Solved on Third Try, Number of Exercises Finished But Not Solved						

Table 9 – Overall significance of model

Coefficients[a]						
		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Std. Error	Beta	t	Sig.
1	(Constant)	12.847	.680		18.880	.000
	Number of Exercises Solved on Third Try	-.351	.142	-.313	-2.462	.015
	Number of Exercises Finished But Not Solved	.225	.085	.342	2.659	.009
	Maximum View Time on Exercise Page	.001	.000	.271	2.933	.004
a. Dependent Variable: Deep subscale Interest in Ideas						

Table 10 – B, Beta, and Sig. values for predictors

Appendices 4.8 – Seeking Meaning Models

Appendix 4.8.1 –Further Justification for inclusion of predictors

Selected Predictor	**Reason for selection**
Number of exercises solved on first try	-Theoretical connections discussed in section 4.8.1 -Enriching further the discussion by allowing useful comparisons to the surface scales, as shown in sections 4.1- 4.5 - Their combination gives a more complete picture as to how students deal with their exercises during their tutorial sessions according to the specific approach to studying.
Number of exercises solved on second try	
Number of exercises solved on third try	
Number of exercises finished but not solved	
Maximum view time on an exercise page	Statistical (see 4.8.2)
Number of pages visited using the TOC	
Number of hyperlinks (concepts links) visited on exercise and reading pages	-It is enriching to know whether those with high a score on the subscale tend to use more such a common ILE feature in order obtain personal understanding for the concepts, compared to those with low scores. -It allows for comparisons with the surface scales, since it has been found to correlate unexpectedly in a positive direction with some of them (see sections 4.1 and 4.2).

Table 1. Reasons for selection

Non-Selected Predictors	**Reason for non-selection**
Metrics related to use of search option mentioned in 3.8.1	Their variation is close to 0, so they cannot be included as predictors.

Table 2. Reasons for not selecting predictors

Appendix 4.8.2 – Further information on development of model

Model 1 –Seeking Meaning - All predictors based on initial selection

Model Summary[b]				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.367[a]	.135	.070	2.675

a. Predictors: (Constant), Number of Exercises Finished But Not Solved, Number of Exercises Solved on First Try, Number of concept links visited in reading and exercise pages, Maximum View Time on Exercise Page, Number of pages visited from TOC, Stratum, Number of Exercises Solved on Second Try, Number of Exercises Solved on Third Try

b. Dependent Variable: Deep subscale Seeking Meaning

Table 1 – Model Summary

ANOVA[a]						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	118.438	8	14.805	2.069	.045[b]
	Residual	758.605	106	7.157		
	Total	877.043	114			

a. Dependent Variable: Deep subscale Seeking Meaning

b. Predictors: (Constant), Number of Exercises Finished But Not Solved, Number of Exercises Solved on First Try, Number of concept links visited in reading and exercise pages, Maximum View Time on Exercise Page, Number of pages visited from TOC, Stratum, Number of Exercises Solved on Second Try, Number of Exercises Solved on Third Try

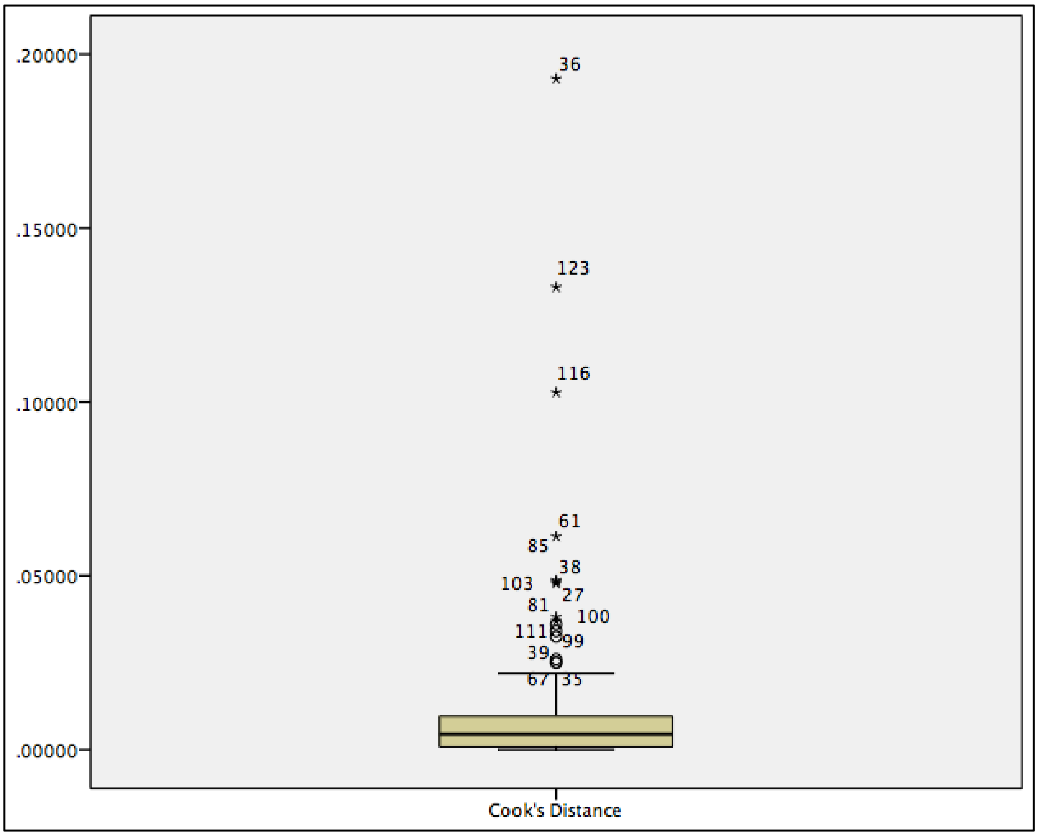
Table 2 – Overall significance of model

Coefficients[a]						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	14.903	1.401		10.637	.000
	Maximum View Time on Exercise Page	.001	.000	.227	2.402	.018
	Number of pages visited from TOC	-.023	.015	-.151	-1.487	.140
	Number of concept links visited in reading and exercise pages	-.202	.115	-.172	-1.755	.082
	Stratum	-.701	2.487	-.030	-.282	.779
	Number of Exercises Solved on First Try	.015	.011	.151	1.427	.157
	Number of Exercises Solved on Second Try	-.150	.075	-.253	-1.995	.049
	Number of Exercises Solved on Third Try	.162	.143	.154	1.139	.257
	Number of Exercises Finished But Not Solved	.056	.083	.091	.676	.501

a. Dependent Variable: Deep subscale Seeking Meaning

Table 3 – B, Beta, and Sig. values for predictors

Figure 1. Cook's Distance Box Plot



	R^2	Adj. R^2	Sig.
Model 1 (all initially selected predictors)	13.5%	7%	0.045
Model 2a (exclusion of case 36)	14.3%	7.8%	0.034
Model 2b (exclusion of cases 36 and 123)	14.5%	7.9%	0.033
Model 2c (exclusion of cases 36, 123, and 116) [Rejected]	12.7%	5.9%	0.073
Model 2d (exclusion of cases 36, 123, and 61)	15.6%	9.1%	0.021
Model 2e (exclusion of cases 36, 123, 61, and 85)	18.2%	11.8%	0.007
Model 2f (exclusion of cases 36, 123, 61, 85, and 38) [Rejected]	18.2%	11.7%	0.008
Model 2g (exclusion of cases 36, 123, 61, 85, and 103)	20.8%	14.5%	0.002
Model 2h (exclusion of cases 36, 123, 61, 85, 103, and 27) [Rejected]	19.7%	13.3%	0.004
Model 2i (exclusion of cases 36, 123, 61, 85, 103, and 81) [Rejected]	19.6%	13.2%	0.004

Table 4. Summary of measures of variance and significance for accepted and rejected models for exclusion of outliers

Model 4 – Seeking Meaning – Recommended version

Model Summary[b]				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.449[a]	.201	.155	2.371
a. Predictors: (Constant), Number of Exercises Solved on Third Try, Maximum View Time on Exercise Page, Number of Exercises Solved on First Try, Number of pages visited from TOC, Number of concept links visited in reading and exercise pages, Number of Exercises Solved on Second Try				
b. Dependent Variable: Deep subscale Seeking Meaning				

Table 5 – Model Summary

ANOVA[a]						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	145.882	6	24.314	4.326	.001[b]
	Residual	578.882	103	5.620		
	Total	724.764	109			
a. Dependent Variable: Deep subscale Seeking Meaning						
b. Predictors: (Constant), Number of Exercises Solved on Third Try, Maximum View Time on Exercise Page, Number of Exercises Solved on First Try, Number of pages visited from TOC, Number of concept links visited in reading and exercise pages, Number of Exercises Solved on Second Try						

Table 6 – Overall significance of model

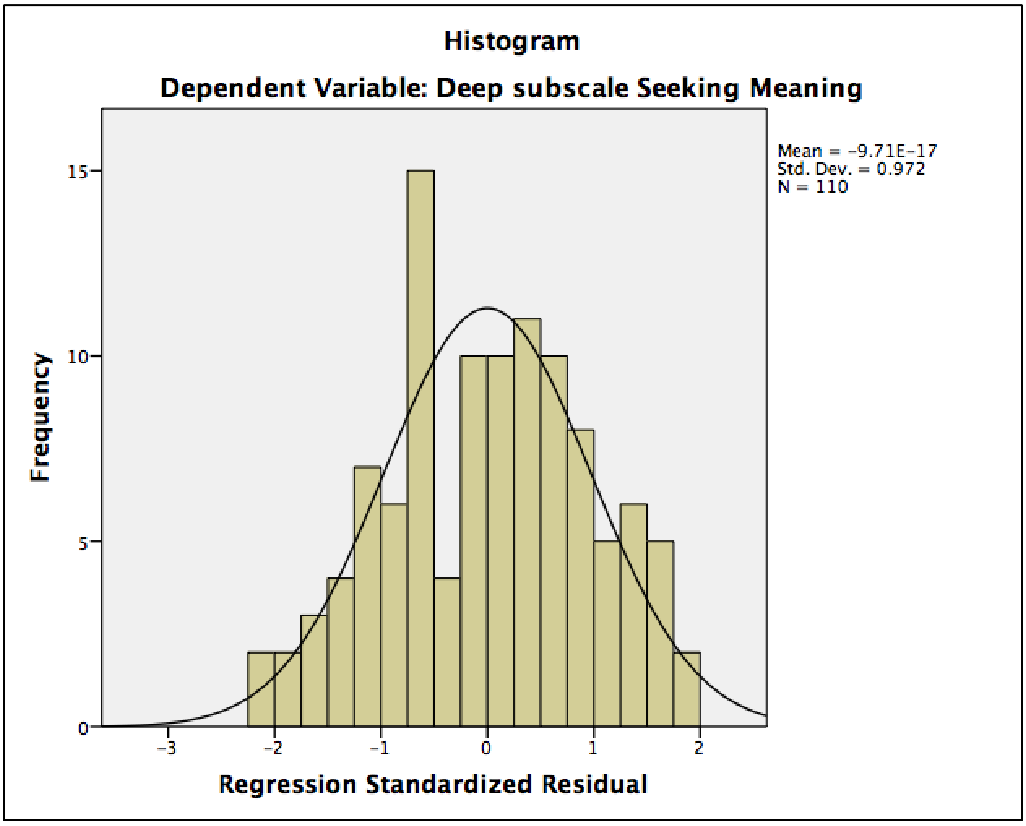
Coefficients[a]						
Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	14.198	.789		17.999	.000
	Maximum View Time on Exercise Page	.001	.000	.297	3.281	.001
	Number of pages visited from TOC	-.028	.013	-.201	-2.156	.033
	Number of concept links visited in reading and exercise pages	-.197	.105	-.177	-1.871	.064
	Number of Exercises Solved on First Try	.021	.009	.227	2.277	.025
	Number of Exercises Solved on Second Try	-.111	.066	-.189	-1.678	.096
	Number of Exercises Solved on Third Try	.226	.126	.197	1.795	.076
a. Dependent Variable: Deep subscale Seeking Meaning						

Table 7 – B, Beta, and Sig. values for predictors

Appendix 4.8.3 – Seeking Meaning – Model 4 – Generalisation - Assumptions

The assumptions regarding: normal distribution of residuals, homoscedasticity of standardised residuals against predicted ones, and the normality of residuals hold well.

Figure 2. Histogram of standardised residuals for the final model



Regarding the assumption about the normal distribution of residuals, as shown in Figure 2, that residuals fit quite closely to a normal distribution.

Figure 3. Plot of standardised residuals

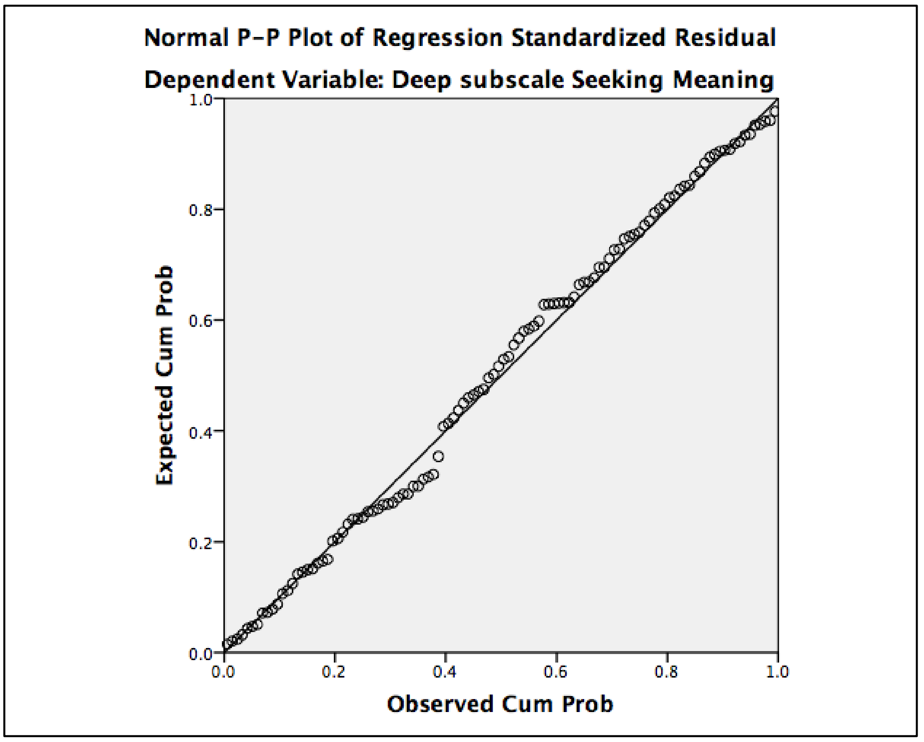
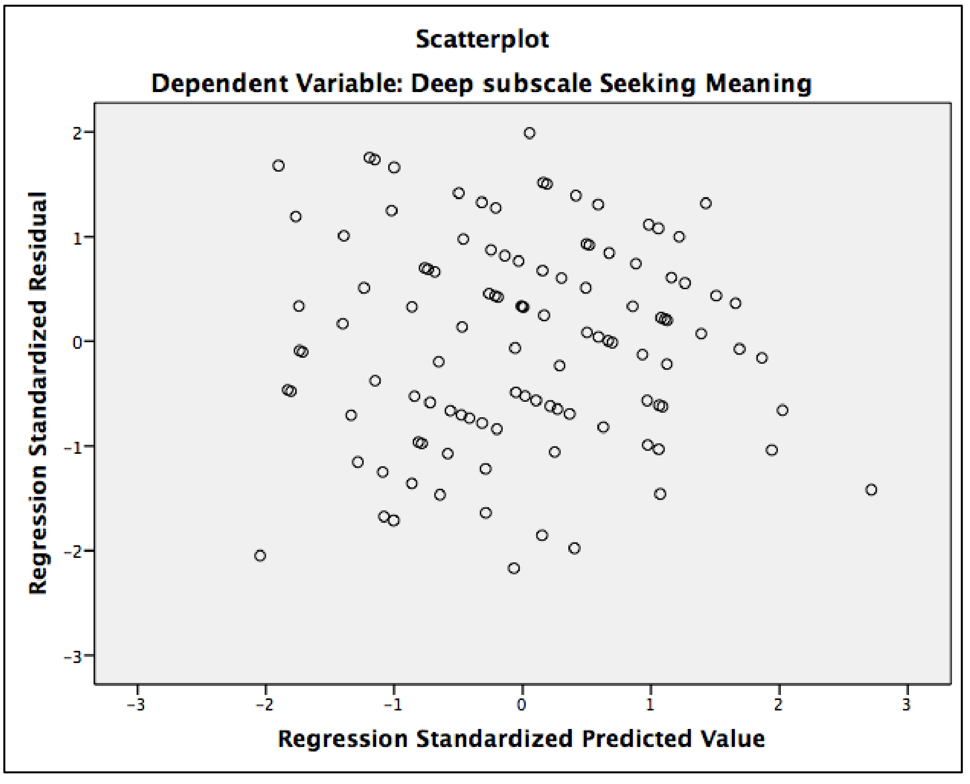


Figure 3 shows that the normality assumption holds since all points are quite close to the straight line.

Figure 4. Plot of the standardised residuals against the predicted ones for the model



The assumption about the residuals is whether the variance of the residuals is constant, in other words there is homoscedasticity. Figure 4 shows that the scatter plot is reasonably random and that most residuals are homoscedastic.

Appendix 4.8.4 – Seeking Meaning subscale –Model 7 – Leanest and Meanest

Exclusion of outliers 36, 123, 61, 85, and 103 and *Stratum and Number of Exercises Finished but not Solved and Number of Exercises on Second Try, Number of Exercises on Third Try, and Number of concept links visited on reading and exercise pages*.

Model Summary[b]				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.392[a]	.154	.130	2.405
a. Predictors: (Constant), Number of Exercises Solved on First Try, Number of pages visited from TOC, Maximum View Time on Exercise Page				
b. Dependent Variable: Deep subscale Seeking Meaning				

Table 8 – Model Summary

ANOVA[a]						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	111.652	3	37.217	6.434	.000[b]
	Residual	613.111	106	5.784		
	Total	724.764	109			
a. Dependent Variable: Deep subscale Seeking Meaning						
b. Predictors: (Constant), Number of Exercises Solved on First Try, Number of pages visited from TOC, Maximum View Time on Exercise Page						

Table 9 – Overall significance of model

Coefficients[a]						
		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Std. Error	Beta	t	Sig.
1	(Constant)	13.728	.734		18.702	.000
	Maximum View Time on Exercise Page	.001	.000	.312	3.408	.001
	Number of pages visited from TOC	-.031	.012	-.224	-2.506	.014
	Number of Exercises Solved on First Try	.016	.009	.174	1.893	.051
a. Dependent Variable: Deep subscale Seeking Meaning						

Table 10 – B, Beta, and Sig. values for predictors

Appendices 4.9 – Relating Ideas Models

Appendix 4.9.1 –Further Justification for inclusion of predictors

Selected Predictor	Reason for selection
Number of exercises solved on first try	-Theoretical connections discussed in section 4.9.1 -Enriching further the discussion by allowing useful comparisons to the surface scales, as shown in sections 4.1- 4.5 - Their combination gives a more complete picture as to how students deal with their exercises during their tutorial sessions according to the specific approach to studying.
Number of exercises solved on second try	
Number of exercises solved on third try	
Number of exercises finished but not solved	
Number of hyperlinks (concepts links) visited on exercise and reading pages	-It is enriching to know whether those with high scores on the subscale tend to use more such a common ILE feature (due to their intention to relate mathematical concepts in an ILE) compared to those with low scores.
Maximum view time on an exercise page	-When tried in Model 1 these metrics contribute to the 'relating ideas' model with both a higher R^2 and a higher Adjusted R^2 compared to pre-models 1a, 1b, 1c and 1d, (see table 3 below). Model 1 is also the only one which is overall statistically significant.
Maximum view time on a content (reading) page	
Average number of 'notes' link is clicked per page	Statistical (see 4.9.2)

Table 1. Reasons for selection

Non-Selected Predictors	Reason for non-selection
Average view time on content (reading) pages	-No strong indications, based on the theory, as to which of these predictors would be the most 'enriching' for the model. -When tried in pre-models 1a, 1b and 1c neither of these metrics contribute to the 'relating ideas' model with both a higher R^2 and a higher Adjusted R^2 compared to Model 1 (see table 3 below). -Note also that there is a multicollinearity issue between *maximum view time on a content (reading) page* and *average view time on content (reading) pages* (so, they cannot be tried in the same model).
Average view time on exercise pages	
Metrics related to use of search option mentioned in 3.9.1	Their variation is close to 0, so they cannot be included as predictors.

Table 2. Reasons for not selecting predictors

Pre-models	**R^2**	**Adj. R^2**	**Sig.**
Model 1d *(Avg number of times a 'notes' link is clicked per page, Number of Exercises Solved on First Try, Number of Exercises Solved on Third Try, Number of hyperlinks (concept links) visited in reading and exercise pages, Number of Exercises Solved on Second Try, Number of Exercises Finished But Not Solved)*	9.9%	4.9%	0.075
Model 1a *(**maximum view time on an exercise page, average view time on exercise pages,** avg number of times a 'notes' link is clicked per page, Number of Exercises Solved on First Try, Number of Exercises Solved on Third Try, Number of hyperlinks (concept links) visited in reading and exercise pages, Number of Exercises Solved on Second Try, Number of Exercises Finished But Not Solved)*	11.9 %	5.3%	0.086
Model 1b *(**Maximum view time on an reading page**, **average view time on exercises pages**, Avg number of times a 'notes' link is clicked per page, Number of Exercises Solved on First Try, Number of Exercises Solved on Third Try, Number of hyperlinks (concept links) visited in reading and exercise pages, Number of Exercises Solved on Second Try, Number of Exercises Finished But Not Solved)*	12.3%%	5.6%	0.076

Model 1c (***average view time on reading pages, maximum view time on an exercise page***, *Avg number of times a 'notes' link is clicked per page, Number of Exercises Solved on First Try, Number of Exercises Solved on Third Try, Number of hyperlinks (concept links) visited in reading and exercise pages, Number of Exercises Solved on Second Try, Number of Exercises Finished But Not Solved)*	12.3%	5.6%	0.076
Model 1 (***Maximum View Time on Content Page, Maximum View Time on Exercise Page***, *Avg number of times a 'notes' link is clicked per page, Number of Exercises Solved on First Try, Number of Exercises Solved on Third Try, Number of hyperlinks (concept links) visited in reading and exercise pages, Number of Exercises Solved on Second Try, Number of Exercises Finished But Not Solved)*	13.4%	6.8%	0.048

Table 3. Pre-models

Appendix 4.9.2 – Further information on development of model

Model 1 – Relating Ideas- All predictors based on initial selection

Model Summary[b]				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.365[a]	.134	.068	2.738

a. Predictors: (Constant), Maximum View Time on Content Page, Avg number an notes link is clicked per page, Number of Exercises Solved on First Try, Number of Exercises Solved on Third Try, Maximum View Time on Exercise Page, Number of concept links visited in reading and exercise pages, Number of Exercises Solved on Second Try, Number of Exercises Finished But Not Solved

b. Dependent Variable: Deep subscale Relating Ideas

Table 1 – Model Summary

ANOVA[a]						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	122.446	8	15.306	2.042	.048[b]
	Residual	794.597	106	7.496		
	Total	917.043	114			

a. Dependent Variable: Deep subscale Relating Ideas

b. Predictors: (Constant), Maximum View Time on Content Page, Avg number an notes link is clicked per page, Number of Exercises Solved on First Try, Number of Exercises Solved on Third Try, Maximum View Time on Exercise Page, Number of concept links visited in reading and exercise pages, Number of Exercises Solved on Second Try, Number of Exercises Finished But Not Solved

Table 2 – Overall significance of model

Coefficients[a]						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	13.785	.963		14.308	.000
	Number of Exercises Solved on First Try	.015	.011	.143	1.383	.170
	Number of Exercises Solved on Second Try	-.224	.078	-.368	-2.878	.005
	Number of Exercises Solved on Third Try	.197	.146	.183	1.353	.179
	Number of Exercises Finished But Not Solved	.117	.082	.187	1.425	.157
	Avg number an notes link is clicked per page	1.873	1.777	.101	1.054	.294
	Number of concept links visited in reading and exercise pages	-.141	.120	-.118	-1.177	.242
	Maximum View Time on Exercise Page	.000	.000	.125	1.305	.195
	Maximum View Time on Content Page	.001	.001	.137	1.411	.161

a. Dependent Variable: Deep subscale Relating Ideas

Table 3 – B, Beta, and Sig. values for predictors

Figure 1. Cook's Distance Box Plot

	R^2	Adj. R^2	Sig.
Model 1 (all initially selected predictors)	13.4%	6.8%	0.048
Model 2a (exclusion of case 112)	15.9%	9.5%	0.017
Model 2b (exclusion of case 112 and 123) [Rejected]	14.2%	7.6%	0.037
Model 2c (exclusion of cases 112 and 85)	18.4%	12.2%	0.005
Model 2d (exclusion of cases 112, 85 and 42)	19.9%	13.7%	0.003
Model 2e exclusion of cases 112, 85, 42, and 98)	20.9%	14.7%	0.002
Model 2f (exclusion of cases 112, 85, 42, 98, and 105)	23.4%	17.3%	0.001
Model 2g (exclusion of cases 112, 85, 42, 98, 105, and 84)	25.8%	19.9%	0.000

Table 4. Summary of measures of variance and significance for accepted and rejected models for exclusion of outliers

Model 3 – Relating Ideas – Recommended version

Model Summary[b]				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.502[a]	.252	.200	2.363

a. Predictors: (Constant), Maximum View Time on Content Page, Number of Exercises Solved on First Try, Avg number an notes link is clicked per page, Number of Exercises Finished But Not Solved, Maximum View Time on Exercise Page, Number of concept links visited in reading and exercise pages, Number of Exercises Solved on Second Try

b. Dependent Variable: Deep subscale Relating Ideas

Table 5 – Model Summary

ANOVA[a]						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	190.022	7	27.146	4.862	.000[b]
	Residual	563.941	101	5.584		
	Total	753.963	108			

a. Dependent Variable: Deep subscale Relating Ideas

b. Predictors: (Constant), Maximum View Time on Content Page, Number of Exercises Solved on First Try, Avg number an notes link is clicked per page, Number of Exercises Finished But Not Solved, Maximum View Time on Exercise Page, Number of concept links visited in reading and exercise pages, Number of Exercises Solved on Second Try

Table 6 – Overall significance of model

Coefficients[a]						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	13.132	.845		15.542	.000
	Number of Exercises Solved on First Try	.019	.009	.200	2.031	.045
	Number of Exercises Solved on Second Try	-.202	.065	-.360	-3.126	.002
	Number of Exercises Finished But Not Solved	.203	.062	.354	3.291	.001
	Avg number an notes link is clicked per page	5.789	2.024	.269	2.861	.005
	Number of concept links visited in reading and exercise pages	-.138	.103	-.125	-1.337	.184
	Maximum View Time on Exercise Page	.000415	.000	.113	1.237	.219
	Maximum View Time on Content Page	.001	.000	.242	2.631	.010

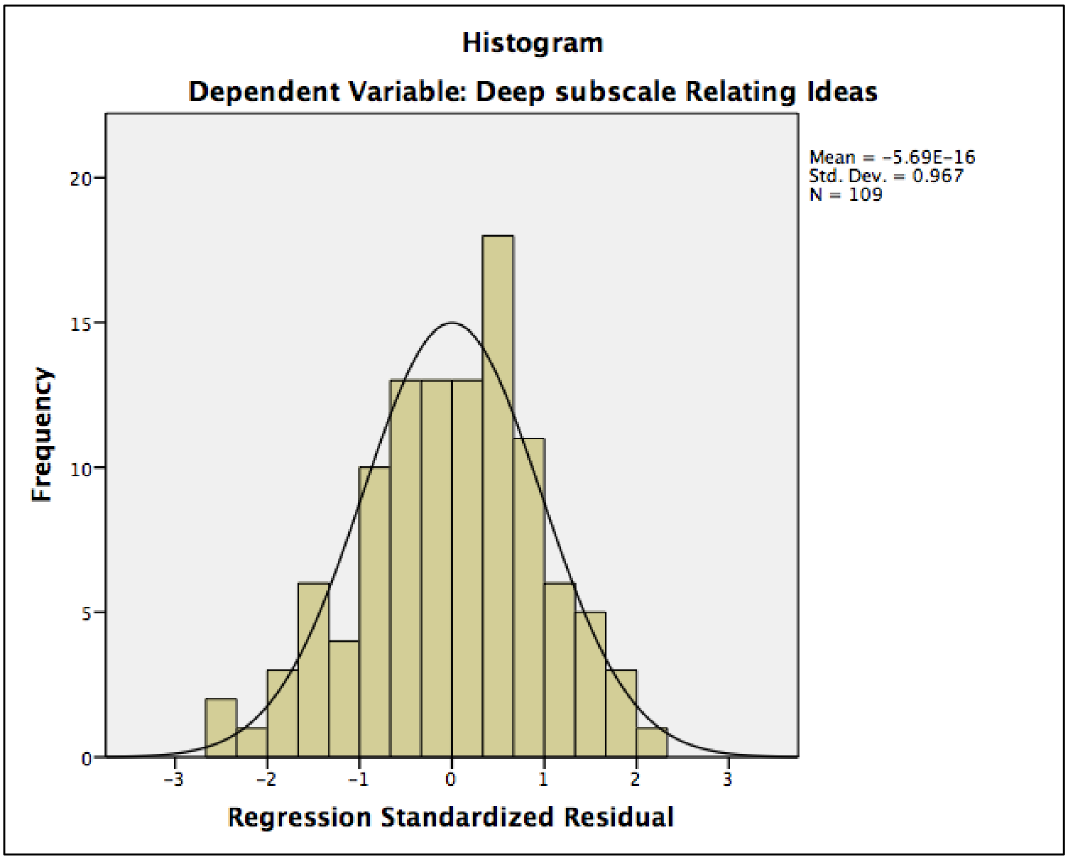
a. Dependent Variable: Deep subscale Relating Ideas

Table 7 – B, Beta, and Sig. values for predictors

Appendix 4.9.3 – Relating Ideas – Model 3 – Generalisation - Assumptions

The assumptions regarding: normal distribution of residuals, homoscedasticity of standardised residuals against predicted ones, and the normality of residuals hold well.

Figure 2. Histogram of standardised residuals for the final model



Regarding the assumption about the normal distribution of residuals, as shown in Figure 2, that residuals fit quite closely to a normal distribution.

Figure 3. Plot of standardised residuals

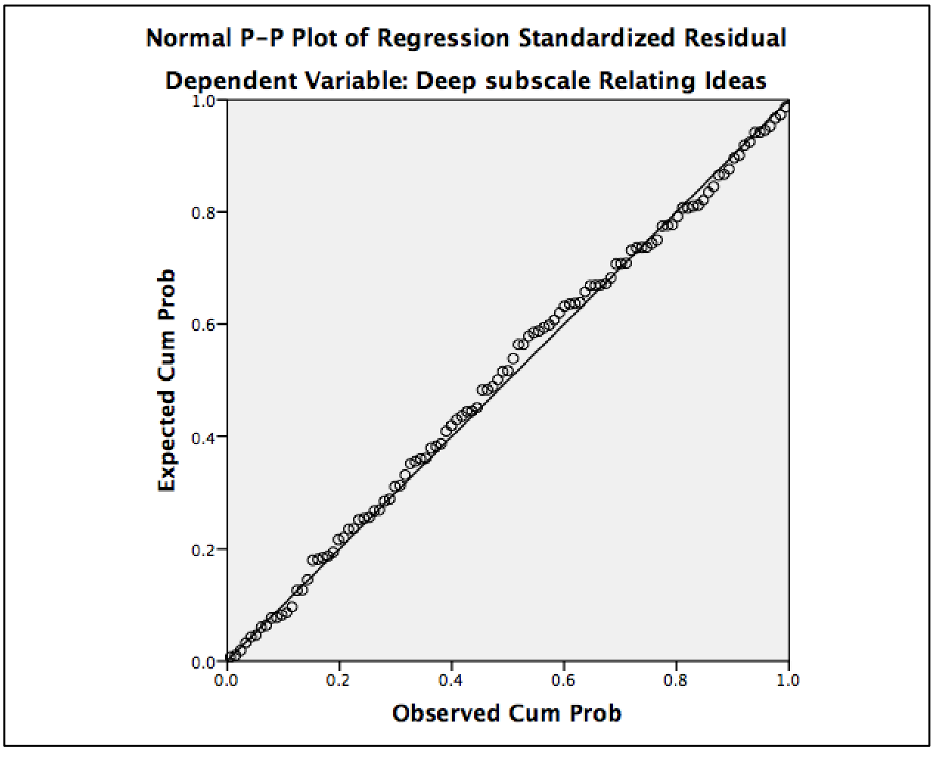
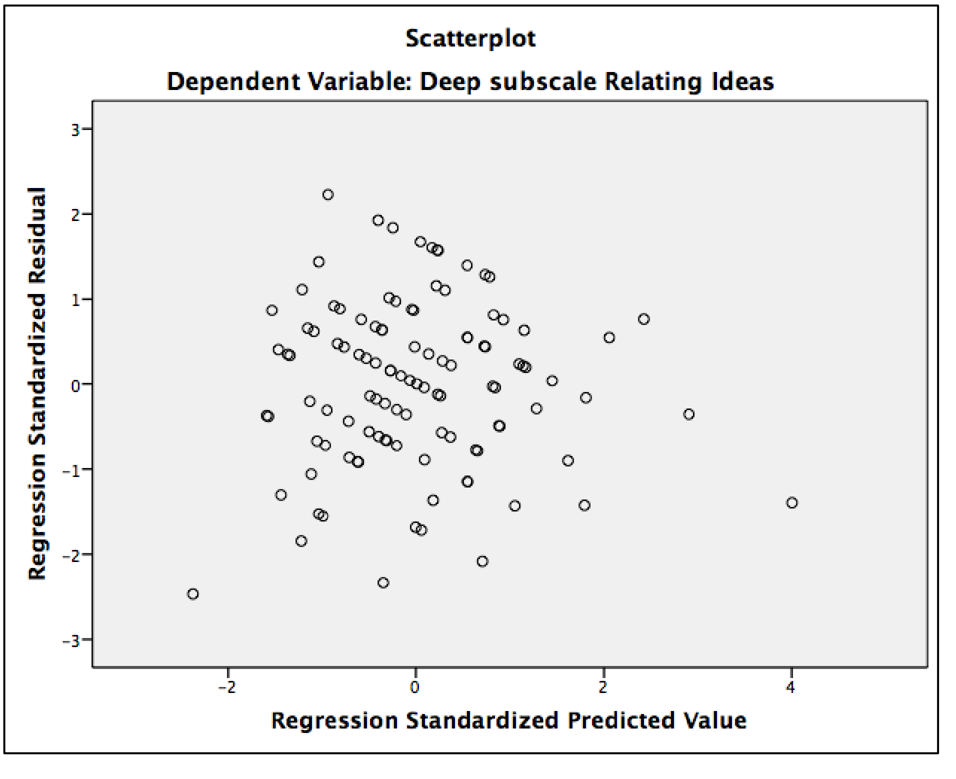


Figure 3 shows that the normality assumption holds since all points are quite close to the straight line.

Figure 4. Plot of the standardised residuals against the predicted ones for the model



The assumption about the residuals is whether the variance of the residuals is constant, in other words there is homoscedasticity. Figure 4 shows that the scatter plot is reasonably random and that most residuals are homoscedastic.

Appendix 4.9.4 – Relating Ideas subscale –Model 5 – Leanest and Meanest

Exclusion of outliers 112, 85, 42, 98, 105, and 84 and *number of exercises solved on third try, and Max View Time view time on exercise page, and Number of concept links visited in reading and exercise pages*

Model Summary[b]				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.474[a]	.225	.187	2.382

a. Predictors: (Constant), Maximum View Time on Content Page, Number of Exercises Solved on First Try, Avg number an notes link is clicked per page, Number of Exercises Finished But Not Solved, Number of Exercises Solved on Second Try

b. Dependent Variable: Deep subscale Relating Ideas

Table 8 – Model Summary

ANOVA[a]						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	169.726	5	33.945	5.984	.000[b]
	Residual	584.237	103	5.672		
	Total	753.963	108			

a. Dependent Variable: Deep subscale Relating Ideas

b. Predictors: (Constant), Maximum View Time on Content Page, Number of Exercises Solved on First Try, Avg number an notes link is clicked per page, Number of Exercises Finished But Not Solved, Number of Exercises Solved on Second Try

Table 9 – Overall significance of model

Coefficients[a]						
		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Std. Error	Beta	t	Sig.
1	(Constant)	13.564	.686		19.767	.000
	Number of Exercises Solved on First Try	.019	.009	.194	1.988	.049
	Number of Exercises Solved on Second Try	-.205	.065	-.366	-3.151	.002
	Number of Exercises Finished But Not Solved	.189	.062	.328	3.054	.003
	Avg number an notes link is clicked per page	5.427	1.918	.252	2.830	.006
	Maximum View Time on Content Page	.001	.000	.228	2.537	.013

a. Dependent Variable: Deep subscale Relating Ideas

Table 10 – B, Beta, and Sig. values for predictors

Appendices 4.10 – Use of Evidence Models

Appendix 4.10.1 –Further Justification for inclusion of predictors

Selected Predictor	Reason for selection
number of exercises solved on first try	-Theoretical connections discussed in section 4.10.1 -Enriching further the discussion by allowing useful comparisons to the surface scales, as shown in sections 4.1- 4.5 - Their combination gives a more complete picture as to how students deal with their exercises during their tutorial sessions according to the specific approach to studying.
number of exercises solved on second try	
number of exercises solved on third try	
number of exercises finished but not solved	
stratum	It is enriching to know whether those with high scores on the subscale tend to master each topic according to the given structure and go through the material in a more linear way compared to those with low scores.
maximum view time on a content (reading) page	-When tried in Model 1 along with the rest metrics contribute to the 'use of evidence' model with both a higher R^2 and a higher Adjusted R^2 compared to pre-models 1a, 1b, 1c and 1d, and 1e (see table 3 below).
maximum view time on an exercise page	Statistical (see 4.10.2)

Table 1. Reasons for selection

Non-Selected Predictors	Reason for non-selection
average view time on content (reading) pages	-No strong indications, based on the theory, as to which of these predictors would be the most 'enriching' for the model. -When tried in pre-models 1b, 1c, 1d, and 1e neither of these metrics contribute to the "use of evidence" model with both a higher R^2 and a higher Adjusted R^2 compared to Model 1 (see table 3 below).
average view time on exercise pages	

Table 2. Reasons for not selecting predictors

Pre-models	R^2	**Adj.** R^2	**Sig.**
Model 1a *(number of exercises solved on first try, number of exercises solved on third try, number of exercises solved on second try, number of exercises finished but not solved, stratum,* and *maximum view time on an exercise page)*	14.4%	9.6%	0.009
Model 1b *(**average view time on exercise pages,** number of exercises solved on first try, number of exercises solved on third try, number of exercises solved on second try, number of exercises finished but not solved, stratum,* and *maximum view time on an exercise page)*	14.4 %	8.8%	0.017
Model 1c *(**average view time on content (reading) pages**, number of exercises solved on first try, number of exercises solved on third try, number of exercises solved on second try, number of exercises finished but not solved, stratum,* and *maximum view time on an exercise page)*	14.5%%	8.9%	0.016
Model 1d *(**average view time on content (reading) pages**, **average view time on exercise pages,** number of exercises solved on first try, number of exercises solved on third try, number of exercises solved on second try, number of exercises finished but not solved, stratum,* and *maximum view time on an exercise page)*	14.5%	8.1%	0.029

Model 1e (***average view time on exercise pages, maximum view time on a content (reading) page,*** *number of exercises solved on first try, number of exercises solved on third try, number of exercises solved on second try, number of exercises finished but not solved, stratum,* and *maximum view time on an exercise page)*	15.3%	8.9%	0.021
Model 1 (***maximum view time on exercise page, maximum view time on a content (reading) page,*** *number of exercises solved on first try, number of exercises solved on third try, number of exercises solved on second try, number of exercises finished but not solved, stratum,* and *maximum view time on an exercise page*	15.2%	9.7%	0.012

Table 3. Pre-models

Appendix 4.10.2 – Further information on development of model

Model 1 –Use of Evidence - All predictors based on initial selection

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.390[a]	.152	.097	2.357

a. Predictors: (Constant), Maximum View Time on Content Page, Number of Exercises Solved on First Try, Number of Exercises Solved on Third Try, Maximum View Time on Exercise Page, Stratum, Number of Exercises Solved on Second Try, Number of Exercises Finished But Not Solved

Table 1 – Model Summary

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	106.892	7	15.270	2.749	.012[b]
	Residual	594.273	107	5.554		
	Total	701.165	114			

a. Dependent Variable: Deep subscale Use of Evidence

b. Predictors: (Constant), Maximum View Time on Content Page, Number of Exercises Solved on First Try, Number of Exercises Solved on Third Try, Maximum View Time on Exercise Page, Stratum, Number of Exercises Solved on Second Try, Number of Exercises Finished But Not Solved

Table 2 – Overall significance of model

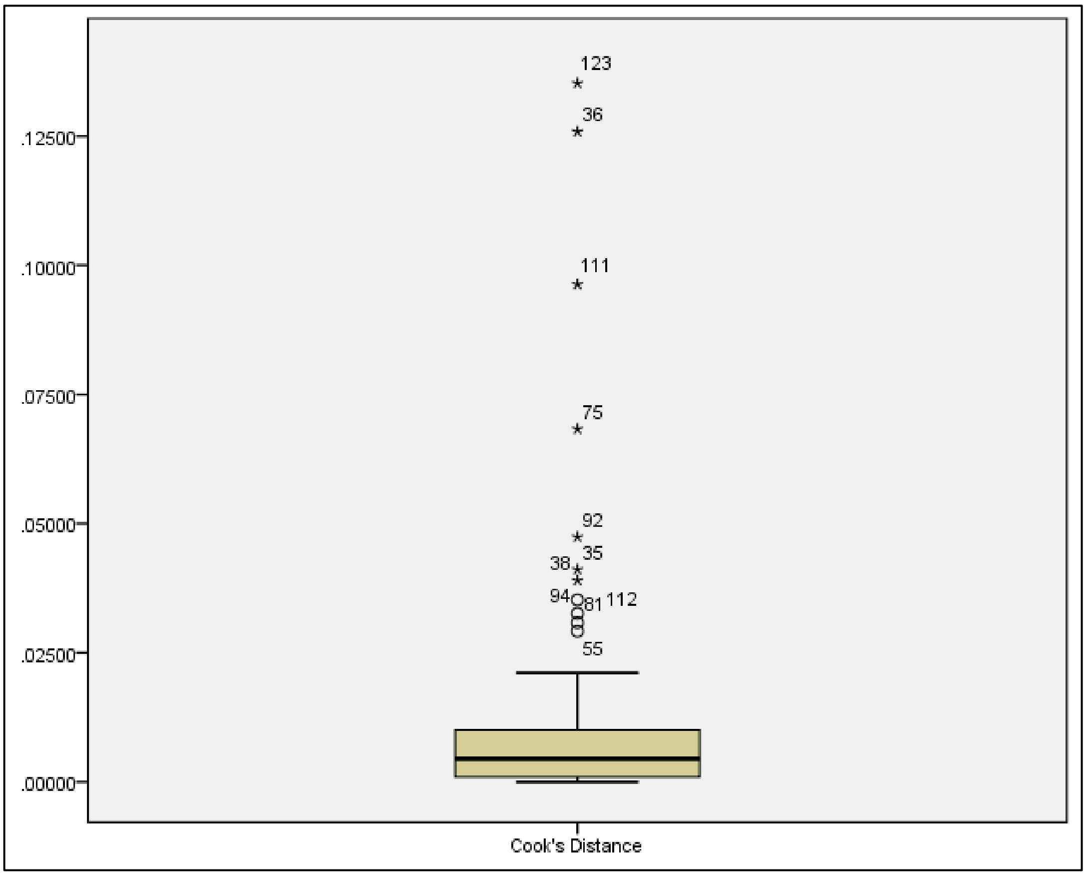
Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	15.927	1.120		14.226	.000
	Maximum View Time on Exercise Page	.001	.000	.186	1.987	.050
	Number of Exercises Solved on First Try	.005	.009	.056	.539	.591
	Number of Exercises Solved on Second Try	-.199	.067	-.375	-2.979	.004
	Number of Exercises Solved on Third Try	.120	.123	.127	.972	.333
	Number of Exercises Finished But Not Solved	.078	.071	.142	1.103	.273
	Stratum	-2.697	2.048	-.128	-1.317	.191
	Maximum View Time on Content Page	.000	.000	.097	1.042	.300

a. Dependent Variable: Deep subscale Use of Evidence

Table 3 – B, Beta, and Sig. values for predictors

Figure 1. Cook's Distance Box Plot



	R^2	Adj. R^2	Sig.
Model 1 (all initially selected predictors)	15.2%	9.7%	0.012
Model 2a (exclusion of case 123) **[Rejected]**	13.8%	8.1%	0.024
Model 2b (exclusion of case 36)	15.4%	9.8%	0.012
Model 2c (exclusion of cases 36, and 111)	17.5%	12%	0.004
Model 2d (exclusion of cases 36, 111, and 75) **[Rejected]**	17.4%	11.8%	0.005
Model 2e (exclusion of cases 36, 111, and 92) **[Rejected]**	16.4%	10.8%	0.008
Model 2f (exclusion of cases 36, 111, and 35)	18.1%	12.6%	0.003
Model 2g (exclusion of cases 36, 111, 35, and 38) **[Rejected]**	17.3%	11.6%	0.006
Model 2h (exclusion of cases 36, 111, 35 and 94)	18.3%	13.3%	0.003

Table 4. Summary of measures of variance and significance for accepted and rejected models for exclusion of outliers

Model 5 – Use of Evidence – Recommended version

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.432[a]	.187	.156	2.285

a. Predictors: (Constant), Stratum, Maximum View Time on Exercise Page, Number of Exercises Solved on Third Try, Number of Exercises Solved on Second Try

b. Dependent Variable: Deep subscale Use of Evidence

Table 5 – Model Summary

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	127.029	4	31.757	6.083	.000[b]
	Residual	553.403	106	5.221		
	Total	680.432	110			

a. Dependent Variable: Deep subscale Use of Evidence

b. Predictors: (Constant), Stratum, Maximum View Time on Exercise Page, Number of Exercises Solved on Third Try, Number of Exercises Solved on Second Try

Table 6 – Overall significance of model

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	17.314	.958		18.081	.000
	Maximum View Time on Exercise Page	.001	.000	.205	2.312	.023
	Number of Exercises Solved on Second Try	-.248	.065	-.405	-3.815	.000
	Number of Exercises Solved on Third Try	.156	.108	.148	1.446	.151
	Stratum	-4.532	2.051	-.207	-2.210	.029

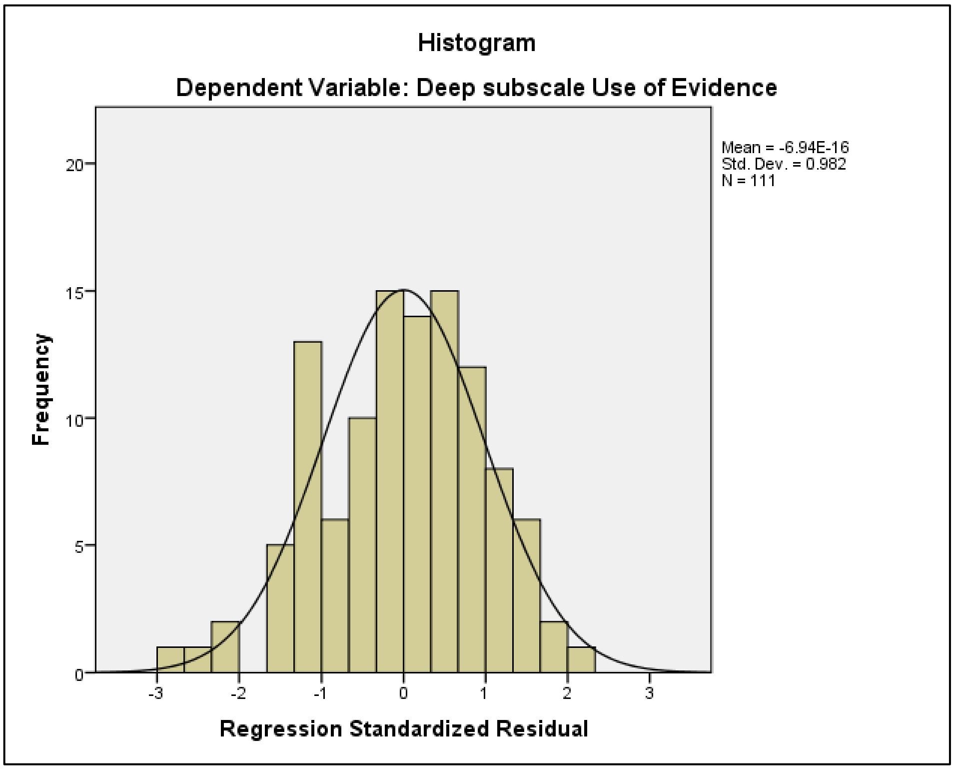
a. Dependent Variable: Deep subscale Use of Evidence

Table 7 – B, Beta, and Sig. values for predictors

Appendix 4.10.3 – Use of Evidence – Model 5 – Generalisation - Assumptions

The assumptions regarding: normal distribution of residuals, homoscedasticity of standardised residuals against predicted ones, and the normality of residuals hold well.

Figure 2. Histogram of standardised residuals for the final model



Regarding the assumption about the normal distribution of residuals, as shown in Figure 2, that residuals fit quite closely to a normal distribution.

Figure 3. Plot of standardised residuals

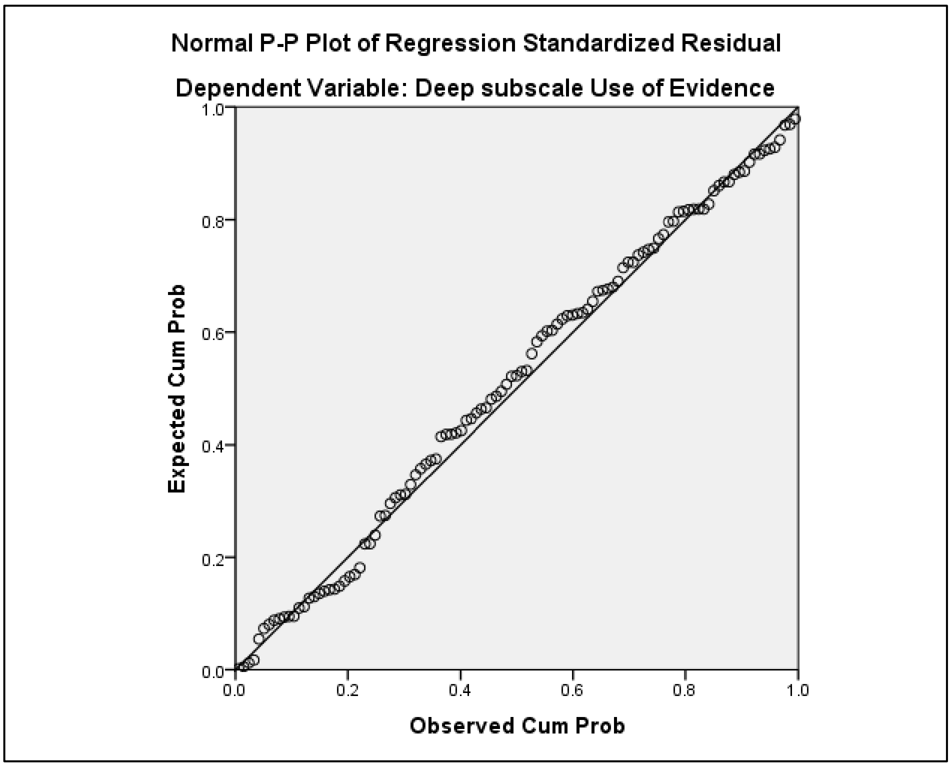
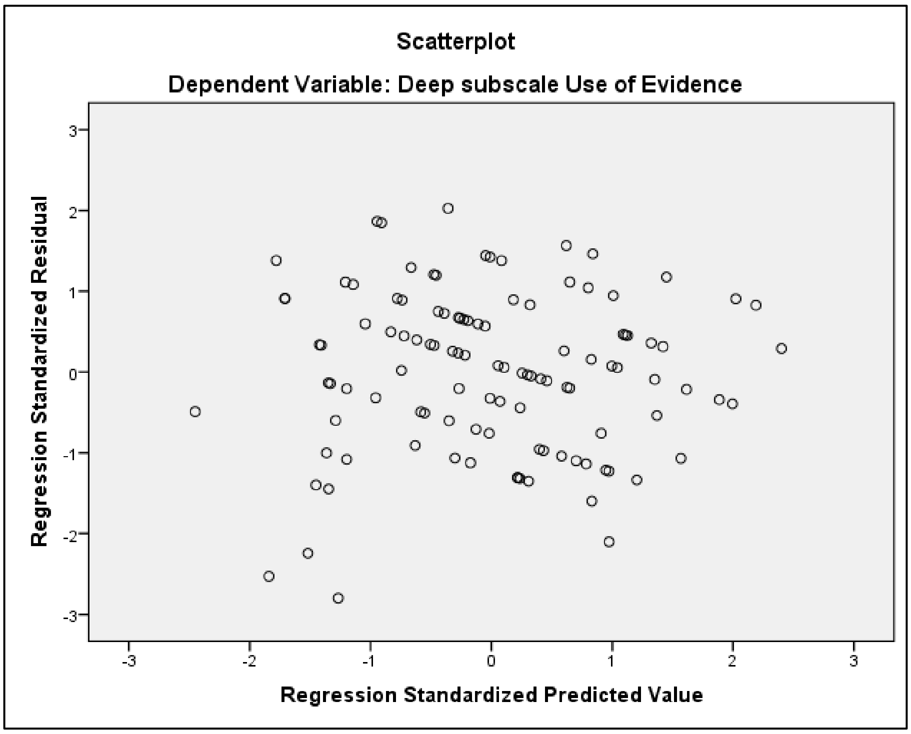


Figure 3 shows that the normality assumption holds since all points are quite close to the straight line.

Figure 4. Plot of the standardised residuals against the predicted ones for the model



The assumption about the residuals is whether the variance of the residuals is constant, in other words there is homoscedasticity. Figure 4 shows that the scatter plot is reasonably random and the residuals are homoscedastic with a few exceptions.

Appendix 4.10.4 – Use of Evidence subscale –Model 6 – Leanest and Meanest

Exclusion 36, 111, 35 and 94, and *number of exercises finished but not solved, number of exercises solved on first try, maximum view time on content page and number of exercises solved on third try*

Model Summary[b]				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.413[a]	.171	.147	2.297
a. Predictors: (Constant), Stratum, Maximum View Time on Exercise Page, Number of Exercises Solved on Second Try				
b. Dependent Variable: Deep subscale Use of Evidence				

Table 8 – Model Summary

ANOVA[a]						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	116.115	3	38.705	7.339	.000[b]
	Residual	564.318	107	5.274		
	Total	680.432	110			
a. Dependent Variable: Deep subscale Use of Evidence						
b. Predictors: (Constant), Stratum, Maximum View Time on Exercise Page, Number of Exercises Solved on Second Try						

Table 9 – Overall significance of model

Coefficients[a]						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	17.447	.958		18.212	.000
	Maximum View Time on Exercise Page	.001	.000	.213	2.397	.018
	Number of Exercises Solved on Second Try	-.204	.058	-.333	-3.533	.001
	Stratum	-4.852	2.049	-.222	-2.367	.020
a. Dependent Variable: Deep subscale Use of Evidence						

Table 10 – B, Beta, and Sig. values for predictors

Appendices – Chapter 5 – General Discussion

Appendix 5.1 – Summary – Effect size R – Variance explained (or accounted for) R^2 and Adjusted R^2 – Significance Sig.

Whole Sample/Recommended Model	**R**	**R^2 / Adj.R^2**	**Effect Size $f^2=1-R^2/R^2$**	**Significance**	**Sample Size / Number of Predictors**
Deep (Model 4 - Suggested)	0.428	18.3% / 14.4%	0.223	Sig. 0.001	109/5 predictors
Interest in Ideas (Model 4 -Suggested)	0.396	15.7% / 11.6%	0.186	Sig. 0.003	110/5 predictors
Relating Ideas (Model 3 - Suggested)	0.502	25.2% / 20%	0.336	Sig. 0.000	109/7 predictors
Use of Evidence (Model 5 -Suggested)	0.432	18.7% / 15.6%	0.230	Sig. 0.000	111/ 4 predictors
Seeking Meaning (Model 4 - Suggested)	0.367	20.1% / 15.5%	0.251	Sig. 0.001	110/6 predictors

Surface (Model 3 -Suggested)	0.675	45.5% / 41.8%	0.834	Sig. 0.000	110/7 predictors
Fear of Failure (Model 4 – Suggested and Leanest and Meanest)	0.645	41.6% / 39.4%	0.717	Sig. 0.000	111/4 predictors
Unrelated Memorising (Model 6 -Suggested)	0.633	40% / 37.7%	0.666	Sig. 0.000	111/4 predictors
Lack of Purpose (Model 4 Suggested)	0.458	21% / 18%	0.265	Sig.0.000	111/4 predictors
Syllabus Boundness (Model 4 - Suggested)	0.465	21.6% / 17.1%	0.275	Sig. 0.000	113/6 predictors

Table 1. Multiple regression effect size is based on Cohen's f^2, given the value R^2 of models

Appendix 5.2 - Summary of Variance of models in Deep and Surface scales (for whole sample and low-high prior knowledge groups) – Meta Analysis

	Whole Sample/Suggested Model **R^2 / Adjusted R^2** **Significance**	**High Prior Knowledge** **R^2 / Adjusted R^2** **Significance**	**Low Prior Knowledge** **R^2 / Adjusted R^2** **Significance**
Deep	18.3%/14.4% Sig. 0.001 (Model 4 - Suggested)	13.3% / 5.7% Sig. 0.137	36.1% / 28.1% Sig. 0.002
Interest in Ideas	15.7%/11.6% Sig. 0.003 (Model 4 -Suggested)	12.2% / 4.8% Sig. 0.162	32.4% / 23.7% Sig. 0.007
Relating Ideas	25.2%/20% Sig. 0.000 (Model 3 - Suggested)	22.3% /12.4% Sig. 0.043	43.1% / 32.6% Sig. 0.002
Use of Evidence	18.7% / 15.6% Sig. 0.000 (Model 5 -Suggested)	19.3% / 13.9% Sig. 0.011%	21.3% / 13.7% Sig. 0.039
Seeking Meaning	20.1% / 15.5% Sig. 0.001 (Model 4 -Suggested)	9% / 0.006% Sig. 0.477	41.3% / 32.3% Sig. 0.001

Surface	45.5% /41.8% Sig. 0.000 (Model 3 -Suggested)	40.7% / 33.6% Sig. 0.000	55.8% / 47.2% Sig. 0.000
Fear of Failure	41.6% / 39.4% Sig. 0.000 (Model 4 –Suggested and Leanest and Meanest)	44.5% / 40.8% Sig. 0.000	33.2% / 26.7% Sig. 0.002
Unrelated Memorising	40% / 37.7% Sig. 0.000 (Model 6 -Suggested)	38.9% / 34.8% Sig. 0.000	37.5% / 31.4% Sig. 0.001
Lack of Purpose	21% / 18% Sig.0.000 (Model 4 -Suggested)	13.1% / 7.3% Sig. 0.072	24.4%/ 17% Sig.0.020
Syllabus Boundness	21.6% / 17.1% Sig. 0.000 (Model 4 - Suggested)	19% / 10.8% Sig. 0.045	33.3% / 23.3% Sig. 0.009

Table 1. Multiple regression models for whole sample, “high prior knowledge” group and “low prior knowledge” group

Note (1): variance of models in high/low prior knowledge groups are highlighted in red when variance is larger than that for the whole sample, while it is highlighted in green when it is higher than that for the whole sample.

Note (2): statistical significance is highlighted in blue when is p>0.05.

Appendix 5.3 – Predictors on Leanest and Meanest Models (compared to recommended models)

Scale **R^2/** **Adjusted R^2**	**Seeking Meaning** **(based on Model 7)** **15.4%/** **13%**	**Relating Ideas** **(based on Model 5)** **22.5%/** **18.7%**	**Interest in Ideas** **(based on Model 6)** **12.2%/** **9.7%**	**Use of evidence** **(based on Model 6)** **17.1%/** **14.7%**	**Deep** **(based on Model 6)** **15.3%/ 12.9%**
Number of exercises solved on first try	(+)	(+)			(+)
Number of exercises solved on third try	(+)		(-)	(+)	
Number of exercises solved on second try	(-)	(-)	(-)	(-)	(-)
Number of exercises finished but not solved		(+)	(+)		(+)
Stratum				(-)	
Maximum view time on exercise page	(+)	(+)	(+)	(+)	(+)
Maximum view time on content page		(+)			
Average view time on exercise pages			(+)		
Number of pages visited using the TOC	(-)				
Number of hyperlinks (concept links) visited in reading and exercise pages	(-)	(-)			
Average number a 'notes' link is clicked per page		(+)			(+)

Table1. It shows which predictors are eliminated in the leanest and meanest versions of the models of deep scales compared to the suggested versions of the (or recommended) models

Note: Eliminated predictors are highlighted in red

Scale R^2/ Adjusted R^2	Unrelate Memorising (based on Model 7) 39.4%/ 37.7%	Syllabus Boundness (based on Model 6) 18.7%/ 15.7%	Lack of purpose (based on Model 6) 19%/ 17.4%	Fear of Failure (based on Model 4) 41.6%/ 39.4%	Surface (based on Model 6) 42.3% / 40.1%
Number of exercises solved on first try	(-)	(-)	(-)	(-)	(-)
Number of exercises solved on third try		(+)	(+)	(+)	(+)
Number of exercises finished but not solved	(+)		(+)		(+)
Number of exercises cancelled		(+)			
Average number a 'notes' link is clicked per page	(+)				
Compactness	(+)	(+)			(+)
Maximum view time on content page				(-)	(-)
Average view time on content pages		(+)			
Maximum view time on exercise page				(+)	(+)
Minimum view time on exercise page		(-)			
Relative amount of revisits			(+)		
Number of hyperlinks (concept links) visited in reading and exercise pages					(+)

Table2. It shows which predictors are eliminated in the leanest and meanest models of surface scales compared to the suggested (or recommended) models (Note: Eliminated predictors are highlighted in red)

Appendix 5.4 – Structure of AM learning material showing whether information is ‘deduced’ or ‘induced’

Chapter 4 -Graphs and Functions

4.0 Preliminaries on functions
Sets (Definition - example) – deduction
Real numbers (definition -example) –deduction
Intervals (definition - example) - deduction

4.1 Introduction to functions

Induction

Introduction (examples of functions)

Definition of Function (theory)

Definition of function – Examples 4.1 and 4.2
(theoretical-generic examples)

4.2 Real Functions

Deduction

Definition of Real functions (theory)

Real functions – Example 4..3

Real functions – Example 4.4

Exercises

Appendix 5.4 – Structure of AM learning material showing whether information is 'deduced' or 'induced' (continue)

Chapter 5 –Graphs

5.0 Introduction to Graphs

Introduction to Graphs
Axes (definition - example) - deduction

Induction

Introduction to Graphs – Example 5.1 (example)

Introduction to Graphs
Example 5.2 (example – definition linear function)
Example 5.3 (example)
Example 5.4 (example)
Definition of variable

Induction

5.1 Linear Graphs

Linear Graph Definition
Linear Graphs – Example 1
Linear Graphs – Example 2
Linear Graphs – Intercept and Gradient (definition)

Linear Graphs – Exercises

5.2 Graphical Solutions of Simultaneous Equations

Graphical Solutions of Simultaneous Equations [induction]

Exercise

Deduction

5.3 Quadratic Graphs

Introduction to Quadratic Graphs (theoretical generic examples)
Quadratic - Special examples (theoretical generic examples)

Quadratic Graphs – Plotting and Interpreting graphs - worked example 5.4.1

Quadratic Graphs – Plotting and Interpreting graphs - worked example 5.4.2

Exercises

Appendix 5.4 – Structure of AM learning material showing whether information is ‘deduced’ or ‘induced’ (continue)

Chapter 8 -Matrices

8.1 Introduction to matrices

Induction

Introduction to matrices
What is a Matrix? - A practical Solution

Definition for Matrices
Examples of matrices
Conventions in Matrices

Introduction to matrices – Exercise 8.1

Deduction

Types of matrices
Types of Matrices – Square Matrix (definition - example) [deduction]
Types of Matrices – Diagonal Matrix (definition - example) [deduction]
Types of Matrices – Unit Matrix (definition - example) [deduction]
Types of Matrices – Zero Matrix (definition - example) [deduction]
Types of Matrices – Symmetric Matrix (definition - example) [deduction]
Types of Matrices – Row Matrix (definition - example) [deduction]
Types of Matrices – Column Matrix (definition - example) [deduction]

8.2 Matrix Operations

Deduction

Induction

Matrix Addition –Definition

Worked Example for matrix addition

Matrix Subtraction –Definition

Worked Example for matrix Subtraction

Exercise

Scalar Multiplication – Definition

Worked example for scalar multiplication

Exercises

Multiplication of Matrices (examples)

Multiplication of Matrices – Process (example)

Induction

Worked example for matrix multiplication

Summary on Matrix Multiplication Process (theory)

Exercises

Appendix 5.4 – Structure of AM learning material showing whether information is ‘deduced’ or ‘induced’ (continue)

Deduction

The transpose of a Matrix (definition - example)

Worked example for the transpose of a matrix

Exercise

Summary of Matrix Operations (theory)

Summary on Addition and Subtraction
Summary on Scalar Multiplication
Summary on Multiplication
Summary on Transpose Matrix

Induction

8.3 Identity and Inverse for 2x2 Matrices

Introduction to Identity and Inverse for 2x2 Matrices (examples - definitions)

Identity Matrix (exercise - definition)

Deduction

Inverse Matrix (general definition)

Exercise

Worked example for inverse matrix

Exercise

Solving Simultaneous equations using the Inverse Matrix

Induction

Intro Matrix Representation of Linear Equations

Worked examples for Matrix Representation of Linear Equations (first example then general formula)

Deduction

Solutions of systems of simultaneous equations (general formula)

Worked examples for systems of simultaneous equations (application of formula)

Exercises

Deduction /Induction

The determinant of Matrix
Introduction to Determinant of Matrix (theory)

Determinant of matrix (example – general conclusion)

Exercises

Appendices – Chapter 6 – Recommendations and Contributions

Appendix 6.1 - Notes – Based on observations and the records on students' notes

The commentary in the 'notes' feature shows that it is used in a number of ways:

1. **To comment on difficulty of exercises, for example:**

   Student 1: "toooo easyyyy"

   Student 2: "i kno... so calm down it's very basic"

   Student 3: "way too easy"

   Student 4: "found this quite tricky."

2. **To record the answers to exercises, for example:**

   Student 5:
   "x = 4"

   Student 6:
   "-3^2 -5(-3)+8
   -9+15+8=32 (--=+)"

3. **To express questions with regards to the learning material, for example:**

   Student 7:
   "wat is this, bk 2 primary?
   (it probably means, book 2 primary for maths)"

4. **To demonstrate how they went about the solution of an exercise and ask support from fellow students, for example:**

   Student 8:
   "i started off by substituting all the x's for 0's which made the equation look like this.

   *02-5*0+8*
   (02=zero squared)

   after doing this is bracketed them off,
   resulting in;

   (02-5)(0+8)*

   i then worked out that 02-5=-5 and 0+8=8 i then multiplied them and the product was 40.

   i typed in this answer and incorrect, i then figured maybe because it was a minus number it would produce a negative result. so i then typed in -40 and was still incorrect, at this time i was frustrated and typed in 20.

   to my surprise the answer was 8.
   can anybody let me know (step by step) where i went wrong and how 2 avoid it in future.

   thanks jme"

5. **To record the logic behind an exercise and not just the actual answer:**

Student 9:
"as complicated as they are the logic sequence is unmistakable
once I know what i need to do - it becomes 'easy'
[a,b] a<=x<=b
(a,b) a "

Student 10:
"we take the values of X nad enter them into the given equation
values x= 0, 1, 2
y=2x-1
for 0 y= -1
for 1 y= 1
for 2 y= 3"

6. **To give advice to other students or share tutor's advice:**

Student 11: "write them as they are x=4, z=3 it is very hard to work out duhh"

Student 12:
"This i did not understand!
Tutor cleared things up
you take the -3 to the other side, negative becomes positive
and bobs your unkle"

7. **To comment or criticise on the way the feedback of exercises is presented by AM:**

Student 13:
"demonstration would have been more useful than just giving the answers"

Student 14:
"a bug ? when you give the right answer (det=0, no inverse) you get a message at the end : ... verify ... multiply ... to get I !!!
(it appears only if it is the first answer)"

Appendix 6.2 – Surface scale – Minimum and Maximum

	Minimum	Maximum
Surface Scale	26	72
Number of Exercises Solved on First Try	4	131
Number of Exercises Solved on Third Try	0	15
Number of Exercises Finished But Not Solved	0	32
Maximum View Time on Exercise Page	314.819000	3750.835000
Maximum View Time on Content Page	108.932000	3118.114000
Compactness	.30	.90
Number of concept links visited in reading and exercise pages	0	11

Appendix 6.3 – A non-profit MOOC - Khan's Academy

Criteria **(as defined in 2.1.1.1)**	**Features**
Multimodality	video lectures, graphics, animated demonstrations, video transcripts
Amplification	calculators
Investigation	Students can investigate freely topics with features such as "explore" and "search"
Collaboration	It allows collaboration amongst students and tutors with features such as "Ask a question", and "Tips and Thanks".
Providing mathematical activity with support.	More specifically, it provides the following elements and features: 1) the element of personalisation, which facilitates appropriate scaffolding at it allows students to complete "missions" (activities) and self-diagnose their level within a chosen topic. 2) the element of formative assessment by providing hints (video and textual hints) during the activities and model answers. 3) the element of summative assessment by providing a summary progress report which shows what skills they have mastered and which ones they need further practice and also directs towards the next most appropriate task. 4) the element of reward (e.g. badges and points according to number of correct answers and levels of skills and knowledge). However the badges can be given for reasons that are not related just to good performance but also for being a good collaborator, helping others, persisting in problem solving, etc. 5) the element of personalisation by providing a "profile" feature which shows recent activities, completed videos, and "discussion" activity.