Do teachers in low-income countries avoid implementing Early Grade Reading Programs? Using a behavioral science lens to study teacher response to educational change in rural Tanzania.

Simon John King

UCL, Institute of Education

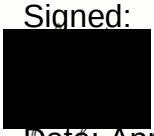
April 22, 2023

This thesis is a partial fulfillment of the requirements for the degree of Doctor in Education (Ed.D.) at the Institute of Education (IOE), University College London (UCL)

Declaration

I, Simon King 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:



Date: April 22, 2023

Abstract

This study explored teachers' responses to introducing an Early Grade Reading (EGR) Program in rural Tanzania. The theoretical framework included a behavioral science perspective to identify behaviors and social norms that restrain effective implementation by teachers. It was found that after six years of the program, the teachers from low-performing schools implemented the program and had a favorable view of its influence on their classroom practice. However, while the teachers had adopted the content of the EGR Program, they had retained many of their default approaches to content delivery. Further, they were not focused on pupil understanding and learning. All the teachers believed that most or all of their pupils would reach the expected proficient reading benchmarks by the end of grade 2. However, external learning outcome data suggests that most pupils fall short of the expected reading benchmark.

In conclusion, when mandated, teacher resistance is not an issue with the adoption of new curricula. However, teachers adapt programming to make aligning with their prior instructional approach a lower mental effort, focusing on content delivery as the observable effect they are accountable for. Similar findings from research in other countries, including the USA and Bangladesh suggest that these observations in rural Tanzania are predominantly human behavioral responses, meaning that if human behavior is predictable, then intervention programming should be able to account for it.

This study contributes value to education research through use of a behavioral science framework to study teacher behavioral barriers to education program implementation.

The study’s findings suggest that typical models of teacher change behavior used for education intervention designs that rely on logic and reason are unrealistic. Therefore, education programming should be designed to account for researched and expected teacher and other stakeholder behavior.

Impact Statement

There are two main potential impacts of this study. First, it presents an approach to education research that leverages a behavioral science research framework on education program implementation. Second, its findings suggest that we need to incorporate models of human behavior into education program design that use principles of behavioral science rather than presume stakeholders such as teachers can "*. . . think and choose unfailingly well*" (Thaler and Sunstein 2009, p.6). This is summarized below.

Using behavioral science theory to research teacher response to educational change

Since the advent of EGR programs, they have been designed using presumptions of teacher change that presume logical and reflective behaviors (Guskey 2002). This study researched evidence of teacher behavior aligned with behavioral economics (Kahneman 2011; Thaler and Sunstein 2009).

Use a research framework that identifies behavioral barriers to program implementation.

A common approach to educational research and recommendations is to focus on *what works* (Innocenti 2019; Gates Foundation, n.d.; Blavatnik School of Government 2023) and efforts to scale up these successes. Instead, this study draws from Kahneman (2011) and Levin (2005), who state we should identify behavioral barriers to implementation and find ways to mitigate them. Therefore, this research framework focuses on data collected in low-performing schools.

Study findings suggest that teachers behave in ways aligned to behavioral economics theories of human decision-making.

This study analyzed secondary data collected in low-performing schools in Tanzania. Key conclusions included that teachers were focused on delivering curriculum and pedagogy, with a misplaced belief that their pupils were performing adequately. The study found no evidence of teacher resistance to change; instead, the teachers were content and often happy to implement the reading program, believing that teaching phonics-based instruction improved

their teaching approach. Teachers used heuristic shortcuts to confirm their quality instructional practice from convenient yet inaccurate sources that did not include practical pupil assessment.

Recommendation: design education programs based on researched and expected human behavior

Finally, EGR programs are often designed around models of teacher change (i.e., Guskey) that rely on rationality and logical decision-making. However, this study's findings suggest are not accurate models of teacher behavior.

Behavioral economics positions human behavior as irrational yet predictable (Ariely, 2009). Given the findings of this study, we need to design programs based on how individuals and groups respond to programming rather than how we hope or expect them to respond.

While no model is perfect, the study recommends that EGR Programs should be designed to account for expected behaviors, rather than idealistic ones. For example, this might mean designing programs that are designed to shift behaviors.

Table of Contents

List of Exhibits

Acronym List

AS Level	Advanced Subsidiary Level
ASER	Annual Status of Education Report
BE	Behavioral Economics
CEU	Continuing Education Units
CoL	Community of Learning
COVID	Coronavirus Disease
DEO	District Education Office
DFID	Department for International Development
DIBELS	Dynamic Indicators of Basic Early Literacy Skills
DOI	Diffusion of Innovation Theory
EdD	Doctor of Education
EGR	Early Grade Reading
GCSE	General Certificate in Secondary Education
IFS	Institution Focused Study
JU	Jifunze Uelewe
LMIC	Low- and Middle-income Countries
MEL	Monitoring, Evaluation, and Learning
MoEST	Tanzanian Ministry of Education, Science, and Technology
MoEVT	Zanzibar Ministry of Education and Vocational Training
PII	Publically Identifiable Information
PO-RALG	President's Office, Regional Administration, and Local Government
PSLE	Primary School Leaving Examination
RFP	Request for Proposal
RISE	Research on Improving Systems of Education
RTI	Research Triangle Institute
SAGES	Strengthening Accountability in Ghana's Education System
SAT	Standardized Assessment Texts
SBCC	Social and Behavior Change Communication
SDG	Sustainable Development Goals
SNA	Social Network Analysis
TaRL	Teaching at the Right Level
TIE	Tanzanian Institute of Education
ToC	Theory of Change
TSLN	Thinking Schools, Learning Nation
UK	United Kingdom
UN	United Nations
USA	United States of America
USAID	United States Agency for International Development
VSO	Volunteer Services Overseas
WEO	Ward Education Officers

Reflective Statement - A Personal and Professional Journey

Over twenty years ago, I was a volunteer teacher in a rural public secondary school in Zambia, working for Volunteer Services Overseas (VSO). The school followed the old colonial boarding school model, with teachers and pupils living on campus. Although most teachers were within a couple of hundred yards of their classrooms, their classroom attendance was low, leaving the pupils sitting in class all day waiting and hoping for their teachers to turn up and instruct them. Concerned about this, I approached the headmaster to explain the issue. He called a faculty meeting and told the teachers to improve their attendance. The next day, most of the teachers were teaching in their classrooms. On the second day, about 70% of the teachers attended. Over a few days, teacher classroom attendance was back to less than 50% level before my conversation with the headmaster.

I worked at the school for two years. Over time, I realized that this behavioral norm of teacher classroom attendance was pervasive and even influenced my conduct. If I had an “important” personal chore to do, such as getting to the Post Office, I felt that this chore could be prioritized over my classroom teaching and attendance. If I set the pupils work to complete, could I prioritize getting the errand done rather than be in class? Several times, I did. This behavior is something I would never have dreamed of repeating when teaching in the UK or the USA.

There was one teacher at the school in Zambia who was different. He had almost perfect classroom attendance and was the go-to for the pupils to arrange extra-curricular activities. I found it hard to keep up with him, and I collaborated with him as much as he was willing. He was the bright spot amongst a faculty who, at times, almost seemed disinterested in the well-being and education of their pupils.

My experience as a volunteer creates a parallel with reading programs, hence the focus of this proposed research; there is evidence that most teachers resist implementing the reading programs as designed due to unwritten social norms that drive acceptable human behavior (Bicchieri and Noah 2017). The

teachers in the minority who defy these social norms are the bright spots, the exceptions, the *positive deviants*.

My experience as a volunteer helps frame my current work as a research analyst supporting Early Grade Reading (EGR) programs implemented in low-income countries. My IFS concluded that implementation fidelity on our USAID reading program interventions was low. My experience as a volunteer creates a parallel with reading programs: there is evidence that most teachers resist implementing the reading program as designed due to systematized social norms. We struggle to look beyond the transfer of knowledge and skills to teachers to provide quality instruction.

My institute, RTI International, heavily focuses on research with a mission statement of “Turning Knowledge into Practice.” My work over the past ten years at RTI has been focused on supporting Early Grade Reading programming and research through monitoring, evaluation, and learning (MEL) support. My work as a statistician has always been in demand. However, I’ve always been unsure of how my prior 14 years of experience as a teacher and school administrator fit into this process. This was not even evident while starting my EdD program. It was not until I was working on the theoretical framework that I was able to connect my experience as an educator and work as an education researcher. The alignment of behavioral economics with Fullan’s New Meaning of Educational Change started to resonate with my earlier career in schools. The theoretical framework of behavioral economics applies human behavior to explain economic decision-making (Thaler and Sunstein 2009), where the process of Type I (fast, emotional, and instinctive) and Type II (slow, logical, deliberate) thinking plus environmental influences influence an individual’s decisions (Kahneman, 2009). This started me on a process of drawing on my teaching experience. How had I responded to change? How had my colleagues responded to change? How had I supported others in the process of change?

My response to change as a teacher has varied. If a school administration or education system-imposed change, I often resisted. I had always been keen to take ownership of my change – taking time to research and change my curriculum and pedagogical practice to meet the needs of my pupils.

Additionally, I had observed my teaching peers as quite varied in their response to change. Some chose to embrace change as inevitable, while others found the process very stressful, choosing to isolate themselves in their classroom, away from the rest of the faculty. This idea of variability in teacher response to change I found very important. There is a tendency in educational research to generalize the findings of statistically significant approaches (such as multi-variate analysis with learning outcomes as a response variable) and then recommend the scaling of these findings. However, my IFS identified different teacher subpopulations based on those gains was critical for me to understand that we needed to look not at those positive deviants who were willing and able to implement, but rather the majority of teachers who did not demonstrate improvement in pupil learning outcomes. This approach contradicts most educational research, which looks for *what works*. My IFS concluded that the typical characteristic of positive deviance is positive personality traits (King 2020) which cannot be easily scaled. However, most educational research attempts to identify and scale behaviors rather than understand the fundamental cause of these behaviors. This is why this EdD research study used secondary data collected from teachers who did not improve pupils' learning outcomes.

As shown in the introduction in the Research Problem section of this study paper, my presumption regarding the subpopulation of teachers with no improvement in earning outcomes is that they resisted implementing the reading program. My mistake was not considering teachers' time interacting with the program. This study was conducted six years after the introduction of USAID reading programs in Tanzania. The initial human response to imposed change can often be one of resistance, then eventual acceptance after a process of change. If teachers are mandated to adopt a new curriculum and eventually adopt it, this new approach becomes part of the teachers' instructional approach. It makes sense that they would endorse the new curriculum. Anything else would be a criticism of their instructional practice.

The teachers implemented the mandated program while minimizing loss, changing curriculum but retaining aspects of their prior default instructional practice, which was primarily focused on the delivery of content, with very little

time spent on pupil learning and understanding. This clearly parallels my observation of my teachers when I was a high school pupil in England when the National Curriculum was introduced in England in the 1980s. Some teachers showed vocal anxiety, while others could accept change and adapt more quickly. However, besides GCSE project work, teaching pedagogies did not significantly change. However, when I entered the teaching workforce ten years later (teaching in Reading, England), I observed no teachers resisting the national curriculum, standards-based reform, and related standardized assessments (e.g., SATs, GCSEs, AS levels). Significantly, while content and assessment had changed, I again noticed no evidence of real change in pedagogical practice from my colleagues. As a high-school teacher of mathematics, I fell into the same gradual release model pedagogical teaching practice as my colleagues. The teacher introduced a new idea, a class discussion took place, then the pupils were given a worksheet to practice and develop their mastery of the new concepts. In other words, the reform efforts had adapted content and assessment, but basic classroom practice remained. Teachers were able to minimize loss by retaining a pedagogical approach. I had little choice but to follow my colleagues as I was expected to assign worksheets to my pupils on a set schedule. Every teacher posted the same work and gave the same unit assessments.

Adoption of new content but the same pedagogical practice is precisely what the teachers observed in Tanzania were able to accomplish. However, the downside was that the pedagogical approach was relatively ineffective for pupils with low literacy rates, the majority of pupils in these Tanzanian classrooms.

The emphasis on how teachers adopted curriculum versus focused on pupil learning is critical. Reflecting on my high school mathematics teaching career, I had adopted a new curriculum and content about thirty times in fourteen years. Still, I had changed pedagogy only twice, both times upon my motivation, and never imposed. This highlights teachers' general comfort with adopting new curricula and content. The mental lift is relatively low compared with changing pedagogy, which demands a fundamental shift in teacher classroom behavior.

Additionally, the education systems in England and Tanzania focus on implementing new curricula as the observable effect of change. In England, it was linked to standards-based assessment. In Tanzania, curriculum implementation was related to classroom observations by head teachers and external coaches. It is a human instinct to focus on the “easy to spot” and accountable observable effect. In my first year of teaching in England, I was responsible for delivering content, following the mathematics department schedule. However, conversations on effective classroom practice and pupil learning were few and far between.

Finally, it has been essential to link my teacher experience, behavioral economics research, and my work in international education in LMICs. It is often the case that domestic-based (USA, UK) research applied in LMICs is criticized due to lack of contextualization. What is critical for me is to create parallels with human behavior, that teacher response to change varies, and individual teacher personality characteristics and the environment mostly explain it. Consequently, integrating my school-based experience into my research has been very fulfilling.

Professional and personal position on EGR Programs

My professional position as an education researcher on EGR Programming is one influenced by my previous career as a mathematics teacher. My first year of teaching as a Newly Qualified Teacher (NQT) was in a school in Reading, Berkshire. During in-school teacher training prior to the first day of teaching, I was introduced to the structure that all the teachers in the mathematics department followed. I was given a schedule by which I used printed worksheets to teach topics aligned to the National Curriculum standards. Each week I would reference the schedule and collect the worksheets I needed for the week. Once I had taught the classes with the respective worksheets, I would return them for other teachers to collect and use. At the end of each half-term, the pupils would be assessed on the curriculum taught, again aligned to the National Curriculum standards. All the teachers used the same assessments. Pupils were ranked and streamed in mathematics classes; grouped according to ability and given work according to their

already demonstrated mathematics skills. While not given explicit instruction to do so, I was encouraged to follow a gradual release model of instruction in the classroom; I would introduce new content, discuss as a class, then pupils would individually practice the new concepts using the worksheets. Indeed, the way the worksheets and timetable were structured it was challenging to use any other approach.

As a new teacher with typical energy and enthusiasm I found this approach to instruction quite rigid. While more able pupils in the higher streams progressed, I found that pupils in the lower ability classes struggled to progress. I observed that many lacked basic numeracy and literacy skills, but we still had to move through the worksheets that delivered curriculum. No remediation was scheduled if pupils scored poorly on unit tests.

I found this experience to be a two-edged sword. Firstly, the pressure on an NQT to prepare six classes per day was overwhelming. That a sequence of worksheets was ready to go helped me to survive my first year of teaching. However, the other side of the coin was that my desire to experiment, learn, and express myself as a teacher was restricted. During my PGCE, I was encouraged to try out different pedagogies and learn. This opportunity was not given to me during my first year of teaching. A few times, I taught my low ability classes some basic numeracy skills, moving away from curriculum worksheets. This was discovered and I was reprimanded by my department chair.

I left the school at the end of the year, moving to a private school that had resources, but no curriculum. The only real benchmark of pupil progress was the GCSE examination they would take. At this school I enjoyed the freedom to experiment with different instructional approaches but spent much more time prepping classes. I would say that during this year, I learned a lot more about my identity as a teacher and developed an idea of what worked in the classroom.

Years later, joining RTI as a statistician, I looked-on with curiosity as structured phonics-based literacy programs were being designed and implemented. All used pre-generated lesson plans, a more structured

approach to the worksheets that had been introduced to me as an NQT. I naturally reflected on my mixed experience with structured education programming. Found it hard to reach a definite conclusion regarding the success of these programs. While there was statistically significant impact, in real terms, there was not much progress with pupils really accelerating in their learning.

In development work, when organizations such as RTI focus heavily on business development, there is a focus on the successful aspects of program implementation, with the understanding that good work is the easiest way to secure more contracts. It was not until embarking on my EdD program that I was able to design and implement my own research and follow a path of objectivity.

Throughout my time supporting EGR programs, I've always reflected on how teachers responded to the introduction of structured programming in the same way I did in my first year of teaching. I have found it difficult to have a definite position on whether structured EGR programs benefit pupil learning given my own relief as an NQT to have worksheets ready to go that helped me survive my first year in teaching. However, I believe that my identity as a teacher and meet the needs of my low-performing pupils was restricted. How do teachers in rural Sub-Saharan Africa respond to EGR programming? Does it restrict their decision-making to support their pupils, or is structured programming welcomed in a system which lacks teaching and learning materials (TLMs) and teacher training?

This is the driver for my EdD. We know many schools in Tanzania lack resources and training compared with the opportunities I had teaching in England.

The EdD process provides an opportunity for objective research that I believe is unfortunately not always available in development work. Without innovation bias and business development as a core consideration, the EdD provides an opportunity to research barriers to EGR Programming, with the understanding that some of these barriers might be insurmountable and question the validity of EGR Programming.

Reflection on EdD program experience

The EdD journey as a working professional has been one where I have often made constant linkages between my professional career and academic research. As someone who made a mid-career jump from the classroom into research, I found that while my quantitative methods were strong due to my Master's degree in statistics, my understanding of the research process was an area where I lacked understanding and experience.

The first assignment, **Foundations of Professionalism**, made me think about a problem-solution process. This was the first time I had conducted a literature review, and I had to think about organizing my research and presenting an argument based on the available literature. Within my institute, my writing at the time was mostly restricted to technical reporting. Therefore, I found this assignment useful in helping me think about the process of designing and organizing my writing.

At the time, most of my work at RTI International centered around monitoring and evaluating EGR Programs. So my starting point was usually establishing research questions followed by a methodology and analysis plan. This unit made me take a new and unfamiliar approach and think about the steps before developing research questions, such as a theoretical framework.

The **Methods of Enquiry - Module One** focused mainly on the research steps preparing for data collection; problem statement development, conceptual framework, theoretical perspective, literature review, and methodology. This module was initially unfamiliar and challenging, highlighting its importance in my research progression. This research process I did not usually do this for my work and consequently, this was a good learning experience to think through the logical steps of presenting my research. I also observed the analysis work my colleagues and myself do, we often forgo the steps before developing the research questions.

Methods of Inquiry - Module Two focused on research aims, analysis, results, and discussion. I could more easily pull from my work experience. The bulk of the work on module two was quantitative analysis, which I have extensive experience with. However, the general objective of how the two

modules linked together was an excellent learning experience in showing me how to think about and link theoretical perspectives and literature reviews to my technical work.

In my time working at RTI, I have observed that there is a tendency for researchers to jump straight into the development of instruments, analysis, and reporting (these are key client deliverables). While research questions are often developed, they are often limited in scope, dealing with the primary questions that need to be answered. Alternatively, we develop instruments and collect too much data because we lose focus on alignment with the research questions. Completing these two modules helped me reflect and apply a more deliberate research process to my technical work.

The biggest challenge in our international education work at RTI International was getting program impact and achieving improved learning outcomes from early-grade reading programs. While the effect sizes are medium or large in terms of educational interventions, there are still high percentages of children who could not recognize a single-word text by the end of grade two (USAID 2022a). There had not been much progress on the percentage of children who can read fluently with comprehension, according to the United Nations Sustainable Development Goal indicator target 4.1.1 (United Nations 2019). Therefore, I applied a straightforward question, "*What are the barriers to program implementation that, if mitigated, would have the greatest impact on learning outcomes?*". To answer this, I looked at measures of fidelity of implementation (FOI). These are measures from classroom observations of teachers that estimate how well teachers implement the reading program. I found that measures of FOI were high, with most programs reporting teachers implementing the reading programs with fidelity levels of 70% or higher. Contrasting this with percentages of grade 2 pupil non-readers, often in the 40%-80% range across our programs, presented what seemed to be a contradiction; it was unlikely that such high levels of fidelity result in such low learning outcomes unless there was a problem with the efficacy of the reading program. However, by researching the issue in Nepal using an explanatory mixed-methods approach, I determined that schools with strong average learning gains were motivated to implement the reading program. Conversely,

schools without improvement in learning gains were only partially implementing the program or not implementing it. This study was conducted after two years of implementing a reading program. This starkly contrasts with the typical FOI measures reported by over 70% of teachers implementing the intervention. This suggested that the FOI measures we and other reading program implementors use to measure fidelity are inaccurate and present an inflated picture of true fidelity. This led to two main conclusions; firstly, if the data on fidelity is inaccurate, how do reading programs use it to make informed decisions? Secondly, why are teachers resisting implementing programs?

These conclusions to my IFS were a logical bridge to my thesis proposal. However, I could take two possible directions for my research; either develop improved measures of FOI or determine the barriers to teachers implementing reading programs. I decided that the former was too challenging. From my previous experience as a head teacher, I was not confident that any classroom observation measure could accurately capture actual (not observed) teacher instructional practice. Additionally, the FOI model I suggested in my IFS that would more accurately measure FOI would be rather challenging and time-consuming to collect data to measure. Therefore, the natural bridge for my thesis and where there was a gap in research was understanding why teachers resisted implementing reading programs.

A key conclusion from my IFS is that positive deviants explained schools that showed improving learning outcomes. Individuals or groups who were motivated to create change. They were resistant to the social norms or status quo. However, these individuals had apparent personality differences (King 2020).

Consequently, I focused my thesis study on researching the teachers and schools with limited or no impact on learning outcomes. My institute, RTI International had collected qualitative program data from Tanzania through qualitative classroom observations, and teacher interviews from a sample of schools where learning outcomes were poor. This created an opportunity for

me to use this data to conduct secondary data analysis using a behavioral science research framework.

Chapter 1: Rationale

1.1 Research Problem

My thesis's research rationale is to explore the gap in research and program implementation that, in its simplest terms, can be expressed as, ***if we train two teachers on a reading program, what makes one teacher decide to implement and the other not?*** This challenge is illustrated in Exhibit 1 below.

Exhibit 1: Neighboring Schools in Ghana

(Photos courtesy of Robin Todd, 2019)

The photos show two classrooms in two separate schools in similar rural environments, where both teachers have access to teaching and learning materials. However, while one teacher leverages opportunities, the other is calcitrant.

My role at RTI International is to support the design and implementation of USAID-funded EGR programs. My EdD research aligns with the challenge my institute is trying to address, as explained below.

United States Agency for International Development (USAID) support of EGR programs evolved as a response to increased enrollment in public schools in low-income countries but a stagnated quality of classroom instruction (UNESCO, 2006). Most USAID reading programs have similar core design characteristics: development, printing, and distribution of classroom learning materials, teacher training, and ongoing teacher support through in-school coaching-type systems (Graham and Kelly, 2018).

While these programs have returned some impressive effect sizes in terms of impact on average learning outcomes (Gove et al., 2017; Graham and Kelly, 2018), the increase in the percentage of children able to achieve a minimum proficiency level in reading according to SDG indicator target, "*4.1.1 Proportion of children and young people: (a) in grades 2/3; (b) at the end of primary; and (c) at the end of lower secondary achieving at least a minimum proficiency level in (i) reading and (ii) mathematics, by sex*" (United Nations, 2019) is still very low (Piper & Stern, 2019). The difference is that effect sizes are reported due to impact evaluation, where learning gains are the focus. Reporting on SDG 4.1.1(a) is a measure of achievement. This lack of impact against the indicator raises questions regarding implementing EGR programs in low-income countries (Piper & Stern, 2019).

1.2 Research Rationale

My thesis represents the logical progression from my Institute Focused Study, which showed that only a relatively small percentage (5% - 35%) (King, 2021) of schools accounted for the impact of USAID Early Grade Reading programs. When conducting a follow-up qualitative analysis, it was determined that the schools demonstrating impact had one common trait, an individual or group within the school who were creating change by ignoring the social norms that reinforced the 'resistance' to education reform or innovation. These schools, individuals, or groups of individuals often go by different assigned names; early adopters, bright spots, and even positive deviants. All these labels suggest an understanding that these individuals behave in ways that the majority do not, at least at a given time.

It is common for education research to focus on challenges of delivery of quality education, such as class size, resources, teacher capacity, socioeconomic-related issues, and teacher support. However, while it is hard to argue against the debilitating challenge a teacher might face in instructing a class of over eighty pupils, there is evidence that pupil learning can improve even when conditions are challenging (Ngware, Oketch, and Mutisya 2014). While these teaching conditions issues are real and challenging, the focus of this research study looks to control for these issues as best as possible through secondary data analysis research design (such as matched schools) and focus on schools with similar characteristics that demonstrate different levels of education outcomes.

There are two possible research avenues to research bright spots. One approach is to understand what actions bright spot teachers, head teachers, and schools are undertaking that create success and scale up these successful approaches (UNICEF Innocenti, 2019). However, my IFS study (King, 2020) identified the common trait of bright spots is not the actions individual(s) undertake. These bright spots have common personality traits akin to early adopters (Rogers, 2003) that cannot easily be replicated or scaled.

Consequently, my research will focus on a second path; understanding the environments of schools where improved learning outcomes remain an issue through a theoretical perspective comprised of three theoretical components:

- Behavioral Economics (BE); pioneered by Nobel prize laureates Daniel Kahneman and Richard Thaler.
- Diffusion of Innovations Theory (DOI), first developed by Everett Rogers in 1962, and
- *The New Meaning of Educational Change*, by Michael Fullan (2015)

My background and literature review will demonstrate that using only one or two of these concepts presents a weak theoretical framework. The motivation of an EdD is that it focuses on the practical reality of applying education research. A lesser framework (using just one or two of the theories above)

would not help my institute better understand and mitigate its challenge in supporting local ministries to implement an EGR Program.

1.3 Summary of Adopted Frameworks

Behavioral Economics (BE) combines economics, psychology, and sociology to understand how individuals deviate from traditional economic behavior models due to environmental and psychological conditions (Thaler and Sunstein 2009). Further, these deviations from how individuals are expected (or desired) to act are said to be irrational yet predictable (Ariely 2009). When researching teachers' mental models in low-income countries, Sabarwal and Abu-Jawdeh (2018) suggested that teachers already believe they are applying maximum effort and do not respond to accountability or incentives. Sabarwal and Abu-Jawdeh recommended further research on teacher mental models by using behavioral economics to understand how teachers act in ways that, while not altogether desirable, can be predicted and accounted for within education systems. Perhaps the most significant potential of behavioral economics is that its solution to challenges is often the removal of obstacles that promote good decision-making (Kahneman 2011). This idea seems contrary to much of the work of international development, which frequently adds intervention components to account for challenges.

The weakness of BE in the context of education is that the vast majority of BE research focuses on civil society, where individuals are free to choose. In public health and personal finance, ‘nudges’ are implemented to help individuals make beneficial decisions. It also seems that the same illustrative examples are used across books and articles on BE. One of these examples is the automatic “opt-in” to an individual’s retirement contributions to get matching employer funds in the USA (Thaler and Sunstein 2009). Individuals are still “free to choose” if they want to opt-in or out of their company’s retirement scheme. However, the automatic option of opting-in changes the *default*, so if the individual does nothing (which is often the case), they opt-in. Thus they are making significant contributions towards their retirement, for which their company provides matching funds.

However, this example also illustrates why using just a singular BE approach would most likely ignore the complexity of our education system context. The special effort by a teacher to change their classroom instructional approach is a considerable mental and emotional effort over time (Fullan, 2015), and an innovative ‘nudge’ approach based on a single decision point seems lacking. If we think about the challenge of getting a teacher to implement a new classroom approach, given their place in an education system – “nudging” seems to ignore the complexity of the environment and challenge.

And so, while BE solutions seem challenging in an education context, using BE to understand the reasoning behind the teacher and other stakeholder behaviors and actions with education seems useful. Aligned with my personal experiences of working in schools, it seems a stretch that most individuals make decisions based on rationality and self-interest (Thaler and Sunstein 2009).

Michael Fullan’s (2015) work on educational change presents a valuable addition to my theoretical framework as it adds the critical educational systems component to the framework. Fullan describes how two education systems can implement education reform with similar components (e.g., standards-based curriculum, professional learning communities, and assessment). Still, only one education system successfully implements and improves learning outcomes. The reason is that this type of reform does not address changing the environment in which stakeholders work (R. F. Elmore 2004). In other words, designing and pushing high-quality reform components need a setting that can be accepted and adopted over time. According to Fullan (2015), three components are required for implementing any new education innovation: the use of new teaching and learning materials, the use of new teaching approaches or pedagogies, and the alteration of teaching beliefs.

The challenge Fullan discusses is with the third component, which requires shifting of individual and shared beliefs and ideals. He references *Loss and Change* by Marrais (1986), who says that all real change involves loss and anxiety.

Fullan summarizes the change process as finding motivation. He says, "*the holy grail of change is to know under what conditions hordes of people become motivated to change*" (Fullan, 2015, p.39). Fullan demonstrates a keen understanding of behavioral economics by saying that individuals focus on costs more than on the benefits of change. This parallels prospect theory (Kahneman and Tversky 1979) in BE, suggesting that individuals place a greater weight on loss over gain. When discussing the challenges of implementing innovations in the public sector, Glor also references "*Motivational dilemmas*" (Glor, 2003, p.14) as the most important issue.

My institute also describes desired characteristics of an education system. *Core Functions* (Bruns 2018) is a systems foundation of (i) setting and communication expectations, (ii) monitoring progress, and (iii) providing targeted support. While Core Functions' tenets are a subset of Fullan's ideas, it suggests an essential focus on practical communication of research and findings from this EdD thesis. However, what might be missing from *Core Functions* is Fullan's (2015) emphasis on the personal and shared meaning of educational change. In other words, Core Functions describe what needs to be achieved by educational systems from an organizational perspective but not the required conditions in which it can be successfully implemented.

Finally, I propose using DOI to think about how EGR Programs, in effect innovations, diffuse over time and are communicated through specific channels among members of a social system (Rogers, 2003). DOI theory paints a complex system of innovation adoption where innovations diffuse laterally, no matter if their genesis is from a central administration or the ground-up. Like BE, most of DOI has been applied to civil society, for example, farmers adopting a new type of seed or uptake of cellular phones (Rogers 2003). However, DOI helps move the focus beyond asserting "what works" in the classroom and onto the environment needed to adopt the innovation. DOI describes how innovations are initially trailed by early adopters, a small percentage (typically less than 20%) of the population. Most of the targeted population wait for feedback from their peers before trying the innovation out for themselves. This theory aligns with my IFS, which stated that only the bright spots, a small percentage of the population (5%-35%) of

the teaching population, accounted for impact. Rogers (2003) states that we should make no assumption that an innovation will diffuse and be adopted by the majority of the targeted population. My IFS suggests this is the case with EGR programs (King 2020).

DOI has had limited application in education. Research has been chiefly used to measure individual attitudes toward educational innovation (Richardson 2011a; Hughes and Keith 1980). We will explore one aspect of DOI theory: how an individual's perception of innovation is critical to their adoption decision (Rogers 2003). For example, the most likely perceptions of an ICT innovation linked to its adoption for trainee teachers in Cambodia were if the innovation was perceived as being mandatory and complex (Richardson 2011a).

1.4 Linking Institution-Focused Study (IFS) to Proposed Thesis Research

The first research question from my Institution Focused Study (IFS) was, "*is presenting a generalized average impact presenting a misleading understanding of program implementation?*" (King, 2020, p. 45). A table produced for my IFS and reproduced below (Exhibit 2) shows that EGR programs report a relatively high percentage of implementation fidelity by teachers.

Exhibit 2: Teacher classroom fidelity versus program impact

Program/country	Fidelity measure	Impact (Effect Size)	Percent of grade 2 pupils unable to recognize a single word
Ghana: Partnership for Education Learning	82% of teachers implemented with near or full fidelity	+6.9 cwpm (0.50)	57%
Nepal EGRP	73% of teachers implement reading program components with fidelity	+5.4 cwpm (0.27)	48%
Nigeria RARA	Observed implementation components between 78% and 100%	+4.9 cwpm (0.71)	82%
Jordan Ed Data II	Observed implementation components between 45% and 66%	+4.2 cwpm (0.30)	4%

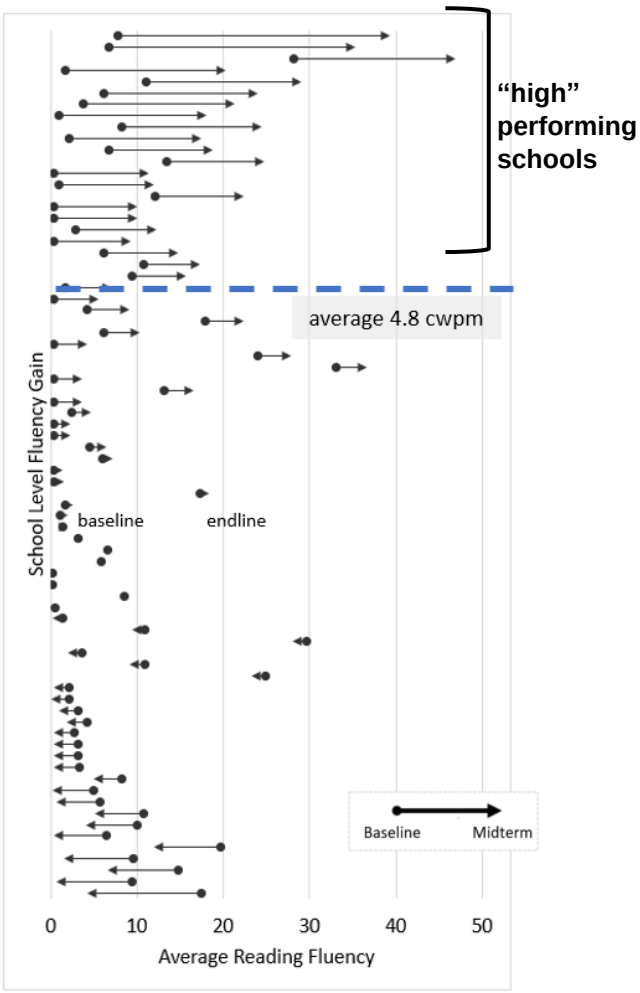
(Figure 20: teacher classroom fidelity versus program impact, King, 2020)

The fidelity measure is an indicator generated from classroom observation measures. Exhibit 2 also displays the percentage of pupils unable to recognize a single word of age-appropriate text. The impact is calculated in terms of learning gains, while the percentage of children unable to identify a single word of text is pupil achievement. Combined, these three measures create contradictions. EGR programs can generally impact in terms of effect size, but in real terms, a large percentage of children are still unable to read (Piper and Stern 2019). For example, the Nepal EGRP reading program reported 73% implementation fidelity, with a small effect size (Cohen 1990) impact of 0.27, and 48% of the pupils could not read a single word.

However, these programs are designed using well-established and tested principles of phonics-based literacy programs (Gove et al. 2017). So my IFS presented evidence that the fidelity of implementation measured through classroom observations is likely inflated (King 2020).

Why are there so many pupils unable to benefit from what are relatively standard phonics-based early literacy programs? It is argued that teachers take little responsibility for enhancing pupils' skills whose learning is low (Sabarwal and Abu-Jawdeh, 2018). In my IFS, I presented evidence that program impact explained by a relatively small percentage (5%-35%) of schools (King 2020), and the qualitative follow-up visits to eight schools in Nepal presented a simple dichotomy; schools with an impact on learning were implementing the reading program, and schools without an impact on learning were simplistic, not implementing. Exhibit 3 shows the school-level learning outcome averages by school for the Nepal Early Grade Reading Program (King, 2020) between baseline (2016) and midterm (2018). The indicator used was the oral reading fluency measure from an Early Grade Reading Assessment (RTI International 2016).

Exhibit 3: Baseline 2016 – Nepal Midterm 2018, Grade 2, Cohort 1



(King, 2020)

The variation in average reading fluency is explained mainly by individual school contexts. Differences by school of teacher education, socioeconomic characteristics, and other key factors undoubtedly play a part. Still, differences in the impact on learning outcomes of schools with similar characteristics were striking. This is evidenced by a lack of correlation (r=-0.081) between the school-level baseline average and school-level gain (King 2020). In other words, while the relationship between socioeconomic status and pupil achievement is well established (Kanyongo and Ayieko 2017; Piper, Jepkemei, and Kibukho 2015), the evidence presented indicates that schools in relatively poor areas can and do implement the reading program.

So, what made the difference in successfully implementing schools? A qualitative follow-up was conducted where four high performing and four low-performing schools were selected for a school visit. In one school, an active

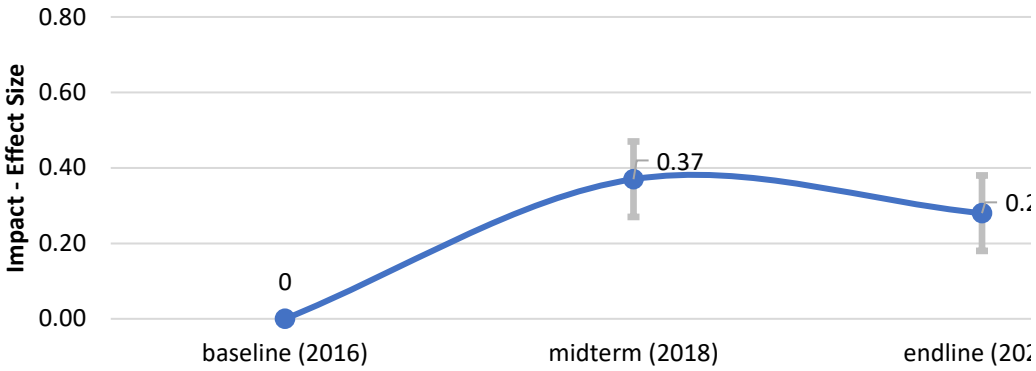
and supportive headteacher spent time watching every teacher providing classroom instruction each day. Another school had a group of excellent teachers supporting each other in learning how to adapt and implement the reading program. Finally, there was a school with an active community supporting teachers and learners (King 2020).

Therefore, the common trait that creates success in these schools is not a single component but rather a common behavior and clear evidence of motivation, the key characteristic mentioned by Fuller (2015). Individuals connected with these schools resist the social norm-driven inactivity exhibited in other schools. In my IFS, I called these individuals (or groups of individuals) "*bright spots*." This is a term also used by Heath and Heath (2011) in the book "*Switch: How to Change Things When Change is Hard*." A term that might also comes from DOI; Rogers (2003) categorized those who adopted an innovation early are referenced as *early implementors*, a group who is keen to get in on the ground floor with early implementation with the understanding that adaption and experimentation are key at this stage (Rogers, 2003).

Pritchett and Honig (2019) mention the importance of *positive deviants* within a lateral accountability setting, highlighting and rewarding those who have performed outstandingly well within an otherwise challenging educational system.

Another important finding for impact evaluations of reading programs is that for evaluations that measured impact beyond just baseline and endline (i.e., also measured at one or more midterms), after the initial impact at a midterm, the impact 'flattens' and no statistically significant gain is achieved at subsequent evaluation timepoints (Exhibit 4).

Exhibit 4: Impact - Nepal EGRP, Grade 2 Oral Reading Fluency



(King, 2022)

Impact evaluations at more than two timepoints show a similar pattern and potentially provide evidence with implementation: programs have a relatively small percentage of teachers implementing early, but few teachers decide to join this group of early adopters of the program at later points in the program. This theory fits well with DOI; adoption of the ready program most likely fails to reach critical mass – when enough stakeholders in the education system have adopted an innovation such that the rate of further adoption becomes self-sustaining (Rogers, 2003). This is often called the tipping point (Gladwell 2001), where change becomes unstoppable. However, the full adoption of new approaches and innovation might be slightly optimistic in an educational context.

The mathematical calculation of how to improve the EGR program's impact on pupil learning outcomes is quite persuasive; getting a higher percentage of teachers to implement the intervention should significantly impact improved learning outcomes and get more children reading. This is conditional on the design of the intervention, when correctly implemented, to increase learning outcomes (i.e., pro-innovation bias).

Subsequently, I propose focusing my thesis research on developing and assessing a theoretical framework that explains the limited impact of ERG Programs from an implementation perspective by the understanding of the social norms that prevent teachers from implementing reading programs and recommend adaptions to programming that mitigate the social norms that prevent them from doing so. This area of focus bridges my IFS conclusion that every school visited, which demonstrated an impact on pupil learning,

relied on *early implementor(s)*: an individual or group of individuals acting as a catalyst for improved classroom instruction and subsequently improved learning outcomes. This person or persons was not limited to being the classroom teacher. Instead, these individuals could be school administrators, parents, or community members. Any school lacking a positive deviant had no improvement in learning outcomes. For this thesis, I focused on understanding the negative social norms from those who mitigate them (*early adopters*) and those for whom these norms impede implementation. This approach is a critical consideration for this research. For example, while it is essential to understand why early adopters mitigate negative social norms, these individuals may *mask* deficiencies and barriers in early-grade reading program design and implementation or the education system. To this end, it is also essential to understand these social norms from those who could not mitigate them.

Critical to this study is using domestic (UK and USA) and international education research and other relevant evidence in LMIC education systems. Research in these contexts informs the majority of policy and implementation decisions. They are often neglected due to a framework that that education systems in LMICs face unique contextual challenges. While environmental considerations are essential, my framework will also postulate that attitudes and behaviors of stakeholders are common when innovations are imposed across any country. In a recent keynote speech at the United Nation's behavioral science week, Samantha Powers, the director of USAID, commented that we must " . . . *learn how the people we hope to serve act or do not act in response to everyday challenges*". (Power 2021). Notably, she highlighted Daniel Kahneman's and others' work, "*Behavioral science is new. It is not old. It is not actually all that familiar. It is not what we have been doing all along."*. This highlights a common misconception within the development community that behavioral science is similar to social and behavior change communication (SBCC), where civil society is persuaded through positive messaging to change individual or group behaviors. However, BE is focused on economic decision-making, leveraging an understanding of the predictable irrationality of humans.

1.5 Core Question for RTI International and International Education

As I write this thesis, it seems that International Education, particularly early-grade literacy, is at a point of reflection. Recent USAID Request for Proposals (RFPs) for early grade literacy programs have focused on building on existing programming; the foundations of access to quality resources and trained teachers are in place. However, achievement levels remain low. The recent "Smart Buys" (The World Bank 2020) publication presents evidence on what works and is cost-effective in the classroom but also mentions, "*Low levels of teacher attendance and low levels of effort (based on what is observable) are pervasive in LICs and MICs, compared to what is observed in high performing systems. This is due to failings in the support and motivation provided by the education system rather than the failings of individual teachers. Still, it is costly to pupil learning nonetheless.*" (The World Bank, 2020, p.17). The key term of "*motivation*" is something that Fullan (2015) also references. In other words, "Smart Buys" (The World Bank 2020) lays the groundwork regarding what types of intervention work, return value for money, and move the international sector into systems and uptake of programming.

Consequently, this is the limitation I have to place on my research. My research will not be concerned with the design of a reading intervention – the quality of classroom teaching and learning materials and teacher training. The assumption I need to make is that the program's design is sufficient to impact pupil literacy rates when implemented as designed significantly. If we look at Exhibit 3, some high-performing schools are showing impressive improvements in learning outcomes, although we cannot entirely attribute these impressive gains to the intervention alone.

The EdD program aligns my professional and institutional goals with research. As an employee of an implementor of over twenty USAID reading programs, there is a fundamental challenge of improving teacher uptake of reading programs. Consequently, the single driving research question that drives this research is, *"If we train two teachers on a reading program, what makes one teacher decide to implement it and the other not?"*.

Chapter 2: Literature Review

This chapter explores the literature on the three theories that underpin the theoretical framework. As explained in chapter 1, it is not sufficient to research a single theory in isolation for my EdD research to be practically useful for my institute. For example, while behavioral economics has the potential to give insights regarding how individuals within an education system respond to change, using a singular nudge approach, as described by Thaler and Sunstein (2009), over-simplifies the complexity of an education system. Most innovative "nudge" approaches focus on a single decision made by an individual in civil society. However, behavioral economics helps reframe how individuals in an education system respond to change; their interaction with innovation is fundamentally emotionally, not logically driven.

Additionally, as this study is designed to be useful for my institute it cannot be just focused on research and understanding. Consequently, the theoretical framework should also be comprehensive enough to be used when considering programming solutions.

This literature review will present the three theories, making connections across the theories and with educational research.

2.1 Diffusion of Innovation

Everett Rogers developed the Diffusion of Innovation (DOI) Theory in 1962. The theory has been updated, and the theory has been reapplied in different settings and contexts, with the fifth edition of the book being released in 2003. This final edition was concerned with applying DOI in more contemporary contexts, such as the advent of the internet, exploring new communication channels (e.g., email, SMS), and how technology innovations (e.g., cell phones, email) have quickly diffused.

Rogers (2003, p.5) defines Diffusion of Innovation theory as *" . . . the process in which an innovation is communicated through certain channels over time among the members of a social system*". He proposed that innovations do not happen at once across the target population. He also defined innovation as "*An innovation is an idea, practice, or object that is perceived as new by an individual*" (Rogers, 2003, p.13). Innovation in education can take different

forms. DOI has been most frequently applied in understanding the introduction of technological innovations and the response of teachers to this innovation, where uncertainties regarding the benefits of ICT in the classroom persist.

The most disruptive type of innovation is systems-level reform, which is ultimately designed to change classroom instructional practices. Illustratively, this could be standards-based reform that impacts all aspects of teaching, including curriculum, instruction, classroom resources, and formative and high-stakes assessments. Teachers, head teachers, or other local stakeholders can also introduce educational innovations. However, my institute designs and implements USAID-funded early-grade reading programs. As such, these programs are usually designed in collaboration with central local education ministries. District education offices or other regional education entities are leveraged to train teachers and provide ongoing support. So, this is a top-down innovation that uses a decentralized education system to provide knowledge and understanding and support the implementation of the innovation to schools.

The trajectories of innovations can vary by design and the system they introduced. How innovation can diffuse from the ground up is perhaps more closely aligned with the original discovery of DOI. An excellent example of this is the 1997 "*Thinking Schools, Learning Nation*" (TSLN) initiative in Singapore, where the Ministry of Education encouraged school leaders to select and adopt pedagogical innovations (Hung et al., 2017). TSLN shows how successful innovations "*. . . have both lateral (decentralized) and vertical (centralized) moves, capitalizing on the affordances of the education system*" (Hung et al., 2017, p.9).

DOI is a complex, layered theory. I will provide a background into DOI, aligning where possible with other literature and the research problem. I will present DOI with the following sections:

- Categories and characteristic of adopters
- Perceived characteristics of innovations
- Innovation-decision process

2.1.1 Categories and Characteristics of Adopters

DOI Theory suggests that innovations gain momentum (or uptake) over time through different personality categories of the population as they relate to the innovation:

1. **Innovators (2.5% of the target population):** Individuals who want to try new ideas early on, even in the development stage. Innovators are aware that the product is in development and might not deliver promised benefits or even get introduced to the entire target population. They are not risk-averse and often support the innovation's improvement or refining.
2. **Early Adopters (13.5% of the target population)**: Individuals comfortable with change, disruption, and adopting new ideas. Both innovators and early adopters tend to be social leaders rather than social followers. Early adopters are said to have a higher social status than their later implementing peers, being social leaders.
3. **Early Majority (34% of the target population):** Individuals who are far more risk averse than early adopters, the early majority will adopt an innovation before the average person does, but the early majority of individuals need evidence that the innovation works and presents a clear benefit. This can often be achieved through interaction with the early adopters.
4. **Late Majority (34% of the target population):** Individuals who look to peers and will adopt an innovation only after the majority (50%) does and social norms favor the innovation. Otherwise, this group is more skeptical of innovation.
5. **Laggards (16%)**: Individuals who are more socially isolated than other adopter categories will be the last to adopt, often as innovation becomes obsolete. Laggards are usually more fixated on the past, preferring to keep hold of traditions and defaults.

(Rogers, 2003)

This idea of categories of adopters appeals to many. The idea has an individual relatability as we consider our adoption of innovations such as

smartphones. The link between these categories of adopters and my IFS (King 2020) drew me to DOI. Observing common personality traits for bright spot schools and aligning with descriptions of innovators and early adopters looked like an excellent fit. However, while many researchers in international education identify “bright spots,” they are sent to focus on understanding their actionable characteristics (Gates Foundation, n.d.-b) for scaling. DOI employs a different framework.

Rogers (2003) states that the personality characteristics of adopters are a scale. The five categories above are created out of convenience for planning purposes, much like socioeconomic quintiles (Filmer and Pritchett 2001). The categories of adopters tend to have generalized characteristics on this scale (Exhibit 5). There are three categories of individual characteristics; socioemotional, personality, and communication behavior.

Exhibit 5: Characteristics of Innovation Adopters

Those from earlier adopter categories tend to have more significant (or more) of the following characteristics:

Socioemotional	**Personality**	**Communication Behavior**
years of formal education	empathy	social participation
socioeconomic status	ability to deal with abstraction	connections through interpersonal networks
degree of social mobility	rationality	likely to be cosmopolite (widely traveled)
	intelligence	
	ability to cope with risk or uncertainty	

(Rogers, 2003)

As the exhibit shows, generally, the earlier the adopter is associated with many more favorable characteristics. Rogers also describes how early adopters tend to be more willing to ignore social norms of behavior (Rogers, 2003), being more likely to implement innovation while those in their social group do not.

Those who most need the benefits of the innovation the most are least likely to adopt it. Additionally, early adopters tend to adopt the innovation over a

much shorter time than later adopters. This highlights the “Innovation Paradox” (Rogers, 2003).

2.1.2 Perception of the Innovation

A critical aspect of DOI theory is how potential adopters perceive the innovation. There are five perceived characteristics of innovations (Rogers, 2003, p.15):

1. **Relative advantage** – the extent to which the innovation is *perceived* as better than the default idea it is intended to replace,
2. **Compatibility** – the extent to which the innovation is perceived as familiar and consistent with existing values and needs of the potential adopter,
3. **Complexity** – the extent to which the individual perceives the innovation as challenging to understand and implement,
4. **Trialability** – the extent to which an innovation may be experimented with,
5. **Observability** – the extent to which the results or benefits of an innovation are **visible**.

Others have expanded on this list of characteristics. Moore and Benbasat (1991) identified the additional characteristics:

- **Image** – the extent to which the innovation is perceived to enhance an individual’s social status (often viewed as an important sub-category of **relative advantage**),
- **voluntariness of use** – the degree to which adoption of the innovation is mandatory

Our application is focused on teacher adoption of innovations within an education system where accountability (perceived or actual) is potentially salient. The latter perceived characteristic seems of particular relevance for an educational system. For example, the main factors determining the adoption of a specific set of ICT skills for teacher trainer’s Cambodian public schools was if the teachers perceived the adoption to be easy (i.e. complexity) and mandatory (i.e. voluntariness of use) (Richardson 2011a). As we will

discuss later in the literature review on behavioral economics, much of DOI is focused on adoption by civil society where individuals are free to choose (Thaler and Sunstein 2009).

Within education research, how the teachers perceive is the most researched. Researchers assess teacher implementation and perceptions of an innovation that has been introduced. The association between implementation and perception of the innovation is then analyzed using Pearson's correlation, and variance is explained using the Coefficient of Determination (r^2) (Hughes and Keith 1980).

Hughes and Keith (1980) found that four out of the original five innovations described by Rogers (2003) were highly correlated with implementation scores:

Exhibit 6: Correlations of Perception of Innovation Ratings with Implementation Scores (n=30)

Attribute	**Pearson's *r***	**p-value**	**Increase in r^2**
Relative Advantage	0.33	<0.05	22.1%
Compatibility	0.41	<0.025	6.7%
Complexity	-0.05	Non-significant	9.5%
Trialability	0.31	<0.05	1.4%
Observability	0.46	<0.01	0.4%

(Hughes and Keith, 1980, Table 3, p.48 and Table 5, p.49)

The variance explained (increase in r^2) is highest for relative advantage (22.1%), and the total variance explained for the five characteristics is 40.1%. While this demonstrates support for the attributes suggested by Rogers (2003), the nature of the relationship between the perceptions of the innovation and its implementation is uncertain (Hughes and Keith, 1980). Additionally, most of the variance in implementation is yet to be explained.

Richards (2009) added the following three characteristics to a framework to understand the factors that explained Cambodian teacher trainers' adoption of an ICT skill set:

- ***Image*** – the extent to which using an innovation enhances an individual's reputation with their peers,
- ***Voluntariness*** – the extent to which the innovation is voluntary
- ***Visibility*** – the extent to which an innovation is visible (e.g., awareness, publicized, etc.)

After controlling for socioeconomic differences, Richards (2009) concluded that the two key perceived characteristics of the innovation that influenced the adoption were voluntariness (i.e., if it was believed that adoption of the innovation was mandatory) and if the innovation was seen as being easy to use.

2.1.3 The Innovation-Decision Process

DOI theory describes the decision to adopt innovation as a process that takes place over time. While every process is unique to the individual, there are five main stages to the innovation-decision process:

1. **Knowledge**

The complexity of the innovation is essential. The individual should have a minimum level of knowledge of the innovation before the trial and adoption of the innovation. Otherwise, the innovation's chance of rejection and discontinuance is significantly increased (Rogers, 2003).

2. **Persuasion**

DOI theory's definition of persuasion is focused on attitudinal formation and change of the individual recipient of the innovation.

The individual

a. decides what aspects of the innovation are credible
b. becomes more emotionally involved with the innovation
c. seeks approval from their peers of their opinion towards and interaction with the innovation
d. can use trialing of the innovation by peers as a substitution for their trial of the innovation

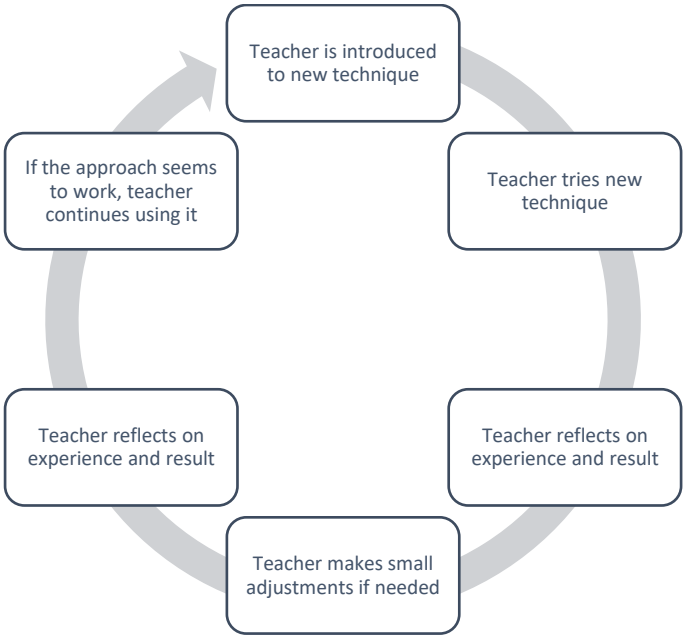
3. **Decision –** the individual decides to use the innovation or reject it. This decision is on a scale from full rejection to full adoption.
4. **Implementation** – forms of implementation include re-adoption, partial adoption (selected components of the innovation), or incorrect implementation.
5. **Confirmation** – the individual looks for reinforcement of their innovation decision.

(Rogers, 2003)

The items above should not be reviewed as steps an individual must complete, nor should they be seen as steps the individual must experience to adopt an innovation. The innovation-decision process is the least researched and applied component of DOI, which often leads to the failure of innovations (Schmidt and Brown 2007). However, this should not be surprising as other components of DOI, such perceptions of innovation and characteristics of implementors, can be more easily assessed at a fixed point in time.

Models of teacher support and adoption have similarities, yet distinct differences between themselves and innovation-decision. For example, RTI's work for the Gates Foundation adapts Guskey's (T. R. Guskey 1986) process of teacher change with the following steps in a learning cycle:

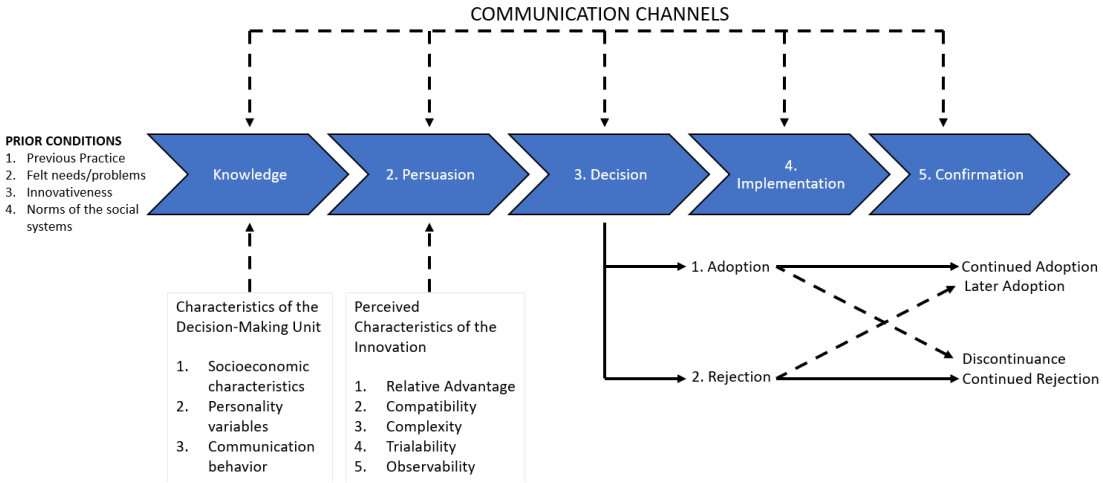
Exhibit 7: Teaching Learning Cycle



(Gates Foundation 2022, Figure 1)

The foundation of DOI is lateral diffusion across social groups of the intended population. How does this differ from DOI? Firstly, the figure presents the teacher's process as an individual. Next, the final step of the repeating cycle suggests that if the approach seems to work, the teacher continues to use it. Rogers (2003) discusses the final **confirmation** stage of the innovation-decision process; post-implementation adoption or rejection is not the final stage of adoption, and that individual seeks reinforcement for the innovation-decision already made, and *". . . may reverse this decision if exposed to conflicting messages about the innovation*" (Rogers, 2003, p.189). The Guskey model describes a continuous loop of teacher reflective practice, adding layers of understanding and competence while implementing the new practice. However, Rogers describes innovation adoption as a far more complex process:

Exhibit 8: A Model of the Five Stages in the Innovation-Decision Process



(Rogers, 2003, p.170)

Guskey's model depends on the meta-cognitive abilities of the teacher, being a logical and reflective practitioner. However, any adoption process relies on the teacher's motivation to be meta-cognitive. Consequently, the DOI innovation-decision process has initial conditions, perceived characteristics of the innovation, and characteristics of the teacher nested within the overall model seems a more realistic landscape, such as negative social normative conditions. The other fundamental difference between DOI and other aspects

presented in this framework (such as Michael Fuller's work) is that it is likely that individual "bright spot" teachers could well act and behave as Guskey describes. Still, later adopters need very different conditions and interactions with their peers. For example, Fuller places a key emphasis on individual and shared goals. Consequently, teacher peer interaction is a necessity on an adoption cycle.

2.1.4 Communication Channels

Communication channels are the final key DOI component discussed. Rogers writes about communication-related to innovation adoption, *". . . most people depend mainly upon a subjective evaluation of an innovation that is conveyed to them from other individuals like themselves who have already adopted the innovation . . . diffusion is a very social process that involves interpersonal communication relationships*." (Rogers, 2003, p.19). Most planned communication within centralized-decentralized education systems where EGR programs occur from a central-level and is "pushed" down through regional education offices and onto schools. In theory, this approach has no issue; it provides a clear understanding of what this type of communication can and cannot achieve. The innovation-decision process begins with knowledge of the innovation. A top-down approach that involves (for example) scheduled teacher training can achieve this step. However, unless there is a true authoritarian mandate that the innovation is to be implemented, top-down communication alone cannot achieve the next four steps of innovation-decision. Rogers (2003) defines two types of communication:

Homophilous Communication – the degree to which two or more individuals interact are similar in specific attributes, such as beliefs, education, socioeconomic status, etc.

Heterophilous Communication – the degree to which two or more individuals interact differently in specific attributes.

Most effective communication for diffusion occurs when two or more individuals are homophilous. When they share common meanings and subcultural language and are alike in personal and social characteristics, the communication of new ideas is likely to have greater effects regarding

knowledge gains, attitude formation and change, and overt behavior change (Rogers, 2003).

However, early adopters are more amenable to cosmopolite communication, defined as *" . . . communication linking an individual with sources outside their social system*" (Rogers, 2003, p.207). Consequently, centralized ministry communication, mass media, and other cosmopolite communication work for early adopters but not late adopters. Centralized training is also a means of cosmopolite communication. This seems to match my IFS bright spot research, where the early implementors were happy to interact with the education innovation at an early stage, not necessarily to link or to need their heterophyllous local network for support. However, lateral networking and communication are generally ad hoc or unstructured within LMIC education systems. Consequently, there is a struggle to implement a heterophyllous communication system, which is a key aspect of DOI theory. Later adopter categories have more dependency on localite communications for implementation. This creates a conjecture that using the current EGR program implementation approach will continue to limit the impact to the early adopters as programs do not consider the characteristics and subsequent needs of later adoption groups.

Who provides communication is also critical. If we focus on the diffusion stage, where early adopters are implementing, and attention turns to the early majority, we have:

- **Early adopters** are homogenous peers who have started implementing the innovation, adapting to their context.
- ***Opinion leaders*** are members of the social system in which they exert influence.
- ***Change agents*** influence innovation decisions in a direction deemed desirable by a change agency. They are usually heterophyllous from their typical clients, thus posing problems for effective communication about innovations they are promoting.
- ***Opinion leadership*** is the degree to which an individual can informally influence other individuals' attitudes or overt behavior in a desired way with relative frequency. This informal leadership is not a function of the

individual's formal position or status within the system. An opinion leader's interpersonal networks allow them to serve as a social model whose innovative behavior is imitated by many other system members. The respect with which the opinion leader is held can be lost if they deviate too far from the system's norms. Opinion leaders can be "worn out" by change agents who overuse them in diffusion activities. Opinion leaders may begin to be perceived by their peers as too much like professional change agents and therefore lose their credibility with their former followers.

Commonly, in EGR programs, the lateral stage of communication is led by change agents in the form of head teachers or district-level coaches who observe and support teachers in schools. These individuals support teachers through the centralized-decentralized education system. However, they are effectively implementing a centrally mandated innovation, and it is uncertain if these individuals are viewed as "authentic" heterophyllous voices. Rogers (2003) views change agents more ideally as facilitators of lateral communication rather than persuading individuals to adopt.

It is far from certain that coaches and head teachers are viewed as pure change agents, and their role should be limited to facilitation rather than direct influence. The World Bank touches on this with their "coach" documentation, suggesting that pedagogical leaders of teachers are not also responsible for evaluating teachers (World Bank, n.d.). However, in many countries, coaches and head teachers evaluate teachers' class instruction while providing instructional support.

Another approach tried by the USAID Kenya Tusome EGR Program in Kenya (USAID 2021c) at this implementation stage was to identify early adopter "bright spot" teachers and arrange for them to discuss implementation and adaption in the same context as the later adopting teachers.

Opinion leaders are the actors that Rogers (2003) identified as the most important actors for lateral diffusions. Rogers (2003, p.319) comments, "*The interpersonal relationships between opinion leaders and followers hang in a delicate balance. If an opinion leader becomes too innovative or adopts a new idea too quickly, followers may begin to doubt their judgment. One role of the*

opinion leader in the social system is to help reduce uncertainty about an innovation . . . To fulfill this role, an opinion leader must demonstrate prudent judgment decisions about adopting new ideas.". However, because social norms influence an opinion leader, they can also negatively influence others if the social norms do not favor change.

This is a challenging adoption stage, as no actors are a perfect fit to lead lateral diffusion. As my IFS demonstrated, it is a hill that EGR programming is yet to climb. Finally, a teacher's social network is critical to supporting their improving classroom practices. An effective teacher network has characteristics such as homophily, reciprocity, and transitivity (Kolleck et al., 2021). However, in many LMICs with large rural areas, schools can often be isolated, with poor transportation and small numbers of teachers in each school. This is very different from domestic education in USA or UK, where schools are larger, located nearby, and transportation is good. This facilitates lateral communication across schools. So, in LMICs, when we consider DOI in an education system, how exactly does a teacher effectively network with peers? Suppose a teacher is in a small, rural school where none of the teachers are implementing the innovation. How should an education system be managed and organized to provide lateral networking that might be critical to support their implementation of the innovation? This seems to be a challenge.

2.1.5 Criticism of Diffusion Research

Pro-Innovation Bias is the implication in diffusion research that innovation should be diffused more rapidly and that the innovation should be neither re-invented nor rejected. The bias leads researchers to ignore issues with innovation design, to underemphasize the rejection or discontinuance of innovations, and to overlook re-invention. Most diffusion research is funded by change agencies, which often have a pro-innovation bias. (E. D. Glor 2003)

It is far from guaranteed that these stages of adoption will occur. Innovations are adopted slowly until the early majority stage and then "snowballs." However, this adoption stage is unlikely unless the management of a planned

innovation considers and responds to the personality characteristics and needs of the early majority (Rogers, 2003).

As stated earlier, champions of innovations can easily ignore the inequality of DOI (Rogers, 2003; Glor, 2003), where opportunity gaps between those who benefit from the innovation and those who do not increase. However, it is reasonable to state that this issue was been better addressed in recent years with more focus on poor rural areas and women.

2.1.6 Contextualizing DOI - Summary

The key aspect of DOI is that it acknowledges different “groups” of stakeholders that require different levels of support to adopt the reading program and that this adoption is a process over time. This alone is a useful reframing of program evaluation which tends to generalize behaviors and actions. However, the science of statistics is to describe variability, and generalization is just one tool available. If the population of interest does have distinct sub-groups with differentiated characteristics, then generalization is most likely not a good or final analysis approach.

Impact evaluation data from multiple EGR programs and this author’s IFS study suggest that the early implementors are social leaders willing to risk adoption with the understanding and willingness to adapt the innovation to the local environment. This common attribute is a different perspective from most prior research, which focused on ***what*** these were doing, rather than a single motivating factor for the action. These actions could be implemented by a headteacher, teacher, or even the whole school. Subsequently, the idea that we need to replicate these bright spots elsewhere is probably misleading because the common characteristic of the bright spots is not how they are implemented but rather the common characteristics that motivate them. The next group of implementors, the early majority, are much more risk-averse and want to know that the innovation will be successful and relatively straightforward to implement. The characteristics of this group demand a different understanding and approach if they are to implement the innovation.

An expected resistance to using DOI as part of this framework is that DOI mostly applies to modern social media (e.g., Instagram “influencers”) and

adoption of modern technology innovations such as hardware (e.g., smartphone, smartwatch, etc.) or software or apps (i.e., Facebook, Twitter, Google Chrome browser, etc.). However, this would be misplaced as modern technology used DOI principles to diffuse the use of modern technology, and the idea of DOI has existed since the mid-20th Century.

It is often a mistaken assumption that adoption of the innovation by different categories of implementors usually eventually occurs, but evidence from other industries suggests that they typically do not. Like EGR programming, they generally struggle to diffuse beyond early adopters (Rogers, 2003). Consequently, innovation fails because it is not being purchased or used by enough of the targeted population. In the case of education in international development in education, the donor funding for the reading program finishes, and the program fades as the impact is usually modest, without being game-changing. (Piper and Stern 2019).

2.2 Behavioral Economics

Behavioral Economics (BE), which *"… incorporates the study of psychology into the analysis of the decision-making behind an economic outcome*" ("What is behavioral economics?," 2017), has gained recent popularity through the book "Nudge" by Richard Thaler and Cass Sunstein (2009). The authors present evidence that humans make decisions in irrational yet predictable ways and that by applying this understanding combined with a better understanding of how the environment shapes teachers' decisions, we can help teachers make more positive and beneficial decisions. Thaler and Sunstein's work bridges the earlier research by psychologist Daniel Kahneman (2011) and his seminal book, "Thinking, Fast and Slow" where he suggested that humans constantly employ two modes of thought:

- **Emotional System 1** - which " *. . . operates automatically and quickly, with little or no effort and no sense of voluntary control*" (Kahneman, 2011, p. 20)
- **Rational System 2** – which *". . . allocates attention to effortful mental activities that demand it, including complex computations.*" (Kahneman, 2011, p. 20)

Illustratively, Kahneman suggests that System 2 thinking is quickly exhausted from the mental effort. Consequently, System 1 often takes over when a decision is required, which wants to make decisions quickly with the short-term benefit of paramount consideration. Jonathan Heidt (2006) created an excellent analogy for how these two systems of thought interact; our emotional side as an elephant and its rider's rational side. Using this analogy, Chip and Dan Heath (2011) described the rider as responsible for the direction. However, whenever there is a disagreement between the elephant and the rider, the elephant wins and decides the direction as the rational rider tires quickly, trying to motivate the elephant. The elephant prefers short-term gratification over long-term planning. Kahneman (2011) describes this process as evolutionary that applies to all humans, and how we make decisions are shaped by our environment, which Heidt's analogy refers to as the path (Heidt, 2006). To support meaningful change, we need to direct the rider, motivate the elephant, and shape the path (Health and Heath, 2011):

- Direct the rider – is the solution to the problem clear and straightforward?
- Motivate the elephant – have we engaged the emotional aspect of decision-making?
- Clear the path – what environmental situations need to change to support positive decision-making?

Kahneman (2011) suggests that to ***nudge good decision-making, we focus more on diminishing the restraining forces, not increasing the driving forces***. This reinforces the counter-intuitive nature of behavioral economics. Kahneman credits much of his theoretical thinking to Kurt Lewin (1890-1947), a German-American psychologist who developed Field Theory (Lewin 1997) which suggests that to achieve a behavior change; there is a good way and bad way to do it. The bad way is to increase the driving forces, such as incentives, while the good way is to diminish the restraining forces. This approach, according to Kahneman, is "profoundly non-intuitive." However, the instinct of development work is often the three-step knowledge-attitude-behavior approach to change. In other words, we think that people have to get people to think differently first if we want to change their behavior. Kahneman

and Lewin suggest that if we want to achieve a behavior change, we must identify and reduce environmental barriers.

2.2.1 The Importance of Applying Behavioral Economics in an Education Setting

Fundamentally, BE can be divided into two considerations: how humans make decisions and the environment that influences these decisions.

BE challenges conventionality about how we think about participants in typical economic models and how the notion that individuals " *. . . think and choose unfailingly well*" (Thaler and Sunstein, 2009, p. 6). Although an education researcher responsible for designing a reading program would most likely resist the mere suggestion that they would develop a reading program with the same parameters an economist would use, I would argue that most reading programs are designed around a term I will refer to as "*knowledge transfer*"; programs are designed to train teachers to implement the program with the new teaching resources provided, and coaching to support the teacher in implementing the program correctly. While the education researcher would maintain that the program design considers local conditions, most programs are still designed with the idea that teachers will be rational decision-makers and adopters when presented with a superior pedagogy and didactic approach.

2.2.2 Positive Deviants

As mentioned earlier, from my IFS, I estimated that the EGR program impact is explained by a relatively small percentage (15%-45%) of schools (King 2020). We call this small percentage of implementing teachers "*positive deviants*." They can resist the prevalent social norms and implement the reading program to varying degrees. Heath and Heath (2011) called these individuals or groups of individuals "*bright spots*" and demonstrated instances where individuals provided a solution to a problem. This paper critiques expected utility theory as a descriptive decision-making model under risk and develops an alternative model called prospect theory. Choices among risky prospects exhibit several pervasive effects inconsistent with the basic tenets of utility theory. In particular, people's underweight outcomes are merely

probable compared to outcomes obtained with certainty. This tendency, called the certainty effect, contributes to risk aversion in choices involving sure gains and to risk seeking in choices involving certain losses. In addition, people generally discard components shared by all prospects under consideration. This tendency, called the isolation effect, leads to inconsistent preferences when the same choice is presented in different forms. An alternative choice theory is developed, assigning value to gains and losses rather than to final assets and in which decision weights replace probabilities. The value function is normally concave for gains, commonly convex for losses, and is generally steeper for losses than for gains. Decision weights are usually lower than the corresponding probabilities, except in the range of low probabilities. Overweighting of low probabilities may contribute to the attractiveness of both insurance and gambling (Kahneman and Tversky 1979) (2011) use this as an example of a *locally* designed and implemented solution to a problem, such as a village in Vietnam solving a nutritional problem for its children; for EGR programs it points to instances where a school has taken ownership of a reading program and found ways to support teachers in its implementation. This is often accomplished by adapting the program and different school community members (administration, teachers, community members) working together that was not part of the program design. Consequently, it becomes important to learn about and understand the strategies of these schools that can successfully implement the reading programs to support the research into how social norms that challenge implementation can be mitigated.

2.2.3 Application of BE in an Education setting

The most common application of BE in education is understanding and nudging pupil behavior in and out of the classroom (Koch, Nafziger, and Nielsen 2015; Levitt et al. 2012). However, there has been little traction with BE in education regarding teachers, social norms, beliefs, and classroom instructional approaches (Jabbar, 2011; Levitt et al., 2012; Stevano, 2019), which is surprising because much of the resistance to applying economics to education problems is that it defines its models too narrowly, ignoring psychology and human behavior (Jabbar, 2011; Levitt et al., 2012; Maxwell, 2012). Heshmat (2017) commented that economic models presume that

individuals will always make the most rational choice upon considering the cost and benefits of their impending decision. In contrast, BE provides insight into how individuals make decisions, not always behaving in their best interests.

One key challenge with applying BE to education systems is that BE is mostly focused on decisions made by individuals in civil society, mainly regarding individual finance (such as pension contributions) and personal health (eating habits and exercise). Thaler and Sunstein (2009) proposed that it be applied using *Libertarian Paternalism* (LP); individuals are free to choose. However, they will be nudged to make good decisions through programming design. LP might not seem a natural fit in an education system where policy is imposed. However, in a low-income country where education systems struggle with accountability and supporting teachers with education reform, LP might be a necessary workaround to nudge teachers into implementing.

Part of the challenge to be addressed with this proposed research will be identifying the causes of teachers' behaviors in low-income countries that resist implementing an early-grade reading program. There are many different behavioral models to consider. Two behavioral models will be researching more due to their possibility of fitting with explanations of why teachers resist implementing EGR programs:

***Habit and preference for the status quo*.** People prefer the familiar option, even when a better alternative is available due to a combination of factors such as loss aversion, transitional costs, social norms, and high stakes (Kahneman and Tversky 1979; Li, Liu, and Liu 2016; Nicolle et al. 2011; Suri et al. 2013). This could be a key consideration for EGR programs when we ask teachers to change their instructional approach and have less autonomy in their classroom instruction.

Framing effects. Choices can be presented to highlight the positive or negative aspects of the decision, thus changing the appeal of certain decisions (Jabbar, 2011). The classic example is presenting a frozen yogurt product to consumers as 80% fat-free instead of 20% fat. In EGR programs, new ideas are presented to teachers with the general presumption that

rational decision-making will occur, and teachers will understand the program's benefits and implement it.

Cognitive Dissonance – The idea that individuals get a feeling of discomfort when their actions or ideas do not align with other beliefs, ideas, or values they hold (Heaton and Quan 2023). This theory has been studied to observe and sometimes to shift teacher behavior towards pupils with disabilities (Emmers, Baeyens, and Petry 2021), pupil diversity (McFalls and Cobb-Roberts 2001), and strategies to counter resistance from supervisor feedback for pre-service pupil teachers in Nigeria (Oyetoro, Adesina, and Eyebiokin 2020).

Peer Influence**.** The idea of convergent behavior seems useful for my study through the perspective of understanding social norms and influences on the group and individual behavior, and also how. This definition not only describes patterns of general behaviors in schools (such as poor teacher attendance) but also suggests a potential place where the solution lies. There are a few different theories of peer influence. The first is **herding**, which is defined as " *. . . .the alignment of the thoughts or behaviors of individuals in a group (herd) through local interaction and without centralized coordination*" (Raafat, Chater, and Frith 2009). Evans (2020) goes further and describes individuals subjugating their own will and thoughts to the behaviors of the majority. Frank (2020) argues that peer influence is often underrated and ignored because of the subconscious aspect of peer influence. We all like to think of ourselves as unique individuals who make out own decisions. Consequently, peer influence on behavior is often seen as unfavorable and as succumbing to pressure to be the same as those around us (Frank 2020). However, most of our behaviors in society are very conforming. Do you queue patiently? Do you wear clothes that, while not necessarily up-to-date fashion, will not embarrass you in public? The likelihood is that most of us are appalled when an individual tries to queue jump, and we have a wardrobe of professional clothes that help us present ourselves in a way that reflects how we want to be perceived by others. In essence, that's why some of us look in the mirror once dressed; the reflection is how others see us.

2.2.4 Social Norms

"*Social norms are rules of behavior. They inform group members how to construe a given situation, how to feel about it, and how to behave in it.*" (Reese, Rosenmann, and Cameron 2019, p.80). As defined, social norms direct our thinking, actions, and behaviors. They create evaluative standards against which individual behaviors are judged (Reese, Rosenmann, and Cameron 2019). Social expectations under the umbrella of social norms theory may also provide insight into understanding the behaviors of teachers and other education system stakeholders. Social norms theory differentiates the perception of expectations of others:

- *"Empirical expectations: what we believe others do*
- *Normative Expectations: what we believe others to think we should do."*

(Bicchieri and Noah 2017, p.6)

Two example statements can illustrate this difference:

- Empirical expectation: "all the other teachers in my school do not want to implement the EGR program."
- Normative expectation: "other teachers in my school think I should not implement the EGR program."

Not to be confused with personal beliefs, conforming to these perceived expectations creates a sense of "*. . . belonging, closeness, and solidarity*" (Reese, Rosenmann, and Cameron 2019, p.73) with peers.

If the implementation is dependent on environmental factors such as social norms, what are reading programs doing about this? Besides *implicit* persuasion from teachers infrequently working with visiting coaches, no single *explicit* component was researched, designed, and applied to change teachers' behaviors and willingness to implement the program. While *implicit* persuasion is an aspect of reading programs' design, the low percentage of schools that account for program impact indicates that this approach, if by design or by accident, is not optimizing fidelity of implementation.

The challenge of applying behavioral economics is that it needs a framework for the context of education systems. Most behavioral economics successes have come from helping civil society make good decisions, such as saving for retirement and public health. BE has yet to be applied to education systems. It

seems reasonable that understanding irrational decision-making by humans helps us understand teacher decision-making for adopting a reading program; however, to think about a teacher's decision being a single point in time seems somewhat limiting.

2.2.5 The Parallel of Education Reform in England and Wales in the 1980s

The Education Reform Act of 1988 in England and Wales is an excellent example of the challenges of change imposed on an education system. Putting aside the opinion of the reform's impact, the implementation of Education Reform created distinct challenges and changes for the teaching profession. The implementation of a complex reform agenda without input from teachers *". . . created a lack of trust and substantial stress, frustration and upset amongst teachers, who were essentially put in a position of 'learning as one goes'*" (Airasian and Gregory 2006) in response to new mandates. A fundamental difficulty for teachers was shifting away from being autonomous professionals to being required to use an instructional approach generally not aligned with their teaching philosophy (Airasian and Gregory 2006). I recently interviewed a reading expert advisor for Primary schools in Norfolk, UK, who supported schools and teachers in implementing the 1980s reform. They commented:

"What became very apparent when the National Curriculum was introduced was that everyone would go through a process of change. I went to loads of meetings with and without speakers. Change is difficult. It was like jumping into the unknown, and suddenly everything you knew and did was wrong. It demoralized teachers, so they got defensive. Lots of teachers just closed their doors and carried on as usual. Lots retired early (good and bad). It was a traumatic time."

There is a close parallel to USAID reading programs; the England and Wale Education Reform and a USAID reading program required the teacher to shift from a subject-centered instructional approach to a child-centered one. This involves a fundamental change to their approach to teaching and loss of autonomy – drawing a line under their previous instructional approach as inferior. The response by many individuals during education reform in England

and Wales was emotional and very much aligned with Kahneman's Type I System thinking. Consequently, while the role of the education system from a central level was to communicate "logical" and "reasonable" changes to teaching and learning, it became the role of the school districts and schools to consider how to implement this reform. A big part of this change is through educational leadership supporting the emotional challenge that change presents to groups and individuals (Fullan 2006).

"Trouble was in the initial stages no one knew how the whole thing would work out in practice. It was like jumping into the unknown; suddenly, everything you knew and did was wrong. It demoralized teachers, so they got defensive. I was running Training days without really knowing what was going to happen. Every school had to find its way through, which took yonks."

The difference between the reform in England and Wales and an EGR program implemented in an LMIC setting comes down to the environment. While the education reform in England and Wales in the 1980s was jarring and unpleasant for many, over time, those who remained in the profession implemented the reform due to their peers' support and an education system that held them accountable. Contrast with a USAID reading program in a low-income country introduced over four or five years through local education systems that struggle to provide teacher support and systems accountability.

Pritchett and Honig (2020) argue that a top-down accountability approach limits the consideration of characteristics and behaviors of "*. . . what you want your child's teacher to be*". Alongside the accountability system, they argue for developing "*strong professional norms among teachers*." They do not comment on how to achieve this. However, BE suggests not adding programming components that train professional norms (e.g., social-emotional training for teachers). Instead, you "clear the path"; make changes to the environment to nudge positive decisions that create the desired behaviors and attitudes.

Sabarwal and Abu-Lawdeh (2018) researched teachers' mental models in low-income countries. They argued that teachers already believe that they are applying maximum effort due to their mental models and how they see

themselves in the environment they inhabit and so do not respond to accountability or incentives. They tested three barriers/themes to quality instruction: that teacher absenteeism is high, teacher accountability for learning is low, and teacher ownership of pupil learning may vary based on pupil ability.

Teacher absenteeism is an issue in many low-income countries (Sow, 2017), with classroom absenteeism averaging between 25% and 55%. Teachers generally justified their absenteeism through assigning work in their absence or completion of the curriculum (Sabarwal and Abu-Jawdeh, 2018), again reinforcing the behavioral norm that teachers see their responsibility as the delivery of curriculum and content, less so pupil learning. This mental model suggests a behavioral norm teachers do not consider their commitment to support pupils of low achievement and focus on pupils who already have the prerequisite literacy skills to build on, a behavior suggested by Oketch et al. (2020).

USAID reading programs impose a new approach to teaching early literacy skills, asking teachers to change their classroom instructional approach. Costa and Kallick (n.d.), in their blog, expressed this challenge well; "*A change in mental models, for most people, implies the unknown . . . people who are invested in their present ways of working believe that if they can just do what they are currently doing better, everything will improve.*".

2.3 The New Meaning of Change

International education, in recent years, has mostly focused on pedagogical practice, followed by a focus on addressing educational systems (The World Bank 2020). The work of Michael Fullan was recommended to me at my VIVA when defending my thesis. Upon researching Fullan's extensive work, I found no application of his work in international education, although Fullan himself has spent time assessing education systems and programs in LMICs. Colleagues involved in systems work who had worked in the education sector in the USA were aware of Fullan's work and generally categorized it as addressing motivational issues. This is probably why Fullan's work has not really been heavily adopted in international education.

Fullan has released many books and other publications, shifting and developing ideas over the years. Consequently, I will focus primarily on his work from 2015, particularly on his fifth edition of "*The New Meaning of Educational Change*."

Through his analysis of education systems throughout the world, Fullan writes a lot about "wrong policy drivers" (Fullan 2015, p.7), including "*punitive accountability, individualistic solutions, . . . and ad hoc policies*" (Fullan 2015, p.42). Fullan (2015) details individualistic solutions focusing on the individual teacher and leadership quality.

2.3.1 Focusing on the Individual within the System

Many local education systems, donors, and NGOs focus on 1:1 coaching and supporting individual teachers (Piper and Zuilkowski 2015; World Bank 2021). This type of coaching within a local education system typically uses a decentralized system to support teacher learning. The coach visits schools from district education centers or Teacher Advisory Centers (TACs) in Kenya and is responsible for visiting 8 – 30 schools (Piper and Zuilkowski 2015). The coach's role in Ghana goes to District Teacher Support Teams (DTST), trained to support schools in classroom instruction (Smiley et al. 2020). Similar models also exist in Malawi, Uganda, and Ethiopia education systems.

External coaching is often employed as a means to address a lack of in-service training provided to teachers in schools. For example, it was reported that greater than 80% of teachers in 72 Kenyan schools reported having no opportunity for in-service training for the previous eighteen months (Ngware, Oketch, and Mutisya 2014). However, while external teacher coaches were provided to address the lack of in-service teacher support and learning, the challenge of external coaching having too many schools to monitor and resources (such as transportation and money for fuel) presents a similar system problem regarding limited teacher support.

However, Fullan's focus is on the limitations of 1:1 teacher coaching support: "*The problem is that no nation has improved by focusing on individual teachers as the driver*" (Fullan, 2015, p.43). By focusing on individuals, Fullan (2015) argues, you fail to address the culture and relationships that exist

within the systems, and this cannot be achieved by focusing on individuals. This aligns with the behavioral economics argument by Kahneman (2011) of changing behaviors by focusing on reducing the restraining forces, not increasing the number of drivers. Kahneman comments on how this approach is not intuitive, and Fullan observes that supporting individual teachers seems "*rationally obvious*" (Fuller, 2015, p.43).

However, there is no denying the evidence that programs involving 1:1 external coaching have had an impact when implemented if a degree of fidelity (Piper and Zuilkowski 2015; Smiley et al. 2020). However, the empirical evidence I have presented points to the fact that implementation is mostly restricted to bright spots (Exhibit 3) and not diffused further than the initial impact (Exhibit 4). The lack of diffusion makes sense when systems focus on individuals to address a lack of in-service teacher support. This creates an interesting perspective on sustainable approaches to improving education. Word Bank (2020) *Smart Buys* describes A Good Buy as "*There is good evidence that the interventions in this category can be highly cost-effective across a variety of contexts*." (Word Bank, 2020, p.7). One good buy is "*Structured lesson plans with linked materials and ongoing teacher monitoring and training*" (World Bank, 2020, p.12). This describes evidence from evaluations from Kenya Tusome and other USAID EGR Programs. In context, the Word Bank (2020, p.12) report comments:
"*In low-capacity settings, this approach may not be too challenging politically because teachers welcome the chance to focus on classroom teaching; however, in other contexts, teachers may resist these programs as infringements on their professional autonomy, so it is important to get their buy-in first (for example, by sharing evidence of effectiveness and including teachers in the implementation process)*"
This interesting statement focuses on resistance due to teacher autonomy, which can be remedied with evidence of effectiveness. While I am unsure of the evidence behind these claims, it emphasizes the constant focus on the teacher as the individual, including their buy-in.

2.3.2 Components of Successful Educational Change

Fullan is diligent in researching evidence for successful change in education. His descriptions of what does not work alongside what works are of particular emphasis. As I will explain later in the research methodology, it is just as important to understand barriers and challenges to implementation alongside examples of success. Fullan lists right versus wrong drivers of change:

Exhibit 9: Right Versus, Wrong Drivers

Right	Wrong
Capacity building for results	External accountability
Collaborative work	Individual teacher and leadership quality
Pedagogy	Technology
Systemness	Fragmented strategies

(Fullan, 2015, p.42 Figure 3.1)

Fullan (2015, p.43) comments that the right drivers must "*. . . get at the motivation and competency development of the vast majority of educators*". He also mentions that the wrong drivers listed are not without merit. However, they do not lead to system change. The most obvious example is education systems acquiring technology for classroom use. As a former teacher who has been on the receiving end of many education technology innovations, it was always a case of solving a problem that had not been defined. Rather than pedagogy being the driver and leveraging available tools to solve the problem, the tool becomes the driver.

Consequently, Fullan criticizes external accountability and focuses on the individual as unable to achieve this aim. He describes how external accountability *". . . assumes that educators have the capacity to provide effective instruction*" (Fullan, 2015, p.43). Fullan has no objection to external accountability per se, but it cannot be a lead driver. However, more often than not, it is the central focus for reform, exampled by No Child Left Behind reform

in the USA (Peterson and West 2003) and recent reform efforts in Ghana with The Ghana Accountability for Learning Outcomes Project (GALOP) (Ghana Ministry of Education, n.d.). Fullan (2015) mentions that when quality managers and leaders are saddled with the responsibility of holding individuals accountable, but the education system is dysfunctional, there is no evidence that large-scale change and transformation occur. (Honig and Pritchett 2019) go as far as stating that the effectiveness of ICT monitoring of education systems to facilitate change "*is an illusion, a self-deception*" (Honig and Pritchett 2019) that usually leads to too much focus and energy being placed on process and data collection, rather than change and improvement.

Fullan argues the correct driver should focus on "*motivation and competency development*" (Fullan, 2015, p.43) of most teachers and other stakeholders.

The evidence presented for the correct drivers is evidence presented by Fullan. His work is comprehensive in presenting not only evidence of the correct drivers contributing to change but also he presents evidence and discusses where incorrect drivers have been used. One example is a study comparing the effectiveness of 100 successful schools against 100 less-successful elementary schools in Chicago (Bryk et al. 2010). The researchers conducting the study found successful schools focused on teachers' individual and collective capacity development, school climate (including safety), community involvement, and a clear mandate for pupil-centered learning with formative assessment. This also creates some similarity and overlap with Effective Schools research (Scheerens 1992), which has the following list of seven correlates for pupil success: instructional leadership, clear and focused mission, safe and orderly environment, climate of high expectations, frequent monitoring of pupil progress, positive home-school relations, and opportunity to learn and pupil time on task (Lezotte 1991). There are many different directions this research can take when looking at Fullan's work on elements of successful change such as instructional leadership, community relations, and school missions. Collaboration and collective responsibility, while not one of the correlates, are key operational objectives for the Effective Schools movement (Glaze 2014; Scheerens 1992). However, circling back to the

alignment of my research with USAID EGR Program design, the teacher collaboration driver is the gap I will continue to focus on.

2.3.3 Internal Accountability

"*If you want to reach a goal faster, you must invest in capacity building and use the group to get there*" (Fullan, 2015, p.44).

Internal accountability within schools is an often-used term. Levine (2005) gives an example of internal accountability of a teacher not wanting to disappoint their pupils, peers, head teacher, or parents. Elmore (1990) defines three theories of accountability:

- Technical accountability – where ministries, districts, and schools make decisions based on evidence of measurements,
- client perspective – exampled by school choice (although not an option in most LMICs),
- professional perspective – achieved through feedback through performance aligned with professional development.

Scheerens (2004) suggests that only the first two definitions of accountability, arguing that accountability requires an element of sanction or rewards. In this third theory of accountability, Fullan (2015) identifies this as social capital, which is defined within three dimensions: "*. . . interconnected networks of relationships between individuals and groups, levels of trust that characterize these ties, and resources or benefits that are both gained and transferred by virtue of social ties and social participation*." (Poteyeva 2018, p.9).

The alignment theoretical perspective with BE and DOE is very relevant, presenting different perspectives to understand the importance of peer interaction. BE presents the idea that as individuals, we make decisions through Type 1 (fast, intuitive) and Type 2 (slow, calculating) thinking (Kahneman 2011), and so to engage Type 2 thinking, we need to change the environment. From a BE perspective, this could be explained through collaborative problem-solving (Thaler and Sunstein 2009) and social norms theory (that an individual responds to what we perceive others will do and what we perceive others think we should do (Bicchieri and Noah 2017)). For

DOI theory, the interaction between intended recipients of an intervention is critical for the innovation to be adopted. However, the key to DOI is the interaction between the early adopters and the majority (Rogers 2003). However, DOI goes as far as to say that opinion leaders can negatively influence the uptake of an innovation (Rogers 2003). Therefore, DOI adds a perspective of the construction of the teacher learning groups. This emphasizes the challenge faced in many LMICs. It is often challenging for teachers in small, rural schools to network and make connections early adopters within the district. In a K-8 rural school, there is typically only one teacher per grade (or sometimes even multigrade classes) and frequently only grades 1-4 teachers are even trained on the EGR program approach. Consequently, if a school does not have *bright spot* catalyst, can it change the school learning culture?

Lant Pritchett adds an interesting reflection on "Isomorphic Mimicry" (Lantt Pritchett 2011) of education systems in LMICS, especially when donors support them to fast-track success:

". . . such strategies produce administrative systems in developing countries that look like those of modern states but that do not (indeed, cannot) perform like them; reforms yield metrics that satisfy narrow bureaucratic scorecards in donor capitals (and thus enable funds to continue to flow and legitimacy to be sustained), but that mask a clear inability to actually implement incrementally more complex and contentious tasks." (Andrews et al., 2017, p.4)*.*

Many EGR programs have components of successful education systems: teachers are trained, receive ongoing support and learning, and many have systems for accountability and improving implementation fidelity (Bruns 2018). However, if the impact of EGR programs is low (Piper and Stern 2019), what is the disconnect? Pritchett (2011) suggests that stakeholders within an education system make daily decisions about compliance with rules and results-driven actions. Compliance with rules is the easy and less risky path. Results-driven actions are more time intensive and have a higher component of risk. Critically, this lens also applies to much of the donor-supported approach. The parallel to the study framework is that Pritchett (2011) sees the

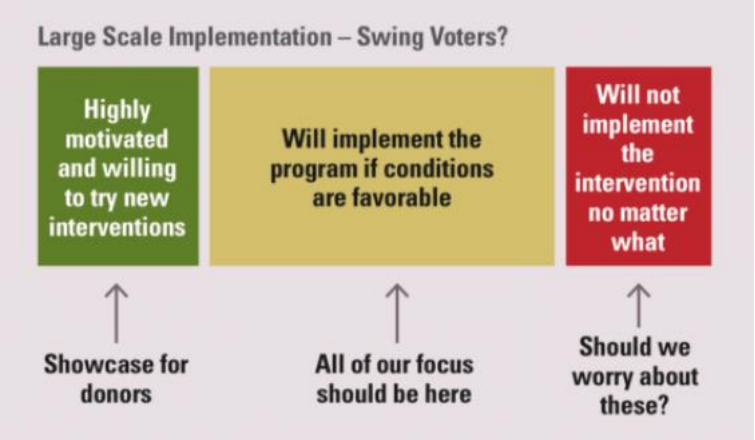
solution as a focus on diffusion, positive deviance, and a focus on results through locally driven problem solving.

2.4 Institutional Gap in Research

My research needs to be useful for my institute, RTI International. However, BE tends to push against traditional economic theory, using a different model of human behavior and taking a new and novel "nudge" approach to solving problems (Thaler and Sunstein 2009; Ariely 2009). Therefore, if my research is to be useful and be used, it needs to complement research already being conducted at my institute. Consequently, the genesis of my research was what I believed was a gap in my institute's understanding of and then being able to change the normative environment of schools and education systems where EGR programming is implemented.

However, my approach will be that current designs of EGR programming are *incomplete* rather than *incorrect*. For example, my institute's research for the Gates Foundation mentions similar groups of teachers implementing EGR programs:

Exhibit 10: The "Swing" Teacher Model



(Gates Foundation, n.d.)

The blog post describes the middle teacher group as not against the program and would implement it if the program was simple and easy to implement (Gates Foundation, n.d.-a). It finishes by saying, "*Supervisors reinforce the need to implement*" (Gates Foundation, n.d.-a). Finally, the blog states that teachers will implement if "*The program is simple, it reduces the amount of*

time it takes teachers to prepare lessons, teachers can identify impact on learning within a few weeks, teachers have the skills to implement the program, Supervisors reinforce the need to implement." (Gates Foundation, n.d.-a).

There are lots of good things with this approach. It aligns with the suggestion of Thaler and Sunstein (2009) that one way to encourage good decision-making is to keep it simple. Next, aligned with DOI, it identifies implementation groups and the teacher's perception of the intervention as critical in the decision to implement. However, the gap in this research is the final step, which suggests the solution lies in supervisors working with this middle group to reinforce that implementation is expected. However, the Gate research misses describing the difference between the implementing group and the next group. The implementing group, akin to early implementors, are more social leaders and unlikely to follow social norms. The next implementing group is much more subject to social norms and sees the majority around them not implementing. Consequently, the approach will likely not impact getting a larger group of teachers to implement. This is supported by Exhibit 4, showing that after the initial midterm impact, we have no evidence of programs getting more teachers to implement and improve the overall impact on learning outcomes.

Sometimes keeping it simple is not enough if the normative environment does not encourage a behavior change. A good example would be my experience as a volunteer teacher in a rural school in Zambia when I asked the headteacher to mandate teachers to come to class in the year 2000. Teacher attendance improved for about three days but quickly returned to the status quo. These teachers lived on the school campus and were no more than a 400-meter walk from their classrooms. Consequently, having supervisors reinforce the need to implement does not address the shift in the social norms needed to change behaviors. That is not to say that supervisors setting expectations is not critical, but implementing a new activity will ultimately fail unless you can change the environment. As Kahneman (2009) puts it, we need to work on reducing restraining forces, not adding new driving

components that will be ineffective in the environment where they are being applied.

Unlike the Gates research, there should be a new group of teachers. The early majority would benefit from interaction with the early implementors for the innovation to diffuse (Rogers, 2003). The Gates research tends to focus on the teacher as an individual. Fullan (2015) discusses the importance of individual and shared values and learning.

The early research in BE focused on using psychology with economics to understand human behavior better. However, understanding the irrationality of human behavior within economics is far from new. John Maynard Keynes (1936) acknowledged human emotions in the form of “animal spirits” driving financial decisions. Consequently, I do not see standard economic theory as incomplete.

However, over time BE has developed to include sociology which includes concepts like the herding of human behavior – how we are influenced by what we believe others will do and our beliefs about what others think we should do (Frank 2020).

2.5 Alignment of Theories and Frameworks

As discussed earlier, my theoretical framework incorporates frameworks and theories from DOI (Rogers, 2003), BE (Kahneman, 2011; Thaler and Sunstein, 2009), and *The New Meaning of Change* (Fullan, 2015).

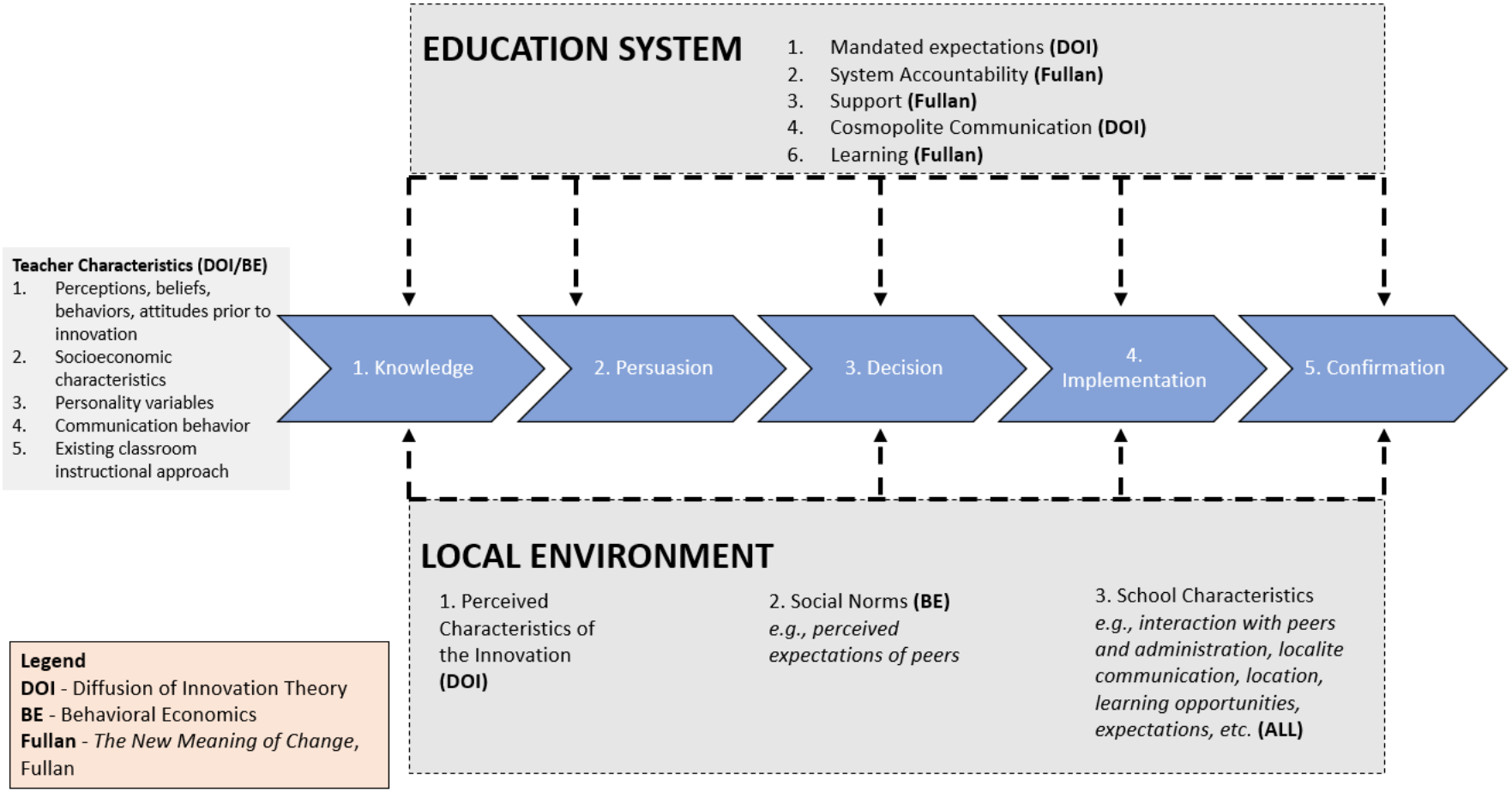
The foundation of my framework is DOI Theory, specifically Roger’s (2003, p.170) “A Model of the Five Stages in the Innovation-Decision Process” (see Exhibit 8). Repeating Roger’s definition of DOI theory (2003, p.5), *“. . . the process in which an innovation is communicated through certain channels over time among the members of a social system*” - DOI is the only theory or framework that is specific about how time and communication plays a factor in the implementation of innovations. Rogers (2003) expects early implementors to be more rational than later adopters but makes it clear that the groups' response should be expected without blame on the individual. This dovetails excellently with BE, where Dan Ariely (2009) says that irrationality of decision-

making should be expected, predictable, and can be accounted for in the design.

Finally, using just BE and DOI to develop a framework ignores the fact that we are implementing it within an education system. Many EGR Programs have similar components to education systems described by Fullan (2009);. Fullan (2015) focuses on the motivational issues for implementation once the components of the educational change are fixed. His approach to contrasting education changes programs with similar designs but often very different results on learning outcomes. However, the country context varies considerably, and Fullan's goals of improved implementation during educational change are the same as in this study. Finally, Fullan explores implementation through interactions of stakeholders within and across schools. This outcome is generally missing or has a low focus on EGR programs.

As will be described in the research methodology, this study will collect data from interviews with 17 teachers. The theoretical framework, adapted from Roger's (2003) model, is shown in Exhibit 11 below. The framework uses the DOI Innovation-Decision process as the framework's foundation, implying the teacher's journey to implement the EGR program. Then, the three inputs to understand a teacher's interaction with the program and the ultimate decision to adopt depend on the teacher's characteristics, the local environment, and the education system.

Exhibit 11: Theoretical Framework



(adapted from *A Model of the Five Stages in the Innovation-Decision Process*, Rogers, 2003, p.170)

The legend in the exhibit shows the source(s) of the source of the contribution to the framework. The three inputs of the theoretical framework are discussed below:

2.5.1 Teacher Characteristics

My IFS and DOI theory suggest that teacher characteristics play a prominent role in how a teacher interacts with and ultimately adopts (or not) the innovation. My thesis study framework uses the DOI categories and characteristics of adopters and BE economics to describe the teacher. For example:

- By understanding a teacher's instructional approach and beliefs before being introduced to the innovation, we know the power of the default (Thaler and Sunstein 2009) and if the EGR program as an innovation is similar to the teacher's existing classroom instructional approach.
- A teacher's communication and social networking (including homophilous and heterophyllous communication) is a critical aspect of how innovations diffuse (DOI) and how individuals learn to implement the new classroom approach with shared goals (Fullan).

2.5.2 Education System

RTI uses Core Functions (Bruns 2018) to describe in simple terms a functioning system. The critical components of Core Functions are expectations, support, and accountability to achieve the expectations. As this study is interviewing teachers, it is limited to understanding if the components of Core Functions influence the teacher's interaction with the innovation. Note that the RTI Core Functions model is effectively a subset of Fullan's framework, which is why Core Functions is not mentioned in the theoretical framework. Alignment with current research at RTI is important to make this a helpful study and for colleagues to connect with the theoretical framework and findings.

A frequently overlooked component of education systems is communication. For a centralized-decentralized system is important to understand how communication reaches teachers, what type of message it is (e.g., expectations, knowledge, etc.), and what action is taken as a result of this messaging. Fullan (2015) discusses how

successful change in education results from learning at all system levels. A critical role of the education system is leading this learning culture (Fullan, 2015).

2.5.3 Local Environment

The local environment input addresses school influences on a teacher's interaction with education innovation. Perceived characteristics of the innovation from DOI address how the teacher perceives the EGR program. These perceptions are formed individually but also influenced by external factors. Social norms address how other teachers affect the teacher's interaction with the innovation. This is from a perception perspective of what the teacher thinks others expect him to do and what he thinks others will do. Finally, the framework explores the local characteristics from different perspectives, such as learning opportunities, communication, interaction with the administration, and teacher peers.

The limitations section of this study will discuss the challenges of this theoretical framework in more detail. A single study could easily take on one of these aspects of the framework. However, just one aspect of this framework has marginal use to my institute. My motivation is not just to research a theoretical framework as an abstract but also to use it to support future EGR programming design.

2.6 Alignment of Thesis Research with Institute activities

This thesis study employs secondary data analysis of data initially collected from operational research conducted by RTI International. The initial investigation was conducted in Tanzania to support the USAID Jifunzu Uelewle EGR Program, which RTI International implements. This operational research study took place in rural Iringa and Morogoro districts.

2.7 Tanzanian Context

This study collected qualitative data from schools in Morogoro and Iringa districts in Tanzania. These districts receive support from USAID Jifunze Uelewe's EGR Program. Therefore, for context, I will describe the Tanzanian Education System, USAID Reading Programs implemented in the country, and relevant assessments of pupil learning.

2.7.1 Tanzanian Education System

The Ministry of Education, Science, and Technology (MoEST) is a Government of Tanzania institution responsible for providing education in Tanzania. The permanent secretary leads MoEST. The primary functions of MOEST are coordinated through different divisions. Key Tanzanian educational organizations that Jifunze Uelewe supports include the Tanzanian Institute of Education (TIE), which is responsible for developing and implementing new curricula and materials, providing teacher training, and providing oversight for quality assurance of all aspects related to classroom instruction. Finally, at the district level, JU works with the President's Office, Regional Administration, and Local Government (PO-RALG). This institute is responsible for decentralizing and improving quality services, including education and public health. A key aspect of PO-RALG's work is providing Ward Education Officers (WEO) as school support. These WEOs visit schools, observe classes and provide feedback to teachers on their classroom instructional practice.

2.7.2 USAID Reading Programs in Tanzania

The first USAID reading program was Tusome Pamoja (2016 to 2021). RTI International implemented Tusome Pamoja in Zanzibar and the districts of Mtwara, Morogoro, Ruvuma, and Iringa (USAID 2018b). It was designed to support MoEST, PORALG, and the Zanzibar Ministry of Education and Vocational Training (MoEVT). The program was designed to build upon existing country initiatives in reading instruction and leverage and improve the existing education system.

The program had three objectives to achieve the overall aim of improved early-grade reading outcomes:

1. Improved classroom instruction
    a. Developing, printing, and distributing teaching and learning materials
    b. Establishment of teacher Communities of Learning
    c. Teacher training on early-grade reading instruction

2. Strengthened Delivery of Skills – focusing on the performance of management throughout the education system, including teacher monitoring, school leadership, and data for decision-making
3. Improved engagement of parents and communities in education through strengthening school governance and management.

The program trained 12,000 grade one and two teachers and developed, printed, and delivered over one million pupil reading books. (USAID 2018b)

The follow-on program to Tusome Pamoja is Jifunze Uelewe (JU), implemented from 2021 to 2025. Implemented in Zanzibar and the same mainland districts, JU has three objectives linked to improved learning outcomes: improved classroom instruction strengthened local and regional ability to support improved pupil learning, and improved community support for safe and inclusive education (USAID 2021d). JU activities included:

- training teachers, headteachers, and WEOs on the phonic-based literacy approach,
- providing schools with inclusive and safe school materials to aid literacy,
- training teachers and parents in strategies to support a safe and inclusive learning community, and
- supporting school communities to establish activities outside of the school.

(USAID 2021d)

Both USAID programs work closely with MOEST and other institutions. Most specifically, collaboration with TIE is critical for phonics-based curriculum, associated teaching and learning materials development, and teacher training. Working side-by-side with PO-RALG is important for delivering educational services at the decentralized district level.

2.7.3. Pupil Achievement in Tanzania

Focusing on pupil achievement in the early grades means that inevitably we focus on assessments that use an Early Grade Reading Assessment (EGRA). USAID or DFID usually funded these assessments, either national or sub-national.

The EGRA tool is an early-grade literacy assessment similar to DIBELS (Dynamic Measurement Group, n.d.). It assesses pupil literacy skills such as letter sounds, familiar words, oral reading fluency, and reading comprehension of a grade-appropriate passage of text (RTI International 2016). For dissemination, a cut-off for a proficiency level is established and the percentage of pupils at or above the proficiency level is reported. The subtask of most interest is the oral reading fluency measure, which can measure SDG indicator 4.1.1.a/b.

We conducted three Tanzania National Early Grade Reading Assessments (EGRA) in 2013, 2015, and 2017. Funded by USAID, these were sample-based National Assessments worked for Standard (grade) 2. A workshop was conducted to set a cut-off for proficient readers. A consensus was reached to set the cut-off at 50 correct words per minute. The percentage of pupils at or above this level was tracked across the three years. 4.7%, 6.5%, and 1.7% of grade 2 pupils achieved proficiency in reading in 2013, 2015, and 2018 respectively (USAID 2018a). While these results are low, they are not unique to Sub-Saharan African Countries (USAID 2022a). Before these assessments, there had not been an assessment of reading in the early grades. The findings from the initial National Assessment were used by USAID to dialogue with MOEST and make a case for supporting early-grade reading. Hence, the introduction of Tusome Pamoja in 2016. Note that these assessments were representative of reading at the national level.

Tusome Pamoja conducted a baseline evaluation of early-grade reading in 2017. This survey was representative at the district level where the program was implemented. For all the districts under Tusome Pamoja, the percentage of pupils at or above the benchmark of reading 50 correct words per minute was 1.5% (±0.5%) (USAID 2017). This value is representative of the entire Tusome Pamoja district and Zanzibar. Note that the schools visited for data collection in this thesis study were purposefully selected low-performing schools in Morogoro and Iringa. While we do not have the grade 2

reading performance for these individual schools, these reported pupil reading proficiency findings nationally and for Tusome Pamoja are consistently low, indicating the results we might expect from the data of the schools visited.

2.7.4 About Jifunze Uelewe

Launched in 2021, the five-year USAID Jifunze Uelewe (“Learn and Understand”) program aims to improve literacy and numeracy outcomes for all pre-primary and primary education pupils in the Regions of Morogoro, Iringa, Mtwara, Ruvuma, and Zanzibar (USAID 2021e). Jifunze Uelewe aims to work through the Government of the United Republic of Tanzania (GOT) systems, providing technical assistance to improve the capacity of government counterparts to improve the effectiveness and efficiency of the delivery of early-grade education programming. The program mostly assists two entities, The Tanzania Institute of Education (TIE) and the President’s Office-Regional Administration and Local Government (PO-RALG). TIE is a public institution under the Ministry of Education, Science, and Technology (MoEST). TIE is responsible for leading the interpretation and implementation of MoEST education policies regarding curriculum programming and instructional materials. PO-RALG manages decentralized regional health and education services, including Local Government Authorities (LGAs). This structure means that Jifunze Uelewe provides technical assistance to education programming design (e.g., curriculum, training, materials) to TIE and implementation support to the LEAs via PO-RALG.

2.8 The Growth and Criticisms of Early Grade Reading Programs

Early Grade Reading (EGR) Programming in international development substantially grew after the introduction of the 2011 USAID Education Strategy (USAID 2011) and later alignment with SDG Indicator 4.1.1 (United Nations 2019). USAID EGR Programs have reached an estimated 246 million primary children since their inception and implementation in 53 countries (USAID 2023b). EGR programs generally supported development of structured teachers’ guides, pupil workbooks and textbooks, and on-going in-service professional development (RTI International et al. 2017). Key to the approach for most EGR programs was that children were to be instructed in their mother tongue or local language of instruction. This meant developing literacy skills in a

local language they could then apply to a second language such as English or French around grade three or four (RTI International et al. 2017). The vast majority of EGR programs used phonics-based reading instruction which was designed to local country and language contexts (RTI International et al. 2017). This approach is almost always combined with the gradual release model, often called “I do, we do, you do” (Killian 2023). This approach starts with the teacher modelling what the pupils need to learn. This is followed by the “we do” stage where they are helped by the teacher to retrieve the new information or concepts. Usually, this stage is completed with the whole class recall. The first first-two teacher centered concepts will then make-way for the “you do” component where pupils will individually or in groups practice the new concepts introduced in class. This structure is repeated on daily basis for literacy instruction, and teachers will use a teacher manual.

Using a synthetic phonics approach (Machin, McNally, and Viarengo 2018), pupils are taught the association between letters and their sounds. This approach supports the concept of learn to read, read to learn (Chall 1996), and espouses the use of the teaching of phonics and the positioning of oral language as central in literacy development (Snow 2021). Pupils are expected to progress from phonics-based reading to comprehending and making-sense of grade appropriate text. Note that all USAID basic education programs are not just pure phonics approaches. Included in the curriculum are from 100 to 200 sight words learned yearly, which teachers are expected to help pupils identify.

Phonics-based instruction is one position in “The Reading Wars” (*The Economist* 2021). The opposite end of the literacy instructional spectrum is the *whole language method,* where children are given sight words accompanied by pictures to help identify the words (Gear 2021). Proponents of this approach argue that it supports learning to read through children engaging with a story, rather than being able to correctly pronounce every word (Scott, n.d.). This approach is often accompanied by levelled texts, where pupils progress through literacy skill appropriate literature. Proponents of this approach believe is that the approach is child centered and immerses them in the meaning of

texts at an early stage, thus keeping the motivation and joy of reading front and center (Froese 1996).

In 2016, the Tanzania education system switched to phonics-based approach from the whole language approach, aligned with the 2014 introduction of the USAID Tusome Pamoja (Let Us Read Together) early grade reading activity. This shift was prompted in part by evidence from the USAID funded National Baseline Assessment for the 3Rs (USAID 2013), introduction of the Tusome program, and the subsequent National Assessment (an EGRA) conducted in 2016. The 2013 report, produced by RTI, mentions that "*children are not learning to decode words effectively and efficiently*" (USAID, 2013, p.67); a conclusion that should have been expected given whole language instruction was being using in Tanzania in 2013. This whole process of using an EGRA as a national assessment and accompanying evidence was used to demonstrate low outcomes and that the country needed to switch to a phonics-based instructional approach was commonly used by USAID (and DFID) in many LMICs. However, the 2013 National Assessment report also discusses issues with a lack of teaching and learning materials, including levelled readers for children to practice and develop their reading skills, poor teacher training and supervision (USAID, 2013). This raises the question that if just these systems issues were addressed, would the whole language instructional approach would have become more effective? However, at that time, USAID would only provide development support and funding for reading programs that included a phonics-based component.

While a phonics-based instruction became front and center in Tanzania in 2016, there was an intent for a more "balanced literacy" (D'Souza 2022) approach in the form of levelled readers in every classroom. However, availability and access to these texts was still known to be inconsistent (USAID 2018a).

The introduction of phonics-based EGR programs in LMICs has aways drawn a degree of concern and criticism. The two key concerns regarding the EGR programs were a focus on a structured approach to literacy instruction and a singular focus on phonic-based instruction. Hoffman was an early critic of the program, stating that EGR programs will demonstrate *". . the inevitable failure of a commitment to a single method*

and the dangers inherent in a narrow, technical version" (Hoffman 2012, p.344). In other words, structured literacy programs provide a lack of flexibility for teachers to make decisions to meet the local conditions, individual pupil needs, and so on. Interestingly, back in 2012 Hoffman predicted that these programs would have an initial short-term success, but then fail to progress due to their inflexible designs. This prediction by Hoffman seems to match the findings shown in *exhibit 4*, where any EGR program with impact at midterm fails to improve at endline. However, any association would certainly require further research to confirm causality.

The impact of EGR programs has been mixed. The Center for Global development conducted a metadata review of USAID EGR Programs, putting the average impact at an improved reading fluency of three correct words per minute (Sandefur et al. 2023). Piper and Stern (2018) point out that although individual EGR Programs have had some successes in improving the average early grade literacy levels, the improved percentage of pupils who can read proficiently has been disappointing. Finally, my own IFS showed that no EGR program has evidence of improved impact beyond midterm and that generally 80% of the overall impact is explained by a smaller number (15%-36%) of schools (King 2020). In conclusion, even if we would want to claim that the generalized impact of some EGR programs have been encouraging, there is no evidence that this progress continues beyond a quick, immediate impact, in a small percentage of programs schools. This conclusion is missed when program impact is only expressed in narrow, generalized terms (Bédécarrats, Guérin, and Roubaud 2019).

The resistance to structured literacy programs which limit teachers' agency to solve their own pupil learning issues have also been expressed in the United States. The Brookings Institute (Loveless 2021) used the same argument as Hoffman to qualify why the implementation of the Common Core (2022) failed to improve learning outcomes, citing the development and implementation of top-down standards-based reform alongside high expectations to address variability in instructional quality. Perhaps the greatest criticism of EGR Programs is the use of the EGRA (RTI International 2016) as the core testing approach (Hoffman 2012; Bartlett, Dowd, and Jonason 2015) and the emphasis on phonics and oral reading fluency. It is important to note that the EGRA as

a generalized assessment of learning pre-dates the introduction of EGR programs in most countries, an unusual sequence for curriculum development; standards and curriculum are usually developed prior to assessment. As a key strategy for local engagement, USAID took the approach using the EGRA as the tool for a National Assessment of early grade literacy in many countries to initially demonstrate the low learning outcomes for early grade literacy (USAID 2022a), and the need for reform for literacy instruction.

USAID is introducing new iterations of EGR programs. These new programs reflected a shift in USAID policy towards “Journey to Self Reliance” (USAID 2020) introduced under the Trump administration, followed by a focus on localization under the Biden adminstration (USAID 2024). Often, the first generation of EGR programs were direct implementation by an NGO funded by USAID. They would develop and print teaching and learning materials, and design, organize and conduct the cascaded teacher training. The emphasis for most EGR programs has since shifted to technical assistance to local governments and supporting local education priorities, policies, and implementation. In many of these countries, the phonics-based local language instructional program has been phased-out in favor of the lingua franca. For example, many countries such as Ghana and Zambia (Zambia Ministry of Education 2023; NaCCA 2020) have created new policy and curricula, focusing on the English as the language of instruction; the mother tongue language becoming a subject area instead. Given overall low impact of EGR programs ((Sandefur et al. 2023; Piper and Stern 2019) and the funds spent, what is the lasting legacy of USAID EGR policy? USAID’s retrospective of the 2011 USAID strategy acknowledges the challenges of implementation, concluding that context matters and change takes time (Martinez 2023). The Center for Global Development’s 2023 report on EGR Programs was inconclusive, noting most EGR programs were expensive, especially considering the unknown sustainability and long-term benefits of these programs (Sandefur et al. 2023). The Global Education Evidence Advisory Panel’s (GEEAP) 2023 Cost-Effective Approaches to Improve Global Learning report (GEEAP 2023) does not call-out EGR programming specifically, but rather an aspect of EGR program design, namely structured pedagogy a process of step-by-step activities that teachers can follow in the

form of lesson plans or other teaching guides. The report does not mention phonics-based instruction, nor early grade literacy instruction in mother-tongue language (GEEAP 2023). Given that USAID basic education policy shift to localization and technical assistance, no matter the local policy on early grade literacy language of instruction, structured pedagogy is the most visible remaining legacy of the EGR programming movement (Gates Foundation, n.d.-a). There are many advocates of structured pedagogy, with many NGOs following its guidance with technical assistance support funded by the Bill and Melinda Gates Foundation. However, given criticisms of structured programming that it limits agency of teachers to make decisions for their own pupils (Olsen and Elliott 2023), there has been a surge in pupil remediation programming, most commonly seen with Teach at the Right Level (Teaching at the Right Level 2022b). Also a GEEAP *smart buy* (GEEAP 2023), remediation programming often comes in the form of low performing pupils being identified and receiving additional instructional support. It is standard nowadays for USAID LMIC education funding and implementation to include some form of remedial support.

Chapter 3: Methodology

3.1 Research Questions

This study is guided by two main research questions developed from the theoretical framework.

Research Question 1: *How does an individual teacher respond to the classroom implementation of an Early Grade Reading Program?*

Research Question 2: *What are the systems-level and social influences on a teacher's response to implementing an Early Grade Reading Program?*

These research questions separate the individual and systems/social factors of the theoretical framework. In this regard, *Research Question 1* explores the following components of the theoretical framework:

1. Teacher Characteristics
    a. Perceptions, beliefs, behaviors, and attitudes drop to the introduction of the EGR Program.
    b. Personality variables – e.g., ability to deal with abstraction and ability to cope with risk or uncertainty
    c. Communication behavior – within the school (e.g., peers and the head teacher) and across schools (e.g., external coach and Communities of Learning (CoL)
    d. Classroom approach before introduction to EGR Program
2. Teacher-perceived characteristics of the EGR Program (i.e., innovation)
    a. Relative advantage
    b. Compatibility of the EGR Program with prior classroom instructional approach,
    c. Complexity to implement the EGR Program,
    d. Trialability
    e. Observability
    f. Mandatory
    g. Visibility (e.g., public awareness, publicized, etc.)

3. Teacher's degree of learning and improvement applied to classroom instructional practice

Research Question 2 explores the social and systems expectations that influence the teacher and their interaction with the ERG Program:

1. Teacher learning and support process
2. Systems expectations for teacher (classroom instructional practice) and pupil expectations (learning outcomes)
3. School administration support and accountability
4. System cosmopolite and localite communication
5. Social norms regarding classroom instructional practice - perception of what others expect the teacher will do and what the teacher perceives others will do

In truth, it will be difficult to fully disentangle some of the data and analysis from research questions 1 and 2 – for example, has an individual teacher adopted a particular instructional practice from their learning process or the influence and practice of peers? Ultimately, this means a certain amount of blending of the first research questions.

The third research question uses the findings from the first two research questions to develop suggestions for program adaption proactively:

Research Question 3: *How can EGR programs adapt existing activities to mitigate individual and systems-level challenges that impede teacher implementation?*

This question is guided by the findings of the first two research questions, which are primarily deductive in nature and focus on identifying themes developed from the theoretical framework. Research question 3 takes a more inductive approach, using the findings from research questions 1 and 2 to develop a hypothesis of how EGR programs could be designed to account for the varied challenges to implementation deduced from the data. Research question 3 will be addressed in the Discussion section of this study, through combining the data analysis, theoretical framework and other supporting literature and evidence.

The three theories used for the research framework were chosen deliberately because they do not just help us understand teacher behavior at a single time point. Still, all have been used as interventions in other studies.

For example, a fundamental idea that underpins behavioral economics is from Lewin's (1997) recommendation that we need to reduce the restraining forces, i.e., change the environment to achieve behavior change. This suggestion is more novel in development work where the tendency is to solve observable problems, according to Kahneman (2011), is a more natural human instinct. Kahneman (2011) suggests that changing the environment is profoundly counter-intuitive, which is why it is not leveraged as an option. However, changing the environment can often be cost-effective compared to adding more intervention components.

DOI (Rogers 2003) was initially developed by observing how innovations diffuse (or not) over time. The DOI framework provides a comprehensive diagnostic to understand how the EGR Program is diffusing and where it falls

Michael Fullan's (2015) work also represents an opportunity to focus on recommendations for improving EGR Programming through his drivers of education change. Fullan (2015, p. 42) discusses how to correct drivers of change, including capacity building for results, collaborative work, pedagogy, and system-ness. Further, Fullan (2015) describes important components of programming that should not drive change, including external accountability and individual teacher and leadership quality. Applying this analysis to the data analysis should provide an interesting comparison of EGR Programming priorities and focus versus evidence provided by Fullan and suggest potential gaps in EGR Programming.

3.2 Research Methods

The research design deployed was an ***qualitative analysis*** of secondary classroom observation and teacher interview data collected from low-performing schools in Tanzania. The motivation behind this approach is two-fold. Firstly, it aligns with behavioral science by exploring behavioral barriers to positive decision-making (Kahneman 2011; Thaler and Sunstein 2009; Ariely 2009). Second, it progresses my EdD research beyond my IFS studies which explored characteristics of individuals at

high performing schools, concluding that they were individuals who possessed positive personality traits and as such were not the typical teacher population (King, 2020).

3.2.1 Research Design

This study deployed a qualitative secondary analysis of data collected from twelve low-performing schools in rural Tanzania. The original source data was seventeen qualitative classroom observations and teacher interviews. The data was originally collected for the purposes of operational research for the JU EGR Program. The intention of this study's secondary data analysis is to answer questions using the research framework. The available data was transcribed teacher interviews and classroom observations.

3.2.2 Research Rationale

There are a number justifications why secondary qualitative data was the appropriate data source for the research framework.

The Importance of Data from Low-performing schools

My IFS study explored characteristics of high-performing schools, concluding that the common characteristic of these schools were individual(s) with positive personality characteristics aligned with Rogers' (2003) definition of early implementors. My IFS conclusion was that most of the education research looks for associations with pupils learning gains. Consequently, typical education research studies the actions of these implementors and looks for ways to scale their actions and behaviors (Innocenti 2019; Blavatnik School of Government 2023). However, a behavioral science framework positions that adding more incentives and components does is often not an effective way to achieve good decision-making (Thaler and Sunstein 2009; Kahneman 2011). Behavioral economist Daniel Kahneman (2011) suggests that to help individuals make good decisions we should focus on understanding and diminishing the restraining forces that prevent them from doing so. This is the counter-intuitive nature of behavioral economics research; looking for barriers to positive decision-making, rather than focus on what works.

Secondary Qualitative Data Justification

The secondary data that I used for this study was qualitative; a classroom observation followed by a teacher interview. The classroom observation was recorded as a text narrative and the interview data consisted of primary questions with follow-up secondary questions. Participant classroom observation data and interviews were linked.

Generally, behavioral data is quantitative. For example, a common data collection tool that focuses on different dimensions of teacher practice including behavior is the knowledge, attitudes, and practice (KAP) tool (Heck, Ormiston, and Husmann 2023). However, a strong argument can be made that researching behavioral science (and in particular behavioral economics) focuses on what causes decision-making is *". . . not accessible to conscious introspection.*" (Gordon 2011, p.179), In other words, individuals may not be aware of why they decide to do what they do.

Behavioral studies, unless observational, require participant recall of events and decisions. Individuals' recall of events can often be unreliable due to multiple psychological reasons such as creating false memories to protect our self-image (Bauer 2014). Additionally, other biases can cloud our ability to accurately recall events such as fading effect bias where individuals introduce a coping mechanism that limits recall of behaviors associated with unhappy emotions (Gibbons et al. 2013; Webster 2023). For example, the impact of education policy on how teachers must change their instructional practice might be unpleasant in the moment, but over-time teachers might be focus on what is good about the change. For this reason, a behavioral economics framework often does not use "*why?"* as a line of enquiry, instead building a picture of behavior based on the *where, how, what, who, and when* of the environment (Gordon 2011). This study is only concerned with the really concerned with participants' reflection on their actions to align with themes such as *cognitive dissonance*.

Given my research focuses on Early Grade Reading programs, I needed secondary data using teachers as participants. Typical quantitative behavioral data such as KAP studies use participant recall as the key mechanism for understanding attitudes and practice and so not a match for this study's theoretical framework. Designing and implementing a study funded by myself was always going to be cost prohibitive.

3.2.2 Assessing the suitability of the available secondary data

A key disadvantage of secondary data analysis is that the data might not have information that adequately answers the research questions. The data that was made available to me was from a qualitative operational research study conducted by JU EGR Program.

It was important for me to determine if the data from the original operational research study was suitably aligned with my proposed study. To understand this issue, I studied the tools used for the JU operational research and contacted the RTI employee who was responsible for leading the data collection of the study. The original study's data collection was qualitative, consisting of classroom observations and follow-up teacher interviews. I was informed that the study deployed a cognitive interview approach, designed to provide quality understanding and opinion from teacher participants. However, there were no measures of reliability or validity collected during assessor training or data collection. Each teacher interview was approximately one hour long. The audio was recorded and then transcribed into English text. Upon review of the fifteen primary questions and associated secondary questions, the tool (Annex 1) used the primary question to asked mostly about teacher actions or behavior and the secondary question helped build a reasonable picture of the environmental influence behind teacher decision-making. To illustrate this, primary question 3 (Annex 1) asked "*Has clear information been expressed to you regarding the level of achievement or progress pupils in your class should show by the end of the year?*". The follow-up secondary questions included:

- *"What are those expectations and whose responsibility is it to achieve these expectations?"*
- *"Do you believe these expectations are achievable? How do you know?"*

These prompts were then cross-referenced with the teacher's use of assessment and interaction with students during the classroom observation. Finally, there was a section of the tool where the assessor used the video recording of the teacher's classroom instruction as a prompt. Consequently, the tool and operational research approach made the text data collected a good fit for the research framework for this study.

I was aware that some of the desired themes aligned to the research questions might lack data and evidence, but given how the research philosophy tries to align multiple behavioral theories to explain the actions of the teachers, this compromise or risk was acceptable. The data was provided to me in the form text files, aligned to the interview questions or a chronological description of the classroom observation.

3.2.3 Research Philosophy

This study employs an positivist approach within a behavioral science framework, aligning different theories of human behavior such mental models and norms to teacher attitudes and behavior. It does not take the teachers' opinion on their actions as causality; rather as evidence of a model of human behavior which might possibly at times contradict teacher opinion; reverting back to the idea that human recall of decision-making is not always reliable (Gordon 2011).

Within EGR Programs or other research funding in development, there is always that pressure to be seen to react to observable issues with some incentive or intervention. To be seen as doing something and having a program deliverable takes off pressure along the management chain and responsiveness to the client. This traditional approach is also quite intuitive. This approach is often seen with RCTs, where attempts are made to resolve observable issues such as teacher attendance (Chen et al. 2001; Banerjee and Duflo 2006). However, no attempt has been made to address the underlying cause, although some effort has been made to understand the underlying teacher mental models for teachers in LMICs (Sabarwal and Abu-Jawdeh 2018).

Prior work has been conducted to quantify teacher perceptions of ICT through quantitative research that includes factor analysis and the coefficient of variation to assign some perceptions as being of more influence than others when it comes to teacher adoption (Hughes and Keith 1980; Moore and Benbasat 1991; Richardson 2011b). However, this approach would have made the teacher's perception of the EGR program the dominant research and limited the usefulness of this research for my institute. Consequently, as a driving factor of the research framework is its utility for EGR programs, there are either the belief or research funds to employ a quantitative approach to the research framework. DeRuyter and Scholl (1998) argued that

qualitative research does not measure; instead, it provides insight into a diagnostic exploratory nature, making it the preferable methodology to understand the complex social and environmental aspects which affect behavior change. The limitations section will discuss the qualitative challenges.

3.3 Data Source used for Secondary Analysis

The sampling methodology of the primary data source was the purposeful selection of 12 low-performing schools in the rural Tanzanian Districts of Iringa and Morogoro. These schools were part of the Jifunze Uelewe (JU) EGR Program, implemented by RTI. JU collected this data as part of its operational research. Schools were selected using the results of the Tanzania Primary School Leaving Examination (PSLE) (The National Examinations Council of Tanzania 2021). Initially, the intent was to observe and interview two teachers per school, one in grade 1 and one in grade 2. However, it seems that only 17 teachers were observed and interviewed due to teachers not being available or present at the school during the data collection visit.

3.4 Instrument Development

Due to this EdD study's secondary data analysis approach, no instrument was developed. The instruments used for the secondary data are described below.

3.4.1 Secondary Data Source – teacher interview instrument

The teacher interview tool (shown in Annex 1 of this study) for the data source had 15 primary interview questions. The interview approach was reported to be semi-structured; secondary interview questions were follow-up options based on the teacher's response to the primary question. The interviews were conducted by local assessors in Swahili or local dialect. The audio from these interviews were recorded, and the scripts were transcribed into English text.

3.4.2 Data source – classroom observation tool

Most EGR Program classroom observation tools are measures of fidelity, assessing how well teachers can implement a program. However, the primary data source designed the classroom observation tool to describe the classroom instructional

approach. The approach was semi-structured, and the text data shows descriptively what the teacher and pupils were doing during each cycle. Within activity ‘cycles’ of up to 15 minutes, the assessors descriptively recorded the classroom activities.

3.5 Secondary Data Collection

The data used for secondary analysis was originally collected in March, 2022. The data was from twelve schools, where seventeen teachers from grades 1 and 2 were observed teaching and then interviewed. The audio for teacher interviews were recorded. I was informed that the interviews were conducted in Swahili or another local language.

3.6 Secondary Data Format

The audio data was translated and transcribed into English and either aligned to the interview tool questions or sequential descriptions from the classroom observation. The data also included teacher gender, age, and grade. Finally, the classroom observation and teacher interview data had a participant ID that linked classroom observation and teacher interview.

3.7 Qualitative Data - Thematic Analysis

The data was initially organized and sorted using a deductive analysis approach. Using the six-stage approach to thematic analysis (Braun and Clarke 2006), a deductive approach was used to align and code the data to the key themes from the theoretical framework that addressed the research questions.

Stage 1: Given the secondary analysis approach, time was taken to become familiar with the tools used for data collection and the data itself. The teacher interviews were conducted in Kiswahili and then translated and transcribed into English. There were a few, but not substantial, issues of quality with the written English. The data was organized by question or sequenced by 15 minute intervals of the classroom observation.

Stage 2: Initial codes for deductive analysis were generated. These codes were aligned with key themes from theoretical framework, aligned with the research questions:

Research Question 1: *How does an individual teacher respond to the classroom implementation of an Early Grade Reading Program?*

Primary coded themes for this research question:

- **Rogers' (2003) characteristics of adopters** – section 2.1.1 describes the different personality characteristics generally associated with different categories of innovation adopters.
- **Power of the default (Thaler and Sunstein 2009)** – section 2.5.1 discusses how teachers may keep many characteristics of their prior instructional approach.
- **Perception of the Innovation (Rogers 2003)** – section 2.1.2 describes how individuals' adoption of an innovation is dependent on how they perceive the innovation through interaction.
- **Mental effort / Type I & II thinking (Kahneman 2009)** – section 2.2 describes how mental fatigue causes individuals to irrational decisions.
- **Innovation-Decision process (Rogers 2003)** – section 2.1.3 describes the decision of an individual to adopt an innovation.
- **Cognitive Dissonance** – section 2.2.3 describes how individuals can have a state of inconsistent thoughts, attitudes, or beliefs related to decisions and actions.
- **Model of Teacher Change** – The literature shows competing theories of how individuals behave. Section 2.1.3 describes how Guskey's (2002) model of teacher change relies on a framework of logical decision-making, while behavioral economics demonstrates how human behavior is fundamentally irrational (Ariely, 2009).

Each theme was also coded for sub-themes as appropriate. For example, Type I & II thinking (Kahneman 2009) included short-cut heuristic behavior as a sub-theme.

Research Question 2: *What are the systems-level and social influences on a teacher's response to implementing an Early Grade Reading Program?*

Primary coded themes for this research question included:

- **Social norm theory (Bicchieri & Noah, 2017)** – section 2.2.4 describes how individuals are influenced by their perception of what they think their peers do and what their peers expect them to do.
- **Collective Efficiency (Fullan, 2015)** – section 2.3.2 discusses the importance of how education systems should focus on group quality
- **Innovation-Decision process (Rogers 2003)** – section 2.3.2 discusses the importance of how peer interaction influences the decision of an individual to adopt and innovation.
- **Focus on results (Fullan, 2015)** – section 2.3.2 discusses the importance of education systems and teachers to focus on results (i.e., pupil learning outcomes).

Sub-themes that were coded as necessary.

Stage 3: Themes were developed from analysis and coding of the data. The teacher interview data came from a qualitative tool with fifteen primary questions (plus subsequent secondary questions), the main process was to explore the data based on teacher response organized by these questions. This made the patterns that formed the themes easier to organize. For example, teachers were asked if they had received clear information regarding expectations for pupil learning achievement which maps to Fullans' (2015) focus on results. Thus, analysis across participant responses using deductive themes was quite efficient.

Stage 4: Themes were reviewed to ensure that they made sense and were supported by the data. Careful consideration was given to subthemes. For example, for Fullan's (2015) focus on results, that teachers knew about learning expectations is a sub-theme connected to other sub-themes such as communication, teacher support and feedback, and classroom behavior. There were some inductive themes that lacked supporting evidence from the data, particularly regarding Research Question 2. This challenge will be addressed in the analysis.

Consideration was also given to an inductive approach; due to the nature of the research framework, the room was given to a 'bottom-up' approach to building new themes that were not necessarily a key component of the framework (Thomas 2006).

For example, behavioral economics and behavioral science have a broad range of psychological and sociological theories to draw from. Therefore while themes and coding followed what the literature indicated there key concepts, space in the analysis was given to inductive theme identification and development (Braun 2022)

Stage 5: Finally, the themes were refined and mapped to each other. The relationship between the themes was explored such that they formed a more coherent response to the research questions.

Step 6: The findings were written-up into the analysis section of this research which included data in the form of teacher interview responses or classroom observations. This evidence took two forms. The first type of evidence used was to collate similar responses by the teachers for certain themes. It is important to clarify for a given theme if the evidence in the qualitative data presents a consistency across participants. However, for certain themes, the evidence was not as consistent or substantial. This was detailed in the analysis. Finally, evidence in the form of teacher quotes were added to illustrate themes.

3.8 Ethical Considerations

This study used secondary data. The original data source was previously funded, collected, and key findings disseminated by RTI International to USAID, MoEST, and TIE. The ethical considerations and steps described below. This section will conclude with the ethics research steps for secondary data undertaken for this EdD study.

3.8.1 Position as Professional and Researcher

As a RTI employee my role is mainly to evaluate and research EGR programs. However, the industry in which I work often requires careful curation of program messaging surrounding impact, due to the reality of my institute desiring to win future contracts based on quality work. USAID country missions desire an EGR program to have positive optics for local governments, stakeholders, and partners. For example, this is often seen with USAID success stories (USAID 2023c) which are often opinion or examples of individual success. However, these stories are never empirical in nature.

However, all EGR programs require reporting of impact on learning outcomes through use of either a Randomized Control Trial (RCT) or Quasi Experimental Design (QED). USAID requires every EGR program to report against a performance indicator, "*ES.1-1: Percentage of learners targeted for USG assistance who attain a minimum grade-level proficiency in reading skills"* (USAID 2023a). Conducting these assessments has been my primary role, adding objectivity to program evaluation. While many EGR programs implemented by RTI have demonstrated statistical significance (USAID 2021a) in terms of an increase in the average oral reading fluency, there is no doubt that many aspects of the program findings in terms of learning outcomes present concern. For example, a program can have statistically significant impact on learning, but still many pupils are unable to recognize even one word of text by end of grade 2 (USAID, n.d.). My IFS presented research to suggest EGR Program impact is explained by around 15% of schools accounting for 80% of the impact (King, 2020), suggesting that generalized impact of programs are often misleading and fail to communicate the large variation of impact on learning across schools.

My research opinion regarding approaches to the teaching of early grade literacy is agnostic. I have never been involved the design of a literacy program and never taught literacy as a teacher, and so choose to follow research regarding the benefits and challenges of different literacy instructional approaches.

Finally, my prior experience as a teacher gives me experience and insight into the benefits and challenges of structured pedagogy versus programs designed with more teacher agency. The research framework assesses teacher behavior and response to a structured pedagogical approach, implemented by the USAID Jifunze Ulewe activity. As a researcher, I do not take the position that this pedagogical design is appropriate and just needs refinement. It is also very possible that the approach is not appropriate, given context. This will be determined by objective analysis of the data.

3.8.2 Primary Research Ethical Considerations

Below is a summary of the data collection approval process undertaken by the primary source. RTI is a non-profit organization registered in the United States of America. Consequently, ethical considerations come under the umbrella of guidance and

regulation for the USA. All proposed RTI data collection activities undergo a review by the RTI Institutional Review Board (IRB), which assures RTI's work adheres to the ethical principles outlined by the Belmont Report ("The Belmont Report | HHS.Gov," n.d.). Created under the US National Research Act of 1974, The Belmont Report identifies the fundamental ethical principles for research involving human subjects, providing guidelines to assure that all research is in accordance with these principles ("The Belmont Report | HHS.Gov," n.d.). Data collection for Tanzania is obtained through the National Research Registration Committee (NRRC).

Before data collection, participants were informed of the purpose of the data collection, and the data collection was described to them. They were given the option of opting out of data collection. They were informed that there was no risk to them participating in the research.

3.8.3 UCL Ethics Approval

Due to the secondary data analysis nature of this thesis study, the UCL Ethics application was made under the *not research with human subjects* classification. A description of the existing data for analysis was provided alongside a description of how the data was stored (behind a secure firewall). UCL Ethics approval was given February 1st, 2023.

3.8.4 EdD Secondary Data Research Considerations

I received permission to use this data for secondary analysis for this EdD study from RTI International. To keep the data secure, all data used for secondary analysis was kept behind the RTI firewall with restricted access by permission only. Analysis of secondary data in international research required consideration and awareness of the context of where data were collected. All sensitivity to the local context was taken before releasing findings. The data used also aligns with UCL's Research Ethics Guidance. PII was not included in the results, so locating participants will not be possible.

Most of the concerns regarding the used of data for secondary analysis center around the potential harm to individual participants (Tripathy 2013). This harm can occur from

identifying study participants in secondary analysis and findings or if data is not adequately kept safe from unauthorized access (Tripathy 2013). The issue of participant safety issue can be particularly pertinent for secondary analysis of qualitative data, where data often shows the personal opinions of a limited number of individuals. Key considerations such anonymizing findings will be discussed later in this chapter.

3.8.5 Pro-innovation Bias

A key consideration is a pro-innovation bias, which implies that the change agent assumes that the innovation should be adopted. Any challenges with implementation lie with the intended recipients of the innovation (i.e., the teachers) (E. D. Glor 2003). Consequently, the primary data collection process and line of questioning might have worked off the assumption that the issue for low-performing schools is centered around challenges with implementation. To mitigate this issue, this thesis study avoids becoming a teacher deficit model. In other words, while some teacher classroom instructional practices might not be optimal methods to implement the EGR Program, this study will focus on how JU and the education system design and implement the program. Findings and recommendations will focus on improved programming based on hopefully a clearer understanding of how teachers and other stakeholders actually respond instead of how it is presumed they will respond.

My research position on pro-innovation bias is that I hold no opinion on efficacy and effectiveness of the JU EGR Program. In other words, I do not position myself with the belief that the program is designed correctly and just needs teachers to correctly implement the program. However, nor do I believe that the program is necessarily poorly designed. These positions will be explored through data analysis.

3.14 Limitations of Methodology

The background section of this study introduced specific biases that present potential limitations to this study. Some of these biases will be referenced in this section.

3.14.1 School and teacher selection

Unlike the IFS, schools were not selected based on their average gain under the EGR Program. Under the direction of the Tanzanian MoE - RTI, Jifunze Uelewe could not use

impact data to choose schools based on average achievement gain. Consequently, for selecting schools for the secondary data, they used the 5th grade PSLE achievement reading data. There are a few limitations this created. Firstly, achievement data cannot necessarily be inferred to have been the direct result of the JU program implementation as it does not incorporate pre- and post-measures. Achievement data is easily confounded with other factors, such as socioeconomic status. The participants in the study are teachers of grades 1 and 2. As PSLE is conducted with grade 5 pupils, it should not be inferred that achievement levels will be similar in grades 1 and 2.

3.14.2 The study is limited to teachers

My IFS demonstrated that positive deviance is not necessarily explained by teachers being the catalyst for change (King 2020). It was found that head teachers and community members can also catalyze school improvement. However, due to the secondary data nature of this study, it was limited to studying the behavioral response of just classroom teachers. One recommendation of this study will be to expand systems behavioral research and potential interventions to look beyond just teachers.

3.14.3 Generalization

This study deliberately uses a sample of schools selected with low learning outcomes to study an identified sub-population. However, these schools are not sampled using randomization or any technique that suggests generalization. Consequently, while this study's findings will guide future research and implementation design directions, replication of the results in different contexts would be recommended.

Chapter 4: Analysis of Data

This chapter focuses on the data findings through qualitative analysis. This section will start with an overview, a general reflection of the assumptions that informed the study’s research questions, and how the findings provided evidence that the main assumption was incorrect and why. Then the chapter will provide an analysis to answer the two research questions. Finally, there will be a deep dive into the classroom observation and interview findings.

4.1 Overview - Reflection on Assumptions

Answering the research questions delivers on the key assumption leading into this research. My central assumption was that the school-level improvement was so poor that it could only have been because teachers were not implementing the reading program. As the response to the research questions suggest, my assumption was incorrect. A second assumption I had going back to my IFS was that effective measurement of implementation fidelity would improve teaching quality. As I will explain later in this section, based on the literature review (particularly Michael Fullan's work), this assumption presents an overly simplified understanding of how educational systems achieve meaningful change. While I am comfortable with effectively rejecting assumptions, this means that the discussion, Implication, and Conclusions will pivot substantially away from these key assumptions.

The qualitative analysis follows a certain form. If the majority of teachers were observed or responded similarly, the findings would present this as a 'majority.' If the results were mixed, then this variability will be described.

This section will initially summarize the findings by responding to the Research Questions. Then, detailed responses will be provided to each primary and secondary teacher interview questions.

4.1.1 Summary of findings aligned to the Research Questions

Research Question 1: *How does an individual teacher respond to the classroom implementation of an Early Grade Reading Program?*

The response to this question is organized by key themes from the qualitative data analysis.

Theme – Cognitive Dissonance: Teachers in low-performing schools believed that most or all of their pupils were proficient readers by the end of the year

This study deployed an explanatory approach, analyzing secondary data from low-performing schools in Tanzania. Teachers were directly asked if most or all their pupils would be proficient readers by the end of the year. All but one teacher responded that most or all of their pupils would be proficient. In comparison, The Tanzania National

Early Grade Reading Assessment (USAID 2018a) reported the % of pupils in grade 2 who can read thirty or more correct words per minute as 36.1%. Given these results and that the study schools were low-performing rural schools, it is likely that the teachers' perception of their pupils' literacy progress is inaccurate.

For this finding to be aligned with cognitive dissonance, then the teachers must have seen evidence that their pupils have been struggling to learn to read which is not clear from the data. The classroom observation data describe how teachers found time for pupil practice, however most of the activity was in the form of pupils copying words from the blackboard by rote. Only a small handful of teachers extended the problem. All the teachers moved around the room and monitored work. This classroom observation is important as checking pupil exercise books was the most common practice teachers indicated they used for assessing pupil progress. Here, the data is inconclusive if the classroom assessments provided teachers with an accurate measure of pupil literacy progress. If the teachers understood that their pupils were struggling to read, then their response that most or all their pupils would be proficient readers is more aligned to cognitive dissonance. If the assessment such as copying by rote created an inaccurate measure of pupil progress, then the issue is more of a teacher capacity issue.

The suggestion that the teachers in this study did not know the reading skill level of performance of their pupils is not a unique finding. These findings are consistent with a study of teachers in Southeast Asia, suggesting that teachers were mostly unaware of the reading level of their pupils, particularly those who were low achieving (Djaker, Ganimian, and Sabarwal, 2022). One teacher commented that their pupils will become proficient readers "*. . . because we have enough teaching and learning materials and teachers we have best approaches to teach and pupils improve their performance*". This will be discussed later, that teacher confirmation (Rogers 2003, Guskey, 2001) of effective instruction is informed not by pupil learning outcomes but rather other sources. For example, the process of teachers is adopting the new program seems evidence enough that their teaching practice is effective. This idea is illustrated by another teacher, who said pupils can become proficient readers as *". . . as teachers we have been trained, we have the resources, and have will so we try our best to achieve these expectations.*"

Theme – Perception of the innovation: Resistance to reading the reading program was neither observed in the classroom nor expressed during the teacher interview.

The assumption leading to this research question comes from Exhibit 1. The initial assumption I had was that many teachers just were not implementing new programming. The general idea was that for the impact to be so low that (Exhibit 3) it was improbable that teachers were implementing. However, this idea turned out to be incorrect. What was first confusing about the data was the level of homogeneity. All the teachers were observed to have implemented the reading program. Noting that these teachers were selected from low-performing schools, how was it possible that all the teachers were implementing the program, however ineffective in improving learning outcomes? This takes a little unpacking.

The teachers expressed a clear sense of contentment with the reading program. For example, one teacher commented on the program training they received, "*How to teach phonetics was a very interesting aspect because before training I was teaching a totally different way. The training on how to use teaching aids in the lessons as well as how to use counters in Mathematics. These were new knowledge to me, and I enjoyed them very much.*"

However, Jifunze Uelewe followed the path of another USAID reading program, Tanzania Tusome. This means that when the teachers were observed and interviewed, some had been interacting with a reading program for over six years. This created a complex problem that the data could not explain, namely how the teachers' attitudes towards the reading program might have changed over time. The theoretical framework for this study adapted the innovation-decision process from DOI theory (Rogers, 2003), as shown in Exhibit 8. The process describes an interaction between the individual and innovation. The key steps are knowledge, persuasion, decision, implementation, and confirmation (Rogers 2003). However, the secondary data was from interviews with teachers who had been interacting with EGR programs and phonics-based instruction for several years, likely making a difference in the perception of the innovation compared with the first few years of interaction. Prior education research into perception of an innovation indicated that the main two factors that made adoption more likely were

the complexity of the innovation and whether it was mandatory (Moore and Benbasat 1991; Hughes and Keith 1980). Rogers (2003) tells us that adoption of an innovation is linked to its perception. Like BE's libertarian paternalism model (Thaler and Sunstein 2009), teachers are not free to choose whether to adopt the reading program or not. What was the teachers' perception of the reading program? This is where resistance was expected. However, again the homogeneity of the response was unanimous. The teachers all believed the reading program was superior to their previous teaching approach they were using prior, which was the whole language literacy approach (see section 2.8). They commented they learned this prior approach at teacher training college in Tanzania. One teacher, when comparing their prior approach (whole language instruction) to the new approach said, "*Training I got from the college influenced me somehow to use some approaches although they were not effective. For example, in the participatory approach I used my songs to teach, but they were not effective compared to what I have now learned.*" Another teacher specified an instructional weakness the reading program helped correct; *". . . I was struggling to pronounce some letters, but after training teaching letter sounds is simple and this led my lesson be effective and as result pupil understanding improved."*

Some teachers appreciated how the new approach was more participatory, commenting that the new reading program made "*. . . learning for these young children more meaningful*".

The teachers were all observed implementing the phonics-based literacy content, mostly demonstrating a competent understanding of the approach. A few teachers struggled a little with sounding the correct phonemes but not to the point of poor delivery of content and something that could not be easily corrected with some support. Only a few teachers said they struggled to adopt the new program's instructional approaches; the teacher interview provided evidence of teacher motivation to learn to implement the new program. As one teacher said, "*You know if something is benefiting you, you will try your best to ensure you use it.*". Another said, "*For me, it was easy because I had the passion to learn new approaches.*"

It is really important to emphasize that the change the teachers focused-on was switching from whole language literacy instruction to phonics-based instructional

practice. While the EGR program training included a focus on student-centered learning (the “you do” final part of the gradual release model), teacher responses to their changed instructional practice focused on the teacher-centered change in curricula and content. Why they did this will be expanded-on in later deductive themes, and is central to answering and understanding research question 1.

Sub-Theme – Perception of the Innovation: Mandatory

Aligned with Rogers’ (2003) perception of an innovation, teachers expressed that implementation was *mandatory*, it was expected of them. However, how teachers defined mandatory focused not only through systems accountability to implement, but also their professional obligation to implement a program they believed improved their instructional approach. One teacher commented, “*I feel good because this new approach makes my lesson easy for pupils to understand. For example, using the I do, we do, and you do model gives pupils a chance to listen, practice together, and later practice themselves. So, these approaches are better than what I had before.*”

The perception that an innovation is complex (Hughes and Keith 1980) can also be a barrier to adoption. Some of the teachers expressed the notion that the phonic-based literacy instructional approach was challenging at first, often because it was unfamiliar. However, if a teacher expressed the new programs as initially challenging, they also said they were able to understand and improve over-time. For example, one teacher commented, “*The phonetic approach initially was challenging as we used to teach differently. And it was challenging because it was something new and it required a lot of practice of letter sounds.*”

What is not clear from the data is how a teacher perceived the program when it was introduced compared with once they had adopted it. The secondary data was collected after six years of implementation of reading programs. So, it is very conceivable that if a teacher has adopted and is using a new instructional approach, they will have a positive outlook on the approach as it is in essence *their* instructional approach. If they criticized the approach, would they effectively be criticizing their own teaching? However, what is clear from the data is that even though the education system made adoption of the reading program mandatory, teachers will over time have a positive perception of the program.

Theme Innovation-Decision Process: Teachers are comfortable with adopting new delivery approaches and curricula

The level of content that the teachers had with delivering the new instructional approach in the classroom suggested that they were likely at the confirmation stage of the innovation-decision process (Rogers, 2003).

However, data analysis draws the conclusion that the teachers in the study adopted the read program but focused more on certain aspects of the new program. They seem able to learn new curricula and introduce new content using new pedagogical practices. In other words, delivery of new content. This perhaps should not be a surprise. From the moment teachers enter the profession, they are learning new content. Often in their careers, they are required to learn more content, perhaps a new subject or a different grade. While this change is not always welcomed, it is a familiar change and as such teachers might be able to manage the mental effort required to achieve this. For the classroom observations, the majority of the teachers were observed competently delivering the phonics-based curriculum. In other words, they demonstrated a basic understanding of the approach and were able to express that approach at the front of the class. There was also a recognition of the challenge with learning new approaches and content. Some teachers were quite open about recognizing the challenge and their struggle of mastery of new content; "*For example, with teaching letter sounds, I was struggling to pronounce some letters but after training all letter sound to me is simple and this led my lesson be effective and confident which as result even pupils their understanding improved.*" The challenge of sounding out phonemes was commented on frequently by teachers and often observed in the classroom.

Adopting new curricula is a considerable mental effort (Kahneman 2009). It seems this challenge was mitigated by the program design delivering new content via a teacher centered approach, thus lessening the mental effort in adopting new curricula. Commonly called "I do, we do, you do," approach, all the teachers followed the gradual release model when introducing new content. The teacher starts as the focal point of the class, introducing a new idea. Then, they will have the entire class repeat speaking

the new sound or word. A teacher centered approach to delivery is likely already familiar to all the teachers.

This type of program design is something espoused in behavioral economics, the idea that if you want someone to do something, make it easy (Thaler and Sunstein 2009). However, behavioral economics mostly focuses on dichotomous decision-making without concern on the quality of the action once the decision is made. In other words, behavioral economics has a lot to say about how individuals can eat better, stop smoking, invest in their retirement (Thaler and Sunstein 2009), but it does not discuss how to help individuals do something better which is a complex activity with many interrelated components and decisions like classroom instructional practice. This speaks to the potential limitation of behavioral economics and *nudging* in education, or perhaps behavioral economics in general.

Theme Power of the Default: Teachers struggled to adopt pupil centered learning

After the teacher centered delivery of new content, pupils work individually, in pairs, or in groups, practicing the new content. The teachers will also call on individuals or groups of pupils to respond to specific parts of the class. However, it was observed that the teachers mostly called upon the same proficient pupils to answer questions. The observed modification to the reading program occurred when the pupils were required to practice the new content or skill. All the teachers modified the approach in a very similar way. Some teachers had the pupils work in groups; others were instructed to work individually or in pairs to practice the new skills and write them into their exercise books. Teachers would move around the classroom, checking and marking exercise books. However, teachers mostly checked that pupils had copied what was on the blackboard into their exercise books. It was observed a few times that if a pupil could not complete the work, they did not ask for support from the teacher; instead waited to see if the teacher would come to them. Many pupils were observed without exercise books and utensils to write with. Either they borrowed from their peers or did not get the opportunity to write. It is not known if this was the only means of pupil assessment. However, assessing pupil literacy skills based on their ability to write words, not sound words or phonemes clearly is insufficient.

Additionally, the classroom observation data indicated that classroom activities were not always well timed. For example, while sufficient or more than the lesson plan scheduled time was provided for the teacher centered components, the pupil-centered practice was often missed or shortened. A good example from the data is when a teacher was asked to comment on a video of their instructional practice: "*Here I was teaching reading and as a teacher was demonstrating how to pronounce na, ne, ni, no and nu and at the end to write. But as you saw most pupils didn't understand very well so I used much time to repeat those syllables.*".The aligned classroom observation noted that the teacher spent most of the class repeating the syllables, giving the pupils any time to practice. This emphasizes how teachers tend to focus on delivery over pupil learning.

The data is unclear exactly why the teachers struggled with to shift their behavior and retain prior instructional practices. There exist a number of possibilities, including power of the default (Thaler and Sunstein 2009), or mental effort (Kahneman, 2009) given the pupil time on task is at the end of the lesson. However, what is clear is that teachers prioritize and can more effectively change their practices which are more procedural (content delivery) than behavioral (focus on pupil learning).

Theme - Mental Effort

Given that the reading program was introduced to most teachers six years ago, it seems reasonable that the teachers have completed the innovation-decision process from DOI. However, completing the process does not necessarily mean adopting the innovation. However, from the responses of the teachers and their corresponding classroom observations where they demonstrated implementing the curriculum, it seems they are at the confirmation stage. It is unclear whether the individual teacher innovation processes can reach the confirmation stage. We can conclude that the knowledge, persuasion, decision, implementation, and confirmation steps are complete. The lesson plans provided to teachers by JU detail the gradual release model, interaction with pupils, and pupil activities. However, the data suggests that while curriculum content was adopted, the gradual release model was only somewhat adopted, and the lesson plans were adopted. This is evident from the classroom observations, where every

teacher took the same heuristic shortcuts to implementation. The "*you do*" aspect of the lesson was the most frequent adaptation. Many of the teachers did not extend the problem, instead requiring pupils to copy what was written on the blackboard into their exercise books. During "*we do,*" the teachers generally asked a small number of competent pupils to provide input. While the teachers moved around the classroom as pupils completed their assigned tasks, the teachers only supported pupils as they chose to do so. In classroom observation, a child raised a hand to ask for help from the teacher. This solution to implementation by the teachers is referenced in DOI Theory, where innovation adoption can be achieved through adaption to context. This adaption can work in two ways; either making the implementation of the innovation effective or ineffective in context. Additionally, using behavioral economics theory, the *mental effort* can often impact positive decision-making. Individuals use Type I thinking to find a quicker but often less-than-perfect solution to a problem.

Consequently, a possible assumption to investigate was the idea that teachers "blended" the new instructional approach with their default approach such that they could be compliant and implement the new approach. However, most teachers commented that they did not modify or adapt the new approach but implemented it as designed. One teacher commented that there was no need to alter the program as the lessons were complete and that the pupils were the problem. The teachers who did modify the program seem to have made minor adjustments. For example, some teachers before the program used songs to engage pupils and help them memorize aspects of foundational literacy. This approach by the teachers of implementing curriculum content but retaining most of their default instructional approach suggests cognitive dissonance. Although they have only partially implemented the lesson plans, they believe they comply with the expectation to implement the program.

Sub-Theme: A familiar type of change?

While behavioral economics suggests our response to change often depends on the mental effort needed to adapt, and the power of the default; it seems that an area that is not explored or discussed in literature is the familiarity with the type of change needed. The teachers in the study were able to implement the phonics-based approach. When

someone starts their professional journey as a teacher either immediately in the classroom or at teacher training college, their priority is learning and mastery of the material they have to deliver. For many teachers, this ***type of change*** is a constant. For example, if a teacher changes grade, subject or textbook there is new content to learn and deliver. It seems teachers are conditioned to focus on new content through some combination of training, accountability and perhaps even human response given that change can be readily achieved through delivery. This idea is emphasized by Ackoff and Greenberg (2008) who argue that many education systems are flawed because they focus on teaching rather than learning. This aligns with Fullan's (2015) driver of the importance of learning at every level of the education system. However, the teachers in the study were squarely focused on change being a function of delivery of the phonics program. Their focus is quite understandable given the Tanzanian shift in policy from whole language literacy instruction to phonic-based instruction (described in section 2.8). We know that the *language wars* and the shift to local language instruction (Gear 2021; John 2023) have taken much of the bandwidth in the discussion regarding how children should learn to read. When the teachers commented on which aspects of the new approach they liked, they mostly commented on the phonics delivery and the interaction with children. For example, one teacher commented about what they liked about the training, *"How to teach phonetics. This was a very good training as teaching phonetics was always a challenge. I used to teach reading without knowing the sounds".* Note that the prior instructional approach did not included teaching phonetics.

Theme - Model of Teacher Change: confirmation of effective practice was an emotional, not logical decision.

Models of teacher change usually mention pupil assessment as how teachers assess if their change in instructional practice is effective (Guskey, 2002). Rogers (2003) calls this stage *confirmation*, where the individual has adopted the innovation and now looks for evidence to confirm that their adoption of the innovation was a good decision. Rogers claims this decision is fundamentally emotional, whereas Guskey's (2002) model relies on rationality. However, given that the majority of the teachers were unaware of their pupils' literacy levels, the idea of confirmation coming from irrational decision-making seems more salient than a logical model of teacher change. It seems

that teachers draw on different stimuli to confirm that their new teaching approach is effective, but not quality pupil assessment. One teacher said, "*I teach my pupils in a way that they all get involved and enjoy my lesson, and by the end of the year, they are all able to read*." This type of response, highlighting interaction with their pupils, was quite common from teachers. As one teacher commented, "*The role of teacher is to teach pupils, to take care of them at school and ensure the learning environment is good for them to learn*". This common teacher response possibly suggests that teacher confirmation of an effective instructional approach is gained from positive interaction with the pupils. However, the data does not provide conclusive evidence regarding the source of teacher confirmation. It is also possible that all the teachers get confirmation of their effective instructional approach from just the fact that they are delivering the new program, as required. It is also very possible that the source(s) of confirmation may vary by teacher. For example, one teacher commented, *"I enjoyed being introduced to new teaching approach of 'I do, we do, and you do' this approach is very useful as it helps me to engage students in the learning process"*. However, this teacher was observed not applying this approach when teaching, rather falling back to a lecture-based delivery. A different model of teacher change will be suggested in the discussion section.

Research Question 2: *What are the systems-level and social influences on a teacher's response to implementing an Early Grade Reading Program?*

Theme introduction of the new program - the system successfully introduced all teachers to the new reading program

All the teachers considered the program mandatory; they were expected to implement it. Teachers in LMICs receive a lot of training and resources from varied NGOs with different levels of collaboration and buy-in from local ministries. For teachers to consider a program mandatory, it should be a program that the education system supports through its structures; in schools this would be in the form of support and accountability from WEOs (coaches) and head teachers.

However, as explained with the analysis for research question 1, the idea of making something mandatory and introduced top-down was greeted with little resistance, at

least after six years of implementation. The classroom observation data suggested that the teachers demonstrated at least a basic understanding of the phonics-based program, indicating that there was a degree of effectiveness with a system implementing a program top-down. The program seemed to have reached all teachers, although some teachers reported that they only received training indirectly from colleagues who attended the training. However, this does not ensure quality implementation.

Theme: Most teachers described formal/informal active support systems

This finding focused on the deductive themes of social norm theory (Bicchieri and Noah 2017), diffusion of effective instructional approaches (Rogers, 2003), and a focus on the collective over the individual (Fullan 2015).

The teachers interviewed could describe many systems and mechanisms to support learning. The WEO and head teacher usually provides a support role within the education system in Tanzania.

However, when asked who supported their teaching instruction without reference to systems structures to achieve this, all the teachers mentioned getting support from many different sources, including the head teacher, WEO, peer colleagues, lead teachers, or COLs. However, every teacher described getting support. All the teachers expressed a reflective process for improvement; they were observed and given goals/pointers to improve on. Given teachers were seeking support using formal (systems) structures or informal, what social influences impacted their instructional approach? There are some suggestions of social norms, and the perception from peers that teachers will implement the program, for example one teacher commented, "*I do communicate with most of teachers who attended trainings and these communications help me to overcome some challenges in the masterly of the approaches were taught during the trainings also when I am asked to support my fellow teachers to master some parts of the training I am encouraged to continue to use the program.*" For teachers meeting with colleagues within their school, the experience of interaction varied depending if the teacher is interacting with teachers who have been introduced to the new reading program. For example, one teacher commented on the instructional approach of their teacher peers, "*No, they are not similar since I am teaching lower*

grades and they are teaching higher grader who are already self-aware of what they are doing.". However, the same teacher also commented on COLs, "*Yes, there is communications among us, these have made a big difference especially in the areas I find to be difficult I involve my fellow teachers from other schools, which helps me to continue teaching smoothly*". Teachers also commented on what they were getting help with. What is perhaps interesting is that they referenced getting help with delivery of the phonics approach. For example, *". . . initially, I did not have a good knowledge of letter sounds until I interacted with other teachers from the ward who helped me on this aspect and viewed it as a simple task. They assisted me and encouraged me to visit YouTube which has helped a lot in improving my teaching.*". What is unclear is what drives the need for assistance. Is it that the teacher acknowledges their lack of mastery when delivering the lesson? Or is it an acknowledgement of pupil mastery through assessment or interaction with pupils? This is unclear and important, given that instruction should be assessment informed.

Although formal and informal teacher works were described by teachers, the data did not provide enough evidence to suggest that teacher behavior was influenced by social norms (Bicchieri and Noah 2017). While teachers did reference recently formed teacher Communities of Practice within and across schools, there is no evidence that they created collective efficiency (Fullan, 2015). The data did had little description of the type of interaction between teacher peers.

Theme - Focus on Results: Expectations for pupil learning is known.

The Jifunze Ulewe program conducted education systems review in 2020. One conclusion was that the system did not set expectations for pupil achievement. Consequently, it was agreed and communicated (through the systems cascade) that by the end of grade 2, pupils should be proficient readers at thirty correct words per minute. Many teachers described how the Teacher Institute of Education (TIE) syllabus lesson plans described the pupil achievement and progress throughout the end of the year. All the grade 2 teachers interviewed correctly recalled these expectations. When asked if their pupils would achieve this learning proficiency by the end of the year, they replied that either most or all would be able to. For example, one teacher commented,

"*By the end of the year every child in grade two should be able to know how to read and write. In order to achieve this target, it the responsibility of teachers, parents, pupils and stakeholders to work together to make sure this is possible*". Only one teacher commented that expectations for pupil progress were unclear to them due to the school's isolation and lack of communication.

In comparison, The Tanzania National Early Grade Reading Assessment (USAID 2018a) reported the percentage of pupils in grade 2 able to read at thirty or more correct words per minute as 36.1%. It is a reasonable conclusion that the teachers were unaware of the achievement levels of many of their pupils. Given that the schools selected were low performing, this expectation by the teachers is quite unlikely.

Teachers all reported opportunities to observe other teachers instructing using the reading program. However, what was uncertain was if they could observe effective classroom instruction. Mostly, they would observe teacher peers in their school, but for a few, it was a teacher as part of their COL. A limitation of the secondary data is that while it indicated that teachers were assessing their pupils, there was no conclusion regarding the quality of those assessments. In other words, if the assessments were inaccurate teachers might have thought pupils were making progress. However, if the assessments did provide a realistic measure of pupil progress, how did the teachers process this?

What does seem evident that, aligned with other research (Loveless 2021), setting high expectations alone does little to improve learning outcomes. This leads to a question that when a program or local ministry set expectations, how do they expect schools to respond to the setting of high expectations? Fullan (2015), says successful change needs to be built around capacity building that is focused on results. Therefore, unless the link is explicitly made between teacher reflective practice and results, high expectations alone will clearly not be effective.

Theme – Innovation-Decision Process Social Norms

Part of the innovation-decision process (Rogers, 2003) is that teachers see if their peers are also implementing the program. This also aligns with social norms theory (Bicchieri and Noah 2017). The data confirms that teachers thought or knew others were implementing the program. Other than one teacher commenting that they did not know if

other teachers were implementing the program due to school distance and isolation, all the teachers commented that they knew other teachers were implementing it. However, it is difficult to conclude from this because the data only provided insights after six years and had limited information on teachers recalling their adoption process of the program. Therefore, it is difficult to conclude how social pressure may or may not have influenced teacher adoption practice. However, what is interesting is how similar each teacher's adaption of the program is. This might also have been influenced by CoLs or WEOs visiting classrooms and looking to observe the same characteristics. One teacher commented, "*I'm encouraged by colleagues in CoL meetings and there is the most experienced teacher in our school that encourages me in so many ways.*" However, the data does not provide detail beyond that teachers commented on how CoLs are useful. Therefore, although Fullan (2015) discusses group quality and learning, the data does not really detail how teachers collaborate other than teachers frequently mention getting support from CoLs over WEOs. For example, one teacher said, "*The CoL helps me to use these approaches and it gives opportunity to improve my teaching day by day. My colleagues encourage me as well because we meet once in every month to discuss the challenges we are facing.*"

Using DOI as part of the theoretical framework makes it necessary to consider the degree of teacher social networking and any influence this might have on teacher implementation of the reading program. Over half the teachers responded that they interacted with teachers from other schools through the CoLs. The frequency of these CoL interactions was reported to be mixed, with some meetings having occurred only once a year, to some regularly occurring twice a month. The support content was generally centered around implementing the reading program and discussing approaches and challenges.

The education system is structured to support teachers through head teachers and WEOs (external coaches) visiting schools. Additionally, the recent addition of teacher Communities of Learning (CoLs) across (Ward-level) and within schools provides another support source. Within schools, a lead teacher often leads in-school support. Key to the support received by teachers would be that they have regular classroom observations conducted with feedback provided. Fourteen of the seventeen interviewed

teachers responded that they had appropriate support, but the source(s) of this support varied. Most of the support described was through CoLs and head teachers, but no teachers responded they received support through WEOs.

This is a surprise because the WEOs are meant to be the primary formal support provided through the cascaded education system design. Many would be polite if the teachers were explicitly asked about instructional support from WEOs, given WEOs are part of the formal instructional support system. This question, therefore, emphasizes where teachers feel they get the support they need instead of the support provided.

Most teachers also describe how they found support in collaboration with groups of colleagues. Some collaboration is described within schools; some is defined as the Ward-level CoL. However, the data only left me asking more questions than providing answers regarding the social aspects of program adoption. The data lacks clarity on the quality and degree of the interactions over the past six years. This is a clear challenge of using DOI Theory within an education system. The system uses clear top-down channels to train, support, and communicate expectations. DOI Theory focuses more on lateral communication to diffuse innovations.

There is conclusion that the data analysis for research question 2 struggled to produce real insight into the question's themes such as Social Norm Theory and Diffusion of effective instructional practice. This will be expanded on in the discussion and conclusion section of this study.

Chapter 5: Discussion

USAID reading programs have been introduced for over six years in Tanzania, and the Tanzanian education system made the new literacy instructional approaches mandatory. This study concluded that teachers are implementing the program but learning outcomes in real terms have failed to substantially improve in most program schools.

This study's findings suggest teachers were able to "swap out" curriculum; replace the prior whole language approach with phonics-based instruction. They have been able to do this because of the training, lesson plans and other teaching and learning materials provided. However, what they have not fully changed is their focus on pupil learning, i.e., the pupil-centered aspect of the instructional approach. This is the behavioral aspect of teaching instruction. However, teachers respond to mandated educational change and act as the program implementation directs them to do so.

There are two key discussion points that focus on this study's contribution to theory, research methods, and the education system in Tanzania. Section 5.1 reflects on the importance of the research design used for this study, which focuses on stakeholder behavior in low-performing schools.

Section 5.2 expands on the evidence from this study that teacher behavior is fundamentally irrational, but we should be designing programs that account for this. The rest of the discussion section expands on these key discussion points.

It is important to recognize that this conclusion of teacher behavior being irrational is not to suggest a deficit with teachers in Tanzania. According to behavioral economics, these behaviors are human in nature and should be expected and accounted for (Ariely, 2009). Therefore, it is dependent on EGR Program implementors to design programs that account for researched and expected human behavior rather than presume that teachers lack capacity or that knowledge changes behavior.

5.1 It is critical to deploy impact evaluations that explore behavioral barriers of program design, not just focusing on "*what works.*"

The general approach of an impact evaluation is to determine if an education program had impact on pupils learning outcomes and then through varied research methods such as mixed-methods, positive deviance, or multi-variate linear regression find the associations between characteristics and actions of actors and improved pupil learning outcomes. For example, UNICEF Innocenti (Innocenti 2019) uses positive deviance research to understand the characteristics and actions of stake holders associated with improve pupil learning and how this can be replicated elsewhere. In 2023, the UK's Foreign Commonwealth and Development Office (FCDO) and other donors invested in a *What Works* Hub for Global Education (Blavatnik School of Government 2023) focused on the science of implementation, building off Gates Foundation's Science of Teaching (n.d.). All these approaches are focused on an optimistic approach, which according to behavioral science (Kahneman, 2009) is quite intuitive.

However, the work completed for my IFS (King, 2020) concluded that what is not often measured in these pockets of success are the positive personality characteristics of the stakeholders, associated with the characteristics of Rogers' (2003) early implementors. This is reinforced by additional IFS (King, 2020) research of EGR Program data that suggested about 80% of the program impact is explained by about 13%-15% of schools demonstrating improved learning outcomes. By scaling these pockets of success, we presume that all individuals need to do is act and behave like others who are successful. However, those who are successful are, by definition (i.e., positive deviance) not normal and have many personality characteristics not associated with the general population. The teachers and other stakeholders who are unable to translate the program implementation into impact on learning are the majority, the typical individuals whom the program should be focusing on understanding better. Therefore, the issue rests with program design and implementation, not the deficiency of stakeholders.

Therefore, this study used a research design of focusing on schools where learning outcomes were poor, which is an approach from behavioral economics. While having

participants who were not just teachers would have been beneficial, the secondary qualitative data available to me was just from classroom observations and teacher interviews. Finally, this study focused on behavioral barriers to positive decision-making, rather than just challenges such as socio-economic status or class-size, given the condition that those schools who did have impact were implemented in similar conditions. Critically, the study focused on the *why* of teacher behavior.

At no point should the findings of this study be interpreted as teacher deficiency and blame; rather we do not design programs that understand and account for *expected* human behavior and how to encourage positive decision-making. By focusing on those pockets of success, we fail to understand this challenge.

So, why are many researchers in international education so keen to focus on what works? One possible issue likely comes from researcher optimism bias, which is defined as when an individual's expectations are better than reality (Sharot 2011). This combined with the need to continually win program implementation awards from donors (such as USAID and FCDO), where a scoring condition for procurement is to look at past evidence of success of bidders before awarding the bid. The need to focus on and highlight success is central to much of the international education sector work.

However, I do not think that looking at what works and behavioral barriers to good decision-making are mutually exclusive when we consider the research design of an impact evaluation. It could be the case that using just one single approach likely presents a limited understanding of education programming. It could be argued that both present a bias by focusing on either optimism or pessimism. It is not difficult to add a mixed-methods component to an impact evaluation that studies both the components of effective implementation and the behavioral barriers where impact is low. For example, Kahneman (2009) recommends that even when projects go well, a postmortem approach is recommended, which still looks at failures and challenges even if a project is an overall success.

5.2 EGR programs should be designed to account for contextually researched models of teacher change, not just idealistic models.

The Guskey (2002) model of teacher change is as discussed in section 2.1.3 uses a model of logical decision-making; the teacher confirms effective instructional practice through improved pupil learning outcomes. In this way, the use of logic is very similar to the standard economic model of human behavior. The Guskey model has been the standard model used for EGR Programs since their inception. However, prior to this study there has not been any research in international education to confirm this model is appropriate for context. Instead, this model of teacher behavior has been borrowed from USA domestic research and presumed to be true. However, after the introduction of EGR Programs using the Guskey model, he has more recently suggested that the approach of "*logic, reason, and philosophical arguments*" (Guskey 2020, p. 19) will fail to change teacher behavior. This is something also emphasized in behavioral economics (Kahneman 2009), which suggests that simple persuasion is a flawed model to change someone's beliefs.

However, persuasion, logic and reasoning are the fundamental approach of the teacher coaching system in Tanzania and many other LMICs. Coaches that visit teachers and focus on classroom observations, not improved learning, as Guskey (2020) suggests will not change teacher behavior. However, the findings of this study additionally suggest that teachers do not even use learning outcomes to improve or confirm instructional practices; rather because the program is mandated, teacher behavior changes due to confirmation from a variety of sources and irrational human behavior is the norm. The analysis suggests that just implementing phonics-based literacy is confirmation of effective instructional practice for many teachers. Given how EGR programs package this approach as the answer to low-literacy rates, why wouldn't they? However, we know that just swapping curricula is not enough (Burtless 2019). Without improving quality time-on-task for pupils, phonics-based literacy instruction will be ineffective.

Behavioral economics suggests that human decision-making is fundamentally irrational (Ariely, 2009). However, it also suggests that human irrationality can be predicted and thus accounted. Given the findings of this study suggesting that irrationality exists with teacher behavior, the conclusion section of this study will suggest future directions for

implementation design. Central to this discussion is the questioning of coaching model of implementation and teacher support against potential alternatives. This study cannot assume that if implemented effectively the EGR program would have much greater impact. However, the exercise of applying literature and this study's research to improve program implementation is a useful contribution.

Finally, although the findings of this study could not fully address the second research question regarding the social aspect of teacher influence and learning, this aspect remains perhaps the area where behavioral science DOI Theory (Rogers, 2003) and social norms (Bicchieri and Noah 2017) and education research on the importance of group learning (Fullan 2015) seem to closely align. However, as systems in Tanzania and other LMICs focus mostly on cascading teacher support through external coaching (e.g., WEOs in Tanzania). As such, literature and best practice in for CoLs and other teacher interactions in LMICs, both formal and informal, is weak. This will be expanded on later in this discussion.

5.3 Did changing literacy instructional approach just create a distraction?

Effective classroom instruction is a complex sequence of interrelated actions, interactions, and activities. Just switching to phonics-based literacy instruction will do nothing to address learning outcomes unless all components of effective classroom instruction are addressed.

It seems reasonable that the introduction of a new instructional approach for literacy is that delivery process becomes the central focus of change. This then aligns with the type of change that is familiar to teachers. When a teacher begins their career, they learn content and how to deliver it. Frequently, teachers will be required to teach a new subject or grade, or receive new textbooks. In each case, the teacher learns new content and curricula. This change, while sometimes imposed and a burden on teachers' mental effort and time, is something they are familiar with and as such are least likely to resist. Focusing on delivery is also something more easily trained and coached. However, changing all components of the instructional approach, including focusing on pupil learning, requires behavior change, a change in instructional philosophy.

If I recount my own fourteen-year teaching career, I learned new curricula or content thirty-four times. However, only twice did I change my instructional approach; moving towards a much more pupil-centered instructional model and moving away from teacher-centered approaches that was central to my teacher training and first few years of teaching. Whenever I changed my instructional approach, it was under my own motivation and never mandated. Still I found this self-imposed behavior change difficult; I was consciously trying to change classroom habits that had been normalized over my first few years of teaching.

When Common Core was introduced in the USA, it failed to meaningfully improve learning outcomes (Loveless 2021), but again there was plenty of evidence that teachers implemented the curricula and used the teaching and learning resources as provided (Loveless 2021). In an interview with the Washington Post, Loveless stated "*Standards-based reform succeeds by changing what schools teach and how they teach it - by changing behaviors, not by writing checks.*" (Strauss 2021). Again, if there is evidence what teachers changed what they taught, is the issue how they taught it? This might be a useful assumption only if The Common Core as designed had a high degree of efficacy compared with what was taught prior.

Detangling this issue is also relevant with basic education reform in Tanzania. Over the past 11 years there have been over $500m of donor investment into basic education in Tanzania from USAID and FCDO alone. While programs have had statistically significant impact, in real terms learning outcomes have been disappointing (Ruddle and Rawle 2020). A question remains; given all the challenges of basic education in Tanzania, was switching to phonic-based literacy instruction just a distraction given all the other aspects of classroom instructional practice that needed change? Was switching from whole language instruction to phonic-based instruction really the low-hanging fruit to improve instructional practice in Tanzania? Unfortunately, we do not have an answer to that question.

5.3 Why did teachers did not know the reading skills of their pupils?

Perhaps the most significant finding from the study is that grade 2 teachers believed that most or all of their pupils would be proficient readers by the end of the year,

aligning with systems expectations that pupils read proficiently (45 correct words per minute) by the end of grade 2. However, we specifically selected schools with low achievement. Nationally and in districts from which the low-performing schools were established, most pupils do not achieve the expected proficiency level (USAID 2018a; 2017). The percentage of pupils reading proficiently at the end of grade 2 is never over 10%. Therefore, it is reasonable to conclude that the teachers in the study do not know the literacy achievement level of many of their pupils. This conclusion is also confirmed by a recent World Bank study, which also stated that teachers in India and Bangladesh overestimate their pupils' performance (Djaker, Ganimian, and Sabarwal, 2022). This paper concluded that teachers in Bangladesh particularly overestimate the achievement of low-performing pupils and underestimate the achievement of high-performing teachers. However, with more low-performing pupils in both Bangladesh and Tanzania, teachers likely overestimated a much higher percentage of pupils in their performance. This thesis study adds the lens of confirming that teachers understood the expectations of pupil performance by the end of grade 2 and the mistaken belief that most or all of their pupils would achieve these expectations. Our natural programming conclusion is that the go-to remedy for this issue is formative assessment. It would serve the purpose of tracking pupil progress. However, formative assessment is a useful tool only if the individual pupil data is useful and is used to adapt pupil learning. Differentiation and remediation immediately come to mind as the means of adaptation. The Tanzanian education system’s core approach to formative assessment is pupils writing into their exercise books during class and teachers moving around the class and checking for understanding. However, in class, pupils mostly wrote down the information on the blackboard if they had an exercise book to write into. Sometimes teachers were observed extending the problem to the pupils and checking for understanding and correcting work in the exercise books. So, why the disconnect? On the surface, essential components of a functioning education system seem in place:

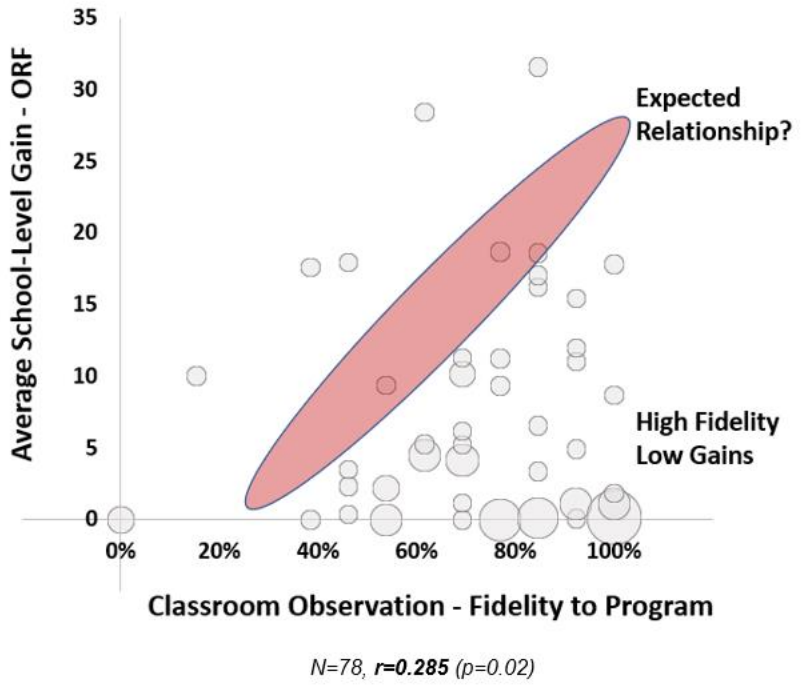
- Teachers are reflective practitioners
- Teachers are observed and set goals
- Expectations for learning are set
- Teachers receive support

- Teachers monitor pupil progress

This might be an example of isomorphic mimicry (Lantt Pritchett 2011). This theory suggests an aspect of organizational survival is *legitimacy*, where the focus on the institute is on "*compliance or, worse, worker characteristics or connections that are irrelevant . . . for performance*" (Andrews 2017, p34). In other words, there are expectations that pupils should be proficient readers by the end of grade 2, and this is a means to achieve and demonstrate success – but it is not used for this purpose. Instead, compliance occurs through classroom observations, which seem reasonable, provided the measures are of quality instruction. When I look back at my IFS, I produced a specific graph that at the time I struggled to qualify. It was a school-level scatterplot of average classroom learning gain (reading fluency) versus classroom observation percent score (Exhibit 13).

Exhibit 12: School-level average school gain versus classroom observation percent score,

Nepal Early Grade Reading Program, 2018



(King et al., 2022)

The diagonal cigar shape shows the positive linear relationship we might expect; as the classroom observation fidelity to the reading program increases, so makes the impact on learning outcomes. Pearson's correlation for the actual data is r=0.285 (p=0.02). However, the exhibit clearly shows the majority of the points in the bottom right-hand corner, with larger bubbles explaining the same scores for multiple data points. This pattern gives high classroom fidelity (i.e., compliance) percent scores for a low or no learning gain. In Nepal, the classroom observation tool used binary indicators to construct the overall index score; each observed behavior carrying the same weight:

- Teaching phonemic awareness
- Teaching graph phonemic awareness
- Fluency modeling
- Pupils read aloud
- Vocabulary
- Reading comprehension
- Writing
- Equity
- Grouping
- Pupil participation
- Feedback
- Monitoring
- Print rich environment

(King et al., 2020)

When a teacher is observed, they are usually more motivated and as a consequence of the accountability (perceived or actual). The challenge with this type of observation is that it is very difficult to consider quality with a binary or likert scale, which tends to confirm compliance. Additionally, we need to consider that teacher behavior might be very different under observed conditions compared with the day-in day-out instructional practice – aka "the observer effect" (Merrett 2006; Raymond G. Miltenberger 2012). These tools for teacher observations are commonplace in EGR Program. USAID's *Classroom Observation Toolkits for Early Grade Reading Improvement* (Pflepsen, Hertz, and Kochetkova 2019) provides similar recommendations for teacher observations, which provide the following illustrative measures:

- Lesson structure
- Classroom management
- Classroom environment (e.g., print-rich)
- Supportive learning environment
- Pupil participation
- Teacher checking for understanding
- Pupil feedback given

(Pflepsen, Hertz, and Kochetkova 2019)

Filmer, Wane, and Ezequiel (2020) highlight the challenge of measuring quality using classroom observation tools and comment that inter-rater reliability can be low, especially regarding subjective measures of quality instruction. They found that common themes were poorly correlated when comparing four different teacher observation instruments. Then, similar to my IFS findings, none of the instruments produced measures predictive of pupil test scores. As discussed later, this might indicate the irrelevance of the data collected compared with the teacher observation, coaching, and dialogue process. This certainly aligns with my experience as a school administrator in a charter school in the USA, where the county where I worked in North Carolina had a classroom observation form I had to complete three times a year for teacher Continuing Education Units (CEUs). The teachers were not interested in the form due to its lack of utility; instead, they wanted a narrative on what I observed and subsequent discussion. However, how the rural teachers interviewed in Tanzania respond to these observations is unknown. They might ignore them as the teachers I interacted with in the USA did or reacted to them as an accountability measure with a "good" score indicating their approach is satisfactory. Most countries (not just LMICs) have teacher classroom observation forms similar to these. Whatever I communicate regarding their utility, it is quite challenging for implementors, donors, and local ministries of education to move away from using them because of the human desire to measure and intervene in the observable effect, as Lewin (1997) commented. Even when presented with evidence of the lack of utility of these types of tools, it would be hard for institutions and local ministries to drop using them and taking-up another

approach. This goes back to Pritchett's suggestion that it is important for individuals working with education systems to ensure everything looks legitimate (Pritchett 2011). These tools certainly add that lens. However, if these tools give "passing grades" to teachers with ineffective classroom approaches, then that would indicate these tools might be harmful and that taking a risk and trying something different might be preferable.

WEOs and head teachers in Tanzania conduct these classroom observations. These individuals periodically travel to schools and observe individual teachers. Although we know these tools are used, we do not see the role these forms play in the coach and teacher dialogue. In my experience in the USA, these forms were mostly ignored and filed for compliance for continuing education credits for teachers to renew their teacher licenses. But they could not drive a quality dialogue between myself and the teacher focused on pupil learning. So, are coaches in Tanzania able to ignore or leverage these forms in such a way that the dialogue between themselves and the teacher? Data or research does not exist to answer this question.

5.4 Is setting high expectations enough?

The expectation for early grade literacy in Tanzania is that pupils should be able to read at 30 cwpm by the end of grade 1 and 50 cwpm by the end of grade 2. These are ambitious goals considering the low literacy rates in Tanzania (USAID 2015). So, the question becomes, is it enough for systems to set high expectations? The Brookings Institute argues that having high expectations is not enough to drive improvement and that systems that set ambitious expectations without considering pupil achievement when entering grade school set the system up for disappointing results (Loveless 2021). However, setting high expectations for pupil achievement is unavoidable as the minister of education is generally a political appointment, and the emphasis on the UN's SDG indicator 4.1.1, "*Proportion of children and young people (a) in grades 2/3; (b) at the end of primary; and (c) at the end of lower secondary achieving at least a minimum proficiency level in (i) reading and (ii) mathematics, by sex*" (United Nations 2019). How are teaching tracking pupil progress? It seems that there are two main methods:

- checking pupil exercise books during class, and

- informal observations of pupil progress and “fluid intelligence” (Djaker, Ganimian, and Sabarwal 2022), which is the individual process of reasoning and problem-solving (Góngora et al. 2020)

Additionally, the USAID reading program has not introduced to teachers an assessment by which they can track pupil oral reading fluency progress to understand if they are on track to achieve the expected reading fluency benchmarks at the end of grades 1 and 2. This seems the obvious solution: provide teachers with formative assessments to track pupil progress and adjust individual pupil learning. However, most education systems in LMICs focus on top-down expectations of teachers progressing through lesson plans and teaching the class in a prescribed manner (Djaker, Ganimian, and Sabarwal, 2022). That we have very little quality evidence about. We have information regarding coach training from RTI International and other implementation institutions. The Jifunze Uelewe program studied coaching effectiveness for operational research, concluding that:

- WEOs spend more time on administrative tasks than supporting teachers on pedagogical issues,
- only one-third of WEOs coach teachers on regular school visit

(CIES 2020)

This Jizunze Ulewe further identified a key challenge of behavior change: shifting the role of coaches from giving instructions to being supportive (CIES 2020). This research emphasizes the understanding of moving beyond coaching as a means of top-down knowledge transfer, highlighting the need for relationship building between coach and teacher that represents a supportive and creative partnership.

5.5 Is the solution as simple as adding formative assessment?

Tanzania's education system relies on teachers checking pupil exercise books as formative assessments. There is an apparent disconnect that the formative assessment is written (mostly copying as observed), but the system's expectation of pupil learning is an oral demonstration of pupil reading skills. However, expectations are that pupils are proficient in oral reading by the end of grade 2. It is evident that the teachers do not

know their pupil's reading skill level. However, there seem to be two capacity issues at play here – that the education system and the teachers are focused on implementing (or delivering) the phonics-based approach with the assumption that this change of content will provide the needed quality instructional process. Nearly all the teachers interviewed in the study were content, believing this new approach was effective.

So, is the answer a formative assessment focusing on the oral assessment of pupil reading skills? This is the approach championed by Teaching at the Right Level (TaRL), developed by the Indian NGO Pratham. The TaRL approach is to assess children using an ASER reading assessment tool, like EGRA, a one-on-one assessment of reading fluency. However, the ASER is a formative, summative, or national assessment tool. It has four subtasks: letters, words, a grade 1 level reading passage, and a grade 2 level reading passage. Pupils are graded according to the level they can complete comfortably (ASER Center 2022). Unlike EGRA, the ASER can also assess individual pupil literacy skills. This is exactly how TaRL leverages the ASER tool. Pupils are assessed using the ASER tool and then grouped by different means to provide the reading skill instructional support they need at the reading skill level they are performing at. This seems like an excellent pivot from the whole-class curriculum delivery observed in this study to individualized pupil learning. However, I would suggest that pivoting to a single solution ignores the complexity of the challenge faced by education systems in LMICs. This argument is made by looking at the impact of TaRL programs and USAID Early Grade Reading Programs. The impact of the TaRL programs range between 0.08 – 0.75 SD (Teaching at the Right Level 2022a), while the impact for Early Grade Reading programs in LMICs range between 0.13 and 0.80 (Piper and Stern 2019). While it is difficult to compare the impact of programs that are often designed in very different ways, with different durations and funding levels, it seems too quick to judge to pivot USAID reading programs to refocus efforts. While there are many issues in LMICs that are barriers to learning, such as gender discrimination and socio-economic factors, I will touch on considerations from the basis of this study's research framework. The first of these is regarding social norms. To refer back to the basic definition of social norms, it is about perception. We adopt behaviors based on . *". . . what we believe we others do, … what we believe others to think we should do*" (Bicchieri and Noah 2017, p.6).

Consequently, if an education system and teacher norms center around curriculum delivery and are believed to be successful, what does it take for an entire teaching population and system to shift its approach to pupil-centered learning? For a social norm to change, it has to be replaced by another social norm. However, can an education system successfully introduce a new social norm from the top down, as is the approach in many LMICs? It is hard to find evidence supporting this. A top-down approach has successfully improved education outcomes in Vietnam; however, many of the factors of Vietnam's progress are sociocultural and hard to replicate in other LMICs (Kataoka et al. 2020). Many top-down approaches to education reform have documented struggles, such as the introduction of Common Core in the USA (Loveless 2021). Fullan (2015) suggests that the issue at play here is a fragmented approach to education change, where many education systems are piecemeal in their implementation and use incorrect drivers of change. If we presume that education change is complex and has many components, most interventions in international education focus on a few explicit components. The first iteration of EGR Programs focused on three; teaching and learning materials, teacher training, and support. However, the second and third iterations of EGR Programs in many LMICs are now focusing on systems strengthening, using the materials developed in the first iteration. The reason is the first iteration of EGR Programs had a moderate impact (Piper and Stern 2019). However, many systems have focused on accountability to enforce the correct use of teaching and learning materials. A good illustration of this is Ghana. Using the Research on Improving Systems of Education (RISE) Education Systems Framework (Kaffenberger and Spivack 2022) and donor support, the Ghana Ministry of Education (MOE) and Ghana Education Service (GES) produced the Education Accountability Framework (EAF) ("Ghana: Accountability for Learning Framework" 2022). This accountability framework predictably uses the premise of holding systems actors (e.g., WEOs, head teachers, teachers, etc.) accountable to a series of indicators and will use an accountability dashboard by which to achieve it. This accounting approach is not new and has been critiqued (Honig and Pritchett 2019; Fullan 2015) as an incorrect focus for education systems. So, how does a systems framework like the RISE turn into an accountability system? This comes back to the idea of isomorphic

mimicry (Lantt Pritchett 2011), where LMIC education systems look like modern states but *“. . . mask a clear inability to actually implement incrementally more complex and contentious tasks.”* (Andrews et al., 2017, p.4)*. Additionally, Fullan (2015) describes accountability as a wrong driver educational change, commenting that “People are rarely motivated by being judged, and impossibly so if the judgment does not contain possible lines of solutions*” (Fullan 2021, p.12). Fullan goes on to argue that the drivers of change should instead be well-being and learning at all levels of education. This is something I will return to later.

It is all too easy to demonstrate the failings of LMIC education systems. However, US researchers should be looking in their backyard regarding failed attempts at education reform. The Common Core (Common Core 2022) is an excellent example. Common Core is effectively a top-down developed and implemented standards-based reform. The newly developed standards guided curriculum development, pedagogy, assessment, and accountability (Common Core 2022). However, *“. . . no convincing evidence exists that the standards had a significant, positive impact on pupil achievement*” (Loveless 2021). So, why? First, like with this study, there is comprehensive evidence that teachers implemented the curriculum. This can be easily seen through online teacher websites, assignments that pupils bring home, and so on. So, over time, schools and teacher resistance was not major factor in the failure. A mirror of one of the findings of this study regarding Tanzania. However, according to Loveless (2021), the approach was top-down and regulatory with *“. . . the illusion of a coherent, well-coordinated system*” (Loveless 2021). And so, although USA and Tanzania have similar educational improvement challenges, the USA is still miles ahead in terms of *achievement*. This is due to stronger inputs such as teacher qualifications and training, resources, and socio-economic factors.

5.6 Teacher perception of the EGR Program after six years of implementation

Perhaps the weakest aspect of this thesis study is the teacher's perception of the EGR program. After six years of USAID program support for early literacy, the teacher feedback regarding the program is very positive. Taken directly from DOI theory, the teacher's perception of the innovation is critical for adoption. The individual teacher

explores and identifies positive characteristics of the EGR program to support their decision to start and continue to interact with it. The teachers had an overwhelmingly positive response to the program. In summary, their perception of the program was that it was superior to their prior approach, easy enough to understand and implement, mandatory, observable (i.e., observing colleagues implement the classroom approaches), compatible with the prior approach, and adaptable for context.
As commented earlier, all the teachers observed were implementing the program, contrary to my earlier assumption that the low impact on learning was due to teacher resistance.

However, it is fair to assume that the teachers' perception of the program in year one of implementation was likely very different from their perception after six years. The likely explanation is that teachers tend to defend their classroom instructional approach. Therefore, as they gradually incorporate the EGR Program content into their classroom approach, they will defend the program as in doing so, they are also defending their classroom approach. If true, this highlights the importance of the approach being perceived as being mandated.

Finally, it also highlights the weakness of my driving assumption that teachers resisted implementing the program. They might well have done in year one, but interaction with the EGR Program changed their perception of the program over time. The mistake I made with the assumption is common; I was thinking about a fixed point rather than describing what might happen over time. Relating this to my experience observing teachers' responses to the introduction of the National Curriculum in England, there was resistance at first. However, acceptance was commonplace years after teachers had adopted the curriculum.

5.7 Revisiting Fullan's drivers of educational change

Fullan's work is interesting because it describes not only drives of effective change but also describes factors that, while important, are frequently misunderstood as critical drivers (Exhibit 14).

Exhibit 13: Right Versus, Wrong Drivers

Right	Wrong
Capacity building for results	External accountability
Collaborative work	Individual teacher and leadership quality
Pedagogy	Technology
System-ness	Fragmented strategies

(Fullan, 2015, p.42 Figure 3.1)

We encounter challenges if we explore the 'right' drivers of educational change. For example, if we examine the capacity building for results driver and define results as an improved instructional practice that directly impacts pupil learning and understanding.

The data from this thesis study implied that teachers and other stakeholders in the system were focused on improving their practice. Teachers in the study reported this. However, what was not reported was that they were working on the aspects of their practice that, if addressed, would provide the greatest return on improved classroom instructional practices.

Additionally, if results improve pupil learning and understanding, teachers assessing this through checking pupil exercise books during class that rote information has been copied will fall short. Not that this approach is accepted practice by the education system.

The teacher support approach was fragmented and limited to one-off observations and 'pointers.' Some teachers mentioned some level of collaboration through interactions with their peers or CoLs. However, while it was good that these collaborations existed, they did not seem to focus on results.

5.8 Cognitive Dissonance

DOI theory suggests that during the confirmation of the adoption of an innovation, human behavior change is partially explained by cognitive dissonance, where individuals seek balance in their state of mind by changing their attitude, knowledge, or behavior (Rogers, 2003). In our study, the teachers were content with the EGR Program, discussing how adopting it improved their instructional approach. Consequently, the teachers achieved equilibrium, aligning their knowledge and attitudes

with their actions. However, they adopted the phonic curriculum, keeping most of their prior pedagogical practices to accomplish this. This approach is described in behavioral economics, where individuals seek a shortcut heuristic to reach an end-point with a reduced mental effort (Kahneman, 2011). We should have expected this in a system that holds individuals accountable for observable classroom behaviors without measuring the quality of classroom instruction.

Consequently, all the teachers in the study prioritized implementing phonics-based content. This is what they were chiefly accountable for through the system. If there observed not using the phonics content, it would show when being observed. However, effective pedagogical approaches are harder to measure and pin down. Teachers are usually adaptable to changing content; a teacher might teach grade 2 one year and grade 4 the next. They tend to replicate their pedagogy while learning new material. However, the successful adoption of pedagogical practices involves a sustained mental effort, and it seems that the teachers kept their default pedagogical approach, which they say they learned at teacher training college. This is what the system holds them accountable to do. Fullan (2015) comments that these compliance measures provide the stakeholders with a system of assurance that individuals are performing in line with expectations. However, he points out that external measures must be aligned with the development of strong internal accountability.

Guskey's (2002) model of teacher change emphasizes the teacher changing their classroom practice, seeing changes in pupil learning outcomes, which then leads to a change in teachers' beliefs and attitudes (Exhibit 15).

Exhibit 14: A Model of teacher change

Guskey (2002, p.383)

Guskey's model is attractive because he suggests that a change in teacher beliefs and attitudes results from improved learning outcomes rather than the cause of the improved learning outcomes. This reordering of attitudes and beliefs after teachers see

evidence of improved learning outcomes is also proposed by Michael Fullan (2006) and DOI's confirmation stage in the innovation-decision process (Rogers, 2003). However, this model is based on research, observations, and interactions with teachers in the USA and Canada, where it was found the correlation between teacher estimations of pupil achievement and actual pupil achievement was r=0.63, compared with r=0.07 for Bangladesh (r=0.07), and r=0.36 for India (Djaker, Ganimian, and Sabarwal 2022). Additionally, the teachers in this study believed that most of their pupils would achieve the reading proficiency benchmark, while external evaluations of pupil learning in Tanzania suggested otherwise (USAID 2015). So, while there might be an argument for Guskey's model of teacher change being at least feasible in the USA and Canada, it becomes much more of a non-starter in LMICs because teachers are consistently overestimating the performance of their lower-ability pupils, who often composed of the majority of their classes. I find there to be an issue with the Guskey model. Firstly, Guskey's model relies on teachers observing improved learning outcomes. However, teachers' implementation of new approaches is often far from perfect.

Consequently, improved learning outcomes might not be that evident, especially if there is no benchmark with which to compare. The support system that needs to be in place or the teacher's perseverance must be robust in understanding that the 'results will come.' However, this contradicts the Guskey process, where observing results is key to improvement. Then, the process of change described by Guskey is fundamentally logical; the teacher will continue if they get results in pupil learning outcomes. However, the theoretical framework for this study suggests that change is an emotional process (Rogers 2003; Fullan 2015; Kahneman 2011). Behavioral economics suggests that humans are quickly exhausted from mental activities that demand complex computations, and often the emotional system of decision-making takes over when a decision is required (Kahneman, 2011, p.20). This suggests that Guskey's logic of teachers looking for quantitative evidence is perhaps not as likely as thought. For example, in my teaching practice, when trying a new classroom instructional approach, I would look for immediate confirmation from my pupils that the new approach was worthwhile.

5.9 Why were the teachers content?

This study showed that the teachers believed that the new approach was superior and would enhance their teaching. However, as noted, they only changed the content, not the pedagogy. This was perhaps my biggest surprise. I had expected some resistance to implementation or even consternation with having a new teaching approach imposed. The key to the teachers being content is when this study was conducted and the fact that they perceived the program as mandatory. There are many different models of change. However, most individual teachers will also respond very differently to change, often due to age and personal orientation to change (Hargreaves 2005, Rogers 2003). However, there is little research on the emotional aspects of teachers' lives (Sutton and Wheatley 2003). Instead, research on teachers focuses on beliefs, attitudes, and practice (T. Guskey 2002). However, psychology has stated for a while that change is fundamentally emotional (Gross 2005), which has then been iterated by behavioral economists (Kahneman 2011; Thaler and Sunstein 2009) and education researchers concerned with change (Fullan 2015) – although the research is not explicitly focused on emotional response. So, to understand why teachers were content with the reading program after six years of implementation, we have to draw from the literature regarding how individuals respond to change.

It is suggested that the best way to get individuals to embrace change is to emphasize what will not change (Venus, Stam, and Knippenberg, 2018). This reduces one of the main anxieties of change, namely loss (Gross 2005). So, the more teachers can retain aspects of their prior approach, the more comfortable they will be as they change aspects of their teaching. The teachers in this study altered the curriculum but mostly retained their pedagogical practice. As described earlier, changing the curriculum is normal for many teachers. It is also what they are accountable for achieving when observed. However, changing pedagogy is far less measurable and understood by those who observe them. It is also a considerably greater mental lift. Consequently, by retaining pedagogy, the teacher could be relieved by the more accessible mental lift and be able to retain something familiar, thus reducing overall loss. Contentment is achieved through cognitive dissonance. There might have been anxiety as teachers were

introduced to the reading program. Teachers are aware of the mandate to change, so finding this compromise of implementation can find a mental balance. When we meet the teachers during this study, there has been a process that led to the implementation of the curriculum. Emotional balance is achieved, and contentment is the result. So, what is the involvement of the WEOs and head teachers in this process? This is beyond the scope of this study, but something important. Speaking as a former administrator who has had the responsibility of facilitating change

5.10 International Education and Barriers to Implementation

Most of the donor-funded work in international education focuses on explicit intervention components. EGR Programs are an excellent example of this with the general approach of material production and distribution, teacher training, and teacher support. When these opportunities are released for bidding by qualified implementors, the Reason for Proposal (RFP) document will almost always include a Theory of Change (ToC) statement. The 2022 solicitation for USAID Ghana Strengthening Accountability in Ghana's Education System (SAGES) has a ToC which states, ***"IF** the enabling environment for accountability is established and **IF** system actors perform at minimum standards and interact effectively, **THEN** accountability will be institutionalized in the Ghanaian primary education system, which will lead to improved education service delivery."* (USAID 2022b). This ToC then leads to a Results Framework (RF) – a cause-and-effect diagram of results that will support the theory of change. Finally, this leads to describing activities that will achieve the intended results. The ToC shown above is very focused on supporting and delivering accountability. While the implementor has some discretion in program design, the ToC, written in the RFP, is the program's cornerstone. The challenge with this type of design is that it is very focused on observable implementation and effects, simplifying barriers to implementation. Returning to Lewin (1997) and Kahneman (2011), this program design is intuitive for humans. If there is a problem to solve, you add driving forces to resolve the issues. However, as this thesis study suggests, there is a social norm of classroom behavior that the teachers have reverted to with support from the system.

5.11 Is the challenge in Tanzania an isolated issue?

The issue uncovered by this study is not necessarily new in other countries and contexts. The issue is not just even isolated to LMIC education systems. We can take the example of the USA. Common Core standards-based reform was released in 2010. As with other standards-based reforms worldwide, it consisted of a top-down implementation of new pupil learning standards that guided curriculum development, pedagogy, assessment, and accountability. According to the Brookings Institute, no substantial evidence exists that Common Core positively impacted pupil achievement (Loveless 2021). Brookings details key findings to conclude that there were no considerable evidence teachers and other stakeholders resisting the implementation of Common Core. Indeed, there is plenty of evidence online that schools and teachers have introduced new curricula, content, and approaches. Like Tanzania, the approach was top-down and regulatory. Brookings main concern with Common Score is that coordinating design and implementation at the top of the system limits the discretion of schools and teachers to tailor instruction to serve their pupils (Loveless 2021). Finally, Brookings comments that with Common Core, there was an illusion of a well-functioning education system. This clearly parallels Pritchett's (2011) Isomorphic Mimicry, which also discusses how education systems strive to look legitimate. However, how can this definition apply to both USA and Tanzania systems with vastly different results in terms of learning outcomes? So, the similarity between the USA and Tanzania's education systems is that both struggle to show improvement in terms of learning outcomes. However, the basic inputs into the education system (e.g., teacher training, teacher qualifications, resources, pupil backgrounds) are very strong for the USA, differentiating itself from Tanzania in terms of pupil achievement. Therefore, when applying the concept of Isomorphic Mimicry to an education system, we should consider it in terms of the system's ability regarding pupil achievement or pupil improvement.

5.12 Reflection on the theoretical framework

The three theories that formed the basis of my theoretical framework (DOI, Behavioral Economics, and The New Meaning of Educational Change) helped add a lens to develop an understanding of different, unseen factors at play. The analysis approach of

searching for counterfactual information, such as teachers believing their pupils were readers at the end of grade 2 when it is quite likely the vast majority of their pupils were not meeting this system expectation. This establishes the motivation behind the behavior of system actors. While education researchers discuss a learning crisis (Piper and Stern 2019), stakeholders within a system are often content, believing they are doing all they can to support pupil learning (Sabarwal and Abu-Jawdeh 2018). This dissonance in perspective is as essential as an EGR Program design. While the term "system strengthening" is frequently used in education research, I would posit that it is more likely a top-down LMIC education system will not be able to produce results based on its design.

Consequently, focusing on system strengthening will attempt to prop up a system that does not work. However, behavioral science and the idea of isomorphic mimicry suggest that leaders and policy makers within an LMIC education system are unlikely to be willing to take huge risks and redesign a top-down education system. The risk to change a system to focus on mitigating unobservable effects such as social norms, seems unlikely. Education systems either introduce new innovations top-down or locally (districts, schools, teachers). Allowing a system introduce innovations (e.g. pedagogy) in schools suggests a higher buy-in, but a very slow diffusion process of the innovation (Hung et al. 2017). The top-down introduction of the innovation suggests quick diffusion but less buy-in and correct usage at the school level (Hung et al. 2017). Therefore, what could be explored is the idea of a top-down introduction of new pedagogy but then employing diffusion once it is introduced.

5.11 The next generation of EGR Programs and research alignment

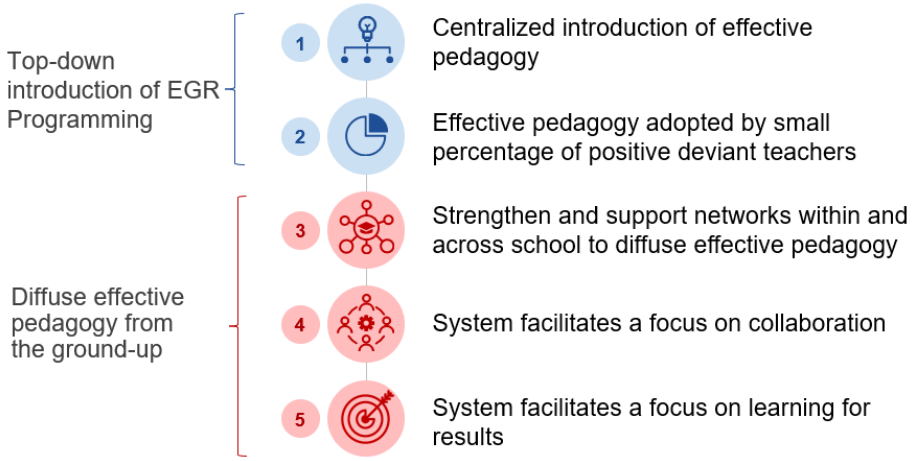
RTI International's main focus in International Education is the implementation of USAID EGR Programming. Therefore, this study is focused on insights and understanding the implementation challenges of EGR Programming; my study contribution to my institute should also focus on supporting the efficiency and efficacy of program implementation. While the first iterations of EGR Programs were focused on supporting the implementation of phonics-based programming (mostly) through the development of teaching and learning materials, teacher training, and teacher support, the next generation of early-grade reading programs is mainly focused on systems support. Most

USAID procurements in early grades education have focused on providing capacity and mechanisms to local ministries to improve the implementation fidelity of basic education. The general concern from researchers at my institute reflects the research from other institutes such as RISE, much of the focus is still firmly on implementation compliance (Honig and Pritchett 2019; Andrews 2017) rather than aligning systems strengthening with a more comprehensive understanding of the drivers of individual and systems norms (Hwa 2022; Sabarwal and Abu-Jawdeh 2018). RTI International also focuses on research that will influence procurements by USAID. This can often be achieved through Institutional Research Development (IRD) funding. The data used for this thesis study was collected through an IRD. RTI was awarded a second IRD to continue the research, which I will describe below.

5.13 New research direction

The new research opportunity builds off the research completed for this study. The research problem is the premise that a top-down education system denies agency and localized learning and decision-making. However, it is very unlikely that top-down education systems in LMICs are going away anytime soon. Indeed, it is unlike that top-down federal and state education systems in the USA will change this approach, either. Consequently, the new research explores building a bottom-up approach to diffusing effective pedagogy and classroom instruction through a ‘pivot.’ Accepting that policymakers in LMIC education systems will control the design and introduction of an education innovation; once introduced, then diffuse the innovation from the bottom-up (Exhibit 16).

Exhibit 15: Pivoting implementation: diffusing a top-down innovation



(King et al., 2022)

Once the innovation is introduced, the evidence from IFS (King 2020) and this thesis study show that only a small percentage of teachers, the positive deviants, achieve effective implementation. Then, rather than rely on WEOs (external district/ward coaches) to support teachers – the idea is to facilitate diffusion across and within schools with the understanding that a small percentage of teachers is already implementing the education innovation.

There are two design stages for this research; the theoretical framework and the research framework. Similar to the thesis study, the key theories used for the theoretical framework are Fullan (2015) and Rogers' DOI theory (2003). While behavioral economics will still feature, I view it as more of a lens to help think through human behavior rather than a central approach.

An important aspect of Fullan's work (2015) is that he focuses on the need for systems change rather than systems strengthening. A top-down approach, such as in Tanzania (as do many other education systems) focuses on Fullan's wrong drivers; accountability, individual, technology, and fragmented (Fullan 2021):

- Accountability – teacher observations on fidelity conducted by WEOs,
- Individual – support for individual teachers through WEOs (although the existence and potential of CoLs need to be acknowledged). However, the teachers in this study indicated their support

- Technology – while many education systems suffer from ICT procurement that frequently raises more challenges than it solves, this was less of an issue in Tanzania than SMS messaging to teachers regarding classroom instructional approaches. We have no data to suggest if this approach is beneficial.
- Fragmented – as shown with the USAID ToC, systems implementation often focuses on a small number of explicit activities, while other factors, such as social norms, push against them. This idea of the need for cohesion is common in other systems frameworks (Kaffenberger and Spivack 2022) and is also a challenge for ERG Programming. Most USAID early-grade procurements focus on a few explicit components, creating a fragmented approach to education change, and limiting the impact of EGR Programming (Piper and Stern 2019).

By pivoting to diffuse education innovation within and across schools, we address to focus on the *right drivers* of educational change:

- Focus on capacity building (vs. accountability)
- Group Quality (vs. individual)
- Pedagogy (vs. technology)
- Systemic (vs. fragmented)

(Fullan 2021)

I've already discussed most of these aspects in the literature review or this section of the thesis. However, the one driver that needs more expansion is pedagogy. Fullan (2015) discusses the importance of deep pedagogy over technology. However, he's referencing countries like the USA or Canada, where technology procurement causes issues, especially when it is a solution looking for a problem. In other words, technology should be viewed as an available tool that may (or may not) provide a solution to an identified classroom instructional challenge. (Dooley 1999). However, technology in LMICs is usually scarce due to a lack of funds and is often implemented without consideration that it can often exacerbate equity issues, especially for children with a lack of access to the internet (UNICEF 2020), something highlighted during COVID and remote learning (USAID 2021b). However, defining what effective pedagogy is in LMIC

education contexts needs more research This study and other evidence suggest what it is not; a focus on curriculum and content pushed from the top-down through accountability systems. However, there needs to be careful detangling of evidence; identifying if the lack of programming impacts the percentage of proficient learners is explained by pedagogical practice or other barriers such as a fragmented systems approach. For example, while this study suggests that teachers did not have a good understanding of the literacy skill level of their pupils, jumping onto formative assessment as the solution needs to be carefully thought through as programs that provide differentiated learning through the initial use of formative pupil assessment do not get better results than EGR Programming (Teaching at the Right Level 2022b).

While this thesis study was able to leverage aspects of DOI theory, it was not possible to measure all aspects of DOI theory. For example, while the study expanded on teachers' perception of the innovation (literacy program), it was unable to make progress in fully understanding diffusion networks and the innovation-decision process over time (Rogers, 2003). For this new research, it will be critical to uncover many more aspects of diffusion including communication channels, content, quality of communication, homophilic and heterophilic communication, formal (e.g., cascaded communications) and informal communication, and so on. To achieve this, the theoretical framework needs to incorporate Social Network Theory (SNT), and the research design employs Social Network Analysis (SNA). This approach acknowledges the importance of relationships and social interaction (Carmichael et al. 2006; Fullan 2015; Daly et al. 2010; Moolenaar and Daly 2012) and facilitates investigating these social networks. This importance has been acknowledged with policy, the development and introduction of teacher-professional learning communities (DuFour and DuFour 2013; Fullan 2015; Stoll and Seashore Louis 2007), and collaborative learning. Communities of Learning (CoLs) have been introduced into the education system in Tanzania in recent years. This study suggested that teachers already look to these groups for instructional support. However, little is known about how the CoLs provides instructional support for teachers. My first assumption would be that these groups focus on delivery of curriculum content; providing capacity to teachers such that they can teach lesson plans and units more effectively. I would also hypothesize that these

groups do not focus much on aspects of tracking and supporting pupil learning such as formative assessment and differentiated instruction. If they did, I would have been likely that these approaches would have been observed or mentioned in teacher interviews. However, these assumptions need to be thoroughly researched. Currently, the established CoLs in Tanzania do not receive guidance regarding the content or topic of their focus. However, that these groups have been established is a clear opportunity and acknowledgement that knowledge and understanding does not just cascade down a system. Finally, the link between effective PLCs and a focus on learning over compliance has been suggested by Daly et. al. (2010, p.379), "*This collaborative learning orientation toward the reform seemed to be present in those grade levels that had more dense connections. In more sparsely connected grade level teams teachers reported a focus on the more "technical" aspects of the reform such as completing minutes or checklists. There seemed to be less of a focus on implementing the reform with depth*."

This suggests that exploring collaborative approaches seems an opportunity to address education systems focused on compliance (Loveless 2021; Lantt Pritchett 2011). However, given that PLCs have existed in the USA for quite a while and the impact of education reform has been limited (Loveless 2021), a quick fix does not seem easy. It seems more likely that the structure and format of PLCs need more scrutiny (DuFour and DuFour, 2013).

5.15 Social Network Analysis (SNA)

Using SNA for the research design for the new study presents a methodology goal of focusing on the diffusion of effective pedagogical approaches. This study will be conducted in three stages: (1) identify teachers who use effective pedagogy, (2) document those teachers' professional social networks, then (3) analyze the structural characteristics of the mapped networks. The initial step would be to identify teachers using effective pedagogy. These individuals will need to be located and observed, and confirmed using to be developed criteria for observing effective pedagogy. These teachers could be:

- **positive deviant teachers** (as identified in Exhibit 3),
- **Exemplar teachers**. Individuals already identified by the education system or

- **opinion leaders**. These teachers (or school administrators) are widely known and trusted within localized areas.

The idea of opinion leadership is essential. Rogers (2003) states that it is incorrect to assume that it is early implementors who diffuse innovations. Instead, Rogers says this role goes to opinion leaders within communities. Consequently, this new study will have to identify these individuals with the understanding that they might be using an effective pedagogy. This also opens up the idea that there might be competing education innovations competing within a social network. An opinion leader might be promoting a pedagogical approach that is not particularly effective. Given that all the teachers in the thesis study were implementing similar classroom instructional approaches, there is some learned or social aspect that explains it.

The analysis of this research would be to map teachers' professional networks. The central challenge of mapping teachers' networks is gaining insight into interactions. Likely, most relevant communication between teachers and their influencers is not taking place via textual media. There are many themes to investigate, including, but not limited to:

- **Size** - The number of individuals in the network. Size influences the structure of relations.
- **Density** – the number of links (i.e., connections or "ties") in the network, expressed as a proportion of the total possible links in a network.
- **Centrality** – Individuals in the network with a relatively larger number of links or ties with other network members.
- **Communication channels** – e.g., social media, in-person
- **Communication messaging** – is it effective, and what is the focus?

The outcome of this research, funded by RTI International, will be a framework that EGR Programs can use to improve implementation through the diffusion of pedagogy. This might involve recommendations for communication (e.g., channels used, who communicates, and what they communicate) aligned with a structure for systems actors (e.g., head teachers and district coaches) to facilitate diffusion. There will be two aspects to this framework;

1. The drivers of the educational system. This provides a purpose to the system and emphasizes that although education innovations should diffuse within an education system, other aspects of the system must also be curated. This is often the challenge with donor funded work such as EGR Programming; it focuses on a few discrete activities, ignoring other factors. Otherwise the system struggles with fragmentation (Fullan 2015) or lack of cohesion (Lantt Pritchett 2011).
2. How social interaction of an education system can and should be managed. This would include the role of structure, focus, and nudging of the social system. Diffusion is generally a slow process, and Rogers (2003) acknowledges the need to support diffusion by linking the innovation's opinion leadership with those who are yet to be persuaded. In the Tanzanian education context, schools can often be isolated in rural areas, so the need to nudge diffusion within and across schools seems an important consideration.

5.2 Study Limitations

There are clear limitations of a relatively small qualitative sample of sixteen teachers selected from low-performing schools. Generalizability beyond these schools is limited. However, had I taken a quantitative survey approach and collected data at more schools with a more empirical approach, the study would have had severe issues with a conclusion due to the assumption of teacher resistance to program implementation being fundamentally incorrect. The qualitative approach facilitated the emergence of new themes during text analysis. While many of the findings would have to be replicated in different environments in Tanzania alone, it was encouraging that at least one of the findings that teachers were not aware of the actual level of performance of their pupils was confirmed in another recent study conducted in Bangladesh and India (Djaker, Ganimian, and Sabarwal 2022).

This study has provided some insights and has influenced the addition of a behavioral lens into programming at my institution. However, a more significant research audience has yet to adopt this research focus. While there has been some recent work research published in behavioral science by the World Bank (Sabarwal and Abu-Jawdeh 2018; Djaker, Ganimian, and Sabarwal 2022) and the Research on Improving Systems of

Education (RISE) Programme (Hwa 2022), there is a limited amount of literature regarding social norms or mental models in international education, especially research that links these observed behaviors to learning outcomes. Consequently, it is hard to claim behavioral science insights as a critical factor in education programming.

This study did not uncover aspects of the theoretical framework that the background section emphasizes as key considerations. For example, the study primarily leveraged DOI Theory regarding teachers' perception of education innovation. This reflects other literature that has applied DOI Theory (Hughes and Keith 1980; Moore and Benbasat 1991; Richardson 2011b). However, given the EGR programs have been implemented for six years, it was not possible to measure changing perceptions over time and who influenced the perception and the steps of the innovation-decision process (Rogers 2003).

Finally, while the study produced good examples of ineffective pedagogical practice, it is not clear precisely what effective pedagogical approach looks like. It has been challenging to define this effective practice and conditions in which the education system effectively improves learning outcomes. As will be discussed later, it is easy to jump to formative assessment and differentiated instruction, but these programs have impacts on learning outcomes very similar to EGR Programs (Piper and Stern 2019; TaRL 2022).

Chapter 6: Conclusions

6.1 Focus on mitigating barriers to implementation rather than just adding or adjusting implementation components.

Most of RTI's work in LMICs has been focused on developing classroom materials, training teachers, and following up with coaching support. All this is to be achieved by working within an education system. The habit of RTI as an implementor reflects the practice of its main client, USAID. The past ten years have focused on introducing new programming and working with local ministries of education to implement it from the top down. Consequently, habits focus on adjusting programming components rather than understanding and mitigating barriers such as harmful social norms and mental models. This recommendation aligns with Lewin's (1994) suggestion of understanding and

diminishing the restraining forces rather than adding more components to address the observable effect. The biggest challenge in international education is the frequent emphasis on accountability. Many voices advocate that accountability systems alone will not address systems issues (Lantt Pritchett 2011; Fullan 2015), and they are often a distraction. This detail is often lost on the next generation of early-grade reading procurements, which have been framed around the tenets of:

- individuals should know their roles and responsibilities within an education system,
- individuals should have the capacity to achieve their roles, and
- individuals should be held to account to complete their role

The challenge with this approach has been well discussed. This study concludes that systems and individuals seek legitimacy and compliance at the detriment of quality instruction (Lantt Pritchett 2011). While colleagues at my institute would agree that pushing accountability systems usually does not result in substantive change, it is challenging to move entirely away from this approach until USAID does. Local ministries and USAID are also very focused on accountability systems, although this study and other researchers have questioned their utility and that they are often a distraction from quality instruction (Fullan 2015; Honig and Pritchett 2019).

A key conclusion of this study is that a bottom-up approach to diffusing effective pedagogy seems more fruitful, with plenty of research support (Fullan 2015; Glaze 2014; Scheerens 1992; Poteyeva 2018; Bryk et al. 2010). However, this approach pushes against the core design of most LMIC education systems, which follow a centralized-decentralized top-down introduction of new educational programming, which is then implemented through a cascade down to districts and schools. However, a bottom-up approach to systems delivery is counter-intuitive, according to Daniel Kahneman 2011, as it does not focus on directly addressing observable effects. Therefore, it is very unlikely that local education ministries would drop their “legitimate” systems approach. The one key benefit of the top-down introduction of educational programming is that new classroom approaches and be introduced and can quickly diffuse (Hung et al. 2017). However, as this study shows – this diffusion soon stalls. It

seems that the best way forward is to provide education leaders with the legitimacy of their top-down approach but then provide a pivot and shift systems to a bottom-up diffusion. However, this would mean a fundamental shift to a system facilitating diffusion by focusing on learning and less on accountability. However, this approach does plenty of support (Lant Pritchett 2013; Fullan 2015; Rogers 2003).

6.2 Establish key drivers of educational change

Something that Michael Fullan does very well is establish key drivers of educational change alongside wrong drivers. His drivers have been developed through a combination of professional experience aligned with research and evidence.

One key driver of change that Fullan (2015) discusses is the focus on improving the collective capacity ahead of the capacity. Fullan also discusses that while individual teacher capacity is essential, it is not a driver of educational change (2015). This idea pushes against the research and implementation of early-grade reading programs. It seems in international education that because much of the research is centered around effective classroom instructional practice, then logically, implementation focused on the capacity of the individual teacher, addressed through teacher training and coaching. Collaborating colleagues have noted that much of the training in early-grade reading focuses on training individual teachers in schools rather than training an entire school simultaneously. For example, some training has been focused on training teachers of specific grades because of curriculum content. So, for example, all grade one teachers from a particular Province in Zambia would be trained at the same time, and at a later date, all the grade two teachers. However, the issue with this approach is that this means the focus of the training is on the transfer of knowledge rather than using training as a means to develop the capacity of a whole school together (the collective) and use training as an opportunity to start the process of developing behaviors of cohesive learning units that can continue these behaviors back at the school. This type of training also focuses on Fullan's second driver of educational change and a key finding of this study, pedagogy. Phonics-based instruction is a very observable change and understandably becomes a central focus of reform efforts in early-grade reading. However, any new curriculum will be ineffective if there is no effective pedagogical

approach to support it. As this study concluded, most teachers will, in time, be able to adopt new curricula and content but are hard-pressed to change pedagogical practices. My institute, RTI International, leads work on “Structured Pedagogy” (RTI International 2021), which is quite positive.

However, “structured pedagogy” incorporates eight components: data systems, curriculum scope and sequence, and materials development (RTI International 2021). Consequently, this approach focuses on implementation and process over aims. Fullan’s drivers are the overarching aims of an education system. Again, it is not to say that the components of structured pedagogy are not necessary; they are not drivers of change. USAID Kenya Tusome is an excellent example of implemented structured pedagogy. However, as evidence will show, after initial implementation, impact plateaus (see Exhibit 4), as does the impact on all early grade reading programs, which have similar components as laid out to structured pedagogy. It seems appropriate that dialogue with colleagues needs to center around the idea that structured pedagogy shows clear objectives of effective systems and teaching but need a small set of focused aims or drivers.

Annex 1: Teacher Interview tool from original data source

Question Number	Primary Question	Secondary Question Considerations
Section 1: Teacher background before Early Grade Reading Program		
1	Why did you originally decide to become a teacher?	How does what you expected [being a teacher would be] compare with your actual experience? How do you feel about that? Who influenced your approach to teaching? What was that approach?
2	What do you believe is the role and responsibility of teacher and of pupils for learning reading/math?	Can you give an example of your teaching and interaction with your pupils that demonstrate your beliefs about your role and responsibility for pupils' learning? Give examples of your pupils' learning behaviors that demonstrate your beliefs about your pupils' role and responsibility for their own learning?
3	Has clear information been expressed to you regarding the level of achievement or progress pupils in your class should show by the end of the year? If yes, by whom?	What are those expectations and whose responsibility is it to achieve these expectations? Do you believe these expectations are achievable? How do you know?
Section 2: Components of the Early Grade Reading Program		
4	What aspects of the program or training did you enjoy when they were initially introduced during training?	Can you describe how you feel about these aspects?
5	What aspects of the program or training did you find confusing or challenging when they were initially introduced during training?	Can you describe how you feel about these aspects?
6	How did you feel the program or training approach compared with the approach to teaching reading/math you were already using?	Are there aspects of your teaching approach that you felt were important but were not part of the Tusome Pamoja Project/training? Are there parts of the Tusome Pamoja Project/training that aligned with your

Question Number	Primary Question	Secondary Question Considerations
		goals or teaching approach? How were they similar?
7	Initially, how easy or challenging was it to apply the program/training? Why?	Do you feel that you started using what you learned in training right away? Did you adapt/modify the Tusome Pamoja Project/training in some way to make it easier to use? How? Have you continued using some of your old approaches even though they were not a part of the Tusome Pamoja Project/training? Can you give examples? Over time, have you continued using the Tusome Pamoja Project training components? Why?
Section 3: Classroom Instruction (using classroom observation tool)		
8	Describe what you are doing and what the pupils are doing in this video clip?	
9	How do you feel about using this approach?	Do you find it easy or difficult to use? Please explain or describe your response. Does this approach reflect anything you knew about teaching prior to Tusome Pamoja Project training? If so, please explain or describe this. Does this approach reflect anything you learned during Tusome Pamoja Project training? If so, please explain or describe this. Do you feel this approach is the same or different from the way you taught before Tusome Pamoja Project training? Does anyone support/encourage you to use this approach? If so, who and can you specify/provide examples of <u>how</u> are they supporting?

Question Number	Primary Question	Secondary Question Considerations
		Does this approach work best despite any contextual/environmental challenges you face in your classroom? How?
Section 4: Accountability		
10	Is there a benefit to implementing the Tusome Pamoja Project/training or a consequence if you don't? How do you feel about this?	Do you think applying the Tusome Pamoja Project/training is mandatory? [IF YES] Where do these expectations [to apply to Tusome Pamoja training] come from? Are they clear to you? When and how often did they do this?
Section 5: Teaching Colleagues & communication channels / networks		
11	Do you think teachers in your school and other schools are implementing the Tusome Pamoja Project/training? Do you feel that a majority of teachers implement the training?	Describe your colleagues. Are they similar to you regarding their teaching approaches? Describe the interactions you have with your colleagues regarding Tusome Pamoja Project/training. Do these interactions make a difference with (impact) your teaching? How?
12	Do you interact with teachers from other schools? If so, how?	Do these discussions influence you to change your own actions or opinions? Give an example.
13	Have you observed any teachers use the training <u>in their classroom</u>? If yes, who?	What was your reflection of the Tusome Pamoja Project/training after observing this teacher?
Section 6: Support		
14	Do you feel you've received the necessary support to apply Tusome Pamoja Project/training? Why?	Describe in what ways the support has been helpful for you to successfully apply the Tusome Pamoja Project/training-consider when the support was available, who gave the support and what type of support it was? **[IF YES]** How does it make you feel when you receive this support? **[IF NO]** Describe the support you think you need to help you successfully apply the Tusome Pamoja Project/training-consider when the support should be available (including frequency), who

Question Number	Primary Question	Secondary Question Considerations
		gives the support and what type of support it was?
Section 7: School Community		
15	Can you describe any interactions you have with your learner's parents? /Can you describe any interactions you have with your learner's parents?	How do you feel about teacher-parent interactions? What responsibility does a teacher have to parents? What do you believe to be the expectations from parents about how you should teach and your responsibility in their child's learning?

Annex 2: References

Ackoff, Russell, and Daniel Greenberg. 2008. *Turning Learning Right Side Up: Putting Education Back on Track*. 1st edition. Hoboken: Pearson Education, Limited.

Airasian, Peter W., and Kelvin D. Gregory. 2006. "The Education Reform Act of 1988." *Assessment in Education : Principles, Policy & Practice* 4 (2): 307–14. https://doi.org/10.1080/0969594970040207.

Andrews, Matt. 2017. *Building State Capability: Evidence, Analysis, Action*. First edition. Oxford Scholarship Online. Oxford: University Press.

Ariely, Dan. 2009. *Predictably Irrational: The Hidden Forces That Shape Our Decisions.* Rev. and exp. Ed. London: HarperCollins.

ASER Center. 2022. "Reading - ASER Centre." Aser Center. 2022. http://www.asercentre.org/p/141.html.

Banerjee, Abhijit, and Esther Duflo. 2006. "Addressing Absence." *Journal of Economic Perspectives* 20 (1): 117–32. https://doi.org/10.1257/089533006776526139.

Bartlett, Lesley, Amy Jo Dowd, and Christine Jonason. 2015. "Problematizing Early Grade Reading: Should the Post-2015 Agenda Treasure What Is Measured?" *International Journal of Educational Development* 40 (C): 308–14. https://doi.org/10.1016/j.ijedudev.2014.10.002.

Bauer, Patricia J. 2014. *Remembering the Times of Our Lives: Memory in Infancy and Beyond*. Psychology Press.

Bédécarrats, Florent, Isabelle Guérin, and François Roubaud. 2019. "All That Glitters Is Not Gold. The Political Economy of Randomized Evaluations in Development." *Development and Change* 50 (3): 735–62. https://doi.org/10.1111/dech.12378.

Bicchieri, Cristina, and Thomas Noah. 2017. "Applying Social Norms Theory in CATS Programming." https://repository.upenn.edu/pennsong/15.

Blavatnik School of Government. 2023. "New What Works Hub for Global Education Will Turn Evidence on Learning into Reality for Millions of Children." July 20, 2023. https://www.bsg.ox.ac.uk/news/new-what-works-hub-global-education-will-turn-evidence-learning-reality-millions-children.

Braun, Virginia. 2022. *Thematic Analysis: A Practical Guide*. London: SAGE Publications Ltd.

Braun, Virginia, and Victoria Clarke. 2006. "Using Thematic Analysis in Psychology." *Qualitative Research in Psychology* 3 (2): 77–101. https://doi.org/10.1191/1478088706qp063oa.

Bruns, Barbera. 2018. "In Bringing a Literacy Project to Scale, Has Kenya Found a Holy Grail?" Center For Global Development. October 26, 2018. https://www.cgdev.org/blog/bringing-literacy-project-scale-has-kenya-found-holy-grail.

Bryk, Anthony S., Penny Bender Sebring, Elaine Allensworth, Stuart Luppescu, and John Q. Easton. 2010. *Organizing Schools for Improvement: Lessons from Chicago*. *University of Chicago Press*. University of Chicago Press.

Burtless, Gary. 2019. "Experimental Economists Win Nobel Prize (and Deserved to Win)." *Brookings* (blog). October 23, 2019. https://www.brookings.edu/opinions/experimental-economists-win-nobel-prize-and-deserved-to-win/.

Carmichael, Patrick, Alison Fox, Robert McCormick, Richard Procter, and Leslie Honour. 2006. “Teachers’ Networks in and out of School.” *Research Papers in Education* 21 (2): 217–34. https://doi.org/10.1080/02671520600615729.

Chall, Jeanne S. 1996. *Stages of Reading Development*. 2nd ed. Fort Worth, Tex: Harcourt Brace.

Chen, Daniel, Paul Glewwe, Michael Kremer, and Sylvie Moulin. 2001. “Interim Report on a Teacher Attendance Incentive Program in Kenya.” https://www.google.com/search?q=Interim+Report+on+a+Teacher+Attendance+Incentive+Program+in+Kenya&rlz=1C1GCEU_enUS819US819&oq=Interim+Report+on+a+Teacher+Attendance+Incentive+Program+in+Kenya&aqs=chrome..69i57j69i60.909j0j7&sourceid=chrome&ie=UTF-8.

CIES. 2020. “Effectiveness of Coaching as a Teacher Professional Development Strategy in Low-Income Countries – CIES 2020.” *Effectiveness of Coaching as a Teacher Professional Development Strategy in Low-Income Countries* (blog). 2020. https://cies2020.org/portfolio/effectiveness-of-coaching-as-a-teacher-professional-development-strategy-in-low-income-countries/.

Common Core. 2022. “Common Core State Standards Initiative.” 2022. http://www.corestandards.org/.

Daly, Alan J., Nienke M. Moolenaar, Jose M. Bolivar, and Peggy Burke. 2010. “Relationships in Reform: The Role of Teachers’ Social Networks.” *Journal of Educational Administration* 48 (3): 359–91. https://doi.org/10.1108/09578231011041062.

Djaker, Sharnic, Alejandro J. Ganimian, and Shwetlena Sabarwal. 2022. “Primary- and Middle-School Teachers in South Asia Overestimate the Performance of Their Students.” In *RISE Annual Conference 2022*. Oxford, England. https://riseprogramme.org/events/rise-annual-conference-2022.

Dooley, Kim E. 1999. “Towards a Holistic Model for the Diffusion of Educational Technologies: An Integrative Review of Educational Innovation Studies.” *Educational Technology & Society* 2 (4): 52–66.

D’Souza, Karen. 2022. “A Movement Rises to Change the Teaching of Reading.” *EdSource* (blog). August 2022. https://edsource.org/2022/a-movement-rises-to-change-the-teaching-of-reading/675989.

DuFour, Richard, and Rebecca DuFour. 2013. *Learning by Doing: A Handbook for Professional Learning Communities at Work TM*. Solution Tree Press.

Dynamic Measurement Group. n.d. “What Are DIBELS? : Dynamic Measurement Group.” Accessed September 16, 2018. https://dibels.org/dibels.html.

Elmore, Richard F. 2004. *School Reform from the inside out: Policy, Practice, and Performance / Richard F. Elmore.* Cambridge, Mass.: Harvard Education Press.

Elmore, -Richard-F. 1990. *Restructuring Schools. The Next Generation of Educational Reform*. 0.

Emmers, Elke, Dieter Baeyens, and Katja Petry. 2021. “The Subjective Norm and Attitudes of Preservice Teachers toward Pupils with a Disability: An Experiment Based on the Cognitive Dissonance Theory.” *Australasian Journal of Special and Inclusive Education* 45 (1): 49–61. https://doi.org/10.1017/jsi.2020.18.

Filmer, Deon, and Lant Pritchett. 2001. “Estimating Wealth Effects Without Expenditure Data—Or Tears: An Application To Educational Enrollments In States Of India*.” *Demography* 38 (1): 115–32. https://doi.org/10.1353/dem.2001.0003.

Frank, Robert. 2020. *Under the Influence: Putting Peer Pressure to Work*. Princeton, NJ: Princeton University Press.

Froese, Victor. 1996. *Whole-Language: Practice and Theory. Second Edition*. Allyn and Bacon, Order Processing, P.

Fullan, Michael. 2006. “Change Theory: A Force for School Improvement.” In . Vol. Seminar Series Paper No. 157. Centre for Strategic Education.

———. 2015. *The New Meaning of Educational Change, Fifth Edition*. New York: Teachers College Press.

———. 2021. “The Right Drivers for Whole System Success.” Center for Strategic Education. https://michaelfullan.ca/wp-content/uploads/2021/03/Fullan-CSE-Leading-Education-Series-01-2021R2-compressed.pdf.

Gates Foundation. n.d.-a. “1. Leadership and Teacher Adoption:Structured Pedagogy Series.” *Science of Teaching* (blog). Accessed February 26, 2022. https://scienceofteaching.site/how-to-guides/learning-outcomes/.

———. n.d.-b. “Structured Pedagogy Series.” *Science of Teaching* (blog). Accessed February 19, 2022. https://scienceofteaching.site/how-to-guides/learning-outcomes/.

———. n.d.-c. “Structured Pedagogy Series: 6. Teacher Professional Development: Ongoing Teacher Support.” *Science of Teaching* (blog). Accessed February 25, 2022. https://scienceofteaching.site/how-to-guides/learning-outcomes/.

Gear, Adrienne. 2021. “Reading Wars and ‘The Science of Reading.’” *Simbi* (blog). August 24, 2021. https://medium.com/readwithsimbi/the-science-of-reading-8631f5614a14.

GEEAP. 2023. “2023 Cost-Effective Approaches to Improve Global Learning - What Does Recent Evidence Tell Us Are ‘Smart Buys’ for Improving Learning in Low- and Middle-Income Countries?” Text/HTML. https://documents.worldbank.org/en/publication/documents-reports/documentdetail/099420106132331608/IDU0977f73d7022b1047770980c0c5a14598eef8.

“Ghana: Accountability for Learning Framework.” 2022. Education Partnerships Group. *Education Partnerships Group* (blog). 2022. https://epg.org.uk/portfolio/ghana-accountability-for-learning-framework/.

Ghana Ministry of Education. n.d. “Ghana Accounability for Learning Outcomes Project.” *Ghana Accounability for Learning Outcomes Project* (blog). Accessed March 5, 2022. https://moe.gov.gh/gallop/.

Gibbons, Jeffrey A., Angela Toscano, Stephanie Kofron, Christine Rothwell, Sherman A. Lee, Timothy D. Ritchie, and W. Richard Walker. 2013. “The Fading Affect Bias across Alcohol Consumption Frequency for Alcohol-Related and Non-Alcohol-Related Events.” *Consciousness and Cognition* 22 (4): 1340–51. https://doi.org/10.1016/j.concog.2013.09.004.

Gladwell, Malcolm. 2001. *The Tipping Point: How Little Things Can Make a Big Difference / Malcolm Gladwell.* London: Abacus.

Glor, E. 2003. “Innovation Traps : Risks and Challenges in Thinking About Innovation.” 2003. https://www.semanticscholar.org/paper/Innovation-Traps-%3A-Risks-and-Challenges-in-Thinking-Glor/92f7ca4c003f32fc12118305b7741a28960a1e50.
Glor, Eleanor D. 2003. “Risks and Challenges in Thinking About Innovation.” *The Innovation Journal: The Public Sector Journal* 8(3): 1–19.
Góngora, D., M. Vega-Hernández, M. Jahanshahi, P. A. Valdés-Sosa, and M. L. Bringas-Vega. 2020. “Crystallized and Fluid Intelligence Are Predicted by Microstructure of Specific White-Matter Tracts.” *Human Brain Mapping* 41 (4): 906–16.
Gordon, Wendy. 2011. “Behavioural Economics and Qualitative Research – A Marriage Made in Heaven?” *International Journal of Market Research* 53 (2): 171–85. https://doi.org/10.2501/IJMR-53-2-171-186.
Gove, Amber, Tracy Brunette, Jennae Bulat, Bidemi Carrol, Catherine Henny, Wykia Macon, Evangeline Nderu, and Yasmin Sitabkhan. 2017. “Assessing the Impact of Early Learning Programs in Africa.” *New Directions for Child and Adolescent Development* 2017 (158): 25–41. https://doi.org/10.1002/cad.20224.
Gross, Richard D. 2005. *Psychology: The Science of Mind and Behaviour*. 5th ed. London: Hodder Arnold.
Guskey, Thomas. 2002. “Professional Development and Teacher Change.” *Teachers and Teaching* 8 (August): 381–91. https://doi.org/10.1080/1354060021000005 12.
Guskey, Thomas R. 1986. “Staff Development and the Process of Teacher Change.” *Educational Researcher* 15 (5): 5–12. https://doi.org/10.2307/1174780.
Guskey, Thomas R. 2020. “Flip the Script on Change: Experience Shapes Teachers’ Attitudes and Beliefs.” *The Learning Professional* 41 (2).
Hargreaves, Andy. 2005. “Educational Change Takes Ages: Life, Career and Generational Factors in Teachers’ Emotional Responses to Educational Change.” *Teaching and Teacher Education*, Emotions, teacher identity and change, 21 (8): 967–83. https://doi.org/10.1016/j.tate.2005.06.007.
Heath, Chip, and Dan Heath. 2011. *Switch: How to Change Thing When Change Is Hard*. London: Random House. https://www.penguinrandomhouse.com/books/77696/switch-by-chip-heath-and-dan-heath/.
Heaton, Rebecca, and Shannon Chan Lai Quan. 2023. “Managing Cognitive Dissonance in Art Teacher Education.” *Cambridge Journal of Education* 53 (4): 473–94. https://doi.org/10.1080/0305764X.2023.2175789.
Heck, Olivia C., Heather Ormiston, and Polly Husmann. 2023. “Utilizing KAP in Schools: An Evaluation of Educators’ and Staff Knowledge, Attitudes, and Practices Related to Trauma.” *Journal of Child & Adolescent Trauma* 16 (4): 889–901. https://doi.org/10.1007/s40653-023-00536-y.
Hoffman, James V. 2012. “Standpoints: Why EGRA—a Clone of DIBELS—Will Fail to Improve Literacy in Africa.” *Research in the Teaching of English* 46 (4): 340–57.
Honig, Dan, and Lant Pritchett. 2019. “The Limits of Accounting-Based Accountability in Education (and Far Beyond): Why More Accounting Will Rarely Solve Accountability Problems.” Working Paper Series. Research on Improving Systems of Education (RISE). https://doi.org/10.35489/BSG-RISE-WP_2019/030.

Hughes, Andrew S., and Joan J. Keith. 1980. “Teacher Perceptions of an Innovation and Degree of Implementation.” *Canadian Journal of Education* 5 (2): 43–51. https://doi.org/10.2307/1494312.
Hung, David, Yancy Toh, Azilawati Jamaludin, and Hyo-Jeong So. 2017. “Innovation Becoming Trajectories: Leveraging Lateral and Vertical Moves for Collaborative Diffusion of Twenty-First Century Learning Practices.” *Asia Pacific Journal of Education* 37 (4): 582–600. https://doi.org/10.1080/02188791.2017.1388213.
Hwa, Yue-Yi. 2022. “RISE Annual Conference 2022.” In *RISE Programme*. https://riseprogramme.org/events/rise-annual-conference-2022.
Innocenti, UNICEF Office of Research-. 2019. “Data Must Speak.” UNICEF-IRC. 2019. https://www.unicef-irc.org/research/data-must-speak/.
John, Philpo. 2023. “Language Teaching and Literacy Learning in Tanzania Primary Education: A Critical Review,” April.
Kaffenberger, Michelle, and Marla Spivack. 2022. “System Coherence for Learning: Applications of the RISE Education Systems Framework.” Research on Improving Systems of Education (RISE). https://doi.org/10.35489/BSG-RISEWP_2022/086.
Kahneman, Daniel. 2011. *Thinking, Fast and Slow*. 1 edition. New York: Farrar, Straus and Giroux.
Kahneman, Daniel, and Amos Tversky. 1979. “Prospect Theory: An Analysis of Decision under Risk.” *Econometrica* 47 (2): 263–91. https://doi.org/10.2307/1914185.
Kanyongo, Gibbs Y., and Rachel Ayieko. 2017. “Hierarchical Linear Modelling of Sixth-Grade Students’ Socio-Economic Status and School Factors On Mathematics Achievement: Case Studies of Kenya and Zimbabwe.” *African Journal of Research in Mathematics, Science and Technology Education* 21 (2): 187–99. https://doi.org/10.1080/18117295.2017.1334368.
Kataoka, Sachiko, Le Anh Vinh, Sandhya Kitchlu, and Keiko Inoue. 2020. “Vietnam’s Human Capital: Education Success & Future Challenges.” World Bank, Washington, DC. https://doi.org/10.1596/34316.
Killian, Shaun. 2023. “The I Do WE Do YOU Do Model Explained - Evidence-Based Teaching.” Https://Www.Evidencebasedteaching.Org.Au/. February 21, 2023. https://www.evidencebasedteaching.org.au/the-i-do-we-do-you-do-model-explained/.
King, Simon. 2020. “Beyond Impact: Fidelity of Early Grade Reading Programs.” London: Universtiy College London, Institute of Education.
King, Simon, Kellie Betts, Alastair Rodd, and Sagar Neupane. 2022. “Is Generalized Impact of Learning Outcomes Explained by Positive Deviance? The Challenge of Social Norms When Scaling of Early Grade Reading Programs.” In . Oxford, England. https://riseprogramme.org/sites/default/files/inline-files/Is%20Generalized%20Impact%20of%20Learning%20Outcomes%20Explained%20by%20Positive%20Deviance.pdf.
Koch, Alexander, Julia Nafziger, and Helena Skyt Nielsen. 2015. “Behavioral Economics of Education.” *Journal of Economic Behavior & Organization*, Behavioral Economics of Education, 115 (July): 3–17. https://doi.org/10.1016/j.jebo.2014.09.005.

Kolleck, Nina, Johannes Schuster, Ulrike Hartmann, and Cornelia Gräsel. 2021. "Teachers' Professional Collaboration and Trust Relationships: An Inferential Social Network Analysis of Teacher Teams." *Research in Education (Manchester)* 111 (1): 89–107. https://doi.org/10.1177/00345237211031585.
Levitt, Steven D, John A List, Susanne Neckermann, and Sally Sadoff. 2012. "The Behavioralist Goes to School: Leveraging Behavioral Economics to Improve Educational Performance." Working Paper 18165. Working Paper Series. National Bureau of Economic Research. https://doi.org/10.3386/w18165.
Lewin, Kurt. 1997. *Resolving Social Conflicts & Field Theory in Social Science*. Washington, DC: American Psychological Association. http://content.apa.org/books/1997-97489-000.
Lezotte, Lawrence W. 1991. *Correlates of Effective Schools: The First and Second Generation*. Okemos, MI: Effective Schools Products.
Li, Jia, Minghui Liu, and Xuan Liu. 2016. "Why Do Employees Resist Knowledge Management Systems? An Empirical Study from the Status Quo Bias and Inertia Perspectives." *Computers in Human Behavior* 65: 189–200. https://doi.org/10.1016/j.chb.2016.08.028.
Loveless, Tom. 2021. "Why Common Core Failed." *Brookings* (blog). March 18, 2021. https://www.brookings.edu/blog/brown-center-chalkboard/2021/03/18/why-common-core-failed/.
Machin, Stephen, Sandra McNally, and Martina Viarengo. 2018. "Changing How Literacy Is Taught: Evidence on Synthetic Phonics." *American Economic Journal. Economic Policy* 10 (2): 217–41. https://doi.org/10.1257/pol.20160514.
Martinez, Rebecca. 2023. "Lessons Learned from a Decade of Early Grade Reading Programs | Education Links." 2023. https://www.edu-links.org/learning/lessons-learned-decade-early-grade-reading-programs.
McFalls, Elisabeth L., and Deirdre Cobb-Roberts. 2001. "Reducing Resistance to Diversity through Cognitive Dissonance Instruction: Implications for Teacher Education." *Journal of Teacher Education* 52 (2): 164–72. https://doi.org/10.1177/0022487101052002007.
Moolenaar, Nienke M., and Alan J. Daly. 2012. "Social Networks in Education: Exploring the Social Side of the Reform Equation." *American Journal of Education* 119 (1): 1–6. https://doi.org/10.1086/667762.
Moore, Gary C., and Izak Benbasat. 1991. "Development of an Instrument to Measure the Perceptions of Adopting an Information Technology Innovation." *Information Systems Research* 2 (3): 192–222. https://doi.org/10.1287/isre.2.3.192.
NaCCA. 2020. "COMMON CORE PROGRAMME (CCP)." NaCCA. 2020. https://nacca.gov.gh/common-core-programme-ccp/.
Ngware, Moses W., Moses Oketch, and Maurice Mutisya. 2014. "Does Teaching Style Explain Differences in Learner Achievement in Low and High Performing Schools in Kenya?" *International Journal of Educational Development* 36: 3–12. https://doi.org/10.1016/j.ijedudev.2014.01.004.
Nicolle, A., S. M. Fleming, D. R. Bach, J. Driver, and R. J. Dolan. 2011. "A Regret-Induced Status Quo Bias." *J NEUROSCI , 31 (9) 3320 - 3327. (2011)*.
Olsen, Brad, and Maya Elliott. 2023. "3 Myths about Teacher Agency and Why They Hold Back Scaling Education Innovations." Brookings. 2023.

https://www.brookings.edu/articles/3-myths-about-teacher-agency-and-why-they-hold-back-scaling-education-innovations/.
Oyetoro, Oyebode Stephen, Bosede Abimbola Adesina, and Tolulope Segun Eyebiokin. 2020. “Personal Epistemic and Learning Approaches as Predictors of Pre-Service Teachers Use of Strategies to Counter Cognitive Dissonance from Supervisor Feedback.” *African Journal of Teacher Education* 9 (2): 62–89. https://doi.org/10.21083/ajote.v9i2.6131.
Peterson, Paul E., and Martin R. West. 2003. *No Child Left behind? The Politics and Practice of School Accountability*. Washington, D.C.: Brookings Institution Press.
Pflepsen, Alison, Ashley Hertz, and Emily Kochetkova. 2019. “Classroom Observation Toolkits for Early Grade Reading Improvement.” Global Reading Network. https://www.globalreadingnetwork.net/resources/classroom-observation-toolkit-early-grade-reading-improvement.
Piper, Benjamin, Evelyn Jepkemei, and Kennedy Kibukho. 2015. “PRO-POOR PRIMR: IMPROVING EARLY LITERACY SKILLS FOR CHILDREN FROM LOW-INCOME FAMILIES IN KENYA.” *Africa Education Review* 12 (1): 67–87. https://doi.org/10.1080/18146627.2015.1036566.
Piper, Benjamin, and Jonathan Michael Bradley Stern. 2019. “Resetting Targets: Examining Large Effect Sizes and Disappointing Benchmark Progress,” April. https://doi.org/10.3768/rtipress.2019.op.0060.1904.
Piper, Benjamin, and Stephanie Simmons Zuilkowski. 2015. “Teacher Coaching in Kenya: Examining Instructional Support in Public and Nonformal Schools.” *Teaching and Teacher Education* 47: 173–83. https://doi.org/10.1016/j.tate.2015.01.001.
Poteyeva, Margarita. 2018. “Social Capital.” Encyclopedia Britannica. December 17, 2018. https://www.britannica.com/topic/social-capital.
Power, Samantha. 2021. “Administrator Samantha Power’s Keynote Remarks at the United Nation’s Behavioral Science Week | Remarks | U.S. Agency for International Development.” Usaid.Gov. 21 2021. https://www.usaid.gov/news-information/speeches/jun-21-2021-administrator-samantha-power-keynote-remarks-united-nations-behavioral.
Pritchett, Lant. 2013. *The Rebirth of Education: Schooling Ain’t Learning*. Washington: Brookings Institution Press. https://doi.org/10.7864/j.ctt1gpccb8.
Pritchett, Lantt. 2011. “One Size Doesn’t Fit All: Lant Pritchett on Mimicry in Development.” *One Size Doesn’t Fit All: Lant Pritchett on Mimicry in Development* (blog). March 14, 2011. https://www.cgdev.org/blog/one-size-doesn%E2%80%99t-fit-all-lant-pritchett-mimicry-development.
Raafat, Ramsey M., Nick Chater, and Chris Frith. 2009. “Herding in Humans.” *Trends in Cognitive Sciences* 13 (10): 420–28. https://doi.org/10.1016/j.tics.2009.08.002.
Reese, Gerhard, Amir Rosenmann, and James T. Cameron. 2019. *The Psychology of Globalization: Identity, Ideology, and Action*. San Diego: Elsevier Science & Technology. https://doi.org/10.1016/C2016-0-03228-X.
Richardson, Jayson W. 2011a. “Technology Adoption in Cambodia: Measuring Factors Impacting Adoption Rates.” *Journal of International Development* 23 (5): 697–710. https://doi.org/10.1002/jid.1661.

———. 2011b. “Technology Adoption in Cambodia: Measuring Factors Impacting Adoption Rates.” *Journal of International Development* 23 (5): 697–710. https://doi.org/10.1002/jid.1661.
Rogers, Everett M. 2003. *Diffusion of Innovations, Fifth Edition*. New York: Free Press.
RTI International. 2016. “Early Grade Reading Assessment (EGRA) Toolkit: Second Edition.” RTI International. https://www.globalreadingnetwork.net/resources/early-grade-reading-assessment-egra-toolkit-second-edition.
———. 2021. “Structured Pedagogy Series.” *Science of Teaching* (blog). 2021. https://scienceofteaching.site/how-to-guides/learning-outcomes/.
RTI International, Jennae Bulat, Margaret Dubeck, Paula Green, Karon Harden, Catherine Henny, Monika Mattos, Aliso Pflepsen, Ana Robledo, and Yasmin Sitabkhan. 2017. “What We Have Learned in the Past Decade: RTI’s Approach to Early Grade Literacy Instruction.” RTI Press. https://doi.org/10.3768/rtipress.2017.op.0039.1702.
Ruddle, Nicola, and Georgina Rawle. 2020. “EQUIP-Tanzania Impact Evaluation.” Oxford Policy Management. https://www.opml.co.uk/files/Publications/8383-assessing-equip-t/1589895669_opm-ie-final-endline-report-final.pdf.
Sabarwal, Shwetlena, and Malek Abu-Jawdeh. 2018. “What Teachers Believe : Mental Models about Accountability, Absenteeism, and Student Learning.” WPS8454. The World Bank. http://documents.worldbank.org/curated/en/804301527601436747/What-teachers-believe-mental-models-about-accountability-absenteeism-and-student-learning.
Sandefur, Justin, Thomaz Alvares de Azevedo, Xiaomin Ju, and Thi Le. 2023. “Phonics and Foreign Aid: Lessons from a Decade of USAID Early-Grade Reading Evaluations.” Center for Global Development. https://www.cgdev.org/event/phonics-and-foreign-aid-can-america-teach-world-read.
Scheerens, J. 1992. *Effective Schooling: Research, Theory and Practice*. School Development Series. London: Cassell.
Schmidt, Nola A., and Janet M. Brown. 2007. “Use of the Innovation–Decision Process Teaching Strategy to Promote Evidence-Based Practice.” *Journal of Professional Nursing* 23 (3): 150–56. https://doi.org/10.1016/j.profnurs.2007.01.009.
Scott, Claire. n.d. “Learning to Read – Why Phonics Isn’t the Be-All and End-All | Learning and Development.” *Teach Early Years* (blog). Accessed February 17, 2024. https://www.teachearlyyears.com/learning-and-development/view/beyond-phonics.
Sharot, Tali. 2011. “The Optimism Bias.” *Current Biology* 21 (23): R941–45. https://doi.org/10.1016/j.cub.2011.10.030.
Smiley, Anne, Yvonne Cao, Wael Moussa, Brian Dooley, and Jane Sullivan. 2020. “Examining ‘Best Practices’ for Literacy Coaching and Monitoring: Evidence from Northern Nigeria and Ghana.” *Social Sciences & Humanities Open* 2 (1): 100014-. https://doi.org/10.1016/j.ssaho.2020.100014.
Snow, Pamela C. 2021. “SOLAR: The Science of Language and Reading.” *Child Language Teaching and Therapy* 37 (3): 222–33. https://doi.org/10.1177/0265659020947817.

Strauss, Valerie. 2021. “Perspective | Why the Common Core Standards Failed — and What It Means for School Reform.” *Washington Post*, April 5, 2021. https://www.washingtonpost.com/education/2021/04/05/common-core-failed-school-reform/.
Suri, Gaurav, Gal Sheppes, Carey Schwartz, and James J. Gross. 2013. “Patient Inertia and the Status Quo Bias: When an Inferior Option Is Preferred.” *Psychological Science* 24 (9): 1763–69. https://doi.org/10.1177/0956797613479976.
Sutton, Rosemary E., and Karl F. Wheatley. 2003. “Teachers’ Emotions and Teaching: A Review of the Literature and Directions for Future Research.” *Educational Psychology Review* 15 (4): 327–58.
TaRL. n.d. “Evidence.” Teaching at the Right Level. Accessed September 4, 2022. https://www.teachingattherightlevel.org/research-and-learning/evidence/.
Teaching at the Right Level. 2022a. “Evidence.” Teaching at the Right Level. 2022. https://www.teachingattherightlevel.org/research-and-learning/evidence/.
———. 2022b. “The TaRL Approach.” Teaching at the Right Level. 2022. https://www.teachingattherightlevel.org/the-tarl-approach/.
Thaler, Richard H., and Cass R. Sunstein. 2009. *Nudge: Improving Decisions about Health, Wealth and Happiness*. London: Penguin.
“The Belmont Report | HHS.Gov.” n.d. Accessed March 25, 2018. https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/index.html.
The Economist. 2021. “American Schools Teach Reading All Wrong,” 2021. https://www.economist.com/united-states/2021/06/12/american-schools-teach-reading-all-wrong.
The National Examinations Council of Tanzania. 2021. “NECTA | PSLE RESULTS.” The National Examinations Council of Tanzania. 2021. https://necta.go.tz/psle_results.
The World Bank. 2020. “Cost-Effective Approaches to Improve Global Learning.” The World Bank. https://documents1.worldbank.org/curated/en/719211603835247448/pdf/Cost-Effective-Approaches-to-Improve-Global-Learning-What-Does-Recent-Evidence-Tell-Us-Are-Smart-Buys-for-Improving-Learning-in-Low-and-Middle-Income-Countries.pdf.
Thomas, David R. 2006. “A General Inductive Approach for Analyzing Qualitative Evaluation Data.” *American Journal of Evaluation* 27 (2): 237–46. https://doi.org/10.1177/1098214005283748.
Tripathy, Jaya Prasad. 2013. “Secondary Data Analysis: Ethical Issues and Challenges.” *Iranian Journal of Public Health* 42 (12): 1478–79.
UNICEF. 2020. “Two Thirds of the World’s School-Age Children Have No Internet Access at Home.” UNICEF. November 2020. https://www.unicef.org/press-releases/two-thirds-worlds-school-age-children-have-no-internet-access-home-new-unicef-itu.
United Nations. 2019. “SDG Indicators.” SDG Indicators. December 2019. https://unstats.un.org/sdgs/metadata/?Text=&Goal=4&Target=4.1.
USAID. 2011. “USAID EDUCATION STRATEGY, 2011-2015.” Washington, D.C.: USAID. https://2012-

2017.usaid.gov/sites/default/files/documents/1865/USAID_Education_Strategy.pdf.
———. 2013. “National Baseline Assessment for 3Rs (Reading, Writing, and Arithmetic) in Tanzania (Summary Brief).” 2013. https://shared.rti.org/content/national-baseline-assessment-3rs-reading-writing-and-arithmetic-tanzania-summary-brief.
———. 2015. “EGRB Results - Tanzania 2013.” Early Grade Reading Barometer. 2015. https://earlygradereadingbarometer.org/results/Tanzania/2013.
———. 2017. “USAID Tusome Pamoja Baseline Findings Report.” https://pdf.usaid.gov/pdf_docs/PA00TC9S.pdf.
———. 2018a. “ABE-ACR Tanzania National Early Grade Reading Assessment (EGRA) Final Report.” USAID. https://shared.rti.org/content/abe-acr-tanzania-national-early-grade-reading-assessment-egra-final-report.
———. 2018b. “Tanzania - Tusome Pamoja.” USAID. https://www.usaid.gov/sites/default/files/documents/1860/2018-10-29_Tusome_Pamoja.pdf.
———. 2020. “The Journey to Self-Reliance.” August 6, 2020. https://2017-2020.usaid.gov/selfreliance.
———. 2021a. “EGRB | Results | Nepal | EGRP 2020.” Early Grade Reading Barometer. 2021. https://earlygradereadingbarometer.org/results/Nepal/EGRP-2020.
———. 2021b. “Resilience in Crisis: First Indications of Why Children’s Reading Scores Increased during the COVID-19 Pandemic in Cambodia.” *Early Grade Reading Barometer* (blog). 2021. https://earlygradereadingbarometer.org/news/cambodia-2021-acr.
———. 2021c. “Tusome Early Grade Reading.” April 30, 2021. https://www.usaid.gov/kenya/documents/tusome-early-grade-reading-activity.
———. 2021d. “Jifunze Uelewe (‘Learn and Understand’).” Usaid.Gov. August 18, 2021. https://www.usaid.gov/tanzania/documents/Jifunze-Uelewe.
———. 2021e. “The United States Government Launches Jifunze Uelewe.” USAID. November 16, 2021. https://www.usaid.gov/tanzania/press-releases/nov-16-2021-united-states-government-launches-jifunze-uelewe-four.
———. 2022a. “Benchmarks - SDG4.” Early Grade Reading Barometer. 2022. https://earlygradereadingbarometer.org/benchmarks/sdg.
———. 2022b. “GRANTS.GOV.” Grants.Gov. 2022. https://www.grants.gov/web/grants/view-opportunity.html?oppId=343336.
———. 2023a. “Compendium of Standard PIRS for Education Programming | Education Links.” Www.Edu-Links.Org. 2023. https://www.edu-links.org/resources/compendium-standard-pirs-education-programming.
———. 2023b. “Ten Years of Early Grade Reading Programming: A Retrospective.” Washington, DC: USAID. https://www.edu-links.org/resources/ten-years-early-grade-reading-programming-retrospective.
———. 2023c. “Success Stories | Humanitarian Assistance.” U.S. Agency for International Development. January 24, 2023. https://www.usaid.gov/humanitarian-assistance/humanitarian-stories/success-stories.

———. 2024. “Localization | What We Do.” U.S. Agency for International Development. January 18, 2024. https://www.usaid.gov/localization.
———. n.d. “EGRB | Benchmarks - ORF Zero Scores.” Early Grade Reading Barometer. Accessed February 24, 2024. https://earlygradereadingbarometer.org/benchmarks.
Venus, Merlijn, Daan Stam, and Daan van Knippenberg. 2018. “Research: To Get People to Embrace Change, Emphasize What Will Stay the Same.” *Harvard Business Review* (blog). August 15, 2018. https://hbr.org/2018/08/research-to-get-people-to-embrace-change-emphasize-what-will-stay-the-same.
Webster, Adrienne. 2023. “Fading Affect Bias & the Impact on Alcohol Recovery.” Monument. February 23, 2023. https://joinmonument.com/resources/fading-affect-bias-alcohol-recovery/.
World Bank. 2021. “Coach.” Text/HTML. World Bank. March 17, 2021. https://www.worldbank.org/en/topic/teachers/brief/coach-helping-countries-accelerate-learning-by-improving-in-service-teacher-professional-development.
———. n.d. “8 Tips to Structure Effective One-to-One Support Systems for Teachers.” Accessed February 25, 2022. https://blogs.worldbank.org/education/8-tips-structure-effective-one-one-support-systems-teachers.
Zambia Ministry of Education. 2023. “THE 2023 ZAMBIA EDUCATION CURRICULUM FRAMEWORK IS PROGRESSIVE – KAMOKO. – Ministry of Education.” 2023. https://www.edu.gov.zm/?p=5334.