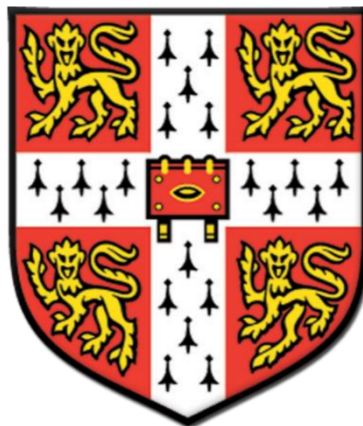# *The relationship between educational television and mathematics capability in Tanzania*

Joseph Watson

Faculty of Education

Homerton College

University of Cambridge

November 2019

This dissertation is submitted for the degree of Doctor of Philosophy

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. It does not exceed the prescribed word limit for the relevant Degree Committee

Joseph Watson

November 2019

Acknowledgements

I would like to express my thanks to my supervisors, Dr. Sara Hennessy and Prof. Anna Vignoles. Without their extraordinary support and patience, this thesis would not have been possible. Thanks also to my advisor, Dr. Benjamin Alcott, whose guidance throughout my PhD has been hugely beneficial.

I am grateful to all those in Tanzania who assisted with or participated in this research. Amongst those who assisted, Mussa provided fantastic guidance as well as playing a vital role in training data collectors and conducting interviews. Titus also played a key part in gathering information and was always so helpful. Thank you as well to Nisha and Lauren, for being so welcoming and answering all my questions about educational television!

Lastly, I would like to thank my parents, Mark and Nichola, for their wonderful help and encouragement.

The relationship between educational television and mathematics capability in Tanzania

Joseph Watson

Abstract

Previous studies have frequently demonstrated that educational television viewing can have a positive effect on learning in low-income country contexts when shows are delivered in controlled settings. However, the consequence of day-to-day viewing in such contexts has scarcely been considered. Additionally, no recent published research has provided any information on the costs of educational television. The lack of research in these areas is striking. Examining educational television viewing in monitored settings provides limited information on the influence of routine television consumption. Further, the broad reach of numerous educational television programmes should provide low per-viewer costs and, resultantly, strong cost-effectiveness findings. This PhD study therefore examined (1) the association between educational television exposure and mathematics capability and (2) the cost effectiveness of educational television interventions. To achieve this, research was carried out that centred on Ubongo Kids – a popular Tanzanian-produced show delivering mathematics-focused content.

Quantitative investigation into the association between educational television exposure and mathematics capability used nationally representative data, collected by Uwezo Tanzania. A household fixed-effects model showed that exposure to educational television was significantly associated with mathematics capability among children aged 7-16, when controlling for age, sex, school enrolment and Kiswahili attainment. Findings from this model were used in cost-effectiveness calculations, alongside cost data and an estimate of the number of Ubongo Kids viewers. Results compared favourably against those for other interventions, with calculations regarding Ubongo Kids' ongoing activities suggesting it to have been more cost effective than any other intervention previously investigated using the same cost-effectiveness approach. These findings indicate that in low-income contexts: educational television programmes can aid learning; and, that directing a greater proportion of available educational resources towards educational television interventions may benefit educational outcomes.

Contents

List of Tables

List of Figures

1 Chapter 1 – Introduction

1.1 Introduction

Educational television has considerable potential to improve learning outcomes in developing contexts. Millions of viewers in multiple nations regularly view shows such as Akili and Me and internationalised versions of Sesame Street. Further, various studies conducted in controlled environments suggest that such shows have a positive impact. It is therefore quite possible that educational television interventions deliver learning at scale. However, there is little research on the effects of day-to-day viewing in conventional settings (Section 2.4). The dearth of research in this area is striking, given that children engage with television in a manner distinct from that which occurs in regulated studies. Accordingly, my PhD addresses the following primary research question: 'What is the association between naturally occurring exposure to educational television and mathematics capability in a developing country context?'.

In answering this question, television is defined broadly to encompass any educationally oriented televised content that may or may not be supported by information on other media platforms. Naturally occurring exposure concerns normal viewing in any conventional setting, which could involve watching on a friend's television or parent's smartphone. Additionally, the primary research question refers to mathematics capability. Use of the term 'capability' reflects the employment of mathematics scores created through item response theory (IRT). These scores are used to give an approximation of a child's underlying level of capability in mathematics, as opposed to their attainment in a particular mathematics test (Section 3.3.1.3.1). The association between television exposure and mathematics capability is examined specifically with regards to Ubongo Kids in Tanzania. Ubongo Kids is a Tanzania-produced television programme for which a limited amount of associated eLearning materials is available (see Section 1.2.2).

In addition to exploration of the primary research question, this thesis investigates a subsidiary research question, namely, 'How cost-effective is educational television in a developing country context?' The subsidiary question is examined using the same definition of educational television. Ubongo Kids is also used again as the vehicle through which investigation into educational television is made. However, the subsidiary research question introduces a new concept, cost-effectiveness. Cost-effectiveness is considered by comparing the relative learning gains attributable to various programmes in developing country contexts for a given cost (see Section 8.1). As with studies concerning the association between educational television and learning outcomes, there is a paucity of literature concerning television's cost-effectiveness. No publicly accessible studies have been published in recent years (Section 2.3.1.3). Given the huge reach of educational television, television-based interventions could be highly cost effective. Investigation into cost effectiveness should

therefore be of value to policy makers seeking to better use available resources to promote educational outcomes.

The structure of the remainder of this chapter is now provided. The chapter begins by providing contextual information concerning the specific country and educational television show with which this research is concerned (Section 1.2). After this, the theoretical perspective of this thesis is considered (Section 1.3). Next, a corresponding philosophical position is adopted (Section 1.4). The chapter concludes with a summary and the provision of information on the overall structure of all remaining chapters (Section 1.5).

1.2   Contextual information

1.2.1   Tanzania and Dar es Salaam

This subsection provides contextual information on (mainland) Tanzania[1] and its de facto capital, Dar es Salaam. It is important to provide information relevant to Dar es Salaam and Tanzania as a whole for numerous reasons. The entirety of Tanzania is a focus of inferential and descriptive analysis in multiple sections (see, for example, Section 4.3.2, 6.1 and 7.1). Conversely, other inferential and descriptive statistics concern data relevant only to two (of five) districts in Dar es Salaam (Ilala and Temeke, see Section 4.3.1 and 5.1). In addition to this, all descriptive and inferential research is supported by qualitative information captured from children and caregivers based throughout Dar es Salaam.

Tanzania is located on the East Coast of Africa, bordering Kenya, Uganda, the Democratic Republic of Congo, Zambia, Malawi, Burundi, Rwanda and Mozambique. Tanzania's most recent census (conducted in 2012 and reported on in 2013 by the National Bureau of Statistics) suggests that of Tanzania's 43,625,354 residents, 4,364,541 are based in Dar es Salaam. Residents of Dar es Salaam are generally wealthier than their counterparts elsewhere in Tanzania. This region possesses the highest levels of computer and phone ownership, as well as the third highest level of radio ownership (calculated using Uwezo 2017 data). These ownership patterns extend to television, with 64 percent of children in Dar es Salaam living in a household with a television (in contrast with a national average of 24 percent: calculated using Uwezo 2017 data).

Differences between Dar es Salaam and other areas of Tanzania also exist when considering information on education. Attainment levels are low across Tanzania, with one in four Standard 7 (age 13) pupils unable to complete Standard 2 (age 8) numeracy tasks (Uwezo, 2015). However, residents of Dar es Salaam (in which Ilala and Temeke are located) score, on average, highest amongst all

---

[1] All subsequent references to Tanzania in this thesis concern mainland Tanzania only (thereby excluding Zanzibar).

Tanzanian regions (ibid).[2] These differences suggest the need for caution when comparing results for Ilala and Temeke with those for the country as a whole. The same could be said when using qualitative data from Dar es Salaam to inform quantitative findings relevant to other regions of Tanzania.

1.2.2   Ubongo

Ubongo Kids is created by Ubongo, a Dar es Salaam-based producer of educational media content. The Ubongo Kids programme was devised for Tanzanian audiences and initially broadcast exclusively in this country. This occurred from the programme's first broadcast in January 2014 up to March 2015, when the programme launched in Rwanda and Kenya. Today, Ubongo Kids is available in 31 countries. While television remains the primary medium through which Ubongo Kids content is delivered, televised material is supported by a selection of other technological resources: DVDs; eBooks; online videos (accessible via YouTube and Ubongo Kids' website); and, one interactive app (which can be downloaded from the Google Play store).[3] These educational resources were created to be accessed primarily outside of school but could also be used in formal education settings to supplement conventional teaching.

From April 2016, Ubongo has also delivered educational material to a younger age group. This material centres on another television programme, Akili and Me. Akili and Me is targeted at pre-primary children (3-6) while Ubongo Kids is aimed at primary-children (7-13). The focus of Ubongo's shows also differs, with pre-primary material seeking primarily to promote school readiness, and primary level material centred on mathematics and science attainment (in addition to other objectives considered below). Thoughts provided by the CEO and co-founder of Ubongo at the outset of this study on the theory of change (TOC) for Ubongo's primary-age show, Ubongo Kids, are presented in Figure 1.1.

---

[2] Whilst the intended ages of pupils in Standard Two and Standard Seven are 8 and 13 years old respectively, age can vary significantly within each Standard.

[3] All Ubongo materials are available without a user fee, with the exception of DVDs and eBooks. Ubongo therefore obtains a limited amount of product revenue, which exists alongside other funding sources (including international donors).
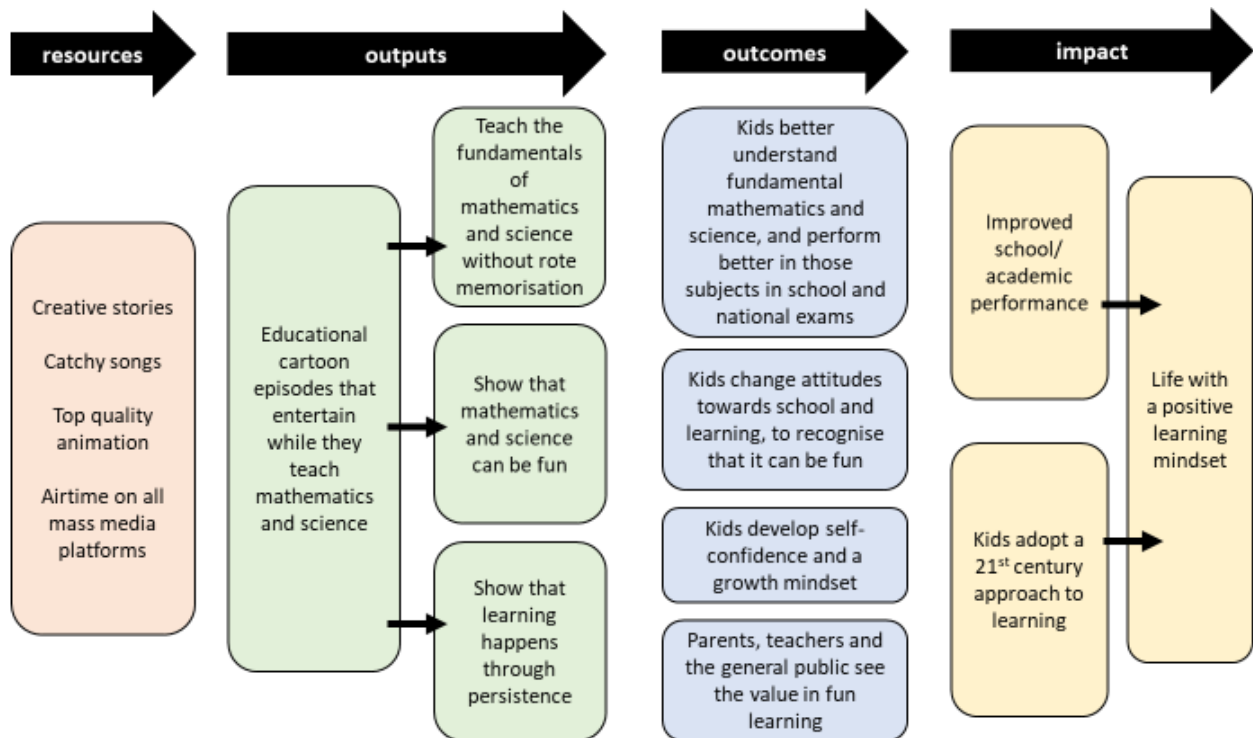
Figure 1. 1: Theory of change

This framework diagram demonstrates both the suitability and complexity of employing Ubongo as a vehicle through which to examine the association between educational television and mathematics capability. While Ubongo's primary-age content seeks to promote the understanding and employment of mathematics, other objectives are present. These include advancing learning in another subject (science), changing attitudes to learning amongst those that engage with the programme (at various stages of their lives) and influencing the perceptions of parents, teachers and even the general public.[4] My project does not examine these additional organisational goals, which underlines the fact that the study should not be considered an all-encompassing programme evaluation. That said, it is hoped that it might support future discussions within Ubongo concerning the development of this organisation's TOC (as well as generating wider messages to those concerned with other educational media initiatives).

1.2.2.1 *Ubongo Kids episode content*

Examination of a curriculum map for Ubongo Kids suggests that it corresponds with this theory of change. Most Ubongo Kids episodes produced thus far (each of which has been aired multiple times)

---

[4] Additionally, the objective concerning the adoption of a '21st century approach to learning' includes the development of skills such as critical thinking and digital literacy. Follow-up conversations with Ubongo also suggested that Ubongo Kids seeks to provide socio-emotional guidance (which is identified in sampled episodes, below), and to promote the longer-term goal of increasing the adult incomes of former child viewers.

address mathematics as their main subject (45 of 52), while the remainder concern health, science, engineering and technology. Certain programme topics listed, such as 'The Joy of Maths', appear to comply with Ubongo's intention of showing that learning can be fun (as stated in the TOC, above; and, as recognised through examination of 'Kibena and the Math Rats', below). Whilst providing information on the subject, topic and programme title, this map does not, however, give details about Ubongo's broader educational objectives or pedagogic approach. To investigate this, I examined two episodes of *Ubongo Kids* available in English on YouTube: 'Kibena and the Math Rats' (Ubongo Kids, 2015a) and 'The Mosquito Army' (Ubongo Kids, 2015b). My approach followed similar investigations into episode content by Linebarger and Piotrowski, 2010, who conducted a content analysis of six literacy programmes available on PBS for primary-age American children, and Singer and Singer, 1994, who reportedly coded a sample of episodes from the US television series, *Barney and Me* (cited in Fisch, 2004, original study unavailable).[5]

1.2.2.1.1   Sample episodes

Both sampled *Ubongo Kids* episodes demonstrate that the series does not provide information solely on mathematics. Socio-economic guidance is common throughout. Characters encourage each other's efforts, saying, for example, "That's a great idea!", in 'The Mosquito Army'. This episode clearly emphasises broader health concerns.[6] Lessons of how one might protect oneself from malaria form a key part of the story: the story opens with the mosquito 'commander' instructing his army not to "touch the treated bed nets". This lesson is reinforced throughout, for example, by a mosquito soldier reporting during an army briefing, "Commander, I [found] a whole group of houses that has no nets – let's attack them in force tonight". In 'Kibena and the Math Rats', there is also material that seeks to promote an enjoyment of mathematics. Songs include lines such as "Mathematics, how I love it". Such attempts to foster enjoyment of an academic subject were also identified to occur frequently in Linebarger and Piotrowski's (2010) investigation, being present in each of their six sampled episodes. This said, basic mathematics concepts are consistently addressed in 'Kibena and the Math Rats', with these employed to solve various problems that arise throughout the narrative. For example, subtraction is used to identify whether or not all rats have been relocated to a new home (see Figure 1.2, below).

---

[5] Analysis of the eBook app or feature phone content is not conducted, as the Ubongo Kids television show is the medium through which most users engage with Ubongo's primary age material, and the format in which most educational content is available.

[6] It is noted that this episode was a 'special', delivered in partnership with Malaria No More, which would explain this episode's divergence from mathematics.

Figure 1. 2: Mathematical concepts addressed

Pedagogic approach

The examination of sample episodes also provides information on the pedagogic approach through which material concerning mathematics, the enjoyment of mathematics, health and socio-emotional guidance is delivered. Pedagogic investigations into other educational shows have reportedly unearthed "oppositional methodologies" such as repetition-based learning and more "constructionist" approaches (Cole, 2016, p. 24). The following examination of Ubongo's cartoons involves the identification of similar methodologies. However, unlike Cole (2016), I do not seek to portray constructionist and repetition-based methodologies as conflicting.

The portrayal of constructionist approaches and repetition-based learning as 'oppositional' threatens to create an artificial dichotomy in which these approaches might be caricatured as 'good' and 'bad'. This overlooks issues of cultural applicability (which might influence a child's progress: Sternberg, 2007). The adoption of constructionist approaches in television-based content might not be hindered by lack of teaching resources, large class sizes and unsuitable teaching environments (all of which were cited by sub-Saharan teachers as being obstructive to the implementation of constructive practices: Schweisfurth, 2013). However, Tanzanian children might simply respond more positively to didactic techniques identified to be utilised frequently amongst East African teachers (Hardman, Abd-Kadir, Agg, Migwi, Ndambuku & Smith, 2009). Indeed, concerns regarding the cultural applicability of pedagogic approaches are particularly pertinent to the implementation of any intervention delivered through technology-based media that are 'foreign' (at least in their conception: Villanueva-Masilla & Olivera, 2012).

In accordance with Ubongo's intention to teach 'without rote memorisation' (see Ubongo Kids' TOC, Figure 1.1), neither sampled episode employs any form of didactic methodology at the outset. In 'The Mosquito Army', for example, it is not immediately stated that mosquitoes might cause malaria. Instead, Kiduchu, the child detective, begins by investigating how people developed various symptoms. Nevertheless, further into Kiduchu's investigation, key information is repeated (albeit in song form, see Figure 1.3). Indeed, having overheard a mosquito army briefing, Kiduchu even writes bullet points in her notepad regarding how to prevent 'attacks': "1. make sure everyone uses their bed nets; 2. fix the protection in homes to prevent the mosquitoes from getting in; and, 3. Clear out the puddles". The following section shall now consider the theoretical standpoint adopted in this thesis.
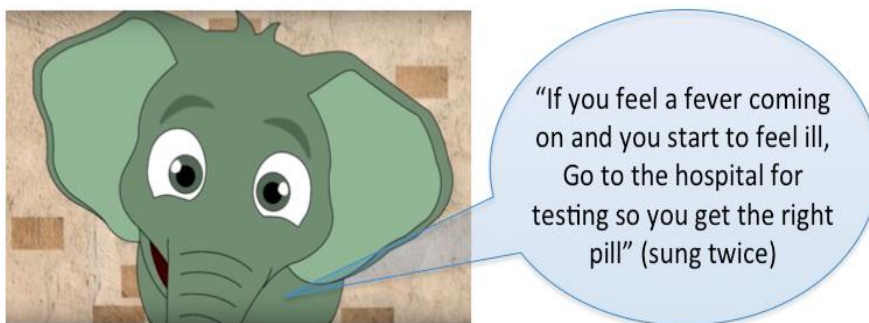
Figure 1. 3: Content repetition

1.3 Theoretical standpoint

1.3.1 Human capital

This thesis is guided by human capital theory (HCT). Under HCT, human capital (HC) can be represented by a stock of skills or knowledge (Acemoglu & Autor, 2011). HCT considers this stock of skills or knowledge as a capital good, "which emphasizes that the development of skills is an important factor in production activities" (Olaniyan and Okemakinde, 2008, p. 479). Indeed, the core tenet of this theory is that "skill development can lead to greater wealth accumulation in the future", with education therefore considered "a financial investment" (Alcott, 2016, p. 16). Following Heckman (2008), the returns to educational investment are understood to be higher when this investment is made in children and younger people. This is because "the old have a shorter time to recoup their investment", and that "Skills acquired early on make later learning easier" (ibid, p, 8). HCT is the core theory within the economics of education sphere (Tan, 2014), with the economics of education itself concerned about returns on investment in education and "how best to allocate scarce resources in education" (Dearden, Machin & Vignoles, 2011, p. 85).

It is acknowledged that HCT is susceptible to criticism. Critics might argue that HCT fails to provide an explanation for how education augments productivity (Marginson, 2019). HCT is also not a theory of learning, i.e. it cannot explain how educational television viewing leads to HC development

(Section 1.3.2). Further, it has been contended that HCT offers limited explanation even within its restricted scope. While the core tenet of HCT is that skill acquisition leads to future wealth, both education and broader measures of skill can "account for only a small fraction of income variation" (Fix, 2018, p. 16).

After acknowledging these criticisms, it should be recognised that the economics of education field offers alternate theories. These include signalling, which refers to the notion that "education serves as a signal to employers of the likely productivity of employees" (Johnes, 1993, p. 18). In some nations, the signal provided through qualifications explains a greater proportion of wages than measured skills (OECD, 2014). If signalling is the dominant reason why those with more education earn more, television exposure may not actually increase later earnings unless viewers becoming more skilled increases their chances of acquiring more or better qualifications.

Yet studies have found higher earnings and employment rates among individuals that demonstrate higher levels of literacy proficiency regardless of educational attainment, and this trend has been found to hold across a range of countries (OECD, 2014). This implies that there is a return on investment in skills and HCT could therefore still provide a valuable framework for this research. Indeed, economic theories centred on qualifications are not directly applicable to this study. Results from the Uwezo assessments used to examine child outcomes cannot be viewed by any future employers (due to Uwezo's anonymisation of study participants), so could not provide any form of 'signal'. Accordingly, HCT must be seen as better suited to the assessment data employed in this thesis.

HCT also appears to be aligned with the educational television intervention under investigation, Ubongo Kids. Correspondence between Ubongo Kids and HCT is evidenced by comparing the theory of change (TOC) for Ubongo Kids (Figure 1.1) with features of the HCT approach. Ubongo Kids' TOC shows that the programme equates to an investment in skills amongst young viewers. Additionally, the final intended 'impact' of Ubongo Kids is to promote the adoption of a 'positive learning mindset' throughout life. Both these points match the aforementioned ideas of Heckman (2000). Further, the low cost of Ubongo Kids per viewer (see Section 10.5) is pertinent to resource allocation, which is key to the economics of education.

This being said, the selection of HCT as the overriding theoretical approach in this thesis was not a result of its alignment with Ubongo Kids. Instead, HCT was chosen due to my theoretical standpoint. That is, the employment of HCT reflects my belief that an efficient allocation of educational resources should be promoted.[7] This theory would therefore remain applicable should I be conducting research

---

[7] It might also be noted that proponents of human capital theory include various prominent international actors. Oliver (2004, p. 120) suggests that the appeal of human capital theory comes partly "from the donor funding that accompanies it" (see further, Tikly & Bond, 2013). Therefore, presenting my argument in a HCT

into an alternate intervention, even if it were the case that characteristics of the intervention conflicted with the ideas of HCT proponents.

For the purposes of this study, HCT is used to create a framework through which the costs and benefits of educational television are considered. While not all costs and benefits that feature in this framework are directly accounted for in subsequent cost analysis (Chapter 9), the framework provides a useful basis for discussion. Indeed, this framework points to limitations in subsequent cost analysis calculations. These calculations follow a predetermined formula that does not account for various costs and benefits recognised in this framework (specifically, psychic costs or any benefits aside from the 'enhanced skill set' attributable to Ubongo Kids).

Costs are borne at the private level through opportunity cost, psychic cost and the cost of materials. Social costs include all private costs, as well as the production costs of Ubongo Kids. The private benefits of educational television viewing include a gain in skills from the short-term onwards, with this leading to longer-term economic gains attributable to the possession of these skills as well as additional psychic income (should non-financial benefit be derived from employment opportunities that would not have been available without gaining skills from cartoon viewing). It should also be recognised that economically valuable skills could be obtained that are non-cognitive, such as the socio-emotional guidance identified to be present in sampled episodes (in a manner consistent with Heckman's interpretation of HCT, which emphasises the importance of recognising non-cognitive indicators). Further, television is likely to have some immediate consumption value to the viewer. Whilst this conception is understood to be highly difficult to quantify (Gullason, 1989), the qualitative component of this study suggests that educational television provides gratification through viewing pleasure (see Section 8.6). Social benefits incorporate all private benefits, in addition to the positive externality of having more workers with higher skill levels in the Tanzanian economy (although it might be noted that little evidence of such "unexploited externalities" has previously been unearthed: Heckman, 2000, p. 7).

---

framework might be a particularly expedient means of guiding donor investment towards more efficient projects (given donors' preference for this theoretical approach).

Table 1. 1: HCT framework

Private costs	Private benefits
Opportunity cost	**Enhanced skill set**
Educational television viewing might incur an opportunity cost if programmes are viewed instead of doing another productive activity. This is more likely to occur if television is perceived to be educational in nature by parents or children and are therefore prioritised over other tasks.	Television viewing may increase the stock of skills possessed by the viewer (including both mathematics-based and socio-emotional skills), with these gains linked to longer-term economic and psychic benefits.
Psychic cost	**Longer-term utility**
Television viewing could have a psychic cost, should the child perceive viewing educational television to require effort.	Possessing higher cognitive and non-cognitive skills could increase future economic and psychic income.
Materials cost	**Consumption value**
There would be a materials-based cost associated with television viewing if a television (or alternate media platform) used to view Ubongo Kids was purchased specifically for that purpose.	Immediate gratification may also be derived from simple viewing pleasure (as a result of the television show being perceived as either enjoyable or even educational).
Social costs	**Social benefits**
Production costs plus all private costs	**Impact on wider economy plus all private benefits**
Production costs cover all costs incurred by Ubongo in delivering Ubongo Kids. Social costs also include all the private costs recognised above.	The wider economy could benefit from possessing a greater number of higher skilled workers. Social benefits also include all identified private benefits.

1.3.2   Transfer of learning

While human capital theory provides a framework to consider an investment in educational media, this theory provides no explanation as to *why* watching television shows might be associated with an 'enhanced skill set' (recognised as a potential private benefit of educational television in the above table). The majority of prior research that might evidence such an explanation has focused on the cognitive factors of 'learning transfer' (see, for example: Fisch, Kirkorian & Anderson, 2005; and, Piotrowski, 2014).[8] Within this literature, the most apt structure through which to consider how television affects the attainment of skills appears to be that created by Fisch (2004). This theory warrants some further explanation owing to its application in studies of educational television (by: Piotrowski, 2014; and, Bonus & Mares, 2015) and frequent recognition in the literature on learning

---

[8] There has also been some investigation into the applicability of social cognitive theories (when contemplating issues of trust and identification with characters: Schleisinger, Flynn & Richert, 2016) and behavioural theories (to consider children's formulation of action schemas based on character behaviours and subsequent use of this "information to guide their own future behaviors": Lavigne & Anderson, 2012, p. 110).

from television (see: Anderson, Lavigne & Hanson, 2013; and Kirkorian, Wartella & Anderson, 2008). There are three key elements of Fisch's (2004) 'transfer of learning' theory. These are listed below (employing the locution used by Fisch, Kirkorian & Anderson, 2005):

1. The viewer's initial learning or comprehension of the educational content in a television program;
2. The nature of the viewer's mental representation of that content; and,
3. The transfer situation – that is, the novel problem or solution to which the content is subsequently applied (p.379)

Not all of these elements could, however, have been examined specifically in this study (which precludes the adoption of this model as an overarching theory). This is due to (limited) data availability. The lack of information captured immediately after viewing precludes attempts to examine the viewer's initial comprehension (component one) or mental representation of knowledge (component two). Available data would only permit legitimate examination of the transfer situation (component three). Indeed, even if the data employed in this project could permit the exploration of Fisch's suggested process in full, concerns regarding the accuracy of applying this model remain (as explained below).

The extent to which information would be learnt from a television programme, abstracted, and then used to answer questions in assessments is clearly questionable. Given the potential length of time between programme viewing and testing (which will often include time at school through which assessed concepts are taught), it might instead be that television viewing enhances learning from other sources – by promoting more regular employment of mathematical concepts in informal day-to-day contexts, or helping children understand concepts that do not feature in Uwezo assessments but which provide a foundation for children to then learn tested concepts in school.[9] 'Improved school performance' was listed as an objective in Ubongo's initial TOC, and the promotion of everyday usage of mathematical concepts commonly features in episodes (see, Figure 1.4) (although sampled cartoons did not feature the setting of specific tasks to be conducted by children before viewing later episodes). It could even be that learning from television differs in nature to both these alternate paths, and the process suggested by Fisch (2004). Any technological artefact "and its modalities of use

---

[9] Indeed, Anderson, Huston, Schmitt, Linebarger & Wright (2001), found that viewing Sesame Street at a young age corresponded with higher grades once children had started school (with controls including parents' education): programme viewing was considered to "help create early success in school as well as foster an interest in learning and academics" (Kirkorian & Anderson, 2008, p. 194). The hypothesis that educational television viewing could facilitate school-based learning is also consistent with findings suggesting that attainment gains from early viewing persist in later years (Wright, Huston, Murphy, Peters, Piñon, Scantlin & Kotler, 2001: Section 2.2.1.1).

become aligned with user concerns and adapted to use settings" (Ruthven, Hennessy & Deaney, 2008, p. 298): Ubongo's educational content might be utilised in unanticipated manners in accordance with viewer requirements (which potentially vary between user groups). Each key element of Fisch's approach is now considered individually, before the employment of theory in this study is summarised.

Figure 1. 4: Encouragement of day-to-day mathematics use

1. *The viewer's initial learning or comprehension of educational content*

Initial learning or comprehension is perhaps the most intuitive requirement for transfer of learning from educational television: "if viewers have not fully understood the material presented in the program, they can hardly be expected to apply it in other contexts" (Fisch, Kirkorian & Anderson, 2005, p. 378). Initial comprehension might be considered as a topic in its own right (as in Linebarger & Piotrowski, 2010; and, Fisch, 2004). Indeed, this component of Fisch's theory can draw upon a specific model of initial comprehension – the capacity model (which has itself been supported in prior research: Piotrowski, 2014).

The capacity model perceives children's capacity to be determined by three elements: "processing of narrative, processing of educational content, and the distance… between the two" (Fisch, 2004, p. 143). Comprehension of educational material is increased where: the distance between narrative and educational content is small (e.g., narrative is intertwined with educational content); narrative processing requires few working memory resources (with this promoted where children have prior knowledge of the subject content or programme, Piotrowski, 2014); and, the "processing demands of the educational content are small (e.g., because it is presented clearly…)" (Fisch, 2004, p. 168). Preliminary investigation into Ubongo Kids sample episodes appears to correspond with these factors

for initial comprehension. Stories appear to be interwoven with educational content. The same key characters return in each episode (thereby reducing narrative processing demands). Additionally, ongoing formative research is conducted to ensure that programme material is presented in a manner that is clear to children.

2. *The nature of the viewer's mental representation of educational content*

Following initial comprehension, the transfer of learning theory next suggests that mental representation must be "abstracted beyond the initial context in which it was encountered" to permit application to alternate situations (Fisch, 2004, p. 169). Learning transfer could therefore be obstructed if educational content can only be considered in its original context. This might occur if children perceive events in educational cartoons to be pure fantasy, or something that merely occurs on a screen (Bonus & Mares, 2015; Schleisinger, Flynn & Richert, 2016). The fact that Ubongo Kids features talking animals could potentially lead to this perception.

Additionally, abstraction could be hindered if educational content is firmly embedded in the story of a show (as was identified to be the case with sampled *Ubongo Kids* episodes). This point could be perceived to highlight an inherent contradiction in the theory utilised: whilst a narrative intertwined with educational content could promote initial comprehension, content that is "too closely tied to its original context" might be difficult to abstract (Fisch, 2004, p. 169). This could, however, potentially be resolved through the presentation of educational content in multiple contexts (*ibid*; Kirkorian, Wartella & Anderson, 2008). Whilst not specifically addressed in Ubongo's theory of change, investigation of the Ubongo Kids curriculum map suggests that the addition and subtraction of integers (which features in the sampled episode, 'Kibena and the Math Rats') arise in alternate contexts (with this material included, for example, in another Series One episode, 'Banana Thief').

3. *The transfer situation*

The final stage of this theory considers the transfer situation, which occurs when "children encounter a problem to which the content might be applied … that is, the processing that allows them to retrieve the appropriate information from memory and apply it to the problem at hand" (Fisch, 2004, p. 170). Following comprehension and abstraction, it is also required that information is perceived as applicable to another task, which, like the process of abstraction, is considered to be more likely if children perceive information as relevant to the "real world" (Bonus & Mares, 2015, p. 8). This might correspond with Fisch's (2004) supposition that transfer is dependent on the initial comprehension and subsequent (real world) application situations being perceived as sufficiently similar. Again, this may depend on the manner in which children perceive the knowledge emitted by various characters

(including talking animals, as mentioned above). The success of a transfer can also be influenced by children's age and motivation (Fisch 2004; Haskell, 2000).

The nature of evidence required to fulfil the final component of Fisch's (2004) transfer of learning theory might, however, require reconsideration. Under Fisch's model, the transfer situation is only evidenced through "the application of knowledge or skills learned in one context… to a new problem or situation that differs from the one that was encountered previously". A "more flexible form of transfer" (Hennessy, 2014, p. 8) could be provided by a broadened conception that includes emphasis on 'preparation for future learning' (PFL) (Bransford & Schwartz, 1999). This approach would consider the effect of prior learning on future ability to learn, which children might demonstrate by "ask[ing] the right questions and seek[ing] appropriate information to help them in approaching a new problem" (Fisch, 2004, p. 177). Further, some conception of 'transfer' could be considered to have taken place before a child demonstrates the application of learnt skills to a new problem (potentially through their performance in a standardised test). That is, knowledge obtained from viewing educational television might be recontextualised in any manner of situations aside from formal assessment (including day-to-day life). Indeed, the likelihood of this occurring with regards to the educational show under investigation is enhanced by programme content that promotes the employment of mathematical concepts in everyday settings (see Figure 1.4 above). Neither prior recontextualisation nor PFL gains would be recognised under Fisch's approach (although it should be noted that existing research has often suggested that educational television produces gains in standardised test scores, see Section 2.2.1).

1.4   Philosophical standpoint

The philosophical considerations relevant to this study are now considered. These philosophical considerations are the product of extensive debate. This section does not, therefore, cover the complete historical development or provide an in-depth account of differing ontological, epistemological or paradigmatic approaches. Instead, this section intends to provide sufficient information on my philosophical position to inform subsequent discussion of the existing literature, methods and results.

1.4.1   Epistemology and ontology

Whist the concepts of epistemology and ontology often emerge simultaneously (and are considered alongside each other in this sub-section to mirror the research process depicted by Crotty, 1998), they are necessarily distinct. Ontology is "the study of being", and is therefore concerned with "what is", the "nature of existence" and "structure of reality" (Crotty, 1998, p. 10). Epistemology is concerned with the nature of knowledge, thereby providing a philosophical basis "for deciding what kinds of

knowledge are possible and how we can ensure that they are both adequate and legitimate" (Maynard, 1994, p. 10, originally cited in Crotty, 1998).

My ontological approach is one of critical realism. I believe there is a reality, yet my position possesses a critical dimension. This critical element arises from the understanding that "any attempts at describing and explaining the world are bound to be fallible, and also because those ways of ordering the world, its categorisations and the relationships between them, cannot be justified in any absolute sense, and are always open to critique and their replacement by a different set of categories and relationships" (Scott, 2005, p. 635). Reality can, therefore, only be approximated (Denzin & Lincoln, 2013). Clarity regarding this position is particularly necessary due to my subsequent use of quantitative methods, which are recognised to provide only estimations dependent on the form of measurement employed. Difficulties concerning the approximation of reality are particularly pertinent to issues later discussed concerning the measurement of exposure to television: whilst methods based upon character recall might be considered superior to measures of stated viewing levels, neither can be considered an entirely accurate portrayal of fact (see Section 3.3.1.3.1).

My epistemological perspective is that objectivity should ultimately be maximised. I do not believe that values are "objectified in the people we are studying and [that] we can discover the objective truth" (Crotty, 1998, p. 8); nor that "meaning, and therefore meaningful reality, exists as such apart from the operation of any consciousness" (ibid). Instead, I understand knowledge to be necessarily value-laden and 'pure' objectivism therefore impossible. The knowledge obtained through my study is clearly shaped by the values of (and relationships between) numerous parties, including study participants, data collection assistants and myself. This does not, however, suggest the pursuit of objectivism to be inappropriate. Indeed, objectivity is promoted wherever possible. This includes, for example, using secondary data featuring standardised mathematics assessment. Prior studies concerning educational cartoons in low- and middle-income countries suggest a similar epistemological position, through their adoption of RCT designs and standardised assessment (see, for example, Borzekowski & Henry, 2011).

1.4.2 Research paradigm: post-positivism

Given my acknowledgement that objectivism is unobtainable, my research could not be considered wholly positivistic. Positivism might consider both belief and enquiry to be limited to "what can firmly be established" (Cohen, Manion & Morrison, 2011, p. 7). This conception is incompatible with my position that an objective understanding of reality can only be approached, as knowledge is value laden. However, whilst I cannot find absolute truth, this neither denies the applicability of scientific methods nor prevents me from seeking absolute truth so far as is possible. As such, my philosophical

position appears to correspond instead with post-positivism, in which the researcher "claims a certain level of objectivity rather than absolute objectivity, and seeks to approximate the truth rather than grasp it in its totality or essence" (Crotty, 1998, p. 29). Knowledge cannot, therefore, be restricted to what is empirically verified. Rather, I consider knowledge (or 'provisional truths') to be created through continual attempts to disprove hypotheses, thereby considering hypothesis rejection as opposed to verification to be "at the heart of the scientific method" (Crotty, 1998, p. 32).

Conceptions of post-positivism are still subject to criticism. Regarding this thesis, these criticisms could be applied to the fact that attempts to isolate the association between educational television exposure and mathematics capability through statistical approaches are fatally flawed (Cohen, Manion and Morrison, 2011). By corollary, it could be contended that the inability to gauge all influences relevant to mathematics capability makes the identification of underlying homogeneity unattainable to the extent that any observed uniformities should be considered mere social constructs or illusions (Gage, 1989). I do not, however, consider this argument to prevent the adoption of a paradigmatic approach that appears to provide the most rigorous means of exploring my overarching and subsidiary research questions (as is reflected in my selection of designs: Section 3.3.1.4).

1.5   Chapter summary and thesis layout

This chapter began by providing contextual information on the locations (Tanzania and Dar es Salaam) and television programme (Ubongo Kids) that are considered in this study. After this, information was given on the overarching theory employed in this study (HCT), which was employed to generate a framework featuring the potential costs and benefits of educational television (Table 1.1). The choice of an economics of education theory was considered important, given the focus within this branch of education on the topic of resource allocation so important to educational policy in low-income contexts. Further, the available (Uwezo) assessment data used in this thesis lent itself to HCT as opposed to qualifications-based economic theories. This is because Uwezo assessment results are not identifiable for individual children, so this assessment data (which forms the basis of subsequent quantitative investigation) cannot not provide a signal of productivity to employers.

Amongst the acknowledged limitations of HCT, it was recognised that HCT does not provide any means of understanding why educational television might produce the private benefit of enhancing a child's skill set (which would itself be responsible for increased future incomes). The structure most suitable to explaining this was considered to be the 'transfer of learning' theory posited by Fisch (2004). Although it was recognised that the process suggested by this theory might not be a wholly accurate portrayal of reality (as well as the related point that this theory concerns only a narrow conception of learning), it is maintained that this approach provides a useful basis on which to consider how television viewing promotes skills. That said, it must be reiterated that this project

cannot explore each of the three elements of Fisch's approach: data are lacking on both immediate comprehension and abstraction (elements one and two). Fisch's theory is merely acknowledged to illuminate possible skill set gains that could be identified to be associated with educational television in this project, and that have (or have not) been found in prior studies.

The overarching theoretical perspective adopted (HCT) corresponds with the philosophical approach of maximising objectivity, and these premises underpin all the following chapters. Their influence is clearly reflected in both the literature search approach and nature of articles discussed in the following ('Literature review') chapter. The methods outlined in the 'Methodology' (Chapter 3) are similarly formulated to approximate reality so far as possible. After presenting 'Descriptive statistics' (Chapter 4), these methods are used to produce results relevant to quantitative datasets concerning Ilala and Temeke district (Chapter 5) and Tanzania as a whole (Chapters 6 and 7) as well as qualitative data from Dar es Salaam (Chapter 8). In the penultimate chapter of this thesis (Chapter 9), the HCT framework is influential in providing cost-effectiveness findings concerning Ubongo Kids. The thesis concludes (in Chapter 10) with a discussion of findings.

2 Chapter 2 – Literature review

2.1 Chapter overview

This chapter presents a review of the literature relevant to this study, which culminates in the creation of two research questions. As stated at the outset of this thesis, the primary research question concerns the association between educational television exposure and mathematics capability while the subsidiary research question concerns educational television's cost effectiveness (Section 1.1). The literature searches presented in this chapter targeted topic areas that were established through initial literature scoping. As a result, the parameters for literature searching were not entirely open ended. However, this literature review shaped the manner in which research questions were both articulated and addressed.

The primary and subsidiary research areas were investigated using different methods. A structured approach was used to retrieve literature relevant to the primary research area. This approach was taken in order to provide a (relatively) complete overview of published material on educational television and attainment or capability. The volume of publications retrieved by this approach permitted a comprehensive investigation involving descriptive quantitative analysis of all identified publications and qualitative assessment of selected publications. Conversely, a more time-efficient approach was adopted to explore the subsidiary research area, since this investigation relied in part on findings regarding the primary research area. The literature search for the subsidiary research area employed a non-systematic approach, with retrieved publications subject to a critical (qualitative) assessment.

This review was conducted in accordance with my theoretical and philosophical standpoints (outlined in Sections 1.3 and 1.4). The presence of these standpoints is evidenced in various manners. Firstly, exploration of the subsidiary research area – cost analysis – corresponds directly with the consideration of education as a financial investment (which is recognised as a core tenet of human capital theory: Alcott, 2016). Further, my philosophical position that objectivity should be maximised is shown by my emphasis on literature search results featuring the use of quantitative methods (which are scrutinised with regards to their rigour: Section 2.2). Lastly, the use of computer-driven literature searching and analysis techniques undoubtedly benefitted the provision of transparent and unbiased information (on, for example, the countries with which identified studies were concerned: Section 2.2.4).

There are four subsequent sections to this literature review. These concern the search methods, review methodology and search findings relevant to the primary and subsidiary research areas (Sections 2.2 and 2.3). After this, a summary of the review and discussion of research gaps are presented (Section 2.4). Finally, this chapter culminates in the provision and consideration of the adopted research questions (Section 2.5).

2.2   Learning from educational television

A search of the literature was conducted using the Educational Resources Information Center (ERIC), Web of Science (WoS) and Google Scholar search platforms. The search terms 'educational television' and 'educational cartoon' were each combined with 12 additional keywords relevant to education and educational media (learn, attainment, effect, outcome, gain, intervention, ability, capacity, capability, impact, Ubongo, Akili), whilst 'Sesame Street' was employed as a stand-alone search term. The search was restricted to English language publications. WoS and ERIC were each successfully searched for all 25 search terms or combinations. However, searching on Google Scholar was restricted to 4 search combinations. Systematic searching using Google Scholar was unfortunately restricted by its security mechanisms, which ultimately led me to be 'locked out' of the search platform entirely. Structured searching was supported by non-structured investigation (involving snowballing from key sources) into topics of relevance to the primary research question. This process uncovered a small number of new (unique) results. All search results were imported to the reference manager, Zotero.

The items imported to Zotero were then extracted to R in order to compile results in a single dataframe. Extraction was performed using the RefManageR package in R. RefManageR was programmed to locate unique Zotero tags applied to Zotero items (with tags corresponding to the search number through which items were identified). However, initial attempts to import items unveiled an apparent cap of 99 items per extract (not currently recognised in RefManageR documentation). As such, additional unique Zotero tags were created for search numbers where there were more than 99 imports. For example, search number 21 featured three separate tags each applied to 99 or fewer items imported from that search. Of the 2679 items imported to Zotero, 2545 items were successfully extracted to R. The loss of 224 items from Zotero to R resulted from RefManageR's extraction process requiring key information fields to contain data (such as item title or year).[10] All separate item imports were then merged into a single dataframe that retained all available meta information. Eleven non-English items were removed from this dataframe, leaving 2534 items. The individual searches, search platform, items found, and items imported to R are presented in the table below (Table 2.1). (Table 2.1 presents search results before the removal of eleven non-English items.)

---

[10] Given that items which were not successfully imported lacked key information, they were abandoned (as opposed to entered manually).

Table 2. 1: Information on search strings, total results and results imported to R

Search number	Search term 1	Search term 2	Search platform	Total results	Imported results
1	Educational television	Learn*	WoS	79	79
2	Educational television	Attainment	WoS	0	0
3	Educational television	Effect*	WoS	17	17
4	Educational television	Outcome*	WoS	8	8
5	Educational television	Gain*	WoS	6	6
6	Educational television	Intervention*	WoS	15	15
7	Educational television	Abilit*	WoS	19	19
8	Educational television	Capacity	WoS	6	6
9	Educational television	Capability	WoS	1	1
10	Educational television	Impact	WoS	34	34
11	Educational cartoon*	Learn*	WoS	3	3
12	Educational cartoon*	Attainment	WoS	0	0
13	Educational cartoon*	Effect*	WoS	7	7
14	Educational cartoon*	Outcome*	WoS	4	4
15	Educational cartoon*	Gain*	WoS	0	0
16	Educational cartoon*	Intervention*	WoS	3	3
17	Educational cartoon*	Abilit*	WoS	0	0
18	Educational cartoon*	Capacity	WoS	0	0
19	Educational cartoon*	Capability	WoS	0	0
20	Educational cartoon*	Impact	WoS	2	2
21	Educational television	Learn	ERIC	215	215
22	Educational television	Attainment	ERIC	26	26
23	Educational television	Effect	ERIC	614	405
24	Educational television	Outcome	ERIC	109	109
25	Educational television	Gain	ERIC	133	133
26	Educational television	Intervention	ERIC	30	30
27	Educational television	Ability	ERIC	249	249
28	Educational television	Capacity	ERIC	74	74
29	Educational television	Capability	ERIC	98	98
30	Educational television	Impact	ERIC	373	358
31	Educational cartoon	Learn	ERIC	0	0
32	Educational cartoon	Attainment	ERIC	0	0
33	Educational cartoon	Effect	ERIC	4	4
34	Educational cartoon	Outcome	ERIC	0	0
35	Educational cartoon	Gain	ERIC	0	0
36	Educational cartoon	Intervention	ERIC	0	0
37	Educational cartoon	Ability	ERIC	2	2
38	Educational cartoon	Capacity	ERIC	0	0
39	Educational cartoon	Capability	ERIC	0	0
40	Educational cartoon	Impact	ERIC	2	2
41	Sesame Street	N/A – no term 2	WoS	233	233
42	Sesame Street	N/A – no term 2	ERIC	384	384
43	Educational television	Ubongo	GS	0	0
44	Educational television	Akili	GS	4	4
45	Educational cartoon*	Ubongo	GS	0	0
46	Educational cartoon*	Akili	GS	2	2
47	Educational television	Ubongo	ERIC	0	0
48	Educational television	Akili	ERIC	0	0
49	Educational cartoon	Ubongo	ERIC	0	0
50	Educational cartoon	Akili	ERIC	0	0
51	Educational television	Ubongo	WoS	0	0
52	Educational television	Akili	WoS	0	0
53	Educational cartoon*	Ubongo	WoS	0	0
54	Educational cartoon*	Akili	WoS	1	1
55	N/A – non-systematic	N/A – non-systematic	All	12	12
			All:	2769	2545

Note. In the above platform, all terms were employed within "quotation marks" and specified as conjunctive (i.e., requiring both terms to be present). Sesame Street was employed without any supporting terms in String term 2, given that the objectives of this programme were inherently aligned with supporting terms. An asterisk signifies the employment of a truncation symbol where permitted by a search platform. This enables the search platform to identify all words beginning

with a specified stem. For example, "cartoon*" would find "cartoon" and "cartoons". Blue text is used to highlight searches that produced no results. Yellow text is employed to emphasise searches for which not all identified items were successfully imported into Zotero. Green text is used to show non-systematic searching.

Deduplication was carried out on the merged imports dataframe based on the contents of a newly created item string containing year and title. To create this string, limited text transformation was conducted on item title information (removing double spaces and spaces preceding and following titles, transforming text to lowercase and removing non-alphanumeric characters) before this was merged with item year information (repeated twice at the beginning of the new item string).[11] For example, the article entitled "Early Childhood Television Viewing and Adolescent Behavior: The Recontact Study" (Anderson et al., 2001) was represented by the following string: "2001 2001 early childhood television viewing and adolescent behavior the recontact study".

The Reclin package for R was used to apply a Jaro-Winkler measure of edit distance between item strings. Jaro-Winkler is a character-based measure of similarity, which determines the likelihood of duplication based "on the number and order of the common characters between two strings" while assigning higher duplicate probabilities "to strings that match from the beginning for a set prefix length" (Gomaa & Fahmy, 2013, p. 14). A commonality between items was assumed for items with Jaro-Winkler similarity of >0.95.[12] Application of Jaro-Winkler deduplication reduced the literature corpus by 26 percent. Results were then scanned manually before individual deduplication errors were corrected manually in the R script. After correcting a small number of errors including both missed duplicates and falsely identified duplicates, 1438 unique items were marked and extracted into a separate dataframe. This dataframe provided the corpus of identified literature employed for descriptive quantitative analysis.

Before proceeding, it must be stressed that the items in this corpus did not represent all publications concerning the relationship between educational television on attainment or capability. Item identification was the product of a restricted number of search strings. Further, these strings were employed only on a small number of search platforms. Indeed, it was also the case that items of varying relevance to the effects of educational television were included. This point is highlighted by Section 2.2.3, 'Beyond attainment and capability', which considers results regarding the effect of television on prejudice reduction (that were identified exclusively through structured searching). Information on this corpus, including visual representation and subsequent deductions, should

---

[11] Year was placed twice (as opposed to once) at the beginning of each item string to facilitate deduplication amongst unique items with very similar titles in different years.

[12] This figure was chosen to obtain duplication information that gave the lowest number of errors (in the form of missed item matches or falsely identified matches) after eyeballing results from various similarity specifications.

therefore be considered to provide a coarse indication of patterns in the literature regarding learning from educational television.

The first descriptive visualisation of items comprising the corpus is a line graph (Figure 2.1). This line graph shows the number of unique items over time (in grey). This is achieved by plotting item count by year of publication. Further, a simple five-year moving average of unique items over time is produced (in black). This line plots the mean item count for a calendar year and for the four years preceding it, thereby producing a smoothed indication of research volume over time.[13] This moving average measure corresponds with that presented below when considering the frequency of item keywords over time (see Section 2.2.1).
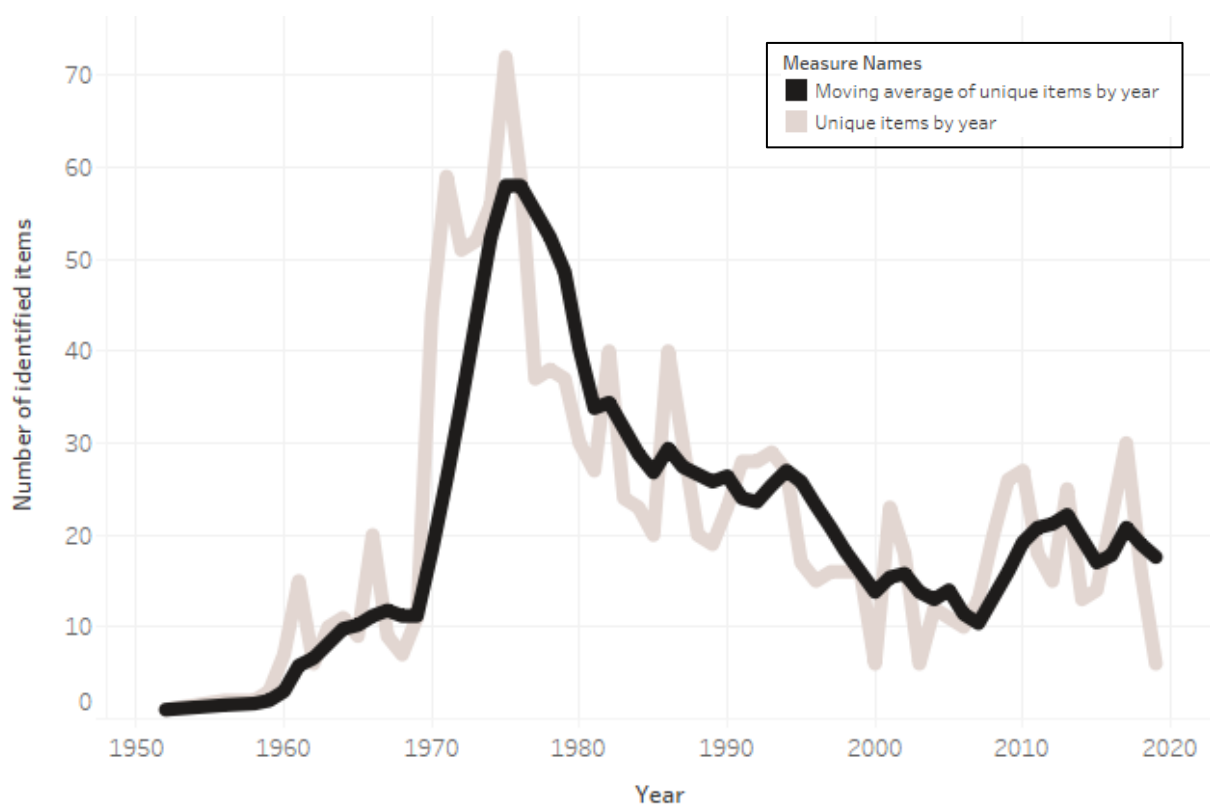
Figure 2. 1: Unique search results by year

The line graph above suggests that research interest in educational television and attainment or capability grew rapidly from 1960 to a peak in 1975. From 1975, decreasing interest in this topic is suggested by a steady decline in the number of identified publications. It is posited that a reduction in

---

[13] At time points where there are fewer than four previous years, the simple moving average includes all available years. For example, the moving average for 1952 is the document count for 1952 only. Further, it is acknowledged that a simple moving average is prone to lag due to its equal weighting of each year included. This method is, however, retained for its simplicity (and resultant transparency and replicability).

research interest from 1975 onwards could be attributed to the subsequent availability of alternate technologies (thus leading to a dilution of television-focused research within the education-technology literature).

Graphics presenting descriptive quantitative analysis are used to support the presentation of a critical (qualitative) engagement with selected items from the corpus. Item selection was based on a logic-based relevance judgement, with items considered to be more relevant should they have:

- Related closely to the primary research question
- Been published after the year 2000
- Concerned a developing country context
- Provided quantitative insight into learning

Information from chosen items are now presented in separate sections on the "Influence of viewer age range on learning from educational television" (Section 2.2.1), "Additional factors influencing impact" (Section 2.2.2), findings "Beyond attainment and capability" (Section 2.2.3) and "Research concerning educational television in developing countries" (Section 2.2.4). Focus on a restricted number of sections permitted the examination of details from individual items from a large body of identified literature.

2.2.1   Influence of viewer age range on learning from educational television

Items identified by the literature review that concerned particular age ranges were divided into the following groups: 'preschool learners', 'school-aged learners' and 'university and adult learners'. These categorisations were supported by visualisations of the corpus, illustrating the frequency of category-specific publications over time. The visualisations used exploited the presence of information relevant to an age-group within item keywords. To produce informative visuals, manipulation of the text strings forming item keyword information was conducted.

First, keywords were transformed to lowercase, and all non-alphanumeric characters, excluding commas, were removed. These cleaning measures were taken to prevent R from identifying a distinction between identical terms in different formats (e.g., "Higher education" and "higher education"). Next, a document term matrix was created which turned all cleaned keywords in the entire corpus into dummy variables for each item (identifying a keyword to begin and finish at a comma). Then, keyword dummies were retained only if they provided information on the age range of study subjects. Certain retained keyword dummies were subsequently removed if a keyword was employed in fewer than fifty items, to prevent graph lines for infrequent keywords cluttering the lower end of y axes. The following keywords were then assigned to each of the following age brackets (with the keyword "children" applied to both preschool and school-aged learner groups):

- **preschool learners:** early childhood education, preschool education, preschool children, young children, children
- **school-aged learners:** elementary education, elementary secondary education, children
- **university and adult learners:** higher education, adult education

To plot keyword frequency in a manner that was not influenced by fluctuations in annual publication frequency (as shown in Figure 2.2), the raw count of a keyword in any given year was divided by the number of corpus items in that year which contained at least one keyword. Thus, the presence of keywords was presented as a proportion of all articles with keyword information. Keyword proportions were plotted over time as a five-year moving average (following the method employed above, Figure 2.1). This plot precedes discussion of the literature relevant to each age bracket, with different lines highlighted according to the age range under review. However, these plots were limited in the extent to which they indicated focus on an age bracket within the educational television corpus (beyond the limitations noted with regards to establishment of the corpus itself). This was due to the following reasons:

- authors might have chosen not to specify an age-related keyword despite its apparent applicability
- articles might have used multiple keywords of relevance to (one or more) age brackets
- article keywords might have been influenced by the keyword requirements of recipient journals
- keywords indicative of age range might have occurred too infrequently (fewer than fifty times) to feature in the subsequent descriptive analysis, which could have arisen due to a lack of keyword correspondence between journals

Figure 2. 2: Moving averages of keywords relevant to preschool learners

The above graphic features all retained keywords of relevance to different age groups, highlighting only those keywords that could concern preschool learners.

The above graphic suggests that research concerning preschool learning has featured consistently in the educational television literature corpus from 1965 onwards. Further, relative interest in this age range appears to have increased from 1995. Search results containing keywords indicative of research pertaining to preschool learning included evaluative research regarding very young children. For such

children, it has been broadly accepted that substantial comprehension of child-focused programmes does not occur until approximately 24 months (Anderson & Subrahmanyam, 2017). Understandably, therefore, exposure to child-directed television has not generally been found to produce meaningful gains in attainment amongst infants or toddlers (see, for example, Barr, Lauricella, Zack & Calvert, 2010; and, Wartella, Richert & Robb, 2010). In fact, exposure to the children's television show, Teletubbies, amongst children up to 30 months was found to have a negative effect on vocabulary and expressive language attainment (Linebarger & Walker, 2005).[14]

However, regarding slightly older preschool children, the consumption of educational television has frequently been identified to have positive effects. This is considered to be a result of children becoming "capable of understanding well-designed educational programs" (while not being "engaged in formal schooling": Anderson, Lavigne & Hanson, 2013, pp. 7-8). For this age group, "The evidence shows positive effects from studies of exposure to single episodes to studies of sustained, repeated viewing of the series" (ibid, p. 9). Such findings have been reported since the very first season of Sesame Street. Ball and Bogatz (1970) found that the most frequent viewers of this programme showed the greatest advances in letter and number recognition. Since this piece of research, numerous studies have confirmed the positive effects of educational programmes on various indicators of preschool readiness, such as programme-specific and normative literacy outcomes (from exposure to Super Why!: Linebarger, McMenamin & Wainwright, 2008) and problem-solving strategy "that was both shown and not shown [during episodes of Blues Clues]" (Crawley, Anderson, Wilder, Williams & Santomero, 1999, p. 636).[15]

Several US-based studies have suggested that gains associated with preschool exposure are retained in subsequent years. Amongst lower income Kansas households, for example, children aged 2-3 who viewed educational programmes more frequently were identified to score higher across a battery of tests (Wright, Huston, Murphy, Peters, Piñon, Scantlin & Kotler, 2001). Furthermore, this "advantage associated with early viewing was still present [at later points]" (ibid, p. 1358). The potential for early viewing of preschool educational cartoons to impact on longer-term success could also be supported by the work of Anderson, Huston, Schmitt, Linebarger and Wright (2001). In a recontact study of 570 adolescents, Anderson and colleagues found that educational television viewing at preschool ages was associated with higher grades (as well as greater reading activity, reduced aggression and more creativity). In common with other cited studies having similar results, the methodology employed by Anderson and colleagues appeared thorough. Anderson and colleagues were able to recontact a high

---

[14] Linebarger and Walker (2005) did however find other child-focused shows (Dragon Tales, Dora the Explorer, Arthur, Clifford and Blue's Clues) to have a positive impact on vocabulary and language learning.

[15] Investigation into the research designs employed in cited studies concerning 'preschool learners' suggested that they possessed sufficient rigour. For example, Linebarger, McMenamin and Wainwright (2008) controlled for differences between control and treatment groups using a measure of pre-test ability as well as child composite and family composite scores (that were based on key child and family characteristics, respectively).

proportion of participants from two prior studies that provided detailed preschool viewing diary information. This diary information contributed to a longitudinal dataset, with post-baseline data including interview responses and school exam transcripts.

2.2.1.2   School-aged learners

Moving average of proportion of corpus items featuring a particular keyword

- Elementary Secondary Education
- Elementary Education
- Children
- Adult Education
- Higher Education
- Early Childhood Education
- Preschool Children
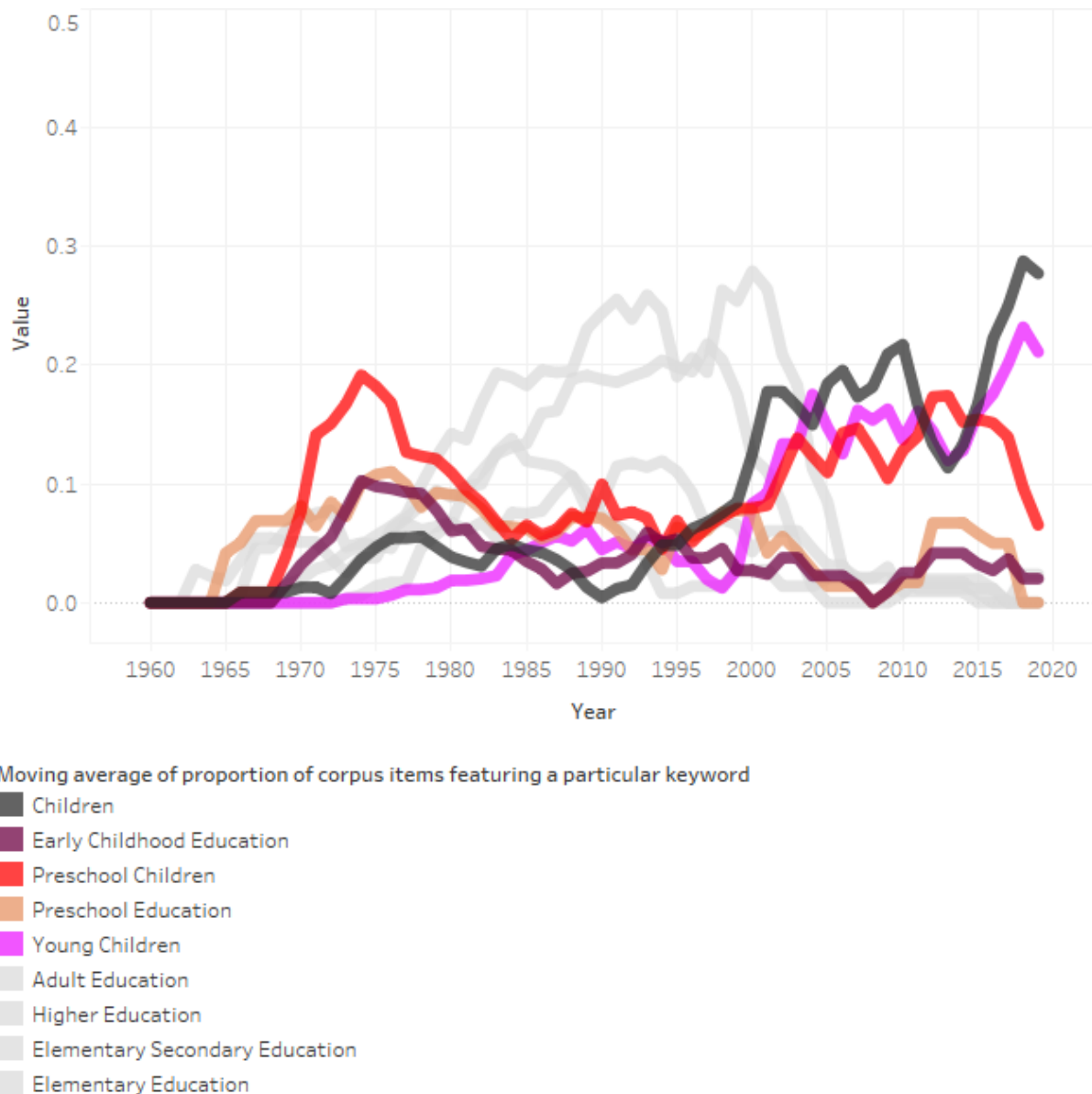- Preschool Education
- Young Children
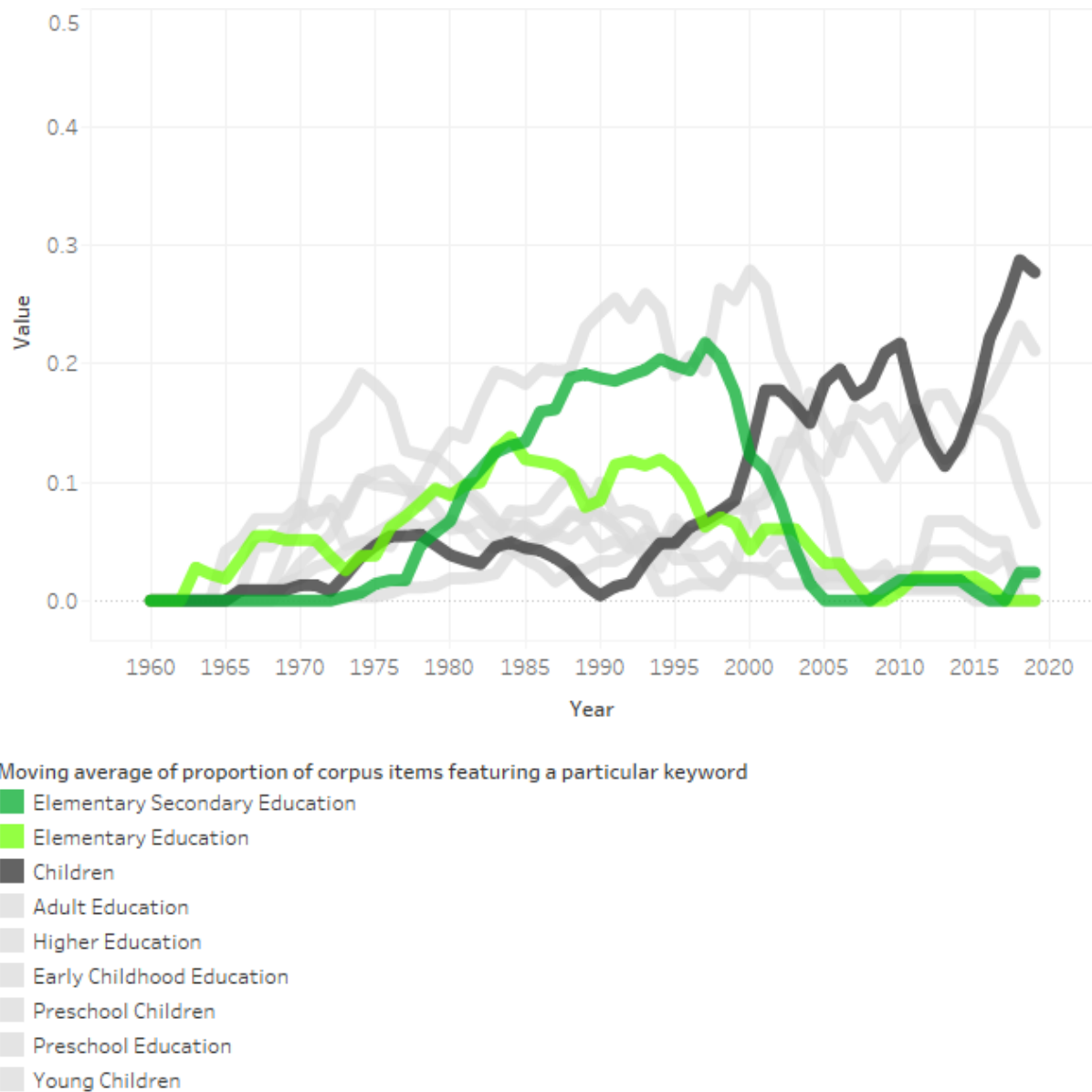
Figure 2. 3: Moving averages of keywords relevant to school-aged learners

The above graphic features all retained keywords of relevance to different age groups, highlighting only those keywords that could concern school-aged learners.

The above graphic shows that fewer keywords related to school age learners appeared in the identified educational television research corpus. There were three keywords of possible relevance to school age children. These were "elementary secondary education", "elementary education" and "children" (with this last keyword also considered to be potentially indicative of research concerning preschool age learners). The limited presence of relevant keywords corresponds with the smaller number of items relating to this age bracket considered to warrant human reading.

Articles selected for reading suggested that viewing amongst primary aged children also had positive effects on learning. In an early study concerning The Electric Company (a reading skills-focused television show), American children in Grades 1-4 who were shown the programme in school over six months performed significantly better in almost all the programme's goal areas (Ball and Bogatz, 1973). The programme was generally perceived favourably by teachers. Further, the programme had similar effects regardless of demographic or gender. However, effects were less pronounced amongst Grade 3-4 children than Grade 1-2 children.

Approximately a decade later, Webb (1982) conducted a summative evaluation of the Let Me See! science-focused television series. This series comprised twelve 15-minute shows on science-based topics. Webb (1982) employed a sample of 59 Grade 1 and 2 classes located in Wisconsin. 48 of these classes formed a treatment group which viewed the entire Let Me See! series and participated in related activities during class. 11 classes formed a control group which received no intervention. An investigation into the effects of the programme on this sample population was conducted using analysis of covariance (ANCOVA), treating class as the N and pretest scores as the covariate. Results suggested that the intervention had a positive and significant effect on science scores for classes in both grades. While the approach adopted by Webb (1982) was certainly more considered than other designs recognised in this review (e.g., Lapinid, Gustilo, Magno, Barrot, Gabinete & Anito, 2017, below), Webb's (1982) methodology still appeared limited. Only pretest scores were used to control for differences between treatment and control groups.[16]

Nevertheless, results from the two papers considered above appeared to correlate with those from studies with more rigorous research designs investigating the effect of television on attainment amongst school age children. These included a paper on the effects of TV viewing amongst Grade 1 and 2 children (6-8 years), which used a randomised controlled trial (RCT) (Borzekowski et al., 2019) (considered in Section 2.2.4.1). Further support was provided by more recent reviews. For example, Anderson, Lavigne and Hanson (2013, p. 14) recognised that science-based programmes including Bill Nye the Science Guy and 3-2-1 Contact had been found to produce "a better understanding of basic scientific knowledge".

---

[16] Further, this "evaluation was conducted by the agency who also produced the series" (albeit with effort to avoid any resultant bias) (Webb, 1982, p. ix).

2.2.1.3   University and adult learners

Moving average of proportion of corpus items featuring a particular keyword
- Adult Education
- Higher Education
- Elementary Secondary Education
- Elementary Education
- Children
- Early Childhood Education
- Preschool Children
- Preschool Education
- Young Children
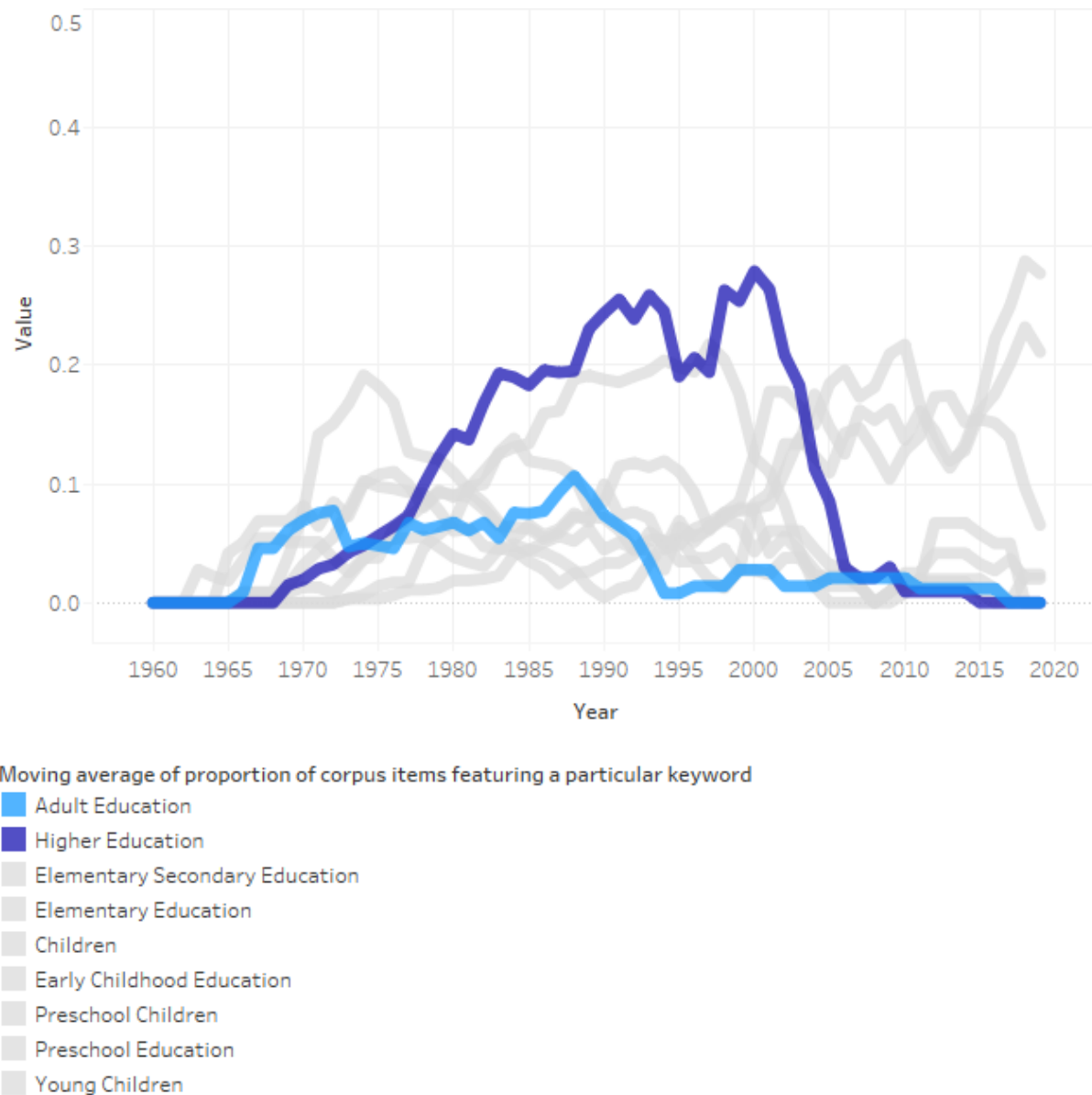
Figure 2. 4: Moving averages of keywords relevant to university and adult learners

The above graphic features all retained keywords of relevance to different age groups, highlighting only those keywords that could concern university and adult learners.

The graphic above suggests that the educational television and attainment or capability literature frequently featured age groups placed in the 'University and adult learners' bracket between approximately 1970 and 2000. However, information concerning higher education students has seemingly become increasingly sparse in more recent years. It is not entirely certain why research

concerning older age groups has become decreasingly prominent within the television-focused literature. As recognised below, television-based studies concerning university and adult learners have included positive findings. It could not therefore be assumed that mere ineffectiveness has led to limited implementation of television education in university and adult learning domains.

This said, it is possible that declining interest in TV for university-level education specifically might reflect the lack of cost-effectiveness gains achieved when employing television for small group learning. As identified through later empirical work (Section 8.7.1), Ubongo Kids' relative cost effectiveness is strongly influenced by its low cost per viewer. Obtaining such low per-person costs would certainly be hindered by the limited audience for (subject-specific) higher education interventions. Additionally, it is possible that there was a shift in adult or university learning research away from educational television towards other forms of educational technology (such as online learning courses for adults).

In the available literature, positive findings are reported by Boulet, Boudreault and Guérette (1998). This study employed a multi-group design to consider the effect of television-based learning on student outcomes in one unit of an undergraduate information technology course. Groups included a television distance education group and a lecture-based group (comprising 93 and 83 students respectively). Both groups followed the same course objectives and reportedly spent equal time on their studies. The study found that students in the television group significantly outperformed their compatriots in the lecture-based group on a test of fundamental unit knowledge. However, it should be noted that television-based learning was not found to be significantly superior to lecture-based approaches when considering its effect on outcomes for other units of the same information technology course (Boulet, Boudreault and Guérette, 1997).

Other studies concerning learners bracketed in this older age group considered the relative effectiveness of various formats of televised learning. Markham (1999), for example, investigated the influence of videotape caption availability on word recognition amongst university-level English as a Second Language (ESL) students. Here, 180 study participants viewed excerpts with or without captions from two differing educational shows. Assessment results showed that caption availability significantly improved ESL students' ability to recognise words that had featured in the video excerpts (Markham, 1999, p. 324). Further, "the effects of captioning remained consistent regardless of the passage content" (ibid, p. 325). The content of this study is of some relevance to the child-directed programme under consideration (Ubongo Kids), which features on-screen numbers to aid learning (see Figure 2.5).

Figure 2. 5: The presentation of numbers on screen during an Ubongo Kids episode

2.2.2 Additional factors influencing impact

Before proceeding to consider research focusing on developing countries, it is recognised that factors beyond viewer age could affect the influence of educational television on learning. One such factor, co-viewing with adults, might enhance the influence of television by "drawing children's attention to crucial moments in the program and answering questions about what they see" (Lavigne & Anderson, 2012, p. 119). The verbalisation of an adult co-viewer's "interpretations and evaluations of television programs and commercials to their children" could assist with the development of critical viewing skills (Corder-Bolz, 1980, p. 117). However, there could be variation in the influence exerted by different adult co-viewers, as "adults' own facility with the academic subject matter" will not be constant (Fisch, 2004, p. 134).

Another factor that could act to increase the effects of educational programmes is whether they (and associated learning materials) are available on multiple platforms. Platform availability is particularly pertinent to this project, as Ubongo's primary age resources are available via TV, YouTube, DVDs and an eBook app. Indeed, "Contemporary children are increasingly able to experience the content of their favorite TV programs on multiple platforms" (Lavigne & Anderson, 2012, p. 117). While there is "little research… on the effectiveness of multiplatform educational programming" (Anderson, Lavigne & Hanson, 2013, p. 16), one US-based study reports on the potential for such a format to improve learning outcomes (Fisch, Lesh, Motoki, Crespo & Melfi, 2010).[17] Amongst a sample of 672 8-11 year olds, the experimental component of this study suggested Cyberchase's positive effects on attainment were more consistent amongst those who were exposed to both videos and online games than either platform alone (ibid). Identification of the potential influences exerted by co-viewing and

---

[17] Another study has been identified which concerns educational programmes when available on multiple platforms, but primarily with regards to the implications of imposing temporal constraints on access (Paulsen & Andrews, 2014).

multiple-platform use led to their inclusion as interview questions in this thesis. Answers to these questions ultimately provided information to support interpretation of the relationship between exposure and capability (see, for example, Section 7.3.2.2). This literature review now progresses to consider literature relevant to developing country contexts.

2.2.3   Beyond attainment and capability

Items selected thus far for human reading concerned only the relationship between educational television and conventional school topics (such as mathematics and literacy). This focus reflected the primary research area, which influenced the choice of search strings. However, it was acknowledged that there was a large body of literature concerning the effect of educational television on outcomes that was not related to concepts such as attainment or capability. Indeed, this was highlighted by the occurrence of literature that did not primarily concern attainment or capability in the final corpus, despite this corpus largely being generated from structured searching based on strings that included terms such as "attainment" and "capacity".

This literature provided further information on the effects of television exposure upon behaviours such as adolescent alcohol consumption (van Leeuwen, Renes & Leeuwis, 2013) and economic measures such as labour market outcomes (Kearney & Levine, 2019). Amongst the available topics, focus was placed on prejudice reduction, because information on this concept was available with regards to adult viewers and child viewers (and was likely of relevance to Ubongo Kids, see below). This topic was addressed by Murrar and Brauer (2018) in their investigation of the effects of on-screen media upon racial prejudice. Experiments revealed that exposure to a sitcom with diverse and relatable Arab and Muslim characters led to lower levels of implicit and explicit prejudice amongst viewers (aged 18-60 and of Caucasian ethnicity). This finding was supported by a further experiment which suggested that exposure to a music video had greater effects on prejudice reduction than other established measures (Murrar and Brauer, 2018).

Similar positive findings regarding prejudice reduction were identified with regards to the destigmatisation of medical conditions amongst preschool children. This was considered with regards to perceptions of HIV positive children in a developing country context (with further information on studies concerning this context provided in Section 2.2.4) (Borzekowski & Macha, 2010). In this study, Borzekowski & Macha (2010, p. 302) reported that children who were more receptive to an international variant of Sesame Street "were more likely to say that an HIV positive child could play with others and that they would invite an HIV positive person into their home to share a meal". This finding is of relevance to Ubongo Kids, which recently introduced an Albino character named Amani. Stigma towards albinism and HIV might differ, but it is certainly possible that Ubongo Kids could positively affect attitudes towards this physical condition. This underlines the limitations of the

primary research question's restricted scope: educational television is likely to have a far broader effect than that estimated when focusing on only mathematics attainment or capability.

2.2.4 Research concerning educational television in developing countries

This section now gives information on learning from educational television with regards to developing country contexts. To show the relative focus on developing countries within the identified corpus, a map graphic is provided (see Figure 2.6). The production of this map graphic required further text manipulation to extract all references to countries within article titles, abstracts and keywords. To perform this manipulation, a string was created that included all title, abstract and keyword information. The following transformations were then performed:

- non-alphanumeric characters were removed
- text was altered to lowercase
- lowercase multi-name countries were temporarily altered to single-word names (e.g., "burkina faso" became "burkinafaso") and countries with differing spellings (or languages) were altered to a common spelling (e.g., "cabo verde" and "cape verde" both became "capeverde")[18]
- words were stemmed, to take a common form (e.g., "complicated" and "complicatedly" became "complicate")[19]

Following these transformations, the tm package in R was employed to create a document-term-matrix of all words in the text string for each item in the corpus. Terms featuring what were originally multi-word country names were then altered back to their initial (lowercase) format (e.g., "saudiarabia" became "saudi arabia". This matrix was then subsetted to retain only terms of interest (i.e., country names), by excluding all column names except for those featuring in a list of all recognised nations. Lastly, the document-term-matrix was converted to a binary format. This conversion was implemented so that findings could give an indication of the number of items that featured a mention of each country. It was therefore assumed that a piece of literature concerned one or more countries if those country names appeared in the item's title, abstract or keywords. Before presenting the graphic itself, shortcomings should be acknowledged (which act in addition to aforementioned limitations in establishing the corpus; see Section 2.2.1). These shortcomings included the following:

- literature items might omit any reference to a country of focus in their title, abstract or keywords
- items might mention a country that did not reflect the country focus

---

[18] This temporary manipulation is required to facilitate the subsequent creation of a document-term-matrix, which distinguished between terms based only on the presence of a space (i.e., " ").

[19] Stemming was applied to reduce the total number of words in all title, abstract and keyword strings. This assisted the subsequent creation of a document-term-matrix, which initially featured all words across all item strings.

- a small proportion of items in the corpus possessed no abstract or keyword information (9.8 percent and 14.2 percent of items respectively)

This said, using computer-driven analysis – to examine which countries identified articles were concerned with – appeared superior to human interpretation. Establishing the country focus of studies has indeed typically been established through human reading (such as in the systematic review conducted by Howe & Abedin, 2013), yet computer-based analysis has numerous advantages. Programming solutions circumvent the multiple errors that might occur in human reading of articles. For example, a human reader could potentially misinterpret whether an article considered a particular country (or numerous countries), especially when examining thousands of articles (as is the case in this instance). Also, establishing the country (or countries) of focus amongst numerous articles would be a hugely time intensive process in the absence of machine reading. Lastly, a programming-based solution meant that the process of deducing whether an article concerned a particular country was transparent (providing that the assumptions and steps of analysis were detailed, as they were previously in this section).
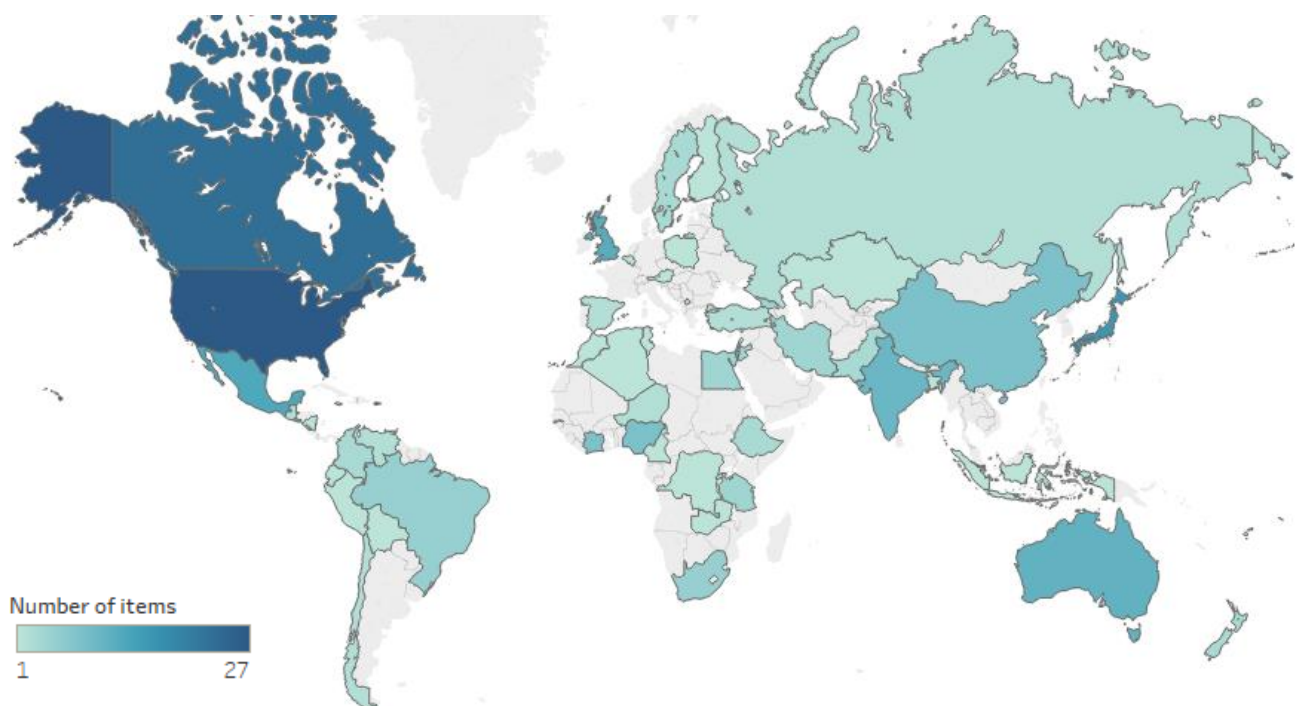
Figure 2. 6: Number of unique items in the entire literature corpus featuring reference to a country

This graphic suggests a relative lack of focus on developing countries. The three nations that were most frequently referenced in items' titles, abstracts or keywords are all located in the Global North: America (27), Canada (23) and Japan (16). Fewer references to developing contexts could potentially

reflect television technology becoming widely available at an earlier stage in developed nations, combined with the greater number of identified literature items published prior to 1990 (see Figure 2.1). The section below features the results of qualitative investigation into key articles with a developing context focus within the identified corpus.

Exposure to international variants of *Sesame Street* have been found to promote school readiness in various developing contexts. A meta-analysis conducted by Mares and Pan (2013) concerning 24 evaluations of *Sesame Street*-based cartoons across 15 countries suggested an average effect size of 0.292. Amongst the studies included, one identified the positive effect of exposure to the Indonesian variant of Sesame Street on outcomes including literacy and mathematics (Borzekowski & Henry, 2011). This study involved a 14-week intervention amongst children aged 3-6, investigated using a randomised controlled trial (RCT) (ibid). These findings corresponded with those from another RCT in Tanzania itself (Borzekowski & Macha, 2010). Here, a six-week intervention involving exposure to the Tanzanian version of Sesame Street was delivered via television, radio and print materials. RCT results also suggested this intervention to have positive effects on numerous outcomes amongst preschool children.

While Mares and Pan's (2013) meta-analysis only concerned the effects of controlled exposure, two articles were found that investigated the influence of normal exposure to international Sesame Street variants amongst pre-primary children. These pieces of research were written by Lee (2009) and Rimal, Figueroa & Storey (2013). Lee (2009) used a measure of exposure intensity (from mother's reports of child viewing frequency) to identify that watching the Bangladeshi adaptation of Sesame Street led to significant improvements in attainment over time. Similarly, cross-sectional regression conducted by Rimal, Figueroa and Storey (2013) showed that exposure to a localised version of Sesame Street in Egypt had a positive effect on child outcomes (in numeracy, literacy and tests on gender-equitable attitudes). It could be noted that these evaluations both possessed potential flaws. For example, Rimal, Figueroa and Storey (2013) measured exposure through a character recognition method (similar to that employed in this study, Section 3.3.1.3.2), yet did not control for recognition of non-Sesame Street characters (thereby allowing for the possibility that media exposure in general produced learning gains). Additionally, Lee's (2009) results were based on relatively coarse measure of viewership (parental reports of child viewership frequency).[20]

Beyond their possible methodological shortcomings, the studies by Lee (2009) and Rimal, Figueroa and Storey (2013) both concerned *pre-primary* children. As such, it remains the case that no developing nations-based investigation into the effects of normal exposure amongst *primary age* children has been conducted. Further, it was reported (by Borzekowski, 2018) that there has been very

---

[20] This said, results were reportedly comparable when a character recognition-based exposure measure was instead employed (Lee, 2009, pp. 26-27).

little literature on the influence of (normal or controlled) exposure to non-Sesame Street television created or employed in developing contexts. With regards to primary school-aged children specifically, this structured review found little evidence to refute Fisch's (2004, p. 71) identification that there has been no "substantive research… conducted outside the United States" on mathematics-based programmes for school-age audiences.

One possible exception, however, was provided by Lapinid and colleagues (2017). Lapinid and colleagues (2017) employed a large-scale cross-section design to examine the effects of educational television amongst 32,768 children in the Philippines. These children attended 516 schools and were in Grades 3 to 6 or the first or second year of high school at the time of testing. 13,095 of the surveyed children attended treatment group schools that received the Knowledge Channel over a three-year period, while the remaining children formed the control group with no Knowledge Channel access. The Knowledge Channel is a Philippines-based channel showing only curriculum-based educational content. Students in treatment schools performed significantly better than their counterparts in control schools in a test covering five subject areas including mathematics.

However, the methodology employed by Lapinid and colleagues (2017) is susceptible to criticism. Differences in achievement between groups were considered through an independent samples t-test only. This method failed to control for any potential differences between children or schools, which were not properly considered by the researchers. There was no indication in the research paper that treatment schools were selected randomly, nor consideration of whether students in control schools (selected in accordance with their proximity to treatment schools and advice from a national authority) provided a sufficient counterfactual. The use of t-tests to identify a relationship between educational television and mathematics performance provided no means of controlling for the observable differences that likely existed between treatment and control groups.

A small number of other articles have been identified concerning the impact of educational television on attainment in non-mathematics subjects amongst primary-age children. Talabi (1989), for example, compared the effects of televised and programmed instruction lasting 40 minutes on geography attainment amongst 600 secondary school students in urban and rural Nigeria. Results from a pre-test-post-test control group design suggested that urban children receiving televised instruction achieved greater attainment gains, but that results were mixed amongst rural children (where only "high achievers gained more from televised instruction", Talabi, 1989, p. 20). Further, limited information on the literacy programme *Al Manaahil*, which aired in Jordan, Tunisia and Morocco in the late 1980s, has been identified. According to Palmer (1993, originally cited in Fisch, 2004), evaluation data suggested that this programme had a positive effect on literacy skills in Modern Standard

Arabic.[21] This review now progresses to consider research that specifically concerns shows produced by Ubongo.

2.2.4.1 Existing research concerning Ubongo

Literature searching unveiled the two existing published research articles on Ubongo programmes. The first of these studies involved a randomised controlled trial amongst 568 children (3-6 years) beginning pre-primary school in Morogoro, Tanzania (Borzekowski, 2018). This sample was divided into two groups, with a treatment group assigned to watch 30 minutes of Ubongo's Akili and Me programme for five days a week over a four-week period. A control group spent the same time viewing non-educational child-directed programmes. All viewing was conducted in designated rooms in the children's pre-primary schools, with child attendance recorded by external monitors. Attainment was measured through an adapted version of the International Development and Early Learning Assessment (IDELA) administered before and after the intervention. Regression models controlling for age, sex and baseline attainment showed that "Exposure to the treatment program significantly improved children's scores, compared to those in the control group, for five of the seven outcomes (drawing skills, shape knowledge, number recognition, counting, and English skills)" (Borzekowski, 2018, p. 57).

These findings corresponded with further research on Akili and Me in Rwanda amongst a slightly older sample (6-8 years) (Borzekowski et al., 2019). Here, the intervention led to significant gains in eight of ten measured competencies (which were again derived from an adaptation of the IDELA test). The research design was again an RCT conducted over a short-time period (two weeks). Children in a treatment group were also shown Akili and Me material in a controlled environment for five days per week, while control group children watched entertainment-based cartoons. This study differed, however, in its measure of exposure. In this instance, a character recognition measure was employed as the treatment variable (with a general media receptivity measure provided by recognition results of other media figures employed as a control) (see Section 3.3.1.3.2 for further discussion).

Additional non-published summative research concerning Ubongo has been conducted.[22] This was again a RCT, but conducted by IPSOS Synovate Tanzania in 2014 over a period of six months. This study involved children from Standard 3 and 5 (in which children are largely 9-11 years of age) in four schools in Arusha, Tanzania. Children were randomly divided into two groups, with one group

---

[21] Here, however, "Palmer did not provide a detailed account of the results of the study, which makes it impossible to evaluate the data in any depth" (Fisch, 2004, p. 64). While the title of numerous other articles appeared relevant to the effect of television on attainment amongst school age children, investigation frequently focused on questions of awareness of, access to, perceptions of, or satisfaction with education media (see, for example, Olumorin, Aderoju & Ononjah, 2018; and, Kim, 2015).

[22] Neither summative nor formative non-published research was reflected in quantitative analyses of the (published) educational television corpus.

spending 60 minutes watching *Ubongo Kids* episodes once each week, and the other spending the same time watching non-educational cartoons. Significant gains were identified for the treatment group on episode-related tests administered immediately after each episode. However, findings from baseline and endline testing of mathematics, literacy and Kiswahili were inconsistent (where pupils' scores in both groups decreased, with this potentially attributable to variation in the timing of tests within the school day).

Lastly, Ubongo also conducts regular formative research, including pre-production testing amongst educators, parents and, primarily, children. Formative research at the pre-production stage with child volunteers (whose parents often work at Ubongo) consists of numerous components, namely: child focus groups and pre- and post-test result analysis at early stages of script development; observation of viewer actions when played content alongside a "distraction" (such as a muted alternative children's TV show or image slideshow); and, A/B testing of differing versions of content. Internal research at the pre-production phase exists alongside other methods of formative data capture post-production, largely relating to programme reach. The following section now considers studies on cost analysis, including all available research that concerns the influence of educational television against its cost (Section 2.3.1.3).

2.3   Cost analysis

In contrast with literature searching related to educational television and learning (the primary research area), the identification of cost analysis material (the subsidiary research area) was conducted using a non-structured approach. Relevant literature was identified through searches using cost analysis-focused keywords on various search platforms, supported by following up key references within identified sources. Literature uncovered through this search process was then assessed critically. This section provides information on this literature in separate sub-sections. These are entitled, 'Cost-effectiveness analysis' (CEA) (Section 2.3.1), 'The limitations of cost-effectiveness analysis' (Section 2.3.1.1), 'Cost-effectiveness analysis in developing contexts' (Section 2.3.1.2) and 'Cost-effectiveness analysis concerning educational television' (Section 2.3.1.3). This information is preceded, however, by consideration of the broader concept of cost analysis.

Cost analysis provides a means of promoting the efficient allocation of educational resources. Without application of this concept, it is possible that "simply allocating more money to education will not necessarily result in increased student achievement" (nor indeed "the reduction in pressing inequities and inefficiencies in the delivery of educational service": Hummel-Rossi & Ashdown, 2002, P. 1). The allocation of resources might instead be the product of "guesswork or politics" (Levin & McEwan, 2001, p. 2). Such decision making would clearly be detrimental to Tanzania's education sector. Despite education expenditure being just under $1.15 billion in 2014 (17.26 percent of total

government expenditure),[23] mathematics attainment remains at concerning levels (see Section 1.2.1). Should cost analysis facilitate an overall efficiency improvement of 5 percent, almost $57.5 million could be available for other purposes. This clearly suggests the application of cost analysis to educational interventions to be worthwhile. The anticipated low cost per person of educational television in developing contexts, attributable to their broad reach, makes cost analysis research a particularly interesting possibility.

Initial literature searching focused on both CEA and cost-benefit analysis (CBA). Each term describes "an approach that can compare the costs and outcomes of educational alternatives" (Levin, 1988, p. 52). However, CEA "refers to the evaluation of alternatives according to both their costs and their effects with regard to producing some [corresponding] outcome" (Levin & McEwan, 2001, p. 10), while CBA involves a "direct comparison of the monetary cost and benefit of a particular intervention where cost and benefit consequences can both be measured in monetary terms" (Levin & Belfield, 2015, p. 402). CBA can therefore "[serve] as an overall guide for making public investments" in and of itself (Levin, 1988, p. 52), with a CBA calculation suggesting an intervention to be worthwhile if the measurable benefits of a programme are greater than its costs.

After preliminary investigation, however, searching ultimately focused exclusively on CEA. This choice of focus resulted from the identification that CBA calculations could only be conducted with very limited accuracy. This identification arose from examination of the identified literature and supplementary investigation into available data. Examination of these materials showed that CBA calculations would be hindered by difficulties in estimating the future effect of a change in test scores on later earning (in addition to other limitations of relevance recognised with regards to CEA, see Section 2.3.1.1).

These difficulties are derived in part from (a lack of) data availability for Tanzania. In British studies, the existence of a longitudinal cohort study of one group from their birth in April 1970 until April 2000 (and beyond) has permitted research into school-based interventions to incorporate a measure of how changes in pupil attainment might affect later earnings (at age 30, with this measure employed to conduct CBA by Machin & McNally, 2008). However, as there is no such dataset for Tanzania, CBA calculations would have to be calculated in a less precise manner. First, learning gains would have to be translated into a coarse measure of additional years of schooling. Next, this additional school years estimate would need to be employed in conjunction with a separate rate of returns to education (RORE) study for Tanzania (e.g., Kerr & Quinn, 2010) to produce a monetary value (with this part of the possible CBA calculation conducted previously by Miguel & Kremer, 2004, and Schultz, 2004,

---

[23] Education expenditure was calculated using figures for total government expenditure and educational spending as a proportion of government spending, obtained from The Global Economy (http://www.theglobaleconomy.com/Tanzania/).

regarding Kenya and Mexico, respectively). The following points highlight the limitations of this possible RORE approach:

- Firstly, "there is no way of knowing what anyone's future earnings will be": estimates could reflect only the current earnings of appropriate individuals in relation to their level of education (Perkins, Radelet, Lindauer & Block, 2013, p. 272).
- Secondly, low- and middle-income countries (including Tanzania) possess workforces encompassing numerous people with informal and grey market incomes that are not reported (including, for example, subsistence farmers). These could not therefore be included in RORE estimations.
- Thirdly, RORE studies are hindered by problems of causality: it is possible that "people who are innately more talented both earn more and receive more schooling" (*ibid*, p. 273).

2.3.1 Cost-effectiveness analysis

CEA permits determination of "the least cost approach to meeting such educational objectives as… raising [attainment]" (Levin, 1988, p. 52), or identifying the relative gains in attainment achieved by differing interventions for a given cost (which is the primary format employed in this study). CEA results concerning educational television in developing contexts could therefore be compared with studies in similar contexts that pursue similar objectives (Levin & McEwan, 2001). This would permit the estimation of whether educational television is a more cost-effective means of raising learner outcomes in developing contexts than, for example, cash transfers conditional on school attendance (Baird, McIntosh and Ozler, 2011), or results-dependent incentives for teachers (Glewwe, Ilias & Kremer, 2010).

2.3.1.1 *The limitations of cost-effectiveness analysis*

The extent to which the objectives of differing programmes are similar is a key issue. While this thesis ultimately concerns mathematics learning, "interventions might be expected to influence achievement in a number of subject areas" (Levin, 1988, p. 56). This could act to limit comparability between results from this investigation with, for example, an evaluation that employed a test in English (as in Baird, McIntosh and Ozler, 2011). Comparison with an evaluation employing one multi-subject test would also be of limited benefit: without the possibility of separating out mathematics outcomes, this would not achieve an accurate comparison. Levin and McEwan (2001) even suggest that reading and mathematics should be considered separate (non-comparable) goals, although CEA comparisons between studies with differing attainment objectives are drawn by the Abdul Latif Jameel Poverty Action Lab (J-PAL, 2014).

Additionally, literature searching suggested practical difficulties in estimating programme costs (Levin & McEwan, 2001). Such issues were later found to be of direct relevance to empirical research concerning educational television. These included challenges with identifying the Tanzania-specific costs of a show that has been broadcast in multiple countries. To create Tanzania-specific estimations, calculations had to account for the proportion of Ubongo Kids viewers that were located in Tanzania (Section 8.5). Additionally, assumptions had to be made regarding the opportunity cost of viewing Ubongo Kids. While opportunity cost was likely a non-zero amount, qualitative information suggested that assuming this to be naught was acceptable. That is, interviewees frequently reported that Ubongo Kids was watched at the expense of playing games or viewing different shows (as opposed to school or monetary work, Section 8.5).

Further, existing research highlighted that no form of cost analysis (or indeed any form of evaluation) lends "[itself] to areas where change may occur rapidly" (Levin, 1988, p. 56). This point is important to this thesis, as cost analysis concerning an educational show might provide limited information about the intervention if, for example, a change in technology were to alter the manner in which programme material was transmitted to beneficiaries. Indeed, such change might have already occurred: while the Ubongo Kids show is conveyed primarily through television, beneficiaries of Ubongo might access educational content through mediums such as YouTube (Section 1.2.2). Changes in the price of technologies over time might also act to diminish the longevity of cost estimations used in cost analyses, with this potentially reflected in the high cost per person found in earlier investigations into educational television (see Section 2.3.1.3).

2.3.1.2   *Cost-effectiveness analysis in developing contexts*

There is a growing but modest number of CEA studies concerning education in developing contexts (Levin & McEwan, 2001). This remains particularly limited in relation to the number of CEA-centred papers focusing on health interventions (see, for example, Hogan, Baltussen, Hayashi, Lauer & Salomon, 2005; Medley, Kennedy, O'Reilly & Sweat, 2009; and, Edejer, Aikins, Black, Wolfson, Hutubessy & Evans, 2005). Evaluative educational studies rarely include a CEA component (Levin & Belfield, 2015), or even provide sufficient information on costs to permit subsequent CEA meta-analyses (as recognised by Dhaliwal, Duflo, Glennerster & Tulloch, 2012). Indeed, when cost information is provided, work to produce CEA-focused meta-analyses is also hindered by the presentation of costs in differing formats (which are also susceptible to inflation, *ibid*). Thus, whilst individual studies have provided comparisons of different educational interventions delivered by the same organisation (Kremer, Miguel & Thornton, 2009; and, Banerjee, Cole, Duflo & Linden, 2007), these findings would still require reconfiguration to be comparable with interventions covered in different studies (which might be conducted in various countries and at different points in time).

As such, this review shall only provide cost information from identified studies calculated using the learner outcomes-based method employed by J-PAL (2014, with cost analyses of some identified studies available in an Excel format, via https://www.povertyactionlab.org/research-resources/cost-effectiveness).[24] This J-PAL approach is undoubtedly imperfect. A ranking of studies in accordance with CEA estimates created using this J-PAL (2014) approach would vary widely if considering the 90 percent confidence intervals around estimates (Evans & Popova, 2016). Further criticism could be ascribed to the J-PAL approach because it takes no account of the fact that some interventions might "pursue multiple objectives" (some of which are not accounted for under the J-PAL approach: McEwan, 2012, p. 192). However, considering only studies that use this J-PAL method promotes compatibility between relevant findings. Additionally, focusing on CEA studies conducted using this approach ensures that all findings concern interventions that possess key similarities. Each intervention was implemented in a developing country context in the relatively recent past (with identified studies being published only after 2000).

A number of cost-effectiveness calculations attempted using the J-PAL (2014) approach focused on interventions that had zero effect on learner outcomes. These included a deworming treatment phased randomly into 75 schools in rural Kenya (Miguel & Kremer, 2004). With regards to educational attainment, this study did "not find any evidence that deworming increased academic test scores" (Miguel & Kremer, 2004, p. 160), which suggested that this intervention was not a cost-effective means of addressing this outcome. This study, however, did identify positive effects on attendance and health (both of which were considered in separate cost-effectiveness calculations). Blimpo & Evans' (2011) evaluation of school committee grants with and without a comprehensive school management-training programme also found no impact on attainment, despite reductions in student and teacher absenteeism of 21 and 23 percent respectively for the grant plus management-training programme (with no reductions in absenteeism achieved through grants alone).

Other interventions for which cost information was available with no significant effect upon learner outcomes included the provision of flipcharts in Kenya (Glewwe, Kremer & Moulin, 2009) – a result which subsequently influenced Kremer, Brannen & Glennerster (2013, p. 299) to contend that the "[e]vidence on nonteacher inputs is… discouraging". Flipcharts were again considered to produce an effect on test scores that "appears to be essentially zero" in a prospective randomised evaluation by Glewwe, Kremer, Moulin & Zitzewitz (2004, p. 263, in contrast with prior findings from retrospective studies on the same subject). The facet of Baird, McIntosh and Ozler's (2011) investigation into cash transfers in Malawi concerned specifically with *unconditional* cash transfers also found no significant effect on educational attainment. In relation to these findings, identifying educational television to be associated with any positive effect on outcomes would suggest it to be a more cost-effective initiative

---

[24] This therefore excludes studies that provide CEA information with regards to objectives such as school absenteeism (Evans, Kremer & Ngatia, 2008).

(regardless of its cost per beneficiary). However, the fact that many of these interventions were deemed to have had positive effects on outcomes ranging from child health to teacher absenteeism acts to highlight the limitations of a CEA result that only accounts for interventions' effects on learner outcomes.

A number of other forms of intervention were, however, identified to have a positive impact. These included a different component of Baird, McIntosh and Ozler's (2011) study, in which test scores were found to increase amongst students for whom cash transfers were provided *conditionally* on school attendance. Whilst this study reported positive impacts on one (of two) mathematics tests, one 'cognitive' test and one English test, the only CEA estimate produced concerns the latter (J-PAL, 2014). This gives the lowest CEA figure identified: a gain of 0.06 standard deviations (SD) per $100 spent (at 2011 USD – the USD year at which all J-PAL style CEA estimates are presented). The highest SD per $100 cost-effectiveness estimate found came from Nguyen's (2008) investigation into the provision of statistics (without a 'role model') on the returns to education to families in Madagascar. Concerning this 'statistics only' treatment group, subsequent J-PAL (2014) estimates suggested an additional 118.34 SD improvement in test scores (which measured "children's competency in three materials: mathematics, French, and Malagasy": Nguyen, 2008, p. 1) per $100 spent.

Further cost-effectiveness information from studies that identified a positive impact included a randomised trial concerning results-based incentives for Kenyan teachers (Glewwe, Ilias & Kremer, 2010). This study identified an increase in student test scores (in tests linked to incentives, but not in others) after 2 years of the intervention, which was subsequently used by J-PAL (2014) to create a CEA estimate of 6.29 SD gains per $100 spent. The importance of this result might, however, be limited as no significant improvement in (incentive-related) test scores was found 1 year after the intervention had concluded. Conversely, a clearer impact was identified by Kremer, Miguel & Thornton (2009) in their examination into the effect of providing two-year scholarships to girls who achieved a score in the top 15 percent of selected schools in Busia and Teso districts, Kenya. Results concerning the intention to treat (ITT) sample suggested that the programme improved girls' test scores by 0.19 SD across both targeted districts. Using this estimation of impact, the scholarship programme was subsequently identified to produce a 1.38 SD increase in academic attainment for each $100 spent (J-PAL, 2014) (with the same dollar figure also achieving 0.27 additional years of student participation: Ganimian & Murnane, 2016).

Other positive results CEA results identified included interventions in Kenya (Kremer, Duflo & Dupas, 2011; and, Glewwe, Kremer & Moulin, 2009),[25] Afghanistan (Burde & Linden, 2013), the

---

[25] This reference to Glewwe, Kremer and Moulin (2009) concerns a separate part of the study previously discussed, that found flipcharts in Kenya to produce no significant impact. This study conversely identified that

Philippines (Abeberese, Kumler, & Linden, 2014), India (Banerjee, Cole, Duflo & Linden, 2007; and, Duflo, Hanna & Ryan, 2012) and Indonesia (Pradhan, Suryadarma, Beatty, Wong, Alisjahbana, Gaduh & Artha, 2011). As with previous CEA results considered, the interventions researched differed in type. The forms of intervention ranged from "tracking students by prior achievement (to place them in sets)" (Kremer, Duflo & Dupas, 2011, p. 1740), to a computer-assisted learning programme with a mathematics focus (one of the interventions considered by Banerjee, Cole, Duflo & Linden, 2007). CEA results in the format of additional SD per $100 spent for each of these interventions fell between those previously identified for Baird, McIntosh and Ozler (2011) and Nguyen (2008).[26]

Lastly, this review identified two recent CEA-focused studies that used the J-PAL methodology but were not included in J-PAL's (2014) comparison of CEA results. One of these papers is a policy brief created by the Cambridge REAL Centre that provided cost-effectiveness results for the Camfed programme in Tanzania (Sabates, Rose, Deprato & Alcott, 2018). The Camfed programme "targets a range of barriers to girls' secondary education at an age when they are at a great risk of dropping out due to factors such as poverty, early marriage and teenage pregnancy" in addition to other educational support for all children in Camfed supported schools (ibid, p. 5). This programme seeks to increase children's learning outcomes and chances of staying in schools. Through employment of a tailored version of J-PAL's cost-effectiveness method, the additional SD gain in English outcomes per $100 spent was estimated at 1.05.[27] The remaining CEA paper concerned a partnership schools programme in Liberia (Romero, Sandefur & Sandholtz, 2017). In this programme, Liberian public schools were managed by various private contractors. While these schools typically outperformed public schools (during the one-year evaluation), costs were high. As such, a J-PAL (2014) CEA calculation based on "the lowest possible cost associated with the program" found the programme to achieve a relatively small additional SD per $100 gain of 0.38 (ibid, p. 49).

To summarise, positive CEA findings in the format of SD gained per $100 spent varied widely. The lowest SD gain per $100 was 0.06 and the highest was 118.34. Given their relative comparability and compatibility, CEA results for all identified studies provide information that could aid the contextualisation of findings concerning educational television in developing contexts (if established using the same method: Section 8.7.1). To further this investigation into CEA, this review shall now consider prior studies concerning educational cartoons or educational television.

---

the provision of textbooks to students in the top quintile had a positive and significant effect on test scores (Glewwe, Kremer & Moulin, 2009), which was subsequently identified to produce an impact of 3.56 standard deviations (SD) per $100 spent (J-PAL, 2014).

[26] Figures for all CEA estimates identified are presented graphically in Section 9.7.1.

[27] As expanded on in Section 9.2, this result could potentially have been biased upwards by the fact that this CEA calculation accounted for reductions in school dropout attributable to Camfed, yet excluded the money spent on school fees from programme costs.

2.3.1.3   Cost-effectiveness analysis concerning educational television

Literature concerning the cost effectiveness of educational television is sparse. Mares & Pan (2013, p. 148) found "no data for the cost-efficiency of creating and broadcasting *Sesame Street* co-productions [– the most prominent international educational cartoons –] relative to other early education interventions". The limited availability of relevant studies within this area in more recent years is underlined by the lack of any evaluation concerning educational television intervention in J-PAL's (2014) cost-effectiveness comparison. Whilst the identification of a worthwhile CEA benchmark from this sector appears highly challenging, this review has found some research concerning the cost of educational television. This is, however, largely school-based, dated, or provides insufficient details for comparison with the present study.

In spite of the anticipated low costs per person for educational television (attributable to its broad reach), the costs identified for earlier television-based interventions were generally high (with this greatly influenced by their smaller scale). These included per child per year estimates of $297.87 for an initiative in Hagersown, USA, and, $133.31 for a Korea-based intervention (reported in Jamison, Klees & Wells, 1978).[28] Highly tentative estimates of Brazil's Telecurso programme for young pre-graduation school leavers seeking to take primary and secondary exams suggested a cost of $17.87 per student (including one corresponding course book costing $5.10) (Wolff, de Moura Castro, Navarro & Garcia, 2002).[29] The longstanding Telesecundaria in-class project in Mexico, on which multiple annual estimates have been identified, had per pupil costs of $892.64 (in 1975), $1,305.73 (in 1981), $621.17 (in 1988) and $793.01 (in 1997) (Perraton, 2005). Perhaps more indicative of the current costs of educational television is an out of school intervention featuring community screenings of *Galli Galli Sim Sim*, an Indian variation of *Sesame Street* (BGM Policy Innovations, 2011, originally referenced in Batada, Banergee & Subramanian, 2016). The paper cited by Batada and colleagues is, however, not accessible through online databases. Thus, whilst costs are stated to be $0.11 per person per screening, insufficient information on educational effect precludes the inclusion of a CEA estimate from this study in the review.

Indeed, there are very few studies in which an actual CEA result was provided (even prior to 2000). The identification of high costs per pupil in earlier interventions had merely been used to draw the general conclusion that "educational television is … on a cost-effectiveness basis, an inefficient way

---

[28] These cost data and all subsequent cost data in this section are translated to 2011 US$ using a CPI information calculator (Bureau of Labor Statistics: http://www.bls.gov/data/inflation_calculator.htm), to promote correspondence with the above findings concerning CEA in sub-Saharan Africa.

[29] In this instance, costs per pupil are not specifically stated as being of any particular year. As such, inflation calculations are conducted assuming that the estimate provided by Wolff, de Moura Castro, Navarro & Garcia (2002) is in US$ at the year of publication. It should be re-emphasised that Wolff and colleagues' (2002, p. 152) cost estimate appears only a rough approximation, having been derived from an assumed cost production (as opposed to an ingredients-based calculation) and an estimate of 3 million users (derived only from figures for book sales, which "indicate[d] that several million students participated in Telecurso somewhat seriously").

to improve the performance of pupils in school" (Carnoy, 1975, p. 247; Klees, 1995). Additionally, some literature concerning educational television that purported to provide cost-effectiveness information did not provide sufficient data on effect (in relation to a control group) to compute effect per amount spent (see, for example, Wolff, de Moura Castro, Navarro and Garcia's [2002, pp. 146-148] discussion of the cost effectiveness of the Telesecundaria programme; and, Batada, Banergee & Subramanian's [2016] discussion of the *Sesame Street* co-production, *Galli Galli Sim Sim*). The only research identified that came closer to providing a complete cost-effectiveness analysis was Carnoy's (1975) investigation into an in-class educational television intervention in El Salvador.

Here, gains over 'reformed classes' (without television) were amalgamated with cost information to produce cost ratios that suggested superior results for educational television (in mathematics and science, but not social studies) (Carnoy, 1975). Even this, however, if converted to a format compatible with the work of J-PAL (2014), would not provide a suitable benchmark for comparison when considering the cost effectiveness of educational television in developing countries. Aside from differences in the intervention's nature (being in-class) and time (concerning attainment data from 1969-1971 and with technology costs from this period unlikely to be reflective of circumstances today), the cost-effectiveness ratio only incorporated evaluation data from the first year of the intervention amongst Grade 7 pupils (when the intervention included pupils from Grades 7-9). Additionally, findings were not produced using a true control group – only reformed classes without television.

2.4   Summary and research gaps

This review has uncovered a large body of literature relating to learning from educational television (Figure 2.1). These publications were identified using a rigorous methodology involving both structured and unstructured searching on multiple platforms. This research suggested the effect of educational television on learning outcomes to be limited amongst younger preschool children (Anderson & Subrahmanyam, 2017). However, studies frequently reported positive effects amongst viewers aged 30 months upwards. Such effects were identified within all adopted age-brackets: pre-primary (Linebarger, McMenamin & Wainwright, 2008); school-age (Webb, 1982); and, university or adult learners (Boulet, Boudreault and Guérette, 1998). Identified studies also suggested that positive effects might be enhanced by factors such as co-viewing (Lavigne & Anderson, 2012) and exposure to educational materials across multiple media platforms (Fisch, Lesh, Motoki, Crespo & Melfi, 2010). Further, the identified literature underlined that educational television could have a positive effect on outcomes beyond attainment or capability (Murrar and Brauer, 2018).

While numerous topics have been addressed in the literature, various research gaps remain. Amongst the most pressing is that there exists no thorough research outside of America concerning the

relationship between educational television intended for school-age audiences and mathematics attainment (Fisch, 2004). One study was identified which purportedly considered the influence of television on mathematics outcomes (in addition to other topics) (Lapinid et al., 2017). However, this study was considered to lack methodological rigour. This is a result of Lapinid and colleagues' (2017) failure to detail measures through which possible differences between control and treatment groups were controlled for (see Section 2.2.4). Further, although various short-term experiments concerning educational cartoons have been conducted in developing contexts (despite a relative lack of focus on developing nations, Figure 2.6), rigorous investigation into longer-term and naturalistic effects amongst primary school children has been lacking.

With regards to cost-effectiveness information, literature searching uncovered CEA findings for non-television interventions in comparable contexts. The presented estimates were each calculated in a manner consistent with J-PAL's approach. Positive CEA estimates ranged from 0.06 to 118.34 additional SD in learning outcomes per $100 spent. However, no cost-effectiveness information following the J-PAL format was identified for any study relating to educational television.

Indeed, cost information in any format concerning educational television was sparse and generally dated. Only one study post-2010 that included cost information has been referenced (BGM Policy Innovations, 2011), although this study could not be accessed online. Older studies provided high cost-per-person estimates for television-based programmes, ranging from $17.87 to $1305.73 (reported in, Wolff, de Moura Castro, Navarro & Garcia, 2002; and, Perraton, 2000, respectively). It is considered that these per-person estimates are not reflective of current costs due to increases in technology availability. Ubongo Kids, for example, is currently watched frequently by 16.68 percent of children (aged 6-16) (estimated from the Uwezo 2017 dataset, see Section 10.4). It is hoped that addressing identified research gaps concerning both cost and effectiveness will be informative for those concerned with educational television targeted at primary-age children (including Ubongo) and policymakers in developing contexts seeking to promote an efficient allocation of educational resources.

2.5   Research Questions

In accordance with this literature review (including its identification of research gaps), the primary research area was explored using the following question:

- What is the association between naturally occurring exposure to educational television and mathematics capability in a developing country context?

In initially devising this question, it was hypothesised that naturally occurring exposure (that is, normal viewing in day-to-day settings) to educational television would be positively associated with

mathematics capability. It was envisaged that this hypothesis would be supported by cross-section models showing that children with greater exposure to televised educational materials possess higher levels of mathematics capability. Further, it was anticipated that difference-in-difference analysis would suggest that children in districts with more exposure to educational television show improvement in mathematics capability relative to children in districts with less exposure.

Additionally, the subsidiary research area was examined using the following question:

-   How cost-effective is educational television in a developing country context?

It was hypothesised that the anticipated low production cost per child of educational television would lead to a strong cost-effectiveness finding. Further, it was expected that cost-effectiveness comparisons with alternative interventions in comparable contexts would largely suggest educational television to be a superior investment.

2.5.1   Consideration of potential factors influencing the development of the research questions

This section acknowledges factors beyond the literature, my theoretical standpoint or philosophical position that have potentially influenced the development of the research questions. Considering the possible presence of additional influences is important, as these might act to undermine the research independence required in PhD research. Amongst possible influences, it is first acknowledged that my primary research question could have been affected by the perspectives of Ubongo stakeholders. This question is complicit with Ubongo's objectives of promoting a 'better understand[ing of the] fundamentals of mathematics', and 'academic performance' (at the outcomes and impact level respectively of Ubongo's framework theory of change, or TOC, Section 1.2.2). Moreover, Ubongo's objectives could themselves be influenced by international donors to the Ubongo programme (with such parties often considered to be predisposed to research featuring standardised attainment measures: Schweisfurth, 2015). Additionally, my subsidiary research question addresses cost analysis, which is similarly associated with international actors (as recognised by, Tikly & Bond, 2013). It could, therefore, be alleged that research question content is related to previous positions of employment within international NGOs or even my current research relationship with Ubongo.

These arguments might be bolstered by the identification that international actors have focused simultaneously on the potentially conflicting concepts of attainment and learner-centred education (Sriprakash, 2012; Schweisfurth, 2015). This thesis could certainly be judged to have done the same, as it examines research questions using standardised outcome measures while identifying the educational content of Ubongo Kids programmes to be learner-centred (see above, 'Pedagogic approach'). Additionally, this approach is maintained despite acknowledgement that an outcomes-

based definition of intervention success could also lead to the failure to acknowledge other benefits provided by exposure to educational television. This point is pertinent, as Ubongo's framework TOC includes goals such as the promotion of positive attitudes towards learning (Section 1.2.2). It has also been recognised that the effect of educational television extends beyond learning outcomes (see Section 2.2.3).

It is not, however, believed that the influence of any international actor, including Ubongo, acts to undermine this thesis. This is supported by the fact that I have no current financial arrangement with, or vested interest in, Ubongo itself. Instead, my research questions have been formulated as an independent researcher in accordance with the identified literature on educational television. These questions also reflect my post-positivist perspective (that an absolute truth should be sought so far as possible) and adoption of human capital theory (through which education is perceived as a financial investment) (see Section 1.3-4).

Further, this study is conducted on the understanding that research in developing countries must prioritise the interests of beneficiaries (Tikly & Bond, 2013). It is contended that the provision of information centred upon standardised attainment measures helps evaluations to be of value to programme recipients, by ensuring some degree of objectivity. In this instance, those involved in educational television will be provided with such information, which could accurately inform programme development to the benefit of beneficiaries.[30] Conversely, prior experience of educational evaluations has suggested that deviation from recognised standardised assessments might produce biased findings. This is a result of alternative outcome measures highlighting positive facets of an intervention (intentionally or otherwise), especially if developed by creators of an intervention. In addition to this, investigating questions focused on cost effectiveness could inform policymakers in a manner that ultimately assists learners. The provision of cost analysis information should support efforts to maximise the extent to which scarce funds are employed to improve educational outcomes. This report now presents the methodology through which the research questions are addressed.

---

[30] Indeed, the aforementioned predisposition of international policymakers to standardised attainment measures could increase the likelihood that the outcomes of this study are acted upon.

3 Chapter 3 – Methods

3.1 Introduction to methods

This chapter details the methods used to examine the primary and subsidiary research questions, which focus respectively on the association between educational television and mathematics capability and the cost-effectiveness of educational television in a developing country context. Research questions were addressed using various forms of analysis applied to multiple data sources. These data sources were both primary and secondary in nature. That is, data were collected first-hand for the purposes of this study (primary) and obtained from existing datasets collected by a third party (secondary).

The primary data used in this study were both quantitative and qualitative. Primary data sources included a survey administered to a large number of children in two Dar es Salaam districts (quantitative) and interviews conducted with a small number of children and their caregivers (qualitative). The secondary data employed in this study, however, were solely quantitative and obtained from Uwezo (https://www.twaweza.org/go/uwezo-datasets).[31] Uwezo is an organisation that conducts household assessments in East Africa that capture child learning outcomes and household demographic information (see Uwezo: http://www.uwezo.net/about-us/). All key data sources used for investigation into both research questions are presented in Table 3.1, before further explanation in Section 3.3.1.1 (regarding the primary research question), Section 3.3.2 (regarding the secondary research question) and Section 3.4 (regarding qualitative information).

---

[31] Uwezo Tanzania data for 2017 is currently pending public release. A draft Tanzania 2017 dataset has been provided for use in this study by Uwezo following personal communication (as opposed to the dataset being downloaded from the Uwezo Data on Learning webpage: https://www.twaweza.org/go/uwezo-datasets).

Table 3. 1: Key data sources

Data	Years which data concern	Geographical region to which data are relevant	Section in which data are directly employed
Primary quantitative data	2017	Dar es Salaam	- Chapter 5 – Matched set findings
Uwezo Tanzania data (2017)	2017	Tanzania	- Chapter 5 – Matched set findings - Chapter 6 – Household set findings - Chapter 7 – Panel data findings - Chapter 9 – Cost-effectiveness analysis
Uwezo Tanzania data (2011-2015)	2011, 2013 and 2015	Tanzania	- Chapter 7 – Panel data findings
Primary qualitative data	2017	Dar es Salaam	- Chapter 8 – Qualitative findings
Ubongo Kids cost data	2014-2017	Tanzania	- Chapter 9 – Cost-effectiveness analysis

This study might be considered as mixed methods research, as it gathered "both quantitative (closed-ended) and qualitative (open-ended) data" (Creswell, 2015, p. 2). These quantitative and qualitative data were integrated to enable a "[thorough investigation of] phenomena of interest" (Teddlie & Tashakkori, 2012, p. 776). This integration involved using interview findings to aid the interpretation of a) the output of statistical models and b) variables within these statistical models (Section 3.4). Qualitative data even influenced the selection of numeric estimates used in cost analysis calculations (Section 8.5). Further, analysis drew on the "combined strengths of [qualitative and quantitative] data to understand research problems" (another characteristic of mixed methods research: Creswell, 2015, p. 2). For example, interviews benefited examination of the primary research question by providing more detailed information on Ubongo Kids episode recall – an indicator of exposure that could not be obtained using quantitative approaches (Section 6.3.2.3).

The mixed methods approach used could most accurately be categorised as an explanatory sequential design (amongst the overarching approaches to mixed methods recognised by Creswell and Plano Clark, 2007). This is because (primary) qualitative data played a relatively minor role in the research, being used chiefly to explain and inform the understanding of quantitative results.[32] It might be noted, however, that elements of the research design conflict with certain features of sequential explanatory approaches. One such feature of this research which conflicts with those typical of sequential

---

[32] A sequential explanatory classification is considered most appropriate, despite elements of the research design conflicting with certain features of this mixed methods approach. One such feature of a sequential explanatory design is that the qualitative phase follows the quantitative phase and may be informed by it. However, delays to the Uwezo data collection meant that the qualitative data were collected before quantitative data in this study.

explanatory designs is that elements of the quantitative phase followed the qualitative phase. Qualitative data were collected before quantitative data in this study, due to delays to the Uwezo data collection. The overall timeline for data collection and analysis is presented below (Figure 3.1).

April to June 2017		
Begin fieldwork (April 2017)	Negotiate and plan for primary quantitative data collection	Negotiate and plan for Uwezo data collection

July to September 2017		
Negotiate and plan for primary quantitative data collection	Negotiate and plan for Uwezo data collection	Primary qualitative data collection

October to December 2017				
Primary qualitative data collection	Backup quantitative data collection (October)	Uwezo data collection (start December)	Primary quantitative data collection (end December)	Finish fieldwork (end December)

January to February 2018
Enter and clean primary quantitative data

March to July 2018
Transcribe and translate qualitative information

August 2018 to March 2019	
Receive Uwezo data (August)	Analyse quantitative data

April to July 2019
Analyse qualitative data

Figure 3. 1: Data collection and analysis timeline

3.2   Ethics

The employment of a mixed methods approach required many ethical considerations. These considerations differed according to the data source, due in part to their differing quantitative and qualitative, as well as primary and secondary, nature. This section details this study's ethical position regarding primary qualitative data, primary quantitative data and secondary (quantitative) data. In eliciting open (and closed) responses from children and caregivers, the qualitative component of this

research required the most significant ethical consideration. While it was recognised before fieldwork that open questioning could potentially lead to the uncovering of sensitive information (given that participant responses were unconstrained), no such information was uncovered from the interviews or during the further consideration of responses. Indeed, no sensitive information arose during any facet of data collection.

Issues of consent were paramount. Given that interviewed children were as young as eight years old, consent was sought from both the children and their caregivers. The means of seeking consent took account of the possibility that respondents were not entirely literate (Liamputtong, 2010). Therefore, children and caregivers were provided with a consent form that was read aloud by the interview assistant (Appendix 1.1). This form provided information on the study, on the lead researcher (myself) and interview assistant, on the nature of questions to be asked and on the intended length of the interview. The consent form also covered interviewee confidentiality, and gave interviewees the option to choose not to participate in a part of the interview, the interview as a whole or indeed the study itself. Further, caregivers were provided with contact details to relay any problems or questions that arose after the interview.[33] Interviews were only conducted after children and caregivers had signed the consent form (thereby signalling their consent, as well as their understanding of the purposes of the study and their right to withdraw from the research at any point).

It was intended that minor concerns that arose during this process would be resolved through discussion with prospective participants (following Thomas & O'Kane, 1998). However, such a dialogue was not needed during the data collection process, since respondents' reactions to questions of consent were generally binary. Caregivers and children who wished to participate in the interview process plainly signalled their intention to do so. Additionally, those who did not wish to engage conveyed their refusal of consent with similar clarity. In the two instances where a prospective family communicated their desire not to participate, this was purportedly a result of the time commitment required for the interview process (which, in Grezani district, would have disturbed the children's school examination preparation).

Consent was also required for quantitative primary data collection and was similarly sought alongside an acknowledgement that children and caregivers had understood the study, the nature of questioning, the approximate length of their role in the data collection, their right to anonymity and their freedom to withdraw from the study at any point. However, due to the scale of this data collection, verbal rather than written consent was sought from children and their caregivers. This said, the adoption of a verbal format did not diminish the importance of ethical requirements. The training of technicians involved in this component of data collection included a thorough discussion of the ethical approach

---

[33] A copy of the consent form including my mobile number was left with interview participants. No follow up queries were made.

and participation in role-play scenarios to ensure they provided the information required for informed consent and were able to act appropriately should consent be refused. No complications arose in this component of primary data collection. Fewer than five potential respondents refused to provide consent and all who did so exercised this right in a manner straightforward for the technician.

Quantitative secondary data were considered to pose fewer potential concerns. This is because the third party that conducted data collection, Uwezo, followed its own ethical procedures.[34] The Uwezo datasets utilised from 2011, 2013 and 2015 are freely available for non-commercial public use (Uwezo Data on Learning: https://www.twaweza.org/go/uwezo-datasets). The only conditions of dataset usage are that any subsequent publications featuring analysis of Uwezo data cite the data source and are sent to Uwezo.

Ethical considerations for the Uwezo data for 2017, however, were slightly more complex. Crucially, this dataset was not available for public use at the time of the study (pending public release as per previous years' datasets, as of 2019, October 23). It must, therefore, be recognised that this dataset was in a draft format. Additionally, the version employed in this thesis originally included private details such as child names and caregiver telephone numbers. As such, this dataset was stored only in password protected locations, was not shared with others and was stripped of any information that could lead to the identification of participants when presenting results. Further, whilst this dataset was broadly secondary, it included responses to a child-directed question that was incorporated specifically for the purposes of this PhD study. In this respect, it must be stated that the content of the question meant it was highly unlikely to elicit information more sensitive in nature than any other question asked of assessed children (for details of this question, see Section 3.3.1.3.2). It should also be recognised that respondents were, as with any other question asked by Uwezo, free to refuse to answer the question I submitted. This dataset was again employed on the assumption that all information within it (including responses to the question that I submitted) was elicited ethically.

The research process fell under multiple agreements. At the highest level, a Tanzanian research visa was granted following approval of this study's design by the Tanzanian Commission for Science and Technology (COSTECH: Appendix 1.2). The National Bureau of Statistics (NBS) also provided approval regarding the collection of my primary quantitative sample, which was collected according to their sample frame (Appendix 1.3). Additionally, Uwezo provided written agreement by email regarding their role in this study. Further, contracts were brokered between Ubongo, the Guluka

---

[34] Publicly available information on Uwezo's ethical procedures can be obtained online (Uwezo, Training: http://www.uwezo.net/assessment/training/). Ethical guidelines for Tanzania are provided in Kiswahili only. However, corresponding ethical guidelines in English language are provided for Kenya (see, for example, Kenya, Training Manual 2013, p. 39: http://www.uwezo.net/wp-content/uploads/2012/08/2013-Kenya-Training-Manual.pdf; and, Kenya, Volunteer Workbook 2013, p. 16: http://www.uwezo.net/wp-content/uploads/2012/08/2013-Kenya-Volunteer-Workbook.pdf).

Kwalala Youth Environment Group (GYEG) and the Tanzania Users and Survivors Psychiatry Organisation (TUSPO) through which teams of technicians were appointed in Ilala and Temeke (Appendix 1.4). It should be noted that this contract involved the provision of financial resources from Ubongo (and The Goodall Foundation) to GYEG and TUSPO, to fund the hiring and supervision of technicians. Despite the role of Ubongo in funding this component of the research, every effort has been made to avoid bias in this thesis.

Additionally, all decisions were made with regards to the BERA Ethical Guidelines for Educational Research (BERA, 2011) and the ethical guidelines set forth by the University of Cambridge's Faculty of Education. Compliance with the latter was inferred by the Faculty's acceptance of a completed ethical clearance form, submitted before fieldwork was conducted. Despite changes to research methods that arose after consideration of this ethics form, the approach to research adopted (detailed above for each data source) underlines that compliance with Faculty ethical requirements was maintained.

A further ethical consideration that arose after completing data collection and analysis related to the use of Google Drive as a method of sharing interview information. Both transcripts and recordings were shared as files via this platform (Section 3.4.2.3). This was not seen as problematic during the transcription and translation process, as Google Drive is recognised as General Data Protection Regulation (GDPR) compliant (in the same manner as alternate cloud-based service providers, such as Microsoft and Dropbox) and classified as suitable for sharing confidential information (by the University of Cambridge: https://help.uis.cam.ac.uk/service/storage/google-drive/google-data-security-guidance). However, it is acknowledged that certain characteristics of cloud-based data services are unsuitable for storing highly sensitive data, including the fact that cloud services "move their customers' data with high frequency from one datacentre to another so as to enable efficient use of storage space" (de Bruin & Floridi, 2017, p. 36). The existence of such characteristics suggests that other precautionary measures should be considered when sharing interview data in the future. These measures could include encrypting files prior to sharing them on Google Drive and training file recipients on how to handle sensitive data.

3.3   Quantitative methods

This section provides information on the quantitative methods used to investigate the primary and subsidiary research questions. The primary research question concerned the association between educational television and mathematics capability. Findings from this question were used in conjunction with numeric cost information to examine the subsidiary research question, which considered the cost-effectiveness of educational television. This methods section begins by considering the quantitative data used to answer the primary research question.

3.3.1 Quantitative methods employed for analysis of the primary research question

Various types of quantitative data could be used to identify the association between educational television and mathematics outcomes. In the identified literature on this topic, more rigorous designs in developing contexts have employed pre- and post-intervention test data from study participants randomised into groups receiving different forms of controlled television exposure (see, for example, Borzekowski & Macha, 2010; and, Borzekowski et al, 2019). However, such data could not be obtained for use in this study. Indeed, the employment of such data might not have been most facilitative of investigating a research question focused on *naturally occurring* exposure to television. This is because research into controlled exposure "under hypothetical and contrived conditions" might not be representative of normal engagement with an intervention (Nagin & Sampson, 2019, p. 140). Further, even if it were assumed that controlled exposure provided a suitable proxy for naturally occurring exposure, a high proportion of children in Tanzania would have most likely been exposed to Ubongo Kids at some point since it first aired in January 2014. It would therefore have been challenging to select a study sample that had zero prior exposure to Ubongo Kids.

Other forms of data, therefore, were identified to permit worthwhile investigation into the relationship between everyday educational television exposure and mathematics capability. Such an investigation required data that featured information on mathematics outcomes and exposure. To obtain data with these characteristics, compatible primary and secondary datasets were employed. Primary quantitative data were used to provide more detailed information on exposure to educational television, which could not be deduced from any available secondary data. Primary data provided additional exposure information for only a subsection of children in secondary datasets, due to budgetary and practical constraints.

Secondary data were obtained from Uwezo Tanzania datasets. This source provided the only nationally representative measure of learning outcomes for both in and out of school children in Tanzania. Uwezo outcome measures were created by experts within the University of Dar es Salaam, the Tanzanian Institute of Education and local teachers (Uwezo, 2012) to provide an accurate assessment of Tanzanian Standard 2 skills. This gave information on mathematics topics that broadly correspond with those taught by Ubongo Kids (see Appendix 2). Further, the employment of Uwezo assessments benefited the quantitative models in this study through their inclusion of various child and household-level demographic measures (many of which have been associated with learning outcomes; see Section 3.3.1.3). Crucially, the 2017 survey also featured a child-level measure of educational television viewing that was included by Uwezo for the purposes of this PhD thesis (see Section 3.3.1.3.2). This measure has supported the estimates of exposure used in all quantitative models.

ACER (2015) found the external reliability of Uwezo assessment to be strong (as noted by Elks, 2016). There were 'high levels of agreement' amongst scores assigned by Uwezo assessors (ACER, 2015, p. ix). When considered alongside an expert rater (for the purposes of ACER research), levels of agreement were similarly positive (ibid, p. x). It should be noted, however, that worse internal (WLE Person separation) reliability results were identified for the main Uwezo mathematics variable (see Section 3.3.1.3.1). Still, the ongoing suitability of this assessment is supported by the fact that it has been administered to hundreds of thousands of children in Tanzania over seven surveys from 2009 to 2017. What is more, employing Uwezo Tanzania datasets in quantitative models also promotes replicability, as data are freely available for public use (via, https://www.twaweza.org/go/uwezo-datasets).[35]

Various designs were permitted by the secondary and primary data used. To give maximum support to findings, multiple approaches were employed. The first approaches considered were cross-section designs amongst individual children. Cross-sectional designs provide a "snap shot of a population at a particular point in time" (Cohen, Manion & Morrison, 2011, p. 267). The application of such an approach permitted investigation into the relationship between exposure (measured differently in primary and secondary datasets) and a measure of mathematics in 2017. Two different cross-section designs were employed, which controlled for fixed but unobserved differences between districts (see Section 3.3.1.4.1) and households (see Section 3.3.1.4.2).

Further, it was possible to create a multilevel longitudinal model using Uwezo surveys from multiple years. This approach gave a means of providing information on the effects of treatment dosage, by comparing change over time in mathematics outcomes across districts with differing exposure levels. Such an approach could not be employed for the analysis of individual children. This is because matching individual children between Uwezo datasets was not possible (as there was insufficient child identification information). As such, it was only possible to consider change in district-level outcomes over time, using districts that were common between pre- and post-exposure time points (see Section 3.3.1.4.3). The following section provides further information on the sources of quantitative data used in cross-section and multilevel approaches.

3.3.1.1   *Data sources and collection methods*

The data ultimately employed in quantitative analysis was derived from sources that could be classified as public, pre-public and primary. The data relevant to each of these three sources are listed below:

---

[35] This excludes the 2017 dataset, which was available for use exclusively in this thesis pending its public release.

- **Public source:** Uwezo Tanzania data for 2011, 2013 and 2015.
- **Pre-public source:** Uwezo Tanzania data for 2017. This dataset was formed of four compatible datasets shared separately by Uwezo. These datasets were collated through the following steps:
    1. A one-to-one merge of datasets sent by Uwezo containing (a) information on child demographics and test results, and (b) child name data
    2. a many-to-one merge of (c) the resultant dataset from Step 1 with (d) household name data
    3. a many-to-one merge of (e) the resultant dataset from Step 2 with (f) district level demographic data
    4. removing clear errors in the resultant dataset from Step 3, by dropping three rows containing non-sampled enumeration areas (EAs).
- **Primary source:** Primary quantitative data giving further information on exposure. The collection of this dataset is detailed in the following paragraphs.

Following a trialling period, the collection of primary quantitative data began by providing information to Uwezo concerning the selection of districts. District selection was followed by the agreement of contracts covering training and data collection arrangements, with GYEG and TUSPO (the data collection teams that had worked on the Uwezo survey in selected districts). The terms of these contracts were created to produce the data most suitable for answering the primary research question, within time and budget limitations. Agreements stipulated that training and data collection would involve thirty technicians and between two and three supervisors per district (from GYEG and TUSPO), all of whom had recently participated in the Uwezo data collection. These data collectors were to visit the households previously sampled by Uwezo individually, in contrast to the Uwezo data collection in which technicians worked in pairs.

Data collection training lasted three days and was led by Mussa and myself (with brief support from Titus in Ilala district). During the training period, assessors began by participating in group activities to promote familiarisation with educational television and Ubongo Kids itself (see Figure 3.2). After these activities, each element of the data collection tool (Appendix 3) and its accompanying question sheet (Appendix 4) were covered sequentially through presentation, examples and role-plays. Explanation of data collection tools was supported by a PowerPoint presentation featuring images of data collection and question sheets. The training was concluded with the assessors implementing the new data collection tool in a location outside of the sample frame before group discussion of any difficulties encountered.

Figure 3. 2: Group activities in Temeke

After training, assessors each spent five days conducting supervised data collection in their assigned district (pictured before departing for data collection, Figure 3.3). Given the geographical extent of Ilala and Temeke, close supervision of technicians posed a challenge. To overcome the difficulties of supervision, Mussa and I worked with the supervisors from GYEG and TUSPO to make impromptu visits to the villages in which technicians were conducting assessments (during which ongoing surveys were observed, Figure 3.4). Additionally, we communicated with technicians to meet them during their work. Further, we ensured that each technician had our contact details and felt able to contact us with queries.

Figure 3. 3: The team of assessors in Ilala

Figure 3. 4: An assessor conducting data collection in Temeke

As a result of the comprehensive instructions provided during training and the inclusion of key instructions on the data collection tools themselves, the number of queries was minimal. When these

arose, they generally focused on procedural matters concerning, for example, whether households outside the sample frame could replace sampled households who were not available. To this question, I responded that new children should not be added to the sample frame. This response reflected the intention that primary data collection should mirror Uwezo's sampling approach. This allowed children who participated in primary data collection to be matched with those assessed by Uwezo. After a process of cleaning, merging and variable transformation, original data (including that from primary data collection) became the datasets used for analysis (see Section 3.3.1.1.1).

3.3.1.1.1   Matched set

The 'Matched set' was created by merging the pre-public data source, Uwezo 2017, with primary quantitative data. The 'Matched set' therefore contained children who were both assessed by Uwezo in 2017 and captured in the primary data collection that immediately followed the Uwezo 2017 survey. Initial manipulation of Uwezo 2017 data was conducted by creating a subset retaining children who had information for any Uwezo mathematics item, complete information for all other variables included in the subsequent regression model and resided in Ilala or Temeke district (see Section 3.3.1.4.1). Next, the primary quantitative data were cleaned to retain children who had information for any Ubongo Kids exposure item and any other media exposure item.[36] A one-to-one merge between Uwezo 2017 and Primary information was subsequently performed. Creation of the 'Matched set' was completed by removing a single child who was located in an EA for which there were no other retained children.[37] Prior to cleaning, the primary quantitative dataset comprised a total of 1,033 children, while the Uwezo subset comprised a total of 1,233. Of these, 793 children possessed the required information and were successfully matched between datasets. Failure to match children between datasets was attributable to the presence of children in the primary dataset that were not captured by Uwezo, and vice versa. Matches were likely also missed because of misreported names/ages/gender/phone numbers in either dataset.

Such discrepancies between datasets impeded an automated matching process initially developed for this purpose, which relied on unique identifiers based on a combination of household telephone numbers and multiples of child ages. All matches captured by this process were verified through manual comparison of child name, caregiver name and head of household name as well as automated checks based on gender. Children within the primary dataset who remained without a match in the Uwezo dataset were classified according to their enumeration area (EA) and considered manually against entries for the corresponding EA in the Uwezo dataset. After completion of this process, matches were again examined through automated checks based on gender and age. Any apparent

---

[36] This process resulted in the loss of one child from the sample, who had no 'exposure to other media' information.

[37] Given that quantitative investigation of the 'Matched set' was ultimately conducted using EA-level fixed effects, this child did not benefit analysis.

match for a child whose age differed by greater than 2 years between datasets was discarded. (Retained matches in the primary sample were then given a fixed value within a 'matching' variable corresponding with the Uwezo dataset to permit matching through a Stata script: Annex I.) In cases where matches were retained despite conflicts between the datasets (e.g. in reported gender or age,), analysis employs values from Uwezo wherever possible. Choosing not to adopt primary data values in cases of conflict permits comparison of my findings with future research based on analysis of the (forthcoming) publicly available Uwezo dataset.

To examine whether this matching process introduced bias into the 'Matched set', children within the Primary sample who *were* (Group 1) and who *were not* (Group 0) matched with those in the Uwezo plus 6 sample were considered through t-tests concerning three characteristics:[38]

- Number of Ubongo Kids characters recognised
- Age (as reported in the primary sample)
- Gender (as reported in the primary sample)

Despite a limited sample size (with Group 0 comprised of between 194-196 children in each test), the lack of any significant differences between Group 0 and Group 1 supported the assumption that matched and non-matched children were not systematically different in terms of observable characteristics.

3.3.1.1.2   Household set

The 'Household set' was created solely from the Uwezo Tanzania dataset for 2017. To create the 'Household set', the following steps were taken:

1. A subset of the Uwezo 2017 dataset was created by selecting those children for whom information was available for any Uwezo mathematics item and all other variables used in the cross-section model (described in Section 3.3.1.4.2).
2. Children were then retained from the subset created in step 1 only if there was information for another child in their household.

Implementing steps 1 and 2 produced a final 'Household set' of 39,717 children.

3.3.1.1.3   Panel set

The 'Panel set' was created from secondary pre-public and public datasets: the Uwezo Tanzania datasets for 2017 (pre-public) and for 2011, 2013 and 2015 (public). It contained children assessed by

---

[38] These t-tests excluded the two 'matched' children who featured in my primary dataset yet were considered outside of the testing age group (6-16) by Uwezo assessors.

Uwezo who resided in districts that featured consistently throughout Uwezo data collections for 2011, 2013, 2015 and 2017. To create the 'Panel set', the following steps were taken:

1   A subset of the 2017 dataset was created by selecting those children who:
- were located in 1 of 45 districts common to the Uwezo 2017, 2015, 2013 and 2011 datasets
- had information for the main mathematics test
- had information for the Kiswahili test

2   Subsets of the 2015, 2013 and 2011 datasets were created by selecting those children who were located in 1 of 45 districts common to Uwezo 2017, 2015, 2013 and 2011 datasets.

3   IRT scores for mathematics capability were produced for children in each subset (see Section 3.3.1.4.3).

4   District-level averages for key variables (including mathematics capability) across each subset were collated into a new dataset (see Section 3.3.1.4.3).

3.3.1.2   Sampling

The Uwezo dataset followed a sample frame produced by the NBS. The specifications of this sample frame were obtained from publicly available documents (cited where available) in addition to information from the NBS sent via email (in the form of shared internal documents and responses to follow up questions). This sample frame remained mostly consistent across each survey round included in this thesis (2011, 2013, 2015 and 2017). In each year, the target population was the entire population of children in Tanzania aged 7-16 years in 2011, 2013 and 2015 and 6-16 years in 2017.

To provide information on these populations, Uwezo employed a sampling frame comprised of three essential layers. Firstly, data were collected in districts (of which there were 133, 131, 159 and 56 for the Uwezo surveys in 2011, 2013, 2015 and 2017 respectively)[39] by separate data collection teams comprising 60 assessors in each district. Next, within each district, 30 enumeration areas (EAs) were selected, for each of which a pair of assessors from a data collection team were responsible. Finally, from each of these EAs, 20 households were selected in which one caregiver and all children within Uwezo's aforementioned age ranges were surveyed by an assessor pair.

Uwezo reportedly selected districts through a simple random sample (Uwezo, 2015b:

https://www.twaweza.org/uploads/files/TZ15-

---

[39] District boundaries were common between Uwezo surveys except for Kinondoni and Temeke. In 2017, Kinondoni was divided into two new districts, (a smaller) Kinondoni and Ubungo. Similarly, Temeke was divided into (a smaller) Temeke and Kigamboni. The numbers provided here reflect the total number of districts only (thereby ignoring these divisions). Analysis involving the 'Panel set' therefore combined Kinandoni and Ubungo as well as Temeke and Kigamboni. Cross-section models using the 'Matched set' and 'Household set' retained the distinction between Temeke and Kigamboni as well as Kinandoni and Ubungo.

Data%20cleaning%20and%20usage%20NOTES%202015.pdf) with the total number of districts selected seemingly determined by the availability of data collection funding in any given year. The selection of EAs was then conducted through a probability proportional to size (PPS) method (Uwezo, 2015: http://www.twaweza.org/uploads/files/UwezoTZ-ALA2014-FINAL-EN.pdf).[40] Lastly, households were selected by a simple random sample conducted by organisations supporting data collection in each district (such as GYEG and TUSPO), using a table of random digits to pick from lists these organisations compiled of all households in each EA.

This sampling approach produced representative information at the district level. However, this survey structure suggested the implementation of weighting to account for the differing probabilities of household inclusion at the national level. Survey weights calculated by Uwezo were applied to create estimates of the number of viewers (see Section 3.3.2). Yet, weights were not applied in any models used to gauge the relationship between mathematics capability and educational television. This is because Uwezo sample weights were applicable only to the entire Uwezo sample, while models concerning the relationship between educational television and mathematics capability employed subsets of Uwezo dataframes. Uwezo survey weights were calculated in the following manner (Uwezo, 2013, p. 3):

> … the inclusion probability [P] for household i located in EA j within [district] k and country l, is the multiple of three probabilities – the probability of selecting that household within the enumeration area in which it is located, the probability of selecting that enumeration area given the [district], and the probability of choosing that [district].

This approach is described by the following equation:

$$P_{ijkl} = \frac{\alpha_j}{E_j} \times \frac{\beta_k E_j}{S_k} \times \frac{\gamma_l}{N_l} = \frac{\alpha_j \times \beta_k \times \gamma_l}{S_k \times N_l}$$

Here, "α_j gives the number of households sampled in EA j; β_k gives the number of EAs sampled from strata k; γ is the number of strata [sampled]; E_j is the total number of households in EA j, S_k is the total number of households in strata k; and N is the total number of strata (districts) in the country" (Uwezo 2013, p. 3).

Primary quantitative data collection for this study followed the sampling approach adopted by Uwezo in 2017 but was restricted to Ilala and Temeke districts (thereby sampling no children from the remaining 54 districts). These two districts were chosen for pragmatic reasons, since Uwezo advised that these were the districts in my city of residence in which organisations would be most likely to

---

[40] Communication from the NBS suggested that PPS was used in 2011, 2015 and 2017, with 2013 following a rotational panel approach that occurred from 2012-2014 in which 20 of the 30 EAs selected in the previous year were re-sampled along with 10 new EAs.

have completed Uwezo data collection before the end of my fieldwork period. In Ilala and Temeke, Uwezo's sampling approach was mirrored precisely. To ensure this, data collectors re-employed the sampling sheets originally provided by Uwezo to GYEG and TUSPO for the Uwezo data collection itself (Figure 3.5).

Figure 3. 5: A household sampling sheet used by assessors in Ilala and Temeke

3.3.1.3   Key primary and secondary measures

This section provides information on the measures employed in analysis of the 'Household set', 'Matched set' and 'Panel set', beginning with an initial theoretical justification for variable selection. Variables concerning mathematics capability (as the dependent variable) and exposure to educational television (as an independent variable) were necessary components of investigation into the primary research question. Further, the inclusion of an independent variable for 'other media exposure' was warranted, as this measure helped control for the possibility that a positive relationship between 'Ubongo Kids exposure' and mathematics capability was attributable to varying character recall levels (or general media exposure, see Section 3.3.1.3.3). In addition, a non-mathematics attainment measure was employed in analysis of the 'Household set' to address the possibility that children with better results across other subjects tested by Uwezo were more likely to view Ubongo Kids (see Section 3.3.1.3.4).

Other independent variables (employed in all models) were selected due to their relationship with child learning outcomes. Previous research involving the analysis of Uwezo datasets had identified numerous measures to have a positive effect on outcomes. The following featured in this study:

- a child's (current) school enrolment status (Alcott and Rose, 2015)
- their age (Jones, 2017)
- their gender (Jones & Schipper, 2015)
- whether their mother attended school (Alcott and Rose, 2016)
- a continuous asset-wealth-index score (Mugo et al., 2015)

Excluded variables of note were participation in private tuition, attendance of a pre-school and the type of school attended. The type of school that children attend is undoubtedly important (Alcott & Rose, 2016). However, the Uwezo set did not provide data of sufficient quality to warrant its inclusion. Type of school information was missing for a large proportion of children in Uwezo data (Section 4.3). Additionally, what information there was frequently conflicted with other information provided by Uwezo (Section 4.3.2). Information on pre-school attendance can also be an important determinant of learning outcomes (Bietenbeck, Ericsson & Wamalwa, 2017), but could not be included because of high levels of missingness (Section 4.3). Lastly, information on private tuition (which has been shown to influence learning outcomes amongst children in India: Alcott & Rose, 2017) was not collected by Uwezo in 2017. This thesis now proceeds to provide further information where required, for variables included in subsequent analyses. This is presented in separate sections for 'Mathematics', 'Exposure', 'Other media exposure' and 'Non-mathematics attainment'.

3.3.1.3.1 Mathematics

The main mathematics variable in Uwezo datasets for 2011, 2013, 2015 and 2017 was initially captured in a level-based format (with six levels, 1-6, in 2017). In Uwezo's 2017 data collection, children were asked five questions sequentially, in order of difficulty, and were assigned a level corresponding with the number of questions that they answered correctly. To demonstrate the level format, three examples are now provided. Level 1 means that a child answered the first item incorrectly, level 6 means that all items were answered correctly and level 5 means that a child answered the first four items correctly and the final item incorrectly.

Distribution of mathematics results in 2017 showed strong floor and ceiling effects (see Figure 3.6). This finding correlates with a prior Kenya-based investigation into the internal (WLE Person separation) reliability of the main Uwezo mathematics variable. A sub-standard internal reliability result was found (<0.100), which was considered to be a result of the Uwezo mathematics items being too easy for some Kenyan children in ACER's (2015) convenience sample. This reliability result was inferior to (lower than) the result found for an alternative mathematics assessment, the Early Grade

Mathematics Assessment (with the reliability of these instruments compared by the ACER Centre for Global Education Monitoring, 2015). Further, the distribution pattern for the main mathematics test pointed to difficulties in treating this measure as an ordinal dependent variable in statistical models. These difficulties were evidenced by application of the Brant test, which suggested that the parallel regression assumption was violated (when the main mathematics variable was employed as the dependent variable in an ordered logit model).
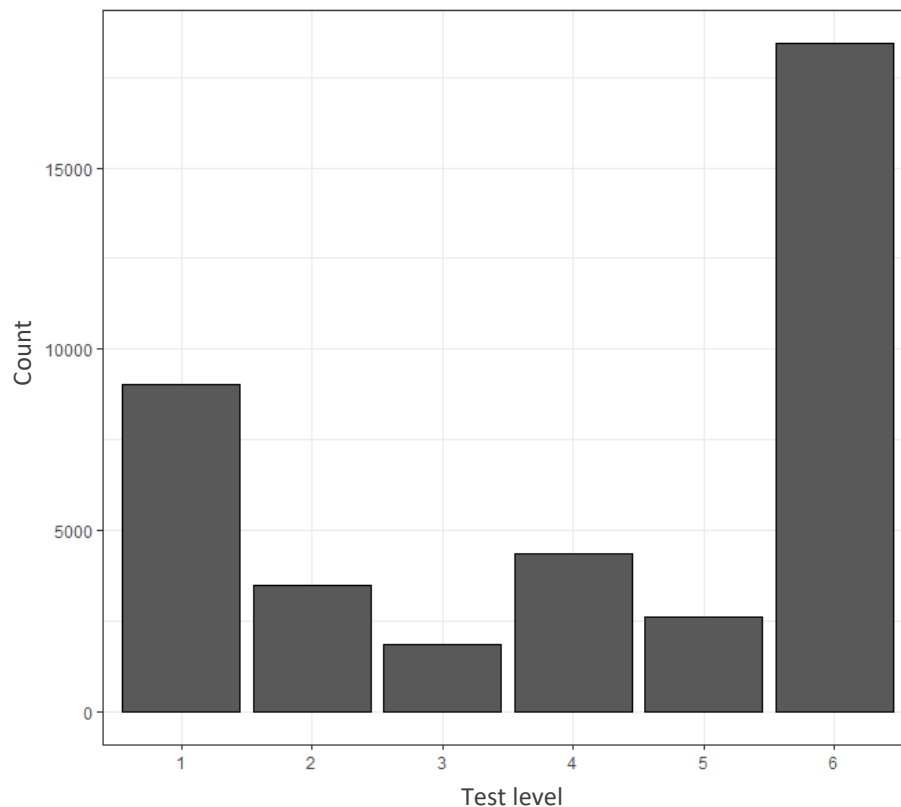
Figure 3. 6: Plot of main mathematics test results for the Household set, excluding missing values

Given the limitations of the main mathematics test administered by Uwezo, a perfectly reliable measure of mathematics outcomes could not be established. However, various measures could be employed to address distribution issues. The main approach taken in this thesis (in the analysis of all datasets) was to create estimates of an underlying latent trait, referred to as mathematics capability.[41]

---

[41] The literature typically considers latent estimates to be measures of an unobservable 'ability' trait (typically denoted as θ) (see, for example, Jones, 2017). However, latent estimates concerning mathematics were instead presented as measures of mathematics 'capability'. The employment of the term, 'capability', followed the criticism of Hart, Dixon, Drummond & McIntyre (2002) regarding the term it replaces, 'ability' (while acknowledging that this criticism applied to a differing employment of ability, i.e., as a moniker for 'fixed' or 'inherent' intelligence). Additionally, latent estimates of exposure to any form of media were merely reported to represent uncertain exposure estimates (see Section 3.3.1.4.2). This decision reflected the inapplicability of the term, 'ability', to measures of unobserved exposure.

These estimates were produced using item response theory (IRT). IRT can provide a framework through which a test takers' level of a trait is derived from responses to multiple items (Jones, 2016).

Creating trait estimates using IRT accounted for the differing spacing between levels in the main mathematics item. This addressed the possibility that, for example, it was relatively difficult for a child to increase their mathematics scores from a level 2 to a level 3, yet relatively easy to progress from level 3 to level 4. Further, the latent estimates employed in this study accounted for varying item discrimination. This would help if, for example, results for a particular test question provided a relatively weak indication of child capability. A test question would perform in this manner if there was limited difference in the probability of a low and high capability child answering a test question correctly.

This work follows previous analysis concerning Uwezo that employed latent modelling (Jones, 2017). However, unlike Jones (2017), this study did not create a multiple-subject measure of capability based on results for the main mathematics test, Kiswahili test and English test. Instead, this thesis drew upon all available mathematics-specific items in each Uwezo dataset employed (2017, 2015, 2013 and 2011). This included the multi-level main mathematics variable in addition to (one or) two further binary variables providing information on "mathematics in everyday life". In the Uwezo 2017 dataset, there were two everyday mathematics items which showed whether a child gave the correct response to verbal questions concerning addition and subtraction. In assessing the application of mathematics concepts to every-day life, these questions sought responses to scenarios such as: "Musa bought an orange for 400 shillings and a papaya for 100 shillings. How many shillings did he spend altogether?" (addition in everyday life, Uwezo 2017 assessment, translated).

Other approaches were used in the previous literature (as outlined by Jones, 2016). These included a basic binary transformation of the main mathematics variable. This approach had the disadvantage of removing detail from a measure already lacking in complexity. However, this simple method promoted replicability and was compatible with prior treatment of Uwezo mathematics information (see, for example, Alcott & Rose, 2016). This transformation was therefore employed to create a binary dependent variable for employment in cross-section models used as robustness checks (to show that results from analysis of the 'Household set' and 'Matched set' were consistent with those from simpler models).

3.3.1.3.2   Exposure

Differing proxies for exposure were employed in various models. Cross-sectional analysis of the 'Matched set' employed a measure formed of five character recognition items. Cross-section models for the 'Household set' and 'Panel set' used a binary self-reported measure of viewership. The employment of these differing exposure proxies reflected data availability. Those in the 'Matched set'

had participated in the primary data collection and Uwezo 2017 data collection, so provided information on character recognition (Rimal, Figueroa & Storey, 2013). Those featuring in the 'Household set' and 'Panel set' had only participated in the Uwezo 2017 data collection (aside from a proportion of those in Ilala and Temeke districts), so provided information only on a binary exposure measure. This section shall now provide information on the character recognition and binary exposure measures separately, before giving information on the construction of exposure variables employed in different models.

Character recognition

Character recognition assessment has been found to generally "provide a reliable assessment of children's exposure" (Rimal, Figueroa & Storey, 2013, p. 605). This is reflected in the employment of this technique for assessment of media interventions within comparable context (e.g., Borzekowski & Macha, 2010). To administer the character recognition items, images were shown sequentially to each child in the shadowing set (via assessor pointing) (Figure 3.7). A 'correct' response was recorded if the child could name the image. In the image below, five characters (1, 2, 3, 8 and 9) all featured in Ubongo Kids. Characters 4, 5 and 6 were used to capture other media exposure (see Section 3.3.1.3.3), whilst character 7 gave supporting information on exposure to another (pre-primary) show produced by Ubongo.

Figure 3. 7: Character recognition image

Self-reported exposure

Child-reported exposure was used as the sole piece of exposure information in the analysis of the 'Household set' and 'Panel set'. A question targeting exposure was submitted to Uwezo for inclusion in their 2017 data collection, in accordance with the following requirements: the question should be child-directed; and, should include just two response options (with the latter requirement attributable to limited space on the data collection form used by Uwezo). The question simply asked each child assessed by Uwezo (in 2017) whether they had watched Ubongo Kids in the past week (i.e., the week prior to being assessed) (see Figure 3.8). Responses were recorded as yes ("Ndiyo") or No/Do not know ("Hapana/Sijui"). These response options were chosen because of problems identified with Uwezo surveys prior to 2017, concerning missing data for certain variables. It was hoped that combining "No" and "Do not know" would avoid a situation in which assessors recorded no response when a child was unsure (which might have occurred when a child was unfamiliar with Ubongo Kids).[42] Further, a short time period (one week) was selected to combat difficulties in recalling

---

[42] Such a situation would have made it impossible to distinguish between situations in which, for example, a child did not respond to the question as they had not heard of Ubongo Kids or where a child was not asked the question at all.

exposure over a longer timeframe. Information on the trialling of this question is provided in Appendix 5.

		Taja idadi	H1705. Hisabati (weka alama ya Vema sehemu zote alizoweza)	
		Utambuzi wa namba		
		Namba zilizokosekana		
		Kujumlisha		
		Kutoa		
		Anaweza	H1	H1706. Hesabu katika maisha
		Hawezi		
		Anaweza	H2	
		Hawezi		
		Anaweza	P1	H1707. Swali la ziada Je, Mtoto ameweza kujibu maswali ya ziada sahihi ?
		Hawezi		
		Anaweza	P2	
		Hawezi		
		Ndiyo	H1708. Je umeangalia kipindi cha TV cha Ubongo Kids wiki iliyopita?	
		Hapana/Sijui		

Figure 3. 8: Binary question on reported viewership

Construction of exposure variables

In the analysis of the 'Household set' and 'Panel set', the construction of an exposure proxy required limited work. For the 'Household set' a cross-section design merely employed results from the binary variable in its original form, as an independent variable (see Section 3.3.1.4.2). For the 'Panel set', results for individual children were aggregated into district level averages to give an exposure measure suitable for investigation into change in capability ranking amongst districts (see Section 3.3.1.4.3).

However, in analysis of the 'Matched set', use of exposure information from child character recognition results required a more involved approach. This was because exposure information was provided by multiple items, to which children were less or more likely to respond 'correctly'. To create a value for exposure, a single latent exposure estimate was created. This followed a similar process to the creation of a latent estimate for mathematics. Information on the IRT model applied is provided below (see Section 3.3.1.4.1).[43]

---

[43] It should also be recognised that, In the simple logit model used to confirm findings from analysis of the 'Matched set', an exposure measure was used which was merely the sum of all positive or correct results for exposure items (giving a 0-6 score).

3.3.1.3.3 Other media exposure

Analysis of the 'Matched set' employed a control for exposure to other media.[44] A non-Ubongo Kids exposure variable for analysis of the 'Matched set' was derived from child responses to characters 4, 5 and 6 from the character recognition tool presented above (Figure 3.7). The recognition of other media figures was assessed in the same way as for Ubongo Kids characters, with a correct response recorded if the child could name an image when it was pointed to by the assessor. Additionally, a single latent estimate of other media exposure was constructed through IRT, following the process employed when creating a proxy for Ubongo Kids exposure.[45]

The use of information from non-Ubongo Kids character recognition items as an independent variable in analysis of the 'Matched set' followed prior evaluations of media interventions (Borzekowski & Macha, 2010). As for previous studies, this variable helped control for differences in mathematics capability being a product of either greater general media exposure or level of recall (which are likely to be correlated with mathematics outcomes). Indeed, with regards to the latter point, the inclusion of a non-intervention character recognition control was considered to counter the possibility that this measure "simply distinguishes children who are good at verbal [labelling] or at remembering media characters" (Cole & Lee, 2016, p. 92). Research by Borzekowski and Macha (2010) suggested that this measure functioned as a worthwhile control: in a RCT trial of preschool children, exposure to other general media captured using character recognition was significantly related to number recognition and arithmetic (albeit not counting) outcomes.

3.3.1.3.4 Non-mathematics attainment

Given the absence of media exposure measures in the 'Panel set' and 'Household set' (see Section 3.3.1.3.3), analysis of the 'Household set' employed an alternative control variable. It was hoped that the inclusion of this variable would help to control for the possibility that children with higher overall (non-mathematics specific) results were more likely to voluntarily engage with Ubongo Kids. Kiswahili was selected as the basis of this non-mathematics measure, as it is the topic assessed by Uwezo that was least likely to be influenced by viewing Ubongo Kids. Results from the Uwezo English test were not included in the non-mathematics attainment control, as performance in this subject could be improved by viewing Ubongo Kids (which also airs in English language). Similarly, 'bonus questions' in the Uwezo assessment were considered inappropriate as they featured elements

---

[44] A measure of other media exposure was not available for analysis of the 'Panel set' and 'Household set'. Information on other media exposure for those in the 'Panel set' and 'Household set' was unavailable, as sampled children had generally not participated in the primary data collection.

[45] As with the Ubongo Kids exposure variable, the simple logit model for analysis of the 'Matched set' employs a total score of all other media exposure items (0-3) as a predictor of mathematics attainment.

of number recognition (and were therefore likely to be correlated with performance in mathematics itself).

It was, however, acknowledged that a Kiswahili-based measure provided an imperfect control. This was due to the spill-over effects that interventions concerning one subject could have on another subject (Machin & McNally, 2008) and the (likely) positive influence on Kiswahili performance of viewing another show produced by Ubongo, Akili and Me.[46] As such, employment of Kiswahili attainment as a control variable might have diminished the apparent relationship between Ubongo Kids exposure and mathematics capability. Conversely, it was also possible that this relationship was now exaggerated through the inability to control for other media exposure in the analyses not involving the "Matched set". Reported viewership may, therefore, have captured exposure to other forms of media that had educational benefits.

Kiswahili test information was presented in the original Uwezo 2017 dataset in a 1-5 format. This format broadly matched information from the main mathematics test, albeit with one less 'level'. Binary transformation dependent on whether a child answered all sequential questions correctly was employed to create a variable for use as a non-mathematics control in analysis of the 'Household set'. This binary transformation approach corresponded with prior research involving Uwezo data (see, for example, Alcott & Rose, 2016).[47]

3.3.1.4 *Analysis*

This section provides information on the modelling approach used for investigation of the primary research question, which concerns the association between educational television and mathematics capability. Modelling was conducted using three quantitative approaches, which were applied separately to each dataset for analysis. Information is now provided for each quantitative model: the cross-section approach applied to the 'Matched set'; the cross-section model used for analysis of the 'Household set'; and, the multilevel model used for analysis of the 'Panel set'.

3.3.1.4.1 Analysis of the Matched set

Analysis of the 'Matched set' used a cross-sectional approach to estimate the association between educational television and mathematics capability in Ilala and Temeke districts. Both fixed and latent measures were employed in this model. This section begins by providing details on estimation of the latent values for mathematics, exposure to Ubongo Kids and exposure to other media that were included in the cross-section model.

---

[46] Akili and Me is targeted at pre-primary children, yet it is conceivable that some viewers of Ubongo Kids would have also watched Akili and Me.

[47] IRT estimates could not be created to provide a measure of non-mathematics capability. IRT required at least two items, yet only one Kiswahili-focused item was available in the Uwezo 2017 dataset.

Matched set: IRT models

The presence of differing question formats across the items relevant to particular latent measures necessitated the application of different IRT models. The mathematics capability measure was based on three items, one of which was polytomous (the main mathematics test item) and two of which were binary (the mathematics in everyday life questions). Conversely, both exposure to Ubongo Kids and exposure to other media were based on binary items only (concerning character recognition).

Given that the estimation of exposure to Ubongo Kids and other media used solely dichotomous data, the available IRT models were a 1PL, 2PL, 3PL or 4PL. These model names reflected the four parameters that could be accounted for, cumulatively, concerning item difficulty, discrimination, guessing (lower asymptote) and fatigue/carelessness (higher asymptote) (Cotton & Baker, 2018). Given the format of questions evaluating exposure (i.e. image recognition, rather than a multiple choice) as well as the limited duration of the assessment, 3PL and 4PL models were considered inapplicable. It was, however, anticipated that test questions would vary in difficulty, since Ubongo Kids characters had featured with varying frequency on the show. Additionally, there was no reason to presume that the discriminatory value of test questions would be non-uniform. A 2PL model was therefore selected.[48]

As items comprising the mathematics capability model were both polytomous and binary in nature, the available models were a Graded Partial Credit Model (GPCM) and a Rasch model (which is equivalent to a GPCM model but with item discrimination constrained to 1). Both models took account of varying item difficulty (and item category difficulty, in the case of the main mathematics test items). However, only the GPCM accounted for varying item discrimination. Since there was no theoretical justification for assuming item discrimination to be uniform, the model that allowed discrimination to vary (i.e. the GPCM model) was selected.[49]

Matched set: Consideration of IRT assumptions

Tests were applied to check the assumptions of (unidimensional) IRT. This involved assessing the assumptions of monotonicity, local independence and unidimensionality. Unidimensionality assumes that responses to all items used to measure a concept provide information primarily on a single dominant latent trait. This was tested through confirmatory factor analysis (CFA). CFA results for items comprising the latent measures of mathematics capability, exposure to Ubongo Kids and exposure to other media suggested that all items loaded onto their respective trait. This was shown by

---

[48] To check the applicability of the 2PL model, it was considered against a 1PL model using ANOVA. ANOVA results suggested that selection of the 2PL was warranted, given its (significantly) better log likelihood findings.
[49] The relative applicability of a GPCM model over a Rasch model was supported by ANOVA comparison of log likelihood information. The GPCM approach produced superior log likelihood results (with this difference being significant).

the factor loading results for the items concerning each trait, which were all above 0.3 (adopted as a cut off point following Kline, 2011). For the assumption of local independence to be satisfied, item responses with each individual trait must be mutually independent at any given theta level. This was assessed using Yen's (1984) Q_3 statistic, via the residuals function in the mirt R package. This test showed that no item residuals for any trait possessed a positive correlation of greater than 0.2 – the cut off at which Yen's Q_3 suggests items to be locally dependent (when adopting the cut off suggested by Chen and Thissen, 1997). Lastly, the assumption of monotonicity requires that the probability of responding positively to an item increases as theta increases. Evidence of monotonicity was provided (descriptively) by the pattern of item characteristic curves (ICC) (see below).[50] Further, application of the check.monotonicity function in the R mokken package showed no evidence of violation for any item comprising any trait (following guidance on the application of this test by, van der Ark, 2007).

IRT also assumes normality of the estimated latent trait. For traits that are inherently skewed (e.g., attention deficit disorder amongst a national population), specification of the trait density form is required to adjust for non-normality. Failure to specify trait density form in such instances could introduce bias (Finch & Edwards, 2016). However, IRT is robust to some level of non-normality (Cotton & Baker, 2018). Given that normality was anticipated with regards to exposure to Ubongo Kids, exposure to other media and mathematics capability (in the 'Matched set', 'Household set' and 'Panel set'), the IRT models applied did not account for trait density form. The presumption that mathematics capability likely possessed a sufficiently normal distribution was influenced by prior employment of IRT to estimate trait levels, using responses to Uwezo test items (Jones, 2017).[51]

Further checks were implemented after model implementation. These included consideration of fit indices for the model concerning exposure to Ubongo Kids. Fit statistics indicated a satisfactory model for Ubongo Kids exposure (see Appendix 6 for fit results and interpretation information). Fit statistics could not be generated for the mathematics capability or other media exposure models as they possessed too few degrees of freedom. Lastly, it might be noted that the discrimination parameters for all items employed to measure each trait were classified as 'moderate' or better (following classification guidance provided by Baker, 2001).

Matched set: Test and item information

Investigation into test information was conducted to consider the amount of information provided across varying theta levels for each trait. This suggested that maximal information was provided at low to average (slightly negative) theta levels for measures of mathematics ability and exposure to

---

[50] For example, an increase in theta corresponded with an increase in probability for each curve representing a binary character recognition item (see Figure 3.12).

[51] The mirt package in R provides no means of plotting theta estimate density. As such, this assumption was not tested empirically.

other media. Conversely, for exposure to Ubongo Kids, most information was provided at a theta level just above 0 (average). These findings on overall test information corresponded with the results of an investigation into the specific items relating to each trait.[52]

The individual items comprising each latent value were considered descriptively using item characteristic curves (ICC). These ICC plots gave information on each mathematics question (see Figures 3.9, 3.10 and 3.11), all Ubongo Kids exposure items (Figure 3. 12) and all other media exposure items (Figure 3.13). In each image, items were plotted on a graph where the x axis represents trait level (theta) and the y axis represents probability (P). For graphs depicting binary items, probability reflects the likelihood of responding positively to an item at any given theta level (Figures 3.10, 3.11, 3.12 and 3.13). For the graph depicting the main mathematics item, probability reflects the likelihood of a child with any given theta (capability) value obtaining a particular 'test level' (Figure 3.9). Specific model parameters are presented infrequently to support explanation of certain graphs. (Parameter statistics were obtained using the coef function in mirt.)

---

[52] Figure 3.13, for example, shows item characteristic curves (ICC) for exposure to other media. The identification that maximal information for exposure to other media was provided for test takers with low to average levels of the latent trait complies with the location and shape of item curves. The theta value at $P(\theta)=0.5$ associated with more discriminating (steeper) curves is towards the left-hand side of the x axis (i.e., such curves represent 'easy' items).
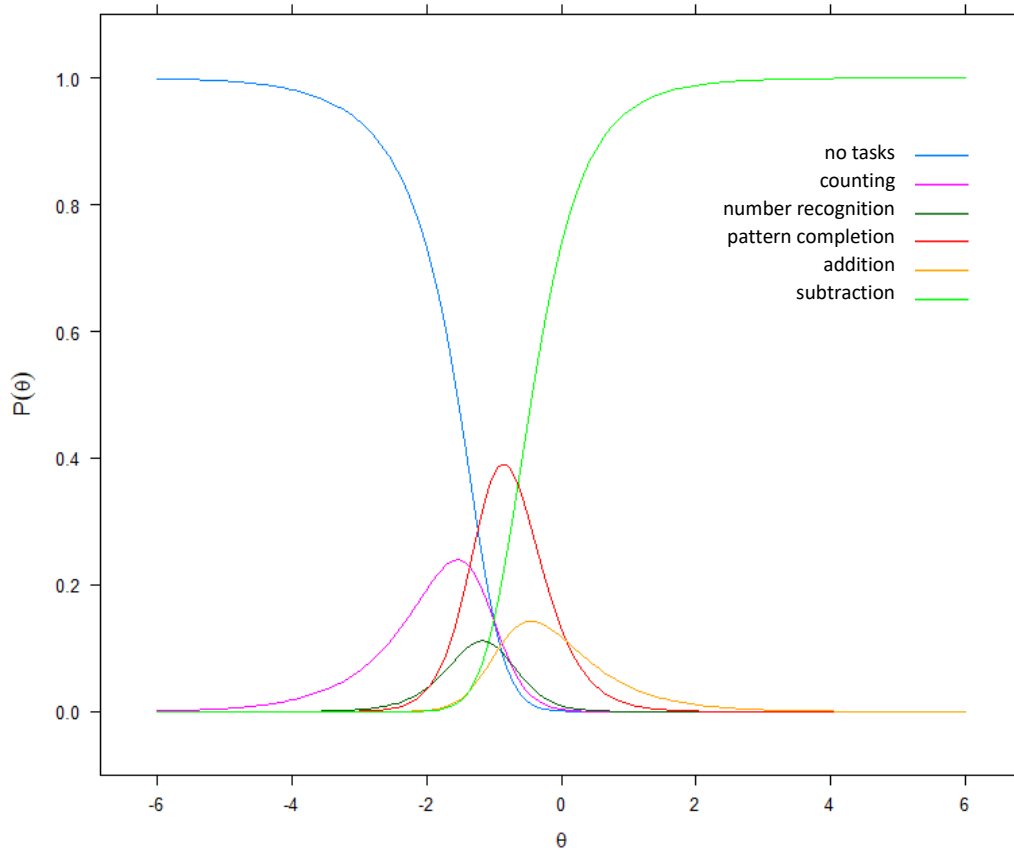
Figure 3. 9: Trace line for main mathematics test

The graph above shows the probability of a child within the 'Matched set' with any given capability (theta) value attaining any given test level. In accordance with the graphic legend, the trace lines represent: no tasks being completed (blue), counting completed only (purple), number recognition and all preceding tasks completed (dark green), pattern completion and all preceding tasks completed (red), addition and all preceding tasks completed (yellow), and subtraction and all preceding tasks completed (bright green).

The above image demonstrates that a test taker with any given theta (capability) value was never most likely to achieve any of the following levels: counting, number recognition or addition. This is shown by the plot lines for counting, number recognition and addition, which do not have a higher theta probability than all other items at any level of theta. There are numerous examples in the literature where the presence of such 'submerged' lines has led to item categories being collapsed, in a (misguided) attempt to improve item trait estimation (see, for example, Hendriks, Fyfe, Styles, Skinner & Merriman, 2012; and, Eklund, Erlandsson & Hagell, 2012).[53] However, all levels were

---

[53] In this instance, this would entail the collapsing of the submerged item levels for counting, number recognition and addition into other non-submerged levels: no tasks, pattern completion and subtraction. To exemplify this process, one stage would involve combining the levels of 'counting' and 'no tasks' by simply recoding 'counting' (level 2) as 'no tasks' (level 1).

retained in accordance with guidance from a simulation-based study, which identified no conditions in which the "collapsing of categories significantly improved the estimation of the latent trait" (Harel, 2014, p. 79).
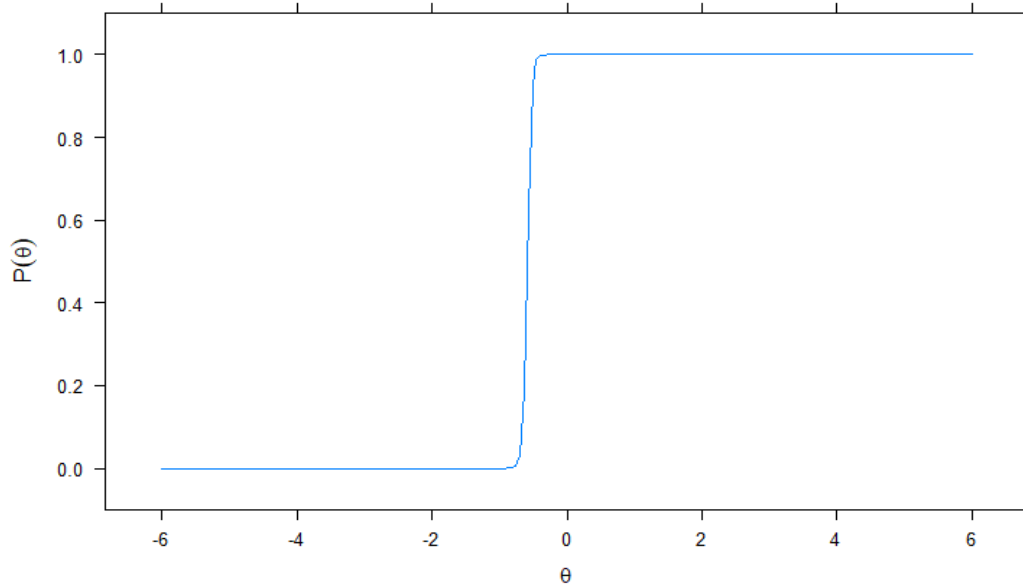
Figure 3. 10: Trace line for addition in everyday life

The graph above shows the probability of a child within the 'Matched set' sample with any given capability (theta) value achieving a correct result for the "mathematics in everyday life" question on addition. The curve shows that children with a theta value of 0 had a high probability of answering the question correctly (approaching $P(\theta)$=1) and children with a theta value of -2 had a low probability of answering the question correctly (approaching $P(\theta)$=1).
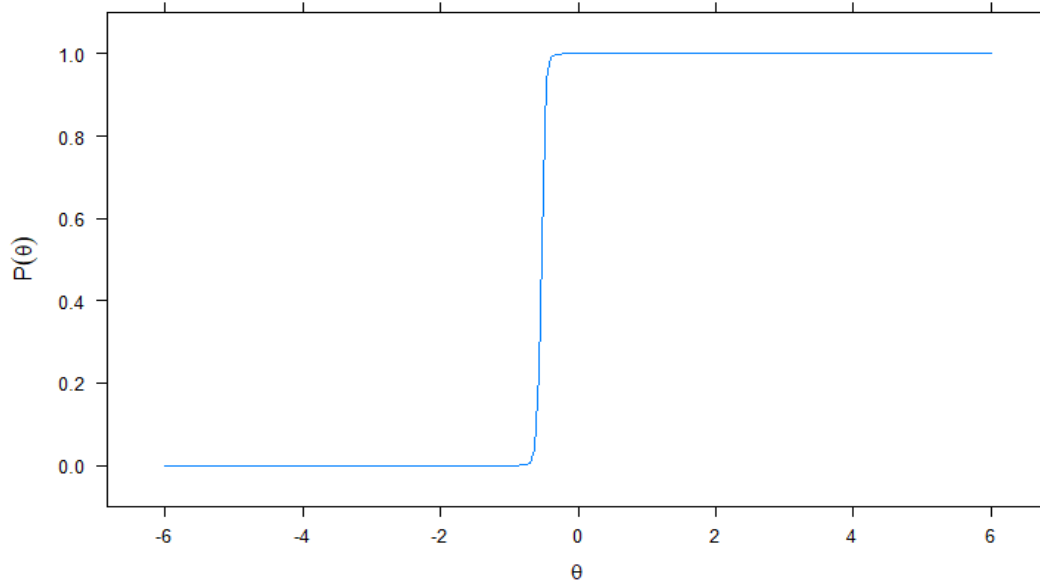
Figure 3. 11: Trace line for subtraction in everyday life

The graph above shows the probability of a child within the 'Matched set' sample with any given capability (theta) value achieving a correct result for the "mathematics in everyday life" question on subtraction. The curve shows very similar properties to that for the above question on addition. However, the curve is slightly further to the right. Curve placement shows that the probability of a test taker with a theta (capability) value of -0.52 answering the everyday subtraction question correctly was 50%, whilst for a test taker to possess the same chance of answering the everyday addition question correctly (Figure 3.14) they required a lower theta value (-0.58).
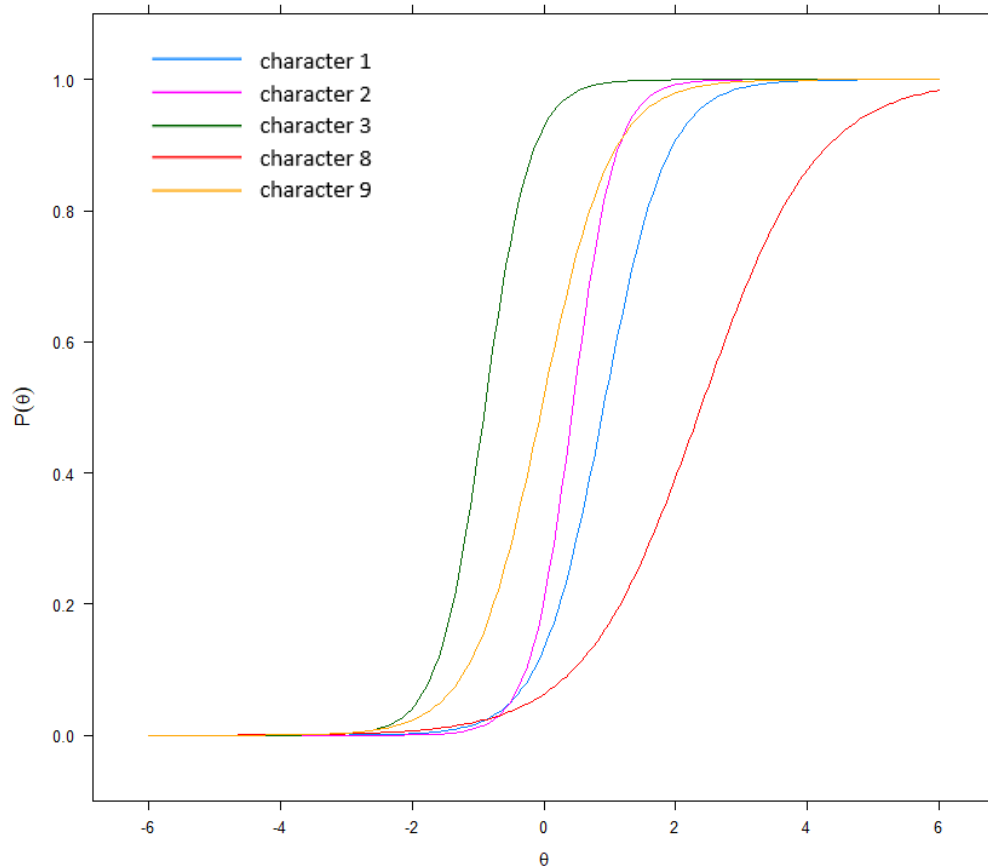
Figure 3. 12: Trace lines for Ubongo Kids exposure

The graph above shows the probability of a child within the 'Matched set' sample, with any given latent exposure (theta) value, achieving a correct result for individual character recognition items.

The presentation of multiple item lines on the same plot (Figure 3.12) provides information on the relative extent to which items discriminate between children with differing levels of exposure to Ubongo Kids as well as the relative 'difficulty' of items. The 'easiest' item ('character 3') is represented by the dark green line that is furthest to the left-hand side of the plot at P(θ)=0.5).[54] Information on the extent to which items discriminate between children who have received differing levels of exposure is shown by the steepness of the curves, with the steeper curves showing greater discrimination between members of the 'Matched set'. The steepest, pink, line represents 'character 2' and the flattest, red, line is 'character 8'.[55]

---

[54] Children possessing a theta level of -0.90 have a fifty percent probability of recognising character 3 but are less likely to recognise other characters.

[55] The contrast between the discriminating properties of 'character 2' and 'character 8' is confirmed by their differing discrimination index values (3.09 and 1.14 for 'character 2' and 'character 8', respectively).
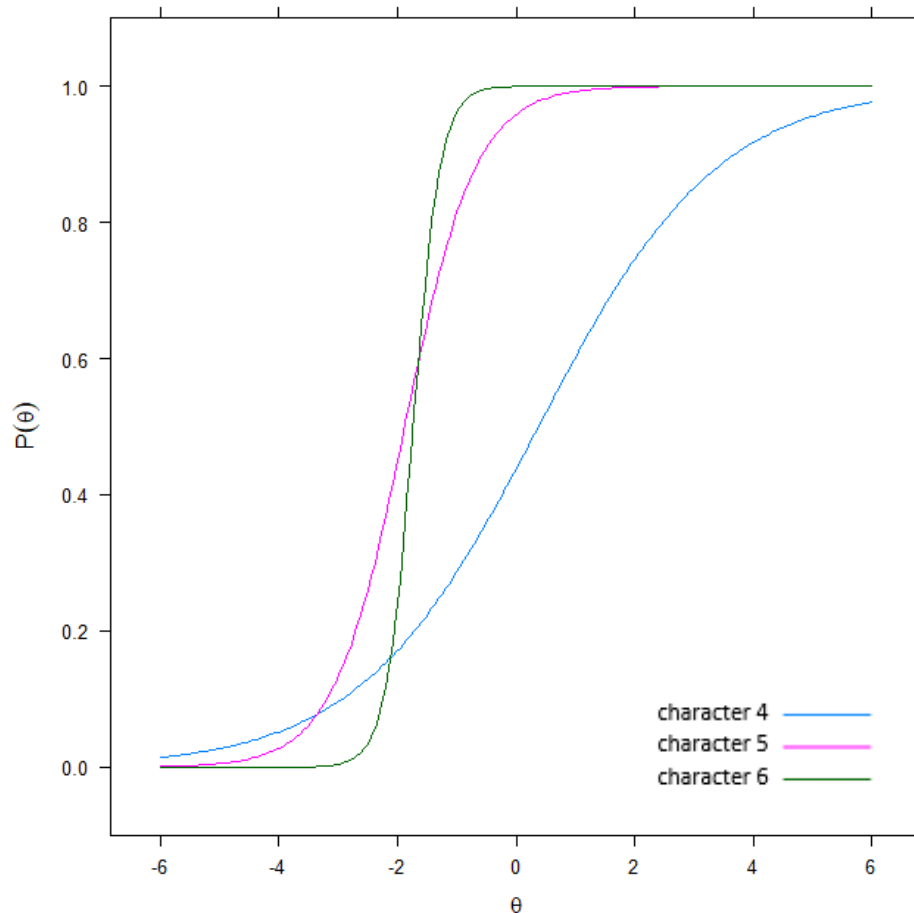
Figure 3. 13: Trace lines for other media exposure

This graph shows the probability of a child with any given non-Ubongo Kids exposure value (theta) recognising character 4, character 5 or character 6.

Matched set: Theta estimation

From the IRT models applied, theta values for each latent trait were predicted for all children in the 'Matched set'. These values were created using an expected a-posteriori (EAP) estimation method in the R package, mirt. This package was employed in a manner to create single theta values per child for each latent concept measured (stored as additional variables in the 'Matched set'), which reflected each child's most likely level of a trait (following Jones, 2017). As such, the theta figures used in analysis did not reflect the varying level of error associated with different estimates. This method could therefore potentially be enhanced by creating multiple plausible values for each latent estimate, all of which could then be treated as multiple imputations in a regression model. However, given the limited amount of mathematics information provided by Uwezo, this approach was not employed

since a very large number of predictions would have been required (to create a scenario in which coefficient results were not strongly influenced by variance amongst individual imputations).[56]

To provide an overview of predicted theta values, subsets of the 'Matched set' were created which retained columns for the items relating to each latent trait. A variable was then added to each subset to provide information on the frequency of each result combination, before all duplicates were removed and the data were ordered by theta estimate (lowest to highest). This information is provided in three tables (Table 3.2, 3.3 and 3.4). The table for Ubongo Kids exposure (Table 3.3) gives information for the lowest and highest predicted theta score only, given the large number of result combinations.

---

[56] Another means of addressing this problem was to exploit the capabilities of the lavaan package in R to estimate capability within regression calculations (i.e., performing both calculations simultaneously). It was not, however, possible to specify fixed effects within lavaan models. This denied the applicability of this approach to the household fixed effects model (see Section 3.3.1.4.2). The number of households in the 'Household set' was too large for fixed effects to be feasibly employed via manually specified dummy variables (which could otherwise have provided a way of circumventing this limitation of the lavaan package). As the household fixed effects model was considered to provide the most accurate estimation of the relationship between television exposure and mathematics capability, all quantitative models in this thesis employed an IRT-based approach.

Table 3. 2: All mathematics results combinations with theta estimates

Main item (1-6)	Everyday addition (0-1)	Everyday subtraction (0-1)	Capability estimate	Frequency
1	0	0	-1.65824	51
1	NA	NA	-1.63278	2
2	0	0	-1.33696	31
NA	0	0	-1.20095	3
3	0	0	-1.12376	15
4	0	0	-0.97821	76
5	0	0	-0.87713	22
6	0	0	-0.80569	18
1	1	0	-0.59735	1
3	1	0	-0.59256	1
4	1	0	-0.58953	6
4	0	1	-0.58612	1
5	1	0	-0.58586	4
4	NA	NA	-0.58357	1
6	1	0	-0.58137	7
6	0	1	-0.57623	2
1	1	1	-0.32555	5
2	1	1	-0.2799	4
3	1	1	-0.20856	6
5	NA	NA	-0.14068	1
4	1	1	-0.08571	37
5	1	1	0.148302	37
6	1	1	0.623686	462

From left to right, the columns in this table present scores for the main mathematics item, the mathematics in everyday life question for addition, the mathematics in everyday life question for subtraction, the capability estimate associated with each score combination and the frequency of each score combination.

Table 3. 3: Results combinations for Ubongo Kids exposure with highest and lowest theta estimates

Character 1	Character 2	Character 3	Character 8	Character 9	Theta estimate	Frequency
0	0	0	0	0	-1.28947	146
1	1	1	1	1	1.64331	30

This table presents two score combinations for character recognition items 1, 2, 3, 8 and 9. The far right and second-right hand side columns present the theta estimate associated with score combinations and the frequency in which score combinations occurred.

Table 3. 4: Results combinations for other media exposure with theta estimates

Character 4	Character 5	Character 6	Theta estimate	Frequency
0	0	0	-2.08221	19
1	0	0	-1.89707	3
0	1	0	-1.61233	17
1	1	0	-1.4256	5
0	0	1	-0.98903	37
1	0	1	-0.68837	17
0	NA	1	-0.13603	2
0	1	NA	-0.1189	2
0	1	1	-0.04795	363
NA	NA	1	0.106595	1
NA	1	1	0.18554	5
1	NA	1	0.396073	2
1	1	NA	0.423848	1
1	1	1	0.450786	319

This table presents score combinations for (from left to right), character recognition item 4, character recognition item 5 and character recognition item 6. The following two columns present the theta value associated with each score combination and the frequency with which this score combination occurred.

Matched set: Cross-section regression

A linear regression model was created using the latent mathematics capability estimate as the dependent variable, which was regressed on the latent measures for Ubongo Kids exposure and other media exposure alongside additional (fixed) control variables. These variables were: child age, child school enrolment status, child sex, the education level of a child's mother and the wealth index figure for a child's household. Information on these variables, including the theoretical justification for their inclusion, is provided above (see Section 3.3.1.3).

Further, fixed effects were employed via EA-level dummy variables. These dummies were employed to control for bias resulting from potential (fixed) unobserved EA heterogeneity correlated with independent variables. Such heterogeneity could include, for example, an EA-specific intervention that had the effect of increasing interest in mathematics. Additionally, heterogeneity between EAs could be a product of the Uwezo data collection format, in which different pairs of data collector teams were responsible for data collection in each EA (see Section 3.3.1.2).[57] The employment of

---

[57] The fact that there might have been omitted variables which were correlated with independent variables suggests the relative applicability of a fixed effects over a random effects approach. A random effects approach would not take account of (fixed) unobserved district characteristics that could be correlated with independent variables (despite this approach accounting for clustering within districts) (Crawford & Vignoles, 2014).

fixed effects to control for location heterogeneity followed prior analyses of Uwezo data (Jones & Schipper, 2015; and, Alcott & Rose, 2016).

The format of the Uwezo sample also influenced the specification of clustered standard errors. It was recognised that data collection in each district was led by one team, who organised 30 unique data collection pairs (each operating in a single EA). This suggested that the data were nested at the EA and district level. Failing to account for nesting structure could have caused standard errors to be underestimated, which would have falsely inflated the chance of identifying significant independent variable effects (on mathematics capability) (Ong, Williams & Lamprianou, 2013). Clustered standard errors were therefore specified at the EA and district levels. While it could be argued that children in the 'Matched set' sample were also nested at the household level, this was not accounted for in the final model. Trialling suggested the specification of additional clustering at the household level had no substantive effect on error values (or p-values). As in Delprato & Sabates (2015), this decision was supported by the fact that a large proportion of households in the dataset used for analysis had only one assessed child (57% of households).

All variables comprising the cross-section model including EA dummies are presented in the formula below:

$$MathsCap = \beta_0 + \beta_1 UbEst + \beta_2 MedEst + \beta_3 Age + \beta_4 Sex + \beta_5 AWI + \beta_6 M + \beta_7 Enr + \beta_8 EA + \varepsilon$$

In this formula, MathsCap represents latent mathematics capability, UbEst denotes the latent estimate of child exposure to Ubongo Kids, MedEst represents the latent estimate of child exposure to other media, Age refers to child age, Sex to child sex, AWI to the household asset wealth index score of a child's household, M to a child's mother's prior school attendance, Enr to a child's current school enrolment status and EA to the enumeration area in which the child was located (thereby containing a vector of dummy variables).

Before estimating the above regression equation, an assessment of normality amongst all non-binary variables was conducted. Variable normality was considered, as the validity of parametric tests including regression requires variables to be normally distributed (Field, Miles & Field, 2012). This investigation into normality included the latent variables created using IRT: while it was assumed that the underlying trait distribution was normal for mathematics capability, exposure to Ubongo Kids and exposure to other media, it is possible that the distribution of EAP estimates for each measure were not. Normality was checked for all IRT estimates and the continuous asset-wealth-index variable by assessing kurtosis and skewness and comparing mean and median values. Mean and median values were similar for all variables.[58] Further results from kurtosis and skewness calculations (using the

---

[58] The largest difference was for mathematics capability, where the mean was 0.02 and the median was 0.62.

"e1701" R Package) fell within broadly accepted boundaries (-2 to 2), except for the kurtosis finding concerning exposure to other media (3.93).[59]

Additional checks were conducted after estimating the above regression equation. These included an investigation of multicollinearity, through consideration of variance inflation factor (VIF). This check was satisfied, as the extent to which the variance of each regression coefficient was increased due to collinearity fell within an adopted cut off point.[60] Finally, the robustness of results was considered using an alternate cross-section model (in a manner comparable to the robustness check performed by Delprato & Sabates, 2015). In this alternate model, all latent estimates were replaced by fixed variables measuring the same concepts. That is, a binary measure of mathematics attainment and a total character recognition score were used in place of mathematics capability and latent exposure, respectively. This model did not produce results that differed substantively.

3.3.1.4.2 Analysis of the Household set

To provide further information on the (primary) research question concerning the association between educational television and mathematics ability, an additional cross-section approach was employed. This cross-section model was used for analysis of the 'Household set'. This model bears a clear resemblance to the cross-section approach discussed previously (see Section 3.3.1.4.1). Both models employed fixed and latent variables concerning a single time point to consider the relationship between exposure and educational television. However, analysis of the 'Household set' differed in three key areas.

The first difference was that 'Household set' analysis used a less granular measure of television exposure. The measure was therefore less informative. In addition to this, the binary measure of television exposure employed could have biased results downwards. This is because some children who had been meaningfully exposed might not have reported viewership in the week prior to the 2017 Uwezo data collection. Further, findings concerning educational television might have been biased in the same direction by child misreporting of viewership occurrence.

The second and third differences, however, suggested analysis of the 'Household set' was superior to that of the 'Matched set'. The second difference related to sampling. Analysis of the 'Household set' was benefited by it being a national data set and having a larger sample size (n = 39,717). The third difference related to the form of fixed effects employed. The large number of households (nationally) from which multiple children were assessed by Uwezo in 2017 permitted the employment of a

---

[59] This result suggested that the distribution of sample data did not tail off appropriately. However, this finding was not considered to warrant the omission or reformulation of the exposure to other media variable. This was because a) kurtosis, mean and median findings were acceptable, and b) it was considered desirable to use a control variable that corresponded in format with the exposure to Ubongo Kids variable.

[60] VIF values were calculated using the vif function from the R package, car. All VIF values fell within the adopted cut off point of 5 (which is commonly adopted as a cut off figure in the literature: Sheather, 2009).

household fixed effects model.[61] Application of household fixed effects should control for a greater amount of unobserved heterogeneity than EA fixed effects (applied for analysis of the 'Matched set'). Additional heterogeneity controlled for could include, for example, caregivers in different households providing varying levels of learning support.[62] On balance, the advantages of the 'Household set' analysis outweighed those of the 'Matched set'. For this reason, it was employed as the preferred estimate of 'effect' in subsequent cost-effective analysis (CEA) calculations (see Section 3.3.2). This section now provides information on the estimation of the latent variable (mathematics capability) employed in the cross-section model, before considering the model as a whole.

Household set: IRT model

Given that the 'Household set' included no character recognition items, mathematics capability was the only latent measure that was modelled. IRT modelling of mathematics capability for the 'Household set' followed the approach taken regarding the 'Matched set' (see Section 3.3.1.4.1). This involved using one polytomous mathematics item and two binary mathematics in everyday life items in a GPCM model. This model was selected in accordance with item format (see Section 3.3.1.4.1). Further, the selection of this model over a Rasch model was supported by a comparison of log likelihood information (using ANOVA).[63] (The GPCM model formula is presented in 'Matched set: IRT models.)

Household set: Consideration of IRT assumptions

As the GPCM model under discussion was applied to a different dataset to that considered previously (the 'Household set' as opposed to the 'Matched set'), the IRT assumptions of monotonicity, local independence and unidimensionality were reconsidered. These assumptions were assessed using CFA (for unidimensionality), Yen's Q_3 (for local independence) and the mokken package check.monotonicity function (for monotonicity). The application of CFA showed that the factor loading for each item was greater than the adopted cut off point of 0.3, with all items having a factor loading value of 0.861 or higher. Employing Yen's Q_3 showed that no items possessed a positive correlation of greater than the selected cut off of 0.2. This suggested that the assumption of local independence was satisfied. Additionally, results from implementation of the check.monotonicity function provided no evidence that the assumption of monotonicity had been violated. This finding was supported descriptively by the shape of binary (Figure 3.15 and Figure 3.16) and polytomous (Figure 3.14) item curves. It might also be recognised that the discrimination values for all items

---

[61] There were a limited number of households in which multiple children participated in both the primary and Uwezo 2017 data collection. This obstructed the application of household fixed effects to the 'Matched set'.
[62] It could be recognised that unobserved differences between children in households could still bias results. Such differences could include individual variation in motivation.
[63] The GPCM model gave superior log likelihood results to the Rasch model. This difference between models was significant.

comprising mathematics capability could be classified as moderate or better (following the classification guidelines proposed by Baker, 2001).

Household set: Test and item information

The mathematics items employed in the GPCM model provided maximal information for test takers with low to average levels of ability (θ = -0.5 to 0). This assertion was supported by ICC investigation into (individual) item information. ICC plots were constructed for each mathematics item, with capability level (theta) on the x axis and probability on the y axis. Probability refers to the likelihood of answering a binary item correctly (for Figure 3.15 and Figure 3.16) or the likelihood of a test taker achieving a particular test category (Figure 3.14). ICC curves supported the assertion that maximal information was provided for test takers with low to average ability, as the midpoint (P(θ)=0.5) of curves for binary items was located just below θ = 0, and levels for the main mathematics test were clustered around a similar theta value.
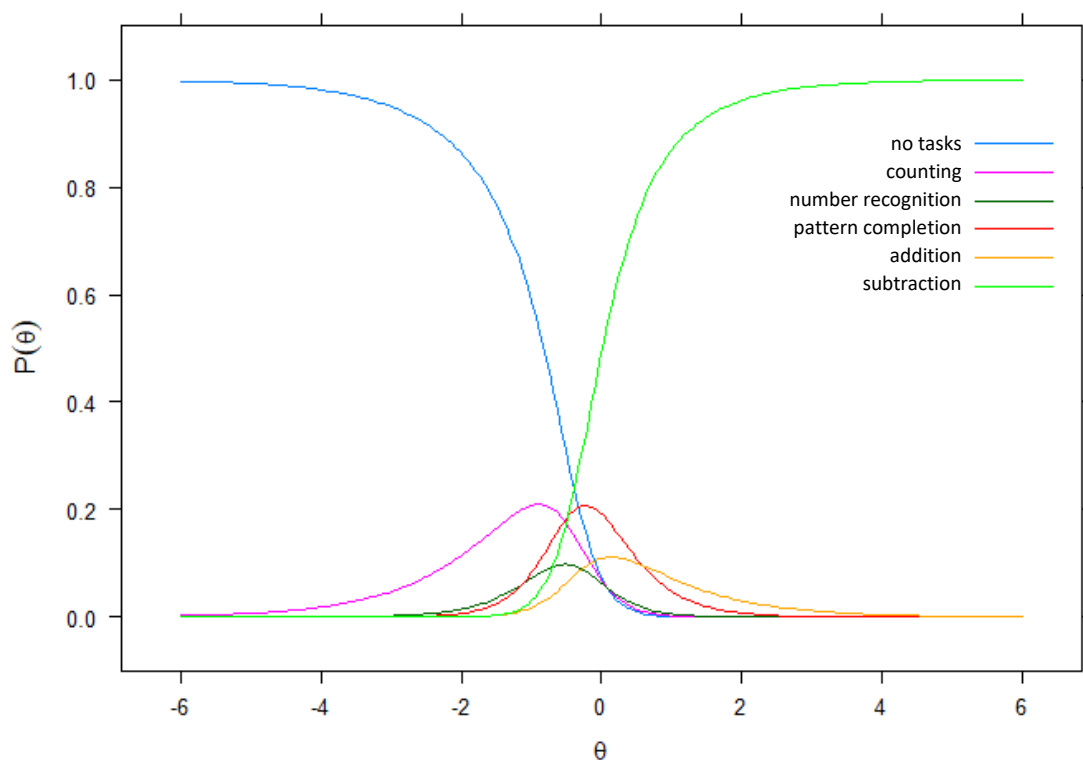
Figure 3. 14: Trace lines for main mathematics item

The graph above shows the probability of a child within the 'Household set' with any capability (theta) level attaining any given test level. Trace lines represent: no tasks being completed (blue), counting completed only (purple), number recognition and all preceding tasks completed (dark green), pattern completion and all preceding tasks completed (red), addition and all preceding tasks completed (yellow), and subtraction and all preceding tasks completed (bright green).
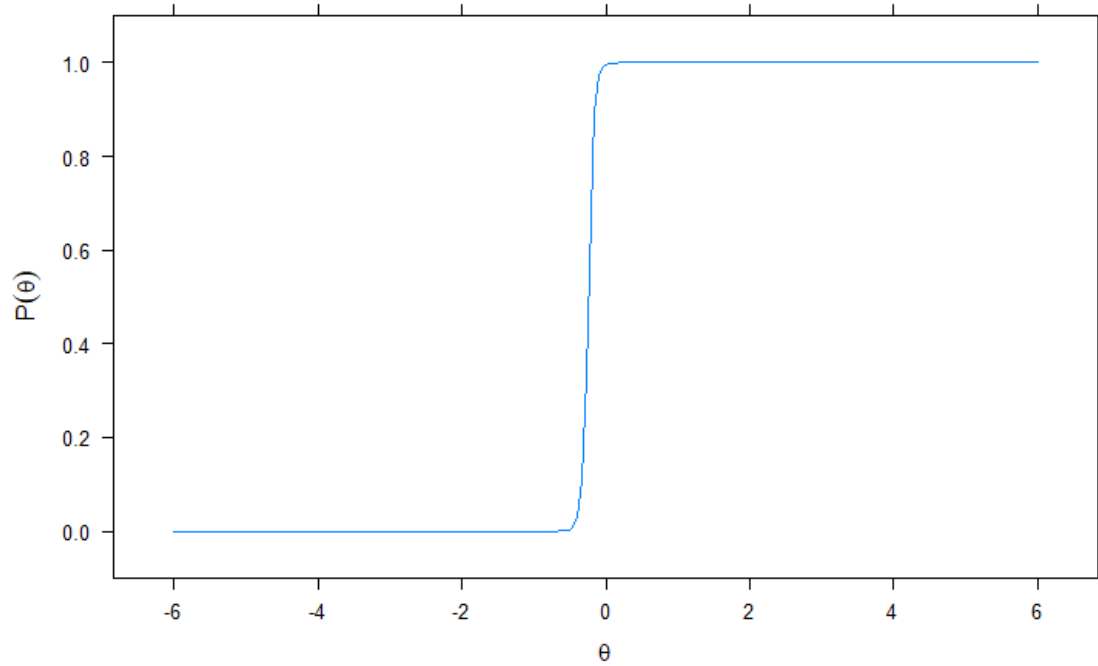
Figure 3. 15: Trace line for everyday addition item

The graph above shows the probability of a child within the 'Household set' sample with any given capability (theta) value achieving a correct result for the "mathematics in everyday life" question on addition.
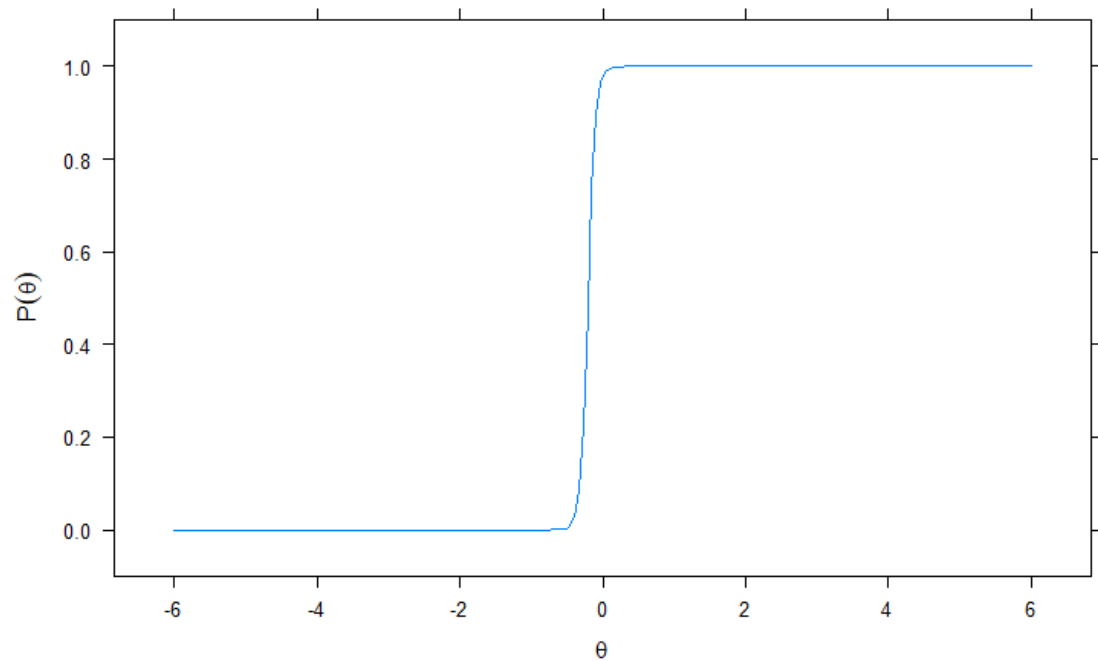
Figure 3. 16: Trace line for everyday subtraction item

The graph above shows the probability of a child within the 'Matched set' sample with any given capability (theta) value answering the subtraction in everyday life item question correctly.

Household set: Theta estimation

Theta values for mathematics capability were predicted from the GPCM model considered earlier in this section (see 'Household set: IRT model'). A single theta value per child was again predicted using an EAP method. To provide an overview of theta estimates, a subset of the 'Household set' was created. This contained columns for the main mathematics item, everyday addition question, everyday subtraction question and theta (capability) estimate associated with each result combination. An additional column providing a value for the frequency of each result combination was also produced. For ease of interpretation, only the rows with the lowest and highest theta combinations are presented (see Table 3.5).

Table 3. 5: Results combinations for mathematics capability with highest and lowest theta estimates

Main mathematics (1-6)	Everyday addition (0-1)	Everyday subtraction (0-1)	Capability estimate	Frequency
1	0	0	-1.20304	7930
6	1	1	0.82638	16576

From left to right columns, this table presents scores for the main mathematics item, everyday addition item, everyday subtraction item, the theta (capability) value associated with the score combination and the frequency of the score combination. Presentation in this table is confined to the rows containing the lowest and highest capability estimates, as 58 score combinations occurred within the 'Household set' (which was considered too many to present in a single table).

Household set: Cross-section regression

In the linear regression model applied, the latent mathematics capability estimate was regressed on multiple fixed control variables. These were: self-reported exposure measure, child age, child school enrolment status and Kiswahili attainment. Information on each of these variables was provided above (see Section 3.3.1.3). This included the theoretical justification for inclusion of these measures (ibid). This cross-section approach included no independent variables for the education level of a child's mother or the wealth index figure of a child's household. This was because this model employed household fixed effects, which accounted for all (fixed) observed and unobserved differences between households. Within the household, there was no variation for mother's education or wealth, so these variables did not need to be specifically included. As with analysis of the 'Matched set', this model accounted for the nesting structure of the Uwezo sample by applying clustered standard errors at the EA level and district level.[64]

---

[64] Clustered errors were again not specified at the household level. This is because, once more, they had no substantive effect on error values or p-values.

The cross-section model for analysis of the 'Household set' is presented in the formula below:

$$(MathsCap_{ih} - \overline{MathsCap}_h)$$
$$= \beta_0 + \beta_1(UbRep_{ih} - \overline{UbRep}_h) + \beta_2(k_{ih} - \overline{k}_h) + \beta_3(Age_{ih} - \overline{Age}_h)$$
$$+ \beta_4(Sex_{ih} - \overline{Sex}_h) + \beta_5(Enr_{ih} - \overline{Enr}_h) + (\varepsilon_{ih} - \overline{\varepsilon}_h)$$

In this formula, *MathsCap* represents latent mathematics capability, *UbRep* denotes self-reported exposure to Ubongo Kids, k represents Kiswahili attainment, *Age* refers to child age, *Sex* refers to child sex and *Enr* concerns a child's current school enrolment status. Subscript *ih* denotes information for an individual, i, in a specific household, h. The bar (e.g., $\overline{k}$) accent is used to show a sample average. In all cases, this is employed at the household level (e.g. $\overline{k}_h$), thereby referring to the mean result of a household in the sample.

Model checks were conducted before and after estimating the cross-section model. Pre-estimation checking involved testing the normality of the only continuous variable included in the model (mathematics capability). Results from kurtosis and skewness tests and a comparison of mean and median suggested distribution to be normal. After estimating the cross-section model described above, multicollinearity amongst all residuals was investigated. Results from this investigation were also satisfactory. Consideration of VIF showed that values for each regression coefficient fell within an adopted cut off point.[65] Lastly, to assess the veracity of results, findings from this household fixed effects model were considered against those from a design featuring an alternative formulation of the dependent variable (following Delprato & Sabates, 2015). In the alternate model, the dependent variable was a binary measure of 'success' in the main Uwezo mathematics test. Results from this model did not differ substantively from those in the model presented.

3.3.1.4.3   Analysis of the Panel set

The final model employed to investigate the primary research question was a multilevel model applied to the 'Panel set'. As with analysis of the 'Matched set' and 'Household set', the model drew on ability estimates created through IRT. However, the analysis was otherwise different in multiple respects. It was not cross-sectional, but instead examined the differential effect of varying treatment dosage (i.e., mean district reported viewership in 2017) on district mathematics capability. This method was *prima facie* superior to the cross-section approaches employed in other models used to consider the primary research question (see Section 3.3.1.4.1 and 3.3.1.4.2). This was (in part) because the multilevel model controlled for time invariant unobserved characteristics at the unit of

---

[65] Results for all coefficients were all below the adopted cut off value of five. By coefficient, VIF statistics were: 1.005 (binary exposure), 1.463 (Kiswahili attainment), 1.417 (age), 1.041 (enrolment status), 1.005 (sex).

analysis employed (i.e., district).[66] However, there were also numerous limitations particular to the multilevel model employed.

The characteristics of the data collected by Uwezo for 2015, 2013 and 2011 prevented the application of a design that could consider change amongst individual children. Since children were anonymised in all publicly available Uwezo data, matching of individual children between multiple timepoints could not be performed. Therefore, the unit of analysis was at the district level (for which Uwezo assessed a representative sample at each data collection). This resulted in a limited sample size (45 districts measured at two time points) and meant that variables included in the multilevel model were merely district averages. Another issue with the multilevel model was that the outcome measure was not average district capability, but district capability ranking. Due to the identified limitations of this approach, results from this model were not believed to provide a sufficiently accurate estimation of effect for employment in CEA calculations (see Section 3.3.2). This section now provides information on the creation of capability estimates and collation of district level averages (the third and fourth steps taken to form the dataset for analysis, initially mentioned in Section 3.3.1.1.1).

Panel set: IRT models

IRT was employed to estimate relative mathematics capability within each round of Uwezo data, before child capability estimates were aggregated at the district level to create a year-specific district capability ranking. The initial application of IRT models to each year gave relative capability scores that permitted a more accurate ranking of districts for a measure of mathematics than using 'raw' attainment results. Mathematics capability was modelled using an IRT approach analogous to that employed with regards to the 'Matched set' and 'Household set'. The mathematics items available for Uwezo 2017, 2015, 2013 and 2011 meant that both binary and polytomous items featured in each year. The presence of binary and polytomous items within each individual year item influenced the selection of a GPCM model for the estimation of capability. In all cases, ANOVA was used to confirm the (relative) applicability of a GPCM model against the only other viable approach: a Rasch model. (For further information on the GPCM approach, see Section 3.3.1.4.1.)

Panel set: Consideration of IRT assumptions

Given that the GPCM models were applied to different datasets, the assumptions of IRT were investigated once again. Models applied to the 2017, 2015, 2013 and 2011 datasets satisfied tests of local independence and monotonicity. These IRT assumptions were again assessed using Yen's Q3

---

[66] Cross-section designs also controlled for district or household heterogeneity through the application of fixed effects. However, in both instances, there were still unobserved differences at the unit of analysis (i.e., individual children) that might bias results.

and the mokken check.monotonicity function respectively. ICC plots provided further support that the assumption of monotonicity was satisfied in all cases.

The assumption of unidimensionality was also considered through employment of CFA. However, problems relating to the limited number of mathematics items were encountered when implementing CFA (via the cfa function in lavaan). Uwezo data for 2013 and 2011 included only two mathematics items. CFA therefore required factor loadings to be held constant between all mathematics items, as the model otherwise possessed too few degrees of freedom to be calculated. Further, calculating factor loadings for the 2017 and 2015 datasets required factor loading values to be held constant between (at least two) mathematics items. When failing to hold certain loadings constant in these calculations a Heywood case error was encountered (specifically, negative variances amongst the manifest variables comprising mathematics capability estimates). Results from calculations where loadings were held constant were satisfactory for data from all years.

Panel set: Test and item information

Mathematics items for each year of data provided maximal information on test takers with theta levels between -0.5 and 0. This identification corresponded with findings from above (regarding the 'Household set' and 'Matched set'). The shape of ICC plots for 2017, 2015, 2013 and 2011 were also clustered around low to average theta values (θ = -0.5 to 0). However, clustering was less pronounced in pre-2017 datasets (and particularly in 2013 and 2011). This was related to variations in item (and item category) difficulty and discrimination, which were themselves partially attributable to Uwezo's employment of varying test formats in different years. Variations in format included there being a greater number of 'main mathematics test' categories for 2015, 2013 and 2011 (albeit with one less 'everyday mathematics' item for 2013 and 2011). The reduction of categories in 2017 resulted from Uwezo's decision to remove a multiplication question, which provided the 'hardest' category in the main mathematics test in 2015, 2013 and 2011. Removal of this test category in 2017 meant that less information was provided on test takers with higher levels of capability. ICC plots for 2017, 2015, 2013 and 2011 for all mathematics items are provided below (see Figures 3.17, 3.18, 3.19 and 3.20).
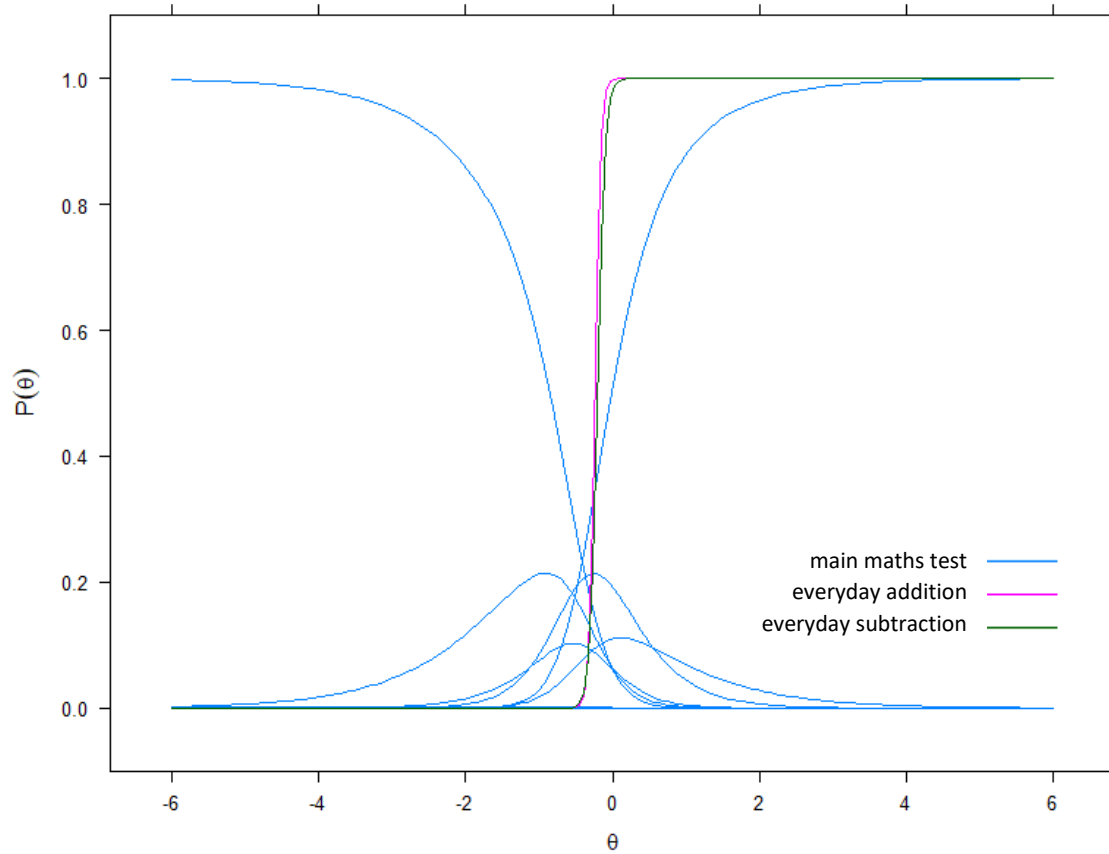
Figure 3. 17: Trace lines for all 2017 mathematics items

The graph above shows the probability of a child in the Uwezo 2017 dataset with any capability (theta) value attaining any given level for the main mathematics test (in blue). The graph also shows the probability of a child with any given theta value answering the "everyday addition" (in purple) and "everyday subtraction" (in green) questions correctly.
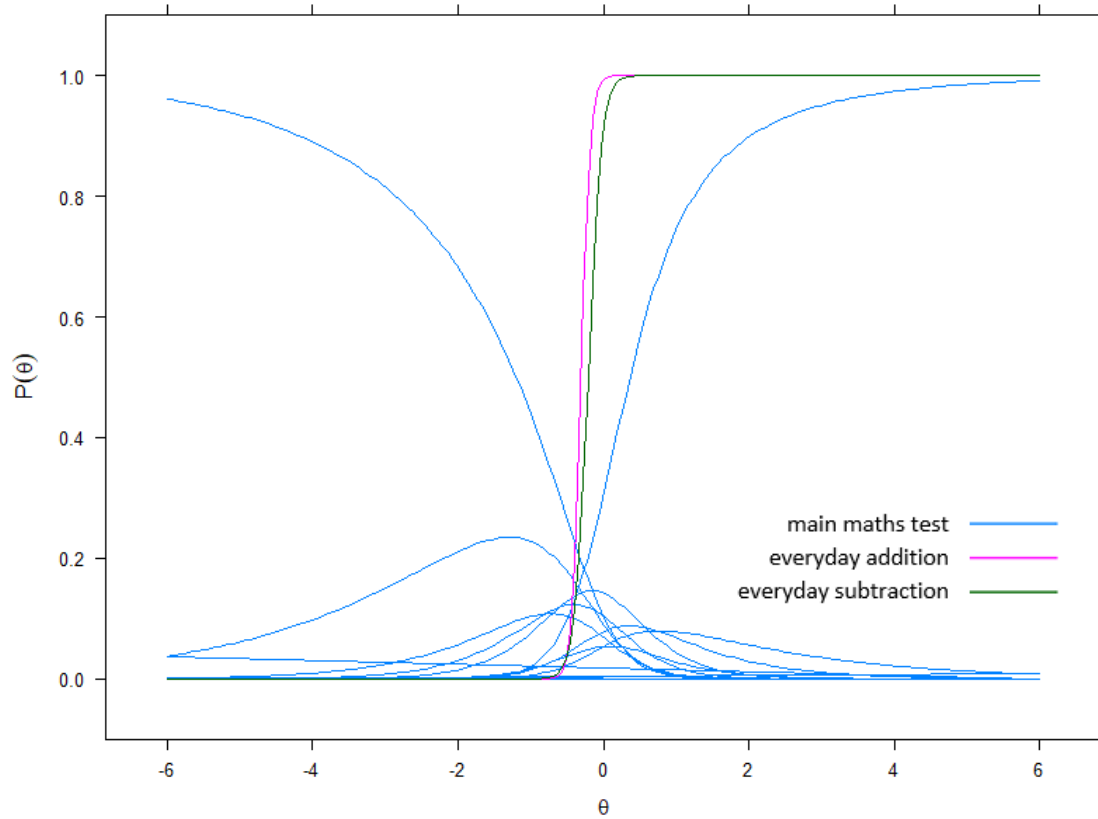
Figure 3. 18: Trace lines for all 2015 mathematics items

This graph shows the probability of a child in the Uwezo 2015 dataset with any capability (theta) value attaining any given level for the main mathematics test (in blue). The graph also shows the probability of a child with any given theta value answering the "everyday addition" (in purple) and "everyday subtraction" (in green) questions correctly.
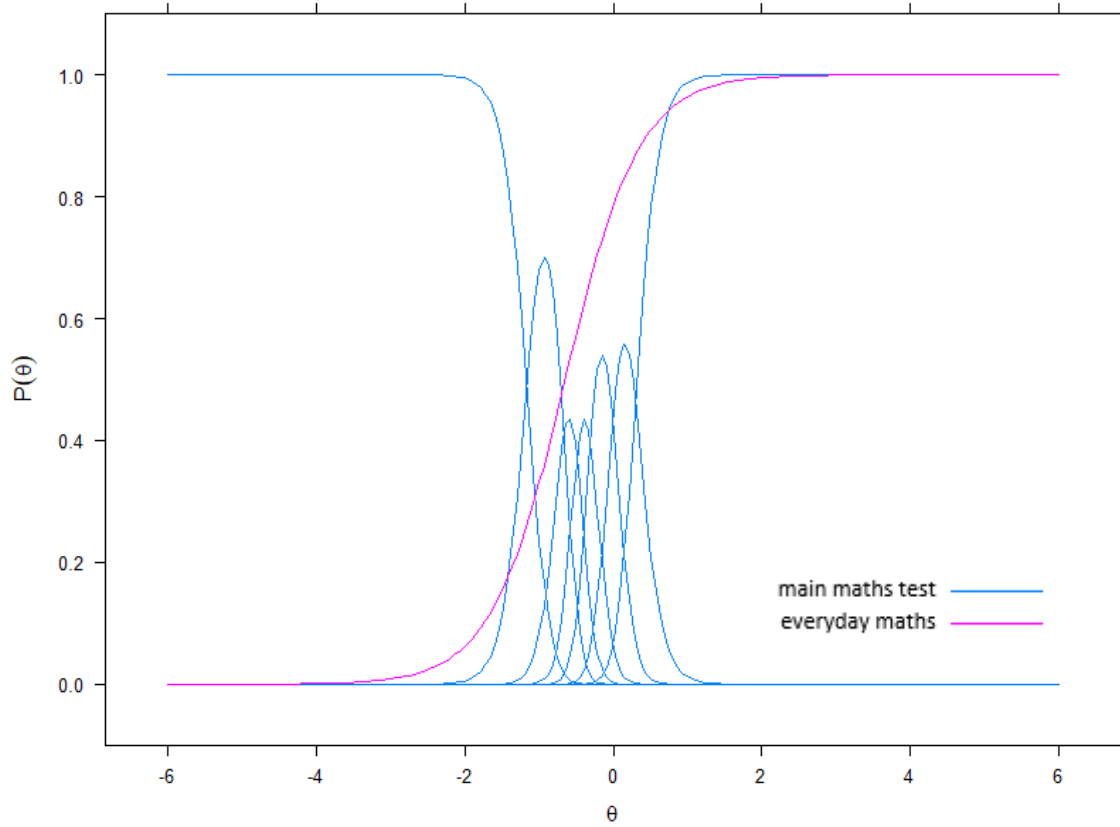
Figure 3. 19: Trace lines for all 2013 mathematics items

This graph shows the probability of a child in the Uwezo 2013 dataset with any capability (theta) value attaining any given level for the main mathematics test (in blue). The graph also shows the probability of a child with any given theta value answering the binary "everyday maths" question (in purple) correctly.
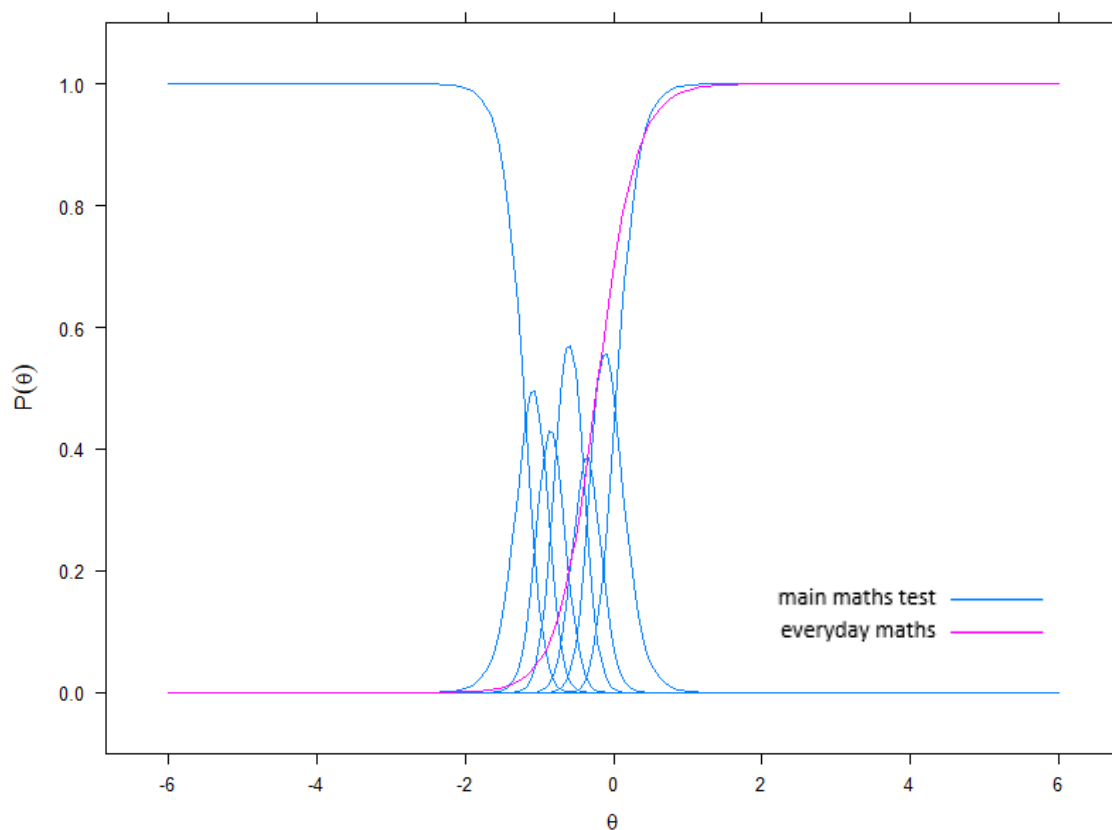
Figure 3. 20: Trace lines for all 2011 mathematics items

This graph shows the probability of a child in the Uwezo 2011 dataset with any given capability (theta) value attaining any level in the main mathematics test (in blue). Additionally, this graph shows the probability of a child answering a binary "everyday maths" question (in purple) correctly.

Panel set: Theta estimation

Theta values for mathematics capability were predicted for each child within each year of data. From the GPCM models calculated, single theta values for each child were created for all years using EAP. Theta estimates were stored as a new variable for each row (child) in Uwezo datasets for 2017, 2015, 2013 and 2011. Capability predictions and other variables of interest were then aggregated at the district level in all years (see below).

Panel set: Aggregation of district-level averages and capability ranking

For each year of data, a new (temporary) dataframe was created to hold district-level averages. These averages were calculated for all variables subsequently included in the multilevel model: mathematics capability, child age, mother's level of education, year of data collection and self-reported exposure (for 2017 only).[67] The temporary datasets for 2017, 2015, 2013 and 2011 were then appended to

---

[67] Mean average was calculated by district for each variable, ignoring any missing values.

create a dataset in long-format – a dataset structure facilitative of longitudinal analysis. In accordance with this format, data for each of the 45 districts included were contained in four rows (corresponding with the four years of Uwezo data employed).

The final stages of data wrangling required to produce the 'Panel set' employed in analysis involved creating exposure tertiles and year-specific capability rankings. Exposure tertiles were created to facilitate examination of the parallel trends assumption (Figure 3.23) and the provision of descriptive information on change over time amongst groups with low, medium or high exposure (see Section 7.2.1). Tertiles were established by placing the fifteen districts with the highest mean level of reported exposure (in 2017) in the upper tertile, the next highest fifteen in the middle tertile and the lowest fifteen in the bottom tertile. While tertiles benefited descriptive analysis and the examination of longitudinal assumptions, they were not used in the final model. This was because the model itself benefited from using the most detailed measure of exposure available – mean district exposure (which varied particularly within the upper tertile: Figure 3.21).
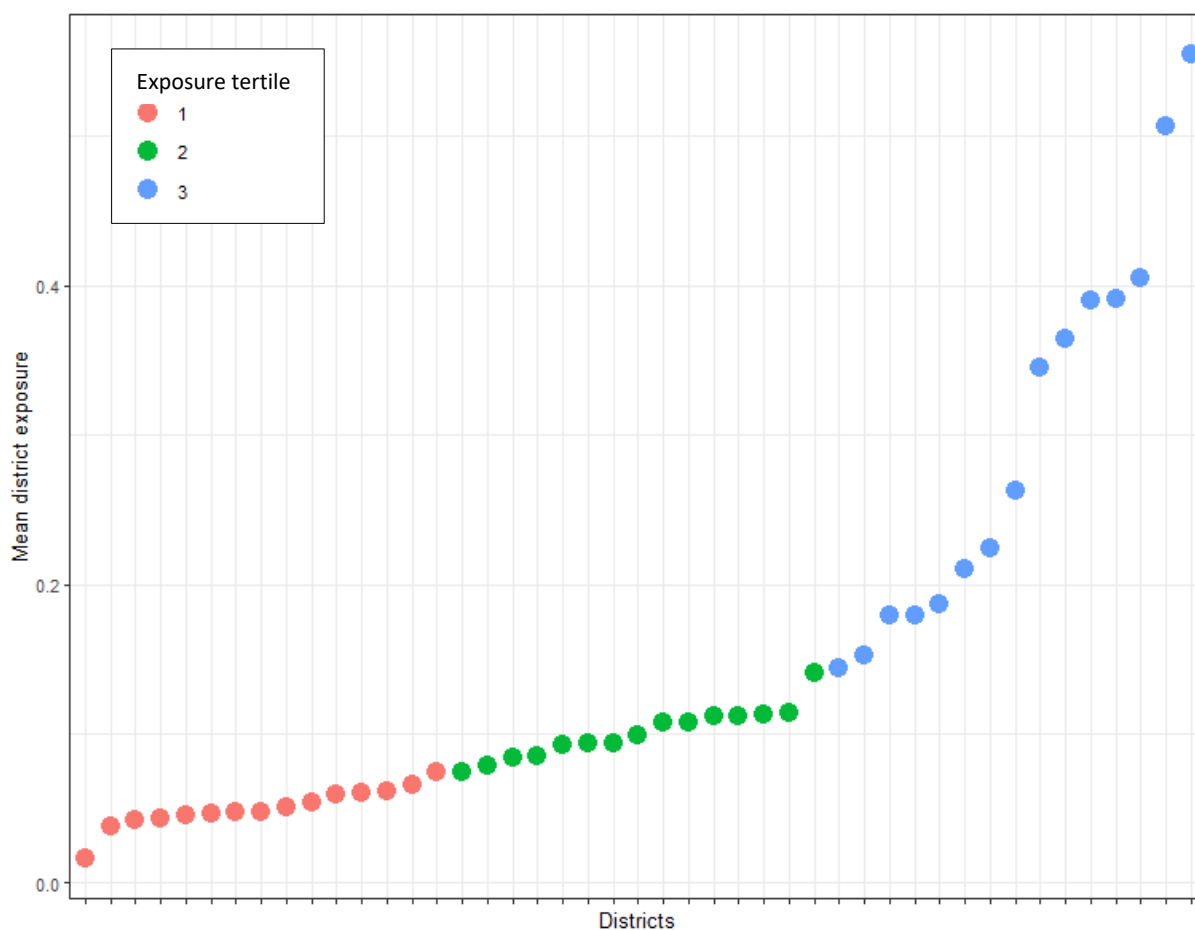
Figure 3. 21: Mean district exposure and exposure tertile

Lastly, capability rankings for each year of data were created by placing each district in a position (1 to 45) depending on their mean mathematics capability value. It initially appeared that it was more appropriate to treat capability in a similar manner to exposure, since district capability means could also be employed directly in this multi-time point study (see, for example, Carlson, et al., 2008, where non-aggregated IRT scores were used).[68] This was because considering a change in ranking instead of a change in mean capability could reduce precision. For example, a large increase (from time point 0 to time point 1) in capability amongst children in a district whose rank did not change (e.g., rank 1 at time point 0 and at time point 1) would not influence results. However, less information relevant to high theta levels was captured in the Uwezo 2017 assessment relative to the 2013 assessment. Retaining 'raw' district mean capability scores could therefore have biased analysis, as variation amongst districts with higher theta values was diminished as a result of changes in test format. This point is supported by a plot of mean theta values by district for 2013 against 2017 (see Figure 3.22).[69] For this reason, IRT rankings were employed (as used in previous analysis: Pensiero, & Green, 2018, who examined change in IRT score ranking quintile).
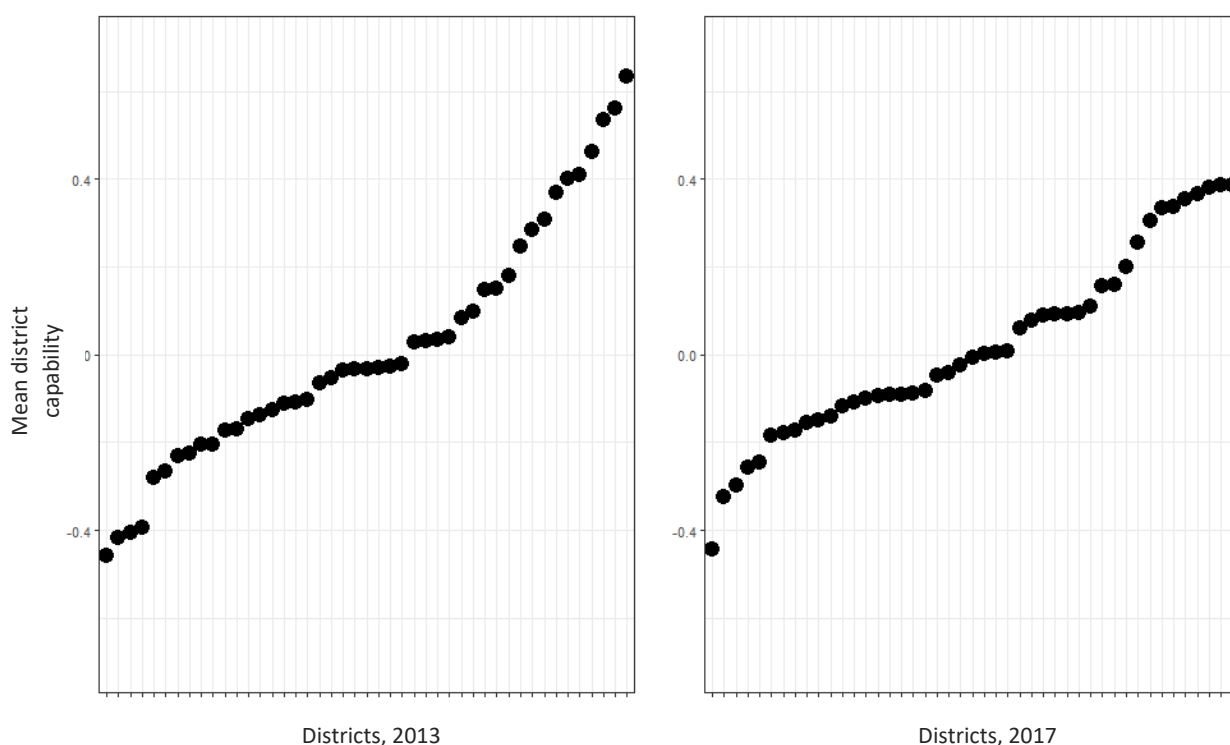
Figure 3. 22: Mean capability scores by district for 2013 and 2017

---

[68] Indeed, a comparison of IRT scores over time would be superior to a comparison of total score in Uwezo assessments over time. The is because the latter would be adversely affected by changes in Uwezo questions and test format by year.

[69] Figure 3.22 also provided support for using district capability rankings as opposed to capability groups (such as ranking quintiles, as used in Pensiero, & Green, 2018). This was because the pattern of mean capability scores by district appeared largely linear.

Panel set: Model creation

A multilevel model was created for analysis of the 'Panel set'. This model used data from 2013 and 2017 only. These years were selected as they included the only point where an exposure estimate was available (2017) and the point immediately prior to the launch of Ubongo Kids (2013). Data for all time points in the 'Panel set' (2017, 2015, 2013 and 2011) were used solely in descriptive tertile trajectory plotting (see Figure 3.23).

In the multilevel model, mathematics capability ranking (discussed above, 'Aggregation of district-level averages and capability ranking') was employed as the dependent variable. District mathematics ranking was regressed on an interaction between time and mean district exposure for 2017. In addition, there were independent variables for time and exposure which were considered individually (i.e., not interacted). Further, additional independent variables were employed to control for time variant group differences. These were mean district measures of age, sex, school enrolment status and mother's education. The inclusion of these variables in the multilevel model followed previous literature (see Section 3.3.1.3), and was supported by comparison with a pared down model in which they were omitted.[70] The model employed featured random group (i.e., district) intercepts only, with all independent variables discussed applied as fixed effects. This multilevel model operated on the same logic as a difference in difference (DiD) model, as it enabled the change in exposure over time and its relationship with the dependent variable to be identified. It was, however, preferred to a DiD model as it provided a truer representation of standard errors (even relative to DiD models with clustered standard errors: Cheah, 2009).

This multilevel model is specified as follows:

$$y_t = \beta_0 + \beta_1 T + \beta_2 UbRep + \beta_3 T * UbRep + \beta_4 Age_t + \beta_5 Sex_t + \beta_6 Enr_t + \beta_7 M_t + \varepsilon$$

Here, the dependent variable, y, is district capability ranking. Subscript t refers to the time point (0 in 2013 or 1 in 2017). T also refers to the timepoint (0 or 1), when included as a dummy variable. UbRep represents mean district self-reported Ubongo Kids exposure. T*UbRep shows an interaction between the time dummy, T, and self-reported exposure, UbRep. Age, Sex, Enr and M refer to the mean district child age, child sex, enrolment status and child's mother's previous school attendance, respectively (at time point t).

---

[70] The model featuring independent variables for age, school enrolment and mother's education possessed a lower AICc, higher R2 and significantly better chi-squared value than a model in which these additional independent variables were excluded. Asset wealth was not included in any model as consistent information was not available for the 2013 and 2017 datasets.

Panel set: Model checks and consideration of longitudinal assumptions

Various model checks were conducted. As for the cross-sectional models created to investigate the primary research question, these checks included an investigation into variable normality and multicollinearity.[71] All variables included in the model that contained aggregated district averages appeared to be normally distributed. This was suggested by kurtosis and skewness results which fell within acceptable boundaries (-2 to 2), and by consistently similar mean and median values for each variable.

VIF statistics for each variable included in the final model were all below the adopted cut off point (5), except for time (6.93) and age (5.12). This suggested that the variables for time and age could be predicted from other independent variables with an unsatisfactorily high degree of accuracy.[72] These results were not, however, deemed to suggest the inapplicability of a longitudinal approach, as they were likely inflated by the limited sample size (45 districts measured at two time points).

Assumptions relevant to longitudinal analysis that did not pertain to cross-section approaches were also considered. These included the parallel trends assumption, which is satisfied if trajectories between groups are constant prior to the intervention (see, for example, Delaney & Kearney, 2015). This was explored by employing data from prior to the launch of Ubongo Kids (from Uwezo's 2011 data collection) and from a mid-intervention time point (2015), to produce a descriptive plot of exposure tertile trajectory over multiple time points. This provided tentative evidence that this assumption was satisfied (see Figure 3.23). However, it also pointed to the potential violation of another assumption: that intervention allocation (i.e. the decision of children in a district to view Ubongo Kids) was unrelated to the baseline dependent variable results. This was suggested by the group with the highest exposure (measured in 2017) having a substantially higher mean capability ranking in 2013.[73]

---

[71] These were supported by plots of model residuals and random effects, which indicated that distributional assumptions of residuals were satisfied (see Appendix 7).

[72] This finding was perhaps least surprising for time, which was interacted with level of treatment to form another variable.

[73] This violation was not, however, confirmed, given the potential influence of other independent variables related to the dependent variable that varied between exposure groups (e.g., mean district school enrolment status).
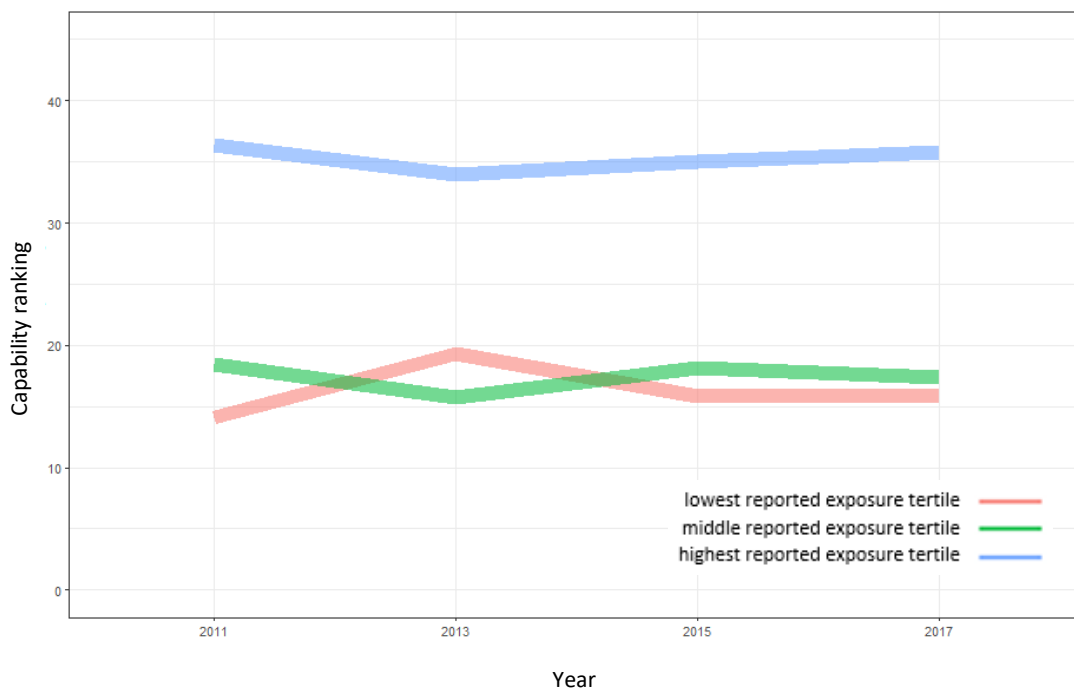
Figure 3. 23: Exposure tertile mean capability ranking over time

3.3.2 Quantitative methods employed for analysis of the subsidiary research question

This section now provides information on the quantitative methods employed to answer the subsidiary research question, which concerns the cost-effectiveness of educational television in developing country contexts. This question was also explored exclusively with regards to Ubongo Kids in Tanzania, and therefore the results provide limited information on the cost-effectiveness of educational television as a whole. However, limiting the scope of investigation to Ubongo Kids in Tanzania permitted a viable means of investigating the subsidiary question since the key measures required for cost-effectiveness calculations were readily available. Information on 'effect' was provided by results from the primary research question and cost information was provided by Ubongo. Further, the number of programme beneficiaries (in Tanzania) could be estimated using self-reported viewership figures from the Uwezo 2017 dataset along with national population estimates.

For the purposes of cost-effectiveness analysis (CEA), this thesis followed the approach advocated by the Abdul Latif Jameel Poverty Action Lab (J-PAL, 2014) (as outlined in Dhaliwal, Duflo, Glennerster & Tulloch, 2012). This ensured comparability with CEA results for numerous other interventions in multiple developing countries (each of which was discussed in the literature review, Section 2.3.1.2). In accordance with the J-PAL (2014) approach, CEA was estimated by calculating the standard deviation gain in attainment per $100 spent on educational television:

Change in standard deviations per $100 spent = 100/(Cost/Influence*Beneficiaries)

In this equation, Beneficiaries refers to the number of children that have received the intervention, Influence denotes the standard deviation association between educational television and mathematics capability for an individual child, and Cost is the total cost of Ubongo Kids. This equation was used to produce two separate CEA estimates, which concerned Ubongo Kids' ongoing operations and Ubongo Kids' activities since inception. The production of these two estimates involved using different influence and cost values.

This chapter now provides a preliminary consideration of influence, cost, the number of beneficiaries and programme duration (which is relevant to the calculation of cost and influence), before each of these concepts is more fully explored when generating cost-effectiveness results (see Sections 8.4-6). CEA results are sensitive to the manners in which cost, influence, the number of beneficiaries and programme duration are calculated (with cost and influence themselves influenced by discount rate calculations, explored in Section 8.7.1).

Programme duration

To calculate CEA results relevant to both the ongoing operations and activities since inception of Ubongo Kids, two programme duration estimations were made. When considering Ubongo Kids activities since inception, the selected duration period was 2013 to 2017. This period covered the point at which Ubongo Kids production began in earnest (July 2013), through to the Uwezo data collection from which influence findings were produced (December 2017). Using this period was justified because it covered all potential times at which a child could have been exposed to the programme. This period was, however, likely produce a result that was strongly downwards biased due in part to difficulties in identifying all viewers who had benefitted from the show (prior to the 2017 Uwezo data collection: see below).

The duration period selected to consider Ubongo's ongoing operations was 2017 only. CEA results concerning Ubongo's ongoing activities are likely to be more accurate, because calculations are now susceptible to bias in both directions. The CEA result would retain some downwards bias relating to difficulties in estimating the number of beneficiaries, since numerous children who did not report viewing Ubongo Kids in the week prior to the Uwezo data collection would have received some exposure. However, results for this duration period could also have been biased upwards, because some reported viewers would have watched the show before 2017. That is, children would have benefited from Ubongo Kids in years prior to the assumed base year. In such cases, a more accurate figure could have been produced by discounting influence back to an earlier base year (in which children were first exposed). Further, the costs incurred in delivering child benefit prior to 2017 would not have been accounted for in cost-effectiveness calculations, leading to an underestimation of costs. The result of an underestimation of costs and overestimation of influence would bias CEA results

upwards. This would counter the downwards bias imposed by the likelihood that various children who did not report exposure in 2017 had received some benefit from the show.

Number of beneficiaries

The next concept discussed is the number of Ubongo Kids beneficiaries. For CEA calculations, a proxy for the 'number of beneficiaries' was derived from an estimated percentage of Ubongo Kids viewership and the approximate number of children (6-16) in Tanzania. The number of children in Tanzania was identified using United Nations (UN) population estimates (https://population.un.org/wpp/Download/Standard/Interpolated/). Additionally, to establish the percentage of child viewers in Tanzania, Uwezo 2017 data on recent exposure was used in conjunction with sample weights. The former was available from responses to the binary item on self-reported viewing (see Section 3.3.1.3.2), while the latter followed sample weighting figures provided by Uwezo. The number of children and proportion of child viewers were then multiplied to approximate the number of beneficiaries.

However, estimates of the number of beneficiaries could have been inaccurate. This was particularly likely to be the case when considering Ubongo Kids' operations since its outset, because children could have viewed Ubongo Kids between the show's first broadcast (January 2014) and the Uwezo data collection (December 2017) without reporting exposure in the week prior to the 2017 Uwezo assessment.[74] This would have resulted in an underestimate of the number of beneficiaries, that would have biased CEA results downwards. That is, findings would have suggested a lower standard deviations gain per $100 than if the number of beneficiaries was established without error. Inaccuracies in gauging the number of beneficiaries in different periods are considered further in Section 8.4.

Cost

Programme duration was also a relevant consideration in establishing the costs to be included in the CEA formula presented above. However, unlike for the number of beneficiaries there was sufficient cost data to create separate estimates relevant to Ubongo Kids activities since its outset (2013-2017) and for ongoing operations (2017 only). All costs were captured using the 'basic costing template' provided by J-PAL (see https://www.povertyactionlab.org/research-resources/cost-effectiveness). This template was completed by Ubongo, in accordance with guidance that I provided that was based on the analysis of interview data. Interviewee responses, for example, suggested that the opportunity

---

[74] The possibility that this could occur is reinforced by subsequent descriptive analysis, where it is recognised that a high proportion of children that did not report recent viewership were able to identify Ubongo Kids characters (see Section 4.4.2).

cost of watching Ubongo Kids was zero. This was because children often watched Ubongo Kids at the expense of leisure activities (see Section 8.5).

Cost estimates were the product of numerous further assumptions. These included those underlying the J-PAL discount rate and the assumptions used to create cost estimates specific to Tanzania. The discount rate used by J-PAL was applied to cost estimates for Ubongo Kids activities since its outset (2013-2017).[75] This discount rate (10 percent) acted to translate costs back to their value at the programme base year (2013), when considering all Ubongo Kids' activities since inception. This rate was adopted by J-PAL following exploration into the discount rate in multiple developing countries (by Dhaliwal, Duflo, Glennerster & Tulloch, 2012). Additionally, assumptions had to be made to create cost estimates specific to Tanzania. This was because a proportion of Ubongo Kids expenditure went towards delivering the intervention across multiple countries. To estimate costs relevant only to Tanzania, total expenditure was multiplied by the proportion of viewers (by year) in Tanzania.

Influence

As with cost, two estimates of influence were created for each period of intervention selected. Both influence estimates were initially based on the coefficient for Ubongo Kids from the household-level fixed effects model. This model was chosen as the basis for the influence estimate because it provided the strongest indication of programme effect (due to its strengths outlined in Section 3.3.1.4.2). To investigate influence when considering Ubongo Kids' ongoing operations (2017 only), the coefficient for Ubongo Kids was multiplied by the standard deviation of the variable for mathematics capability. However, when producing an influence figure for Ubongo Kids' activities since inception (2013-2017), this figure had to be adjusted for the adopted discount rate (so that benefits were "discounted back to the base year of the programme", Dhaliwal, Duflo, Glennerster & Tuloch, 2012, p. 37).

Cost-effectiveness analysis and comparison

The estimates of influence, cost and number of beneficiaries described above were used to create two CEA results, relating to Ubongo Kids' operations since inception and to its ongoing activities. These results were then compared to other educational interventions for which CEA results were available (and which had been examined using the same approach, see Section 2.3.1.2). Comparison with other CEA results provided information on whether Ubongo Kids produced a greater or lesser increase in learning outcomes per $100 spent than other interventions in developing contexts. The findings produced could therefore inform policy makers seeking to achieve a more efficient allocation of educational resources.

---

[75] No discount rate is required when considering Ubongo Kids ongoing costs, as estimates concern costs incurred up to twelve months prior to the Uwezo data collection only.

3.4   Qualitative method: Interviews including Ubongo Kids viewing

Interviews provided information that primarily assisted in the interpretation of quantitative findings related to the primary and subsidiary research questions. To gather this information, interviews solicited responses from children and caregivers concerning their television viewing habits and preferences, as well as their perspectives on Ubongo Kids specifically. Regarding the primary research question, interviews informed and explained quantitative investigation into the association between educational television and mathematics capability. Interviews performed this function by giving further information on the possible effects of educational television exposure in general, as well as specific variables included in regression models (namely, exposure to Ubongo Kids and mathematics capability: Section 8.4). Furthering understanding of how quantitative measures functioned was crucial to interpreting regression outcomes. Indeed, interviews provided information on concepts that could not have been obtained through larger scale surveys. With regards to Ubongo Kids exposure, for example, researcher questions building upon interviewees' responses permitted investigation into longer instances of content recall (see Section 7.3.2).

While the outcomes of thematic analysis are provided in Chapter 8 after all quantitative results chapters concerning the primary research question, qualitative work also aided cost-effectiveness analysis (Chapter 9). Interviews aided the interpretation of cost-effectiveness findings by providing additional information on the benefits of educational television viewing. This highlighted a concept that was not accounted for in CEA calculations: consumption value.[76] Section 8.3.2 features child reports of being entertained by Ubongo Kids, with these reports suggesting that CEA estimates based on learning did not reflect all benefits received by viewers. Additionally, analysis of interview data informed assumptions made regarding the value of a user cost for which no quantitative information was available. This user cost was the opportunity cost of viewing (recognised as a programme cost under the J-PAL CEA framework, see Section 9.5). Interview responses led to the assumption that this cost should have a zero value, because children and their caregivers typically watched Ubongo Kids at the expense of leisure activities.

The justification for using interviews instead of other qualitative approaches was also considered (as recommended by Silverman, 2011). Interviews were considered more suitable than alternate means of gathering qualitative data, such as an immersion in the practice of educational television viewing. This was partly because interviews provided a less time-consuming means of gathering qualitative data, which was beneficial in a finite fieldwork period dominated by quantitative data collection. Additionally, more intensive periods of viewer observation were considered ill-suited to examining engagement, since bias attributable to the presence of the researchers could have been introduced

---

[76] Indeed, the consumption value of viewing Ubongo Kids might not even be considered "amenable to quantification", which provides justification for the use of qualitative data to support quantitative models (see Sabates-Wheeler, Sabates and Devereux, 2018, p. 1080).

(Cohen, Manion & Morrison, 2011).[77] That is, researcher presence might have altered viewer behaviour to the extent that it no longer reflected normal practice.

3.4.1 Interview sampling

The interview sample was selected from a backup quantitative dataset. This dataset had been collected as a failsafe to ensure that at least some primary quantitative data would have been obtained, should the planned larger scale data collection in Ilala and Temeke not have gone ahead. The decision to draw data from the backup sample was made because delays in Uwezo data collection failed to permit subsequent qualitative work within the fieldwork timeframe. The selection of individual families to participate in the qualitative data collection was made purposively, to include one family from each of the six villages participating in the backup sample, whose children's character recognition results suggested at least some degree of prior exposure to Ubongo Kids. Conducting interviews in households where children appeared to have had at least some exposure to Ubongo Kids was considered necessary to obtain informative responses to questions regarding the programme. This sample was not, however, restricted to children who reported Ubongo Kids viewership in the week before the Uwezo data collection. This characteristic of the sample permitted exploration into the meaning of positive and negative responses to this binary exposure item.

Households included in the backup data collection – from which the interview sample was drawn – fell broadly into three types of setting. Those in Somangira and Pugu were from sparsely populated villages in rural locations (see Figure 3.24a and Figure 3.24b, respectively); Hananasif, Mabibo and Makangarawe provided more compact groups of houses (see Figure 3.24c, Hananasif), whilst Grezani was decidedly urban and densely populated (see Figure 3.24d). Almost all dwellings identified within these locations were permanent brick or concrete structures (excluding very few informal mud-walled structures in Somangira), with those in Grezani predominantly comprising flats in high-rise buildings. Limited availability of household members subsequent to backup data collection, however, acted to limit the types of settings represented in interviews.

---

[77] It should be noted that a portion of interviews was devoted to watching Ubongo Kids with the child (see Section 3.4.2.1). However, the chief purpose of watching Ubongo Kids was to support further interview discussion.

Figure 3. 24: Interview settings

3.4.1.1   *Unintended limitations to sample*

Limited household member availability reduced the qualitative sample from six families to four families, with only Hananasif, Makangarawe, Pugu and Somangira represented in the final sample. In Grezani, the caregiver declined to participate in the interview due to the children's exam schedule. No suitable replacement family was identified amongst the Grezani sample – captured from an industrial area in which very few children reside. In Mabibo, there was apparently no child or caregiver availability during any of the three attempts to visit the selected household. On the third attempt, it also became apparent that the caregiver did not wish to participate in the qualitative component of data collection. No alternate family could be identified from Mabibo at this stage, due to overall fieldwork time restrictions. The eventual sample therefore comprised four families, from each of which one caregiver and all children aged 6-16 were interviewed, giving eleven interviewees in total. The table below gives an overview of this sample organised by district and including key quantitative characteristics of child interviewees from the backup data collection.

Table 3. 6: Interviewee characteristics

Name	Age	Child or Caregiver	District	Gender	Current grade	Watched UK last week	Watches TV where	Character recognition, Ubongo Kids	Character recognition, all media
Hamica	8	Child	Hananasif	M	2	1	family home	3	5
Naima	23	Caregiver	Hananasif	F	NA	NA	NA	NA	NA
Eric	9	Child	Makangarawe	M	3	1	family home	2	5
Zena	33	Caregiver	Makangarawe	F	NA	NA	NA	NA	NA
Collins	11	Child	Pugu	M	5	0	family home	3	7
Agatha	14	Child	Pugu	F	9	0	family home	3	6
Alan	46	Caregiver	Pugu	M	NA	NA	NA	NA	NA
Zainab	13	Child	Somangira	F	Stopped after 3	0	relatives' home	0	2
Ramadhani	10	Child	Somangira	M	1	0	family home	1	4
Ibrahim	13	Child	Somangira	M	4	0	family home	4	5
Amina	32	Caregiver	Somangira	F	NA	NA	NA	NA	NA

3.4.2 Interview data collection

3.4.2.1 *Interview format*

Child interviews amongst the qualitative sample were conducted with a research assistant (see Figure 3.25). Children were interviewed separately from their caregivers, as children are likely to be "the best sources of information about themselves" (Docherty & Sandelowski, 1999, p. 177, originally cited in Cohen, Manion & Morrison, 2011). Indeed, the attendance of adults important to the children's lives (such as their caregivers) could have "prevent[ed] children's full participation in the research" (Thomas & O'Kane, 1998, p. 337). Obtaining strong information from children was considered key, as 'Children's voices bring unique perspectives' that could act to both challenge and complement the views of other household members (Kuchah & Pinter, 2012, p. 284).

Another format decision concerned whether to conduct one-to-one child interviews or (child only) group interviews. One-to-one child interviews were preferred. While both format options have advantages, one-to-one interviews are considered to elicit a greater range of ideas (Heary & Hennessy, 2006) and might also reduce the chance that data are contaminated by the presence of others (not taking into account the influence exerted by the interviewer and translation assistant: Section 3.4.3). Perhaps most importantly, however, the characteristics of the interview sample meant that multiple children were included from only two families. As such, a group interview format could

not have been used in all cases. One-to-one interviews therefore provided greater consistency between child interviews, which broadly aligned with my philosophical perspective that objectivity should be maximised (Section 1.4).

The literature suggests that there can be difficulties in negotiating child privacy. These difficulties might arise as "adults do not consider children's need for private space for an interview" or even feel that "children do not [possess these rights to privacy] at all" (Mauthner, 1997, p. 18). Further, it is possible that caregivers might limit the interaction between child and researcher as a result of their distrust of an 'outsider' interviewer (Shah, 2004). Nevertheless, caregivers were in fact not present for extended periods during every child interview conducted (nor indeed was any other extended family member, as has been reported in Indian contexts: Militades, 2008).

The restricted presence of caregivers during child interviews could have resulted from multiple factors. These included emphasising the importance of independent child responses prior to the interview; reassuring the caregiver that they would have the opportunity to share their views in a separate interview; and, the slight geographical distance from child interviews to caregiver homes (in certain settings). This being said, a child's friends or siblings were present for some parts of interviews in multiple locations. It did not appear that this influenced child responses to interview questions, but this did have a noticeable effect on how children watched a sample episode of *Ubongo Kids* played during the interview (which fortuitously provided some insight into co-viewing practices).

It was also intended that caregivers were interviewed alone. As with child interviews, the intended format was generally achieved. The only others present during some sections of caregiver interviews were their children. This was not, however, considered to have affected the responses provided by caregivers.

Figure 3. 25: Interview assistants

Each interview was conducted in Kiswahili. The selection of this language was necessary to communicate with both children and caregivers, all of whom were more capable of expressing views in Kiswahili than English. The use of this language might also have provided some evidence to interviewees that I was employing a form of "culturally appropriate communication" (Eide & Allen, 2005, p. 4; originally cited in Liamputtong, 2010). However, as I cannot converse fluently in Kiswahili, there were challenges. I sought to circumvent the limitations imposed by my lack of language knowledge by preparing questions and even responses to anticipated interviewee answers. Further, in preparing for interviews with my research assistant, pilot and actual interviews were considered at length to consider points where my research assistant could ask further questions to encourage the interviewee to provide additional information about a specific topic.

All child and caregiver interviews were conducted near to the interviewee's home. This provided a comfortable environment for interviewees that was maintained throughout all interviews. In the less populated areas of Somangira and Pugu, it was decided that child interviews should occur a slight distance from the interviewee's main building (see, for example, Figure 3.24a). As noted previously, this arrangement might have acted to limit potential caregiver intervention in child interviews. All interview content obtained from each interviewee was recorded on a portable Dictaphone.[78]

3.4.2.2   Interview content

Like the interview format, interview content remained largely consistent. The content was based on predetermined questions and video stimuli. It could be argued that this was to the detriment of the data collected, as obtaining the most valuable information from each individual child or adult depends on using "a varied repertoire of verbal and non-verbal techniques" (Thomas & O'Kane, 1998, p. 342). Additionally, using pre-written questions that did not account for all possible interviewee responses might have led to the interviewee believing that their response was not being properly acknowledged in some instances. Should this have occurred, my research assistant and I might have been perceived to have lacked interest in the interviewee's perspective (Partington, 2001). Yet in spite of these drawbacks, a consistent approach was preferred. A consistent approach increased comparability between responses (in a manner consistent with my desire to promote objectivity). Further, using fixed content had practical benefits. This permitted the establishment of an approach that functioned in spite of my limited language and cultural knowledge (Shah, 2004) (see Section 3.4.2.1).

---

[78] Video recording was not employed, on the assumption that this would make the environment more uncomfortable for the interviewee (with even the Dictaphone itself being identified to provide some distraction if left facing upwards, due to the display of audio levels and recording time on its digital screen). Whilst the lack of video footage undoubtedly left some non-verbal aspects of the interview uncaptured, limited notes of key events that would not be captured by audio recording were taken.

Pre-established interview questions and associated probes were generally open-ended in nature.[79] The employment of pre-written and unplanned probes and questions therefore created a semi-structured format. Open-ended questions catered well for the differing cognitive and linguistic abilities of the child sample, which ranged from 8 to 14 years. Further, the employment of open-ended questioning helped elicit responses with greater depth. In particular, it was considered a great benefit that open-ended questions and probes allowed interviewees to provide *specific examples* on topics, in both their initial and follow-up answers. In doing so, claims of learning or viewership could become more compelling. In Section 8.4.1, for example, reports of recent viewership that were supported by details of an episode storyline, were considered more likely to be accurate (than reports not supported by an example).[80]

Open-ended probing was also employed to promote the capture of perspectives that suggested something 'negative' regarding Ubongo Kids. For example, children who reported that they did not learn from Ubongo Kids were encouraged to say *why* this was the case (see Section 8.3.2). Indeed, if my interview assistant and I merely sought confirmation of the study's hypotheses – that educational television is both positively associated with mathematics capability and a cost-effective intervention – then knowledge would not have greatly been advanced (Partington, 2001). Responses featuring examples (from both initial answers and probes) and 'negative' statements were also beneficial to analysis, as they helped to counter the bias potentially attributable to Titus, Mussa and myself (see Section 3.4.3).

A restricted number of draft questions were written prior to fieldwork. The number of established questions was limited to reduce interviewee fatigue.[81] These questions were influenced by research aims as well as prior literature concerning educational television. Caregiver question 4 employed locution similar to a scale item used by Beyens, Eggermont & Nathanson (2016) in evaluating caregivers' attitudes towards television. Additionally, caregiver question 2 used a timeline form viewed by the caregiver before they provided verbal feedback, which was derived from the same source (ibid).

Video stimuli were employed in each child interview. This involved showing children an *Ubongo Kids* mini-episode, on which subsequent questions were focused (after a set of initial questions had

---

[79] Closed questions were employed, with their presence required to provide categorical information at points. Those included generally corresponded with categorical and closed items employed in previous studies concerning educational TV: Rimal, Figueroa & Storey, 2013 – for caregiver interview question 4 and child interview questions 1 and 2; and, Borzekowski & Henry, 2011 – for caregiver interview questions 1, 2 and 3.

[80] Reports of learning also became more informative. For example, in Chapter 8, learning claims featuring educational content embedded within programme narrative were even be employed to support the idea that Ubongo Kids *should* lead to improvements in mathematics capability.

[81] Child interviews took an average of 52 minutes, with a maximum of 66 minutes and a range of 33 minutes. Caregiver interviews took an average of 37 minutes, with a maximum of 60 minutes and a range of 36 minutes. (All interview lengths provided in this footnote are given to the nearest minute.)

already been asked). Showing an Ubongo Kids mini-episode had multiple benefits. This was expected to help maintain child engagement throughout the interview, which could otherwise have waned after initial questioning. Additionally, showing the episode could have facilitated discussion by making interviewees more comfortable, given that children "are often more comfortable when pictures, toys, or other props are used during a discussion with an adult" (Brenner, 2006, p. 365). Lastly, playing a mini-episode to children provided the opportunity to capture information on something I could be certain that they had watched (as opposed to something they reported viewership of). This may have acted to increase the credibility of responses and counter children's limited powers of recall (in relation to adult interviewees: Arksey & Knight, 1999). As noted previously, this component of the interview also provided some insight into viewing or co-viewing practices, when children were joined by their friends or siblings. A similar idea was employed by Jellison & Wolfe (1999) to consider an educational song, and could be considered an extension of the guidance provided by Cohen, Manion & Morrison (2011, p. 433) that interviewers might "show a picture or set of pictures, then ask children for their responses".

Figure 3. 26: The use of video stimuli during child interviews

3.4.2.2.1 Piloting of content

Piloting was conducted to establish honed translations of interview questions and probes, identify suitable Ubongo Kids mini-episodes to show child interviewees and improve interview technique. The first stage of piloting focused on translation only. Initial translations were created through a cyclical process of in-person translation discussions with staff members at the Ubongo office, attempts to ask translated questions to other staff members, then using findings to inform further revisions. Having established translations that were broadly understood, individual questions were asked to children and adults living near to the Ubongo office. This led to further revisions in multiple cases. Individual translated questions were then combined and put alongside various mini-episodes (for children's interviews only), to conduct complete pilot interviews. Throughout these interviews, translated

questions were continually refined.[82] Additionally, complete pilot interviews provided guidance on interview procedure and the employment of video stimuli.

Child pilot interviews suggested that two mini episodes should be used. This number of episodes provided some degree of consistency between interviews, whilst ensuring that siblings from sampled families were not continually shown the same episode (if children chose to watch a show played during another child's interview). Experience from conducting complete pilot interviews also influenced interview technique. For example, piloting showed that I should take a written copy of questions and probes in Kiswahili to all actual interviews (to assist with language during the interview itself). Additionally, piloting helped to promote a good understanding between the interview assistant and myself, which allowed us to progress smoothly through interview topics. Further, by listening to pilot interview recordings with my translation assistant, we worked on identifying areas where interviewee responses could be examined further by the translation assistant through additional probes (in instances where I lacked understanding of the interviewee's response).[83]

3.4.2.3 Interview Transcription and translation

While transcription and translation of interviews might be considered merely procedural, some information on this process is provided in this section. These tasks were both hugely time consuming and were potentially conducted inefficiently. Transcription first involved a document being produced in Dar es Salaam (in Kiswahili), by Mussa, Titus or Timoth (a Ubongo intern) from a recording of the interview shared on Google Drive. I then reviewed this transcription alongside the recording, before suggesting edits. Edits were then made before I re-checked the document. This process could certainly have benefited from the employment of speech recognition software to create a first draft, although available programmes could not interpret Kiswahili satisfactorily.

The re-checked transcriptions were then translated. This process was initially approached in the 'conventional' manner, with Mussa adding an English translation to the Kiswahili transcript (on a Google Docs file). The production of an initial draft was time-intensive, taking approximately 30 hours per 1 hour of transcribed text. After this, I would check through the translated document. This check focused on the English and used Google Translate to re-translate (and flag) particular sections, taking approximately 6 hours per 1 hour of transcribed text. After this, flagged sections were reconsidered by Mussa before we agreed on a final document. After completing three documents, I sought more efficient means of translation. The method employed from then onwards is outlined below:

---

[82] The questions and translations ultimately adopted are presented in Appendices 8 and 9.
[83] Adding 'reactionary' probes (that were not pre-written) during the interview itself clearly conflicted slightly with my desire to adopt a consistent approach. However, doing so was considered important to achieving expansion on particularly interesting topics raised by the interviewee.

- I performed an initial translation of the interview transcription on the Google Docs file using Google's inbuilt translation features (saving this as a duplicate document)
- I copied both English and Kiswahili versions of the file to Excel, using red text for English (see Figure 3.27a)
- I did initial work to improve translations and flagged difficult passages
- At this stage, the file was again shared in a Docs file (copying the entire table) before edits by Timoth (a Ubongo intern, with Mussa conducting data collection outside of Dar es Salaam at this point)
- Following Timoth's edits, I re-checked the document and further alterations were made where necessary, leading to an English and Kiswahili document (Figure 3.27b), from which I could extract all Kiswahili text using the filtering function on Excel, leaving an English only document in a format suitable for import to NVivo (Figure 3.27c)

Figure 3. 27: Interview transcription and translation process

3.4.3   Thematic Analysis

Qualitative data was examined using thematic analysis. Thematic analysis "is a method for identifying and interpreting patterns of meaning across qualitative data" (Braun & Clarke, 2014, p. 6626). Following Braun & Clarke (2006), thematic analysis was applied in a recursive six-phase process. This involved

1. Generating familiarity with the data

2. Creating initial codes

3. Searching for themes

4. Reviewing themes

5. Naming and refining themes

6. Producing findings

These phases are described below, clustering stages two and three as well as stages four and five. This clustering reflects the recursive nature of thematic analysis. Further evidence of this recursiveness is provided by the contents of subsequent descriptions. For example, the application of initial codes and themes involved using codes that were increasingly tailored in accordance with the data.

Generating familiarity with the data

Thematic analysis began by establishing strong familiarisation with the qualitative data. Knowledge of interview data was initially formed through the transcription and translation process. As recognised above (Section 3.4.2.2), transcription and translation were hugely time-consuming endeavours that involved many hours of listening to, writing and reading material in both Kiswahili and English. After the completion of translation and transcription, written interview transcripts were re-examined before tentative notes were made on features of the data relevant to the research questions and theoretical approach (in accordance with guidance from Braun & Clarke, 2014). In line with the fact that thematic analysis is a recursive process, familiarity was strengthened through all subsequent stages of qualitative investigation.

Creating initial codes and searching for themes

Gaining familiarity was followed by applying initial codes to the interview transcripts. These codes were labels applied to the text data that acted to index or categorise sections of the dialogue (Gibbs, 2007). Coded data were then assigned to broader themes. These themes represented broader meaning

patterns across all interview data (Braun & Clarke, 2006). Codes and the themes they comprised prioritised the research questions (Section 2.5) and pre-selected theoretical lens for this study – human capital (Section 1.3.1).[84] As such, initial codes and themes were generally established in a deductive manner (Patton, 2002; Braun & Clarke, 2006).

Still, it should be recognised that the creation of themes and codes was inductive in parts. Inductive analysis "involves discovering patterns, themes and categories in one's data" (Patton, 2002, p. 453). Themes centred on the adopted theoretical framework or a key facet of the primary research question, yet the codes underlying these themes and even theme names themselves were tailored in accordance with the data. This point suggests the occurrence of a cyclical deductive and inductive process (Teddlie & Tashakkori, 2009), which was continued by the application of the tailored code and theme structure to the data – another predominantly deductive step.

The form of thematic analysis used in this study was distinct from a Grounded Theory (GT) approach due to the initial application of an existing theory (Grbich, 2013), yet the inductive processes described above suggests some complicity with the GT characteristic of "theory construction" (or adaption: Charmaz and Bryant, 2011, p. 292). The occurrence of theory construction was suggested by the development of codes that extended beyond the preconceived HCT structure (see Section 1.3.1). These included unanticipated private costs of viewing media (other than Ubongo Kids). An example of a private cost falling outside the HCT structure was the negative impact of viewing material which could be inappropriate for children. Such material might include people wearing "short clothes" or fictional beings that could frighten children (citing a caregiver in Makangarawe district). Reports of such material were labelled under a new code, entitled, 'Costs of viewing UK cartoons and some, non-UK, inappropriate content for children'.

Reviewing, naming and refining themes

After working through several transcripts, codes and themes were reviewed together to establish final names and permit further refinement. This iterative procedure included combining some low frequency codes, given that applying large numbers of separate codes with low frequencies would not have been "helpful in finding patterns in the data" (Friese, Sorrato & Pires, 2018, p. 15). Content tagged under codes with similar definitions was also examined, with a view towards creating unifying codes. The production of progressively more focused codes and themes also involved re-categorisation. Re-categorisation was intended to improve the usability of the entire codebook, which

---

[84] Indeed, the data itself were influenced by the content of questions produced in accordance with the notion of HCT.

it was hoped would facilitate subsequent discussions involving analysis and inter-rater reliability (IRR) testing.

Qualitative analysis discussion and IRR testing began with sharing my codebook with Lisa, a Cambridge PhD student. We held ongoing discussions in which we considered the location of codes within the overall structure and possible means of improving code descriptions. With regards to code location, my coding partner considered codes related to co-viewing to be more suited to location under the top-level 'Exposure factors' code (alongside codes such as 'viewing environment') than in its original location within 'Private benefits of media viewing'. After consideration, I believed the suggestion to be beneficial and moved codes accordingly.[85] Advice on improving code description provided prior to and during IRR testing was similarly heeded. In accordance with this advice, descriptions related to character recall were supported by including a list of key *Ubongo Kids* and other programme characters. Further, this code description was clarified by noting the requirement for 'any spontaneous mention' of a character name. Other descriptions elucidated included 'Viewing environment', which was updated to include reference to the 'physical (viewing) space'. Lastly, descriptions for the codes 'Specific programme preference' and 'Specific programme viewed' were reconsidered to address potential overlap.

Lisa and I commenced IRR by applying the code structure to an example section of interview text during a Skype call. This suggested that the codebook functioned suitably, and we therefore progressed to verification of its application to one caregiver interview and one child interview. After my PhD colleague had coded these documents in accordance with my codebook, I estimated IRR through an association coefficient commonly applied to gauging agreement between multiple coders: Cohen's Kappa (Warrens, 2014). Kappa results were calculated using NVivo and exported to Excel to produce average values.[86] Results gave an average Kappa value of 0.8545 (not weighted to reflect differing child/caregiver interview length) or 0.8563 (weighted to reflect interview length) for both the caregiver and child interviews, for all codes applied to either document by either myself or my colleague.[87] Kappa results on average and for individual codes are provided in Appendix 10.

---

[85] Still, it remains the case that co-viewing, alongside other 'Exposure factors', will influence the level of benefit attained through media exposure.

[86] Average values were calculated in accordance with the advice provided by NVivo: http://help-nv11.qsrinternational.com/desktop/procedures/run_a_coding_comparison_query.htm#MiniTOCBookMark10

[87] This average is comprised of results for all codes featuring in either coder's coding of either document. Codes that did not feature in either coder's coding were therefore excluded (as in, for example, Hennessy, et al., 2016). However, codes referenced with a low frequency (one or more) were included. Whilst low frequency codes have been excluded from code averages in prior literature (ibid), the amount of codes with lower frequencies across the two documents used to consider inter-coder agreement suggested their inclusion was warranted. However, Cohen's Kappa results for individual low frequency codes were volatile (and their inclusion in Kappa averages that were unweighted, or weighted only according to interview length, could cause misleading results).

When considering results by individual code in the child and caregiver interviews, there was excellent agreement (a Kappa value of over 0.75)[88] in 73 instances (of 86 total instances), fair to good agreement (a Kappa value of 0.40 - 0.75) in 6 instances, and poor agreement (a Kappa value below 0.40) in 7 instances (covering only 6 of the 43 individual codes applied). This was the case for:

- Type of programme viewed or listened to (in both the child and caregiver interviews)

- Educational content, UK, socio-emotional, with specific example (in the caregiver interview only)

- UK or other media viewing alone (in the child interview only)

- Educational content, UK, non-specific (in the child interview only)

- Educational content, UK, motivation to learn (in the child interview only)

- Viewing or listening frequency, general, not UK specific (in the child interview only)

Each of these codes was examined individually to identify the root of poor inter-rater agreement. Following consideration and further discussion with my coding partner, the descriptions for four of these six codes were updated (see Appendix 11). Largely, these alterations provided additional information – such as giving specific examples of different *types* of TV programme or non-*Ubongo Kids* shows – and incorporating short hypothetical examples (to support the longer extract example that featured in the codebook). Further, a reconsideration of coding choices led to some revision of my own decisions, particularly in cases where updating the code description was not required. For the codes 'Educational content, UK, motivation to learn' and 'Educational content, UK, socio-emotional, with specific example', poor Kappa results were attributable solely to my coding errors (with each code applied once, in error, across both documents). The upper level codes and themes established upon completion of IRR testing are provided below (Figure 3.28).

---

[88] Interpretation of Kappa results is in accordance with the guidelines provided by NVivo (at: http://help-nv11.qsrinternational.com/desktop/procedures/run_a_coding_comparison_query.htm).
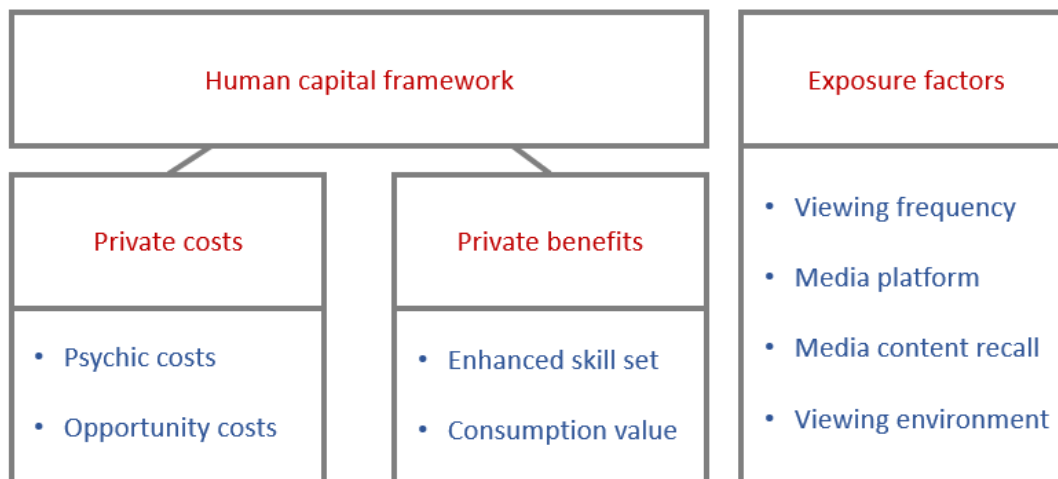
Figure 3. 28: Theme and code structure

Producing findings

Braun and Clarke (2006, p. 93) describe the final stage of thematic analysis as telling "the complicated story of your data in a way which convinces the reader of the merit and validity of your analysis". In accordance with Braun and Clarke's (2006) guidance, interview data were used to evidence the presence of final themes and codes. This involved selecting interview excerpts for integration into a qualitative results chapter (Chapter 8). Excerpts were chosen to be as vivid as possible (Braun & Clarke, 2006). However, interviewee responses were often somewhat limited in detail (particularly when provided by younger children), despite attempts to encourage engagement by communicating in Kiswahili and employing probing questions (Section 3.4.2). Further, it is possible that child and caregiver responses were constrained by the framework imposed by the pre-written questions and probes used during interviews (Partington, 2001; Section 3.4.2.2).

This section now considers an additional limitation to qualitative findings, which is pertinent to my epistemological perspective that objectivity should be maximised while acknowledging that knowledge is necessarily value laden and 'pure' objectivism is therefore impossible (Section 1.4.1). The values of the interviewee, interviewer and research assistant were key considerations in qualitative analysis. Whilst the 'sense' contributed by researcher and interviewee might commonly be referred to as the "double hermeneutic" due to the existence of a "process of double sense-making" (Usher, 1996, p. 19), the additional presence of a research assistant during the interview added another sense-maker, thereby suggesting a 'triple hermeneutic'. The sense contributed by the interviewee was likely to have been particularly influenced by differences in our personal characteristics regarding race, ethnicity and, on occasion, gender (Denzin and Lincoln, 1994). Responses might also have been influenced by the relationship that my research assistant and I were perceived to have had with

Ubongo: it was possible, for example, that the educational benefits of watching Ubongo Kids were exaggerated due to interviewees' belief that this was our desired response.

Although the influence of interviewers on interviewees likely biased the data collection process, parts of the qualitative data were considered less prone to misrepresentation. The 'sense' contributed by Mussa, Titus or myself appears unlikely to have affected the more 'objective' exposure-related information to have emerged from interviews (such as a spontaneous mention of Ubongo Kids characters: Excerpt 8.12). In fact, interviews were particularly helpful in providing information on this topic. This is because interviewees were drawn from a backup quantitative sample that included numeric exposure information (which corresponded in format to that available in the 'Matched set': Chapter 5).

Further, thematic findings were produced in a manner that reduced the limitations imposed by the apparent 'sense' of interviewers. Objectivity was advanced in part by placing greater focus on excerpts supported by examples (which were considered to provide more credible evidence: e.g. Section 8.4.1) and seeking counterpoints (e.g. Section 8.3.2). Additionally, objectivity was pursued by promoting triangulation (Anfara, Brown & Mangione, 2002). To do this, responses from individuals were corroborated with data from alternate methods and sources. This included comparing statements from children with those from their caregiver (when, for example, seeking to identify what children might be doing instead of viewing Ubongo Kids, see Section 8.3.1). Child claims of learning from educational television were also considered alongside responses from other children in different districts (including all available reports that learning had not occurred, see Section 8.3.2).

The extent of triangulation was, however, limited. This was due to two key factors. Firstly, quantitative and qualitative samples were generally distinct (notwithstanding the availability of quantitative exposure data for all interviewees from the backup dataset). That is, no family in the qualitative sample were known to have featured in the quantitative sample used in any of the preceding results chapters (Chapters 5, 6 and 7). Quantitative information could not, therefore, be employed to directly consider the veracity of respondent claims regarding mathematics learning. Secondly, an important purpose of qualitative information was to garner information that went beyond quantitative approaches. For example, interviews gave longer instances of child recall that were far more detailed than what was captured through the binary self-reported viewership measure used in analysis of the 'Household set' and 'Panel set' (see Section 8.4.2.1). This demonstrated the coarseness of quantitative approaches, which could not be employed to prove or disprove qualitative statements.

3.5   Summary of methods and consideration of mixed methods approach

This chapter has provided information on the quantitative methods used to address both the primary and subsidiary research questions. With regards to the primary question, three models were presented which each considered the relationship between educational television and mathematics capability. These models were applied to three datasets and used both cross-sectional and longitudinal approaches. While each design possessed relative strengths, it was contended that the strongest finding was provided by the cross-section model applied to the 'Household set'. The relative strength of this model was attributed to various distinguishing features. Key amongst these was sample size, the application of household fixed effects and use of a continuous capability measure as the dependent variable. These characteristics were, respectively, identified to provide a close approximation of the population under consideration, account for a large amount of unobserved heterogeneity and facilitate a more accurate estimation of coefficient for Ubongo Kids exposure (relative to the multilevel model).

Given these advantages, the cross-section model applied to the 'Household set' was used as the basis of the influence estimate in CEA calculations. The section concerning cost effectiveness explained how the 'Household set' exposure coefficient estimate was employed in a formula alongside measures of programme duration, cost and number of beneficiaries to calculate the approximate SD gain in mathematics capability per \$100 spent on educational television. Further, it was identified that comparison with other CEA results for alternate interventions would provide information on the relative cost effectiveness of educational television.

Following the provision of information on the quantitative approaches used to investigate the primary and subsidiary research questions, qualitative methods were considered. Qualitative data were employed primarily to aid the interpretation of quantitative findings. Regarding the primary research question, interviews performed this function by a) providing additional information on exposure measures used in statistical models, and b) giving data concerning mathematics learning from television viewing that could be considered against overall model findings. The qualitative findings presented in Chapter 8 also provided limited information of relevance to cost analysis (Chapter 9), as interview responses assisted the interpretation of important numeric estimates used in CEA calculations. Further, the role of qualitative data in considering cost effectiveness extended to informing the quantitative estimates (such as viewer opportunity cost) used in CEA models (Section 9.5).

4   Chapter 4 – Descriptive statistics

This chapter presents descriptive statistics for the datasets used to explore the primary research question, which concerns the association between educational television and mathematics capability. Three distinct quantitative data sets were used to address this research question, which were described in some detail in Chapter 3. Here, further information on the features of these data sets are discussed before presenting the three models used in the analysis in Chapters 5-7. The purpose of this chapter is to give the reader a clear overview of the dimensions of the data used, to highlight issues with missingness and discuss the key variables used in the analysis. This thesis also addresses a subsidiary research question relating to the cost effectiveness of the intervention, yet discussion of the data used to address this subsidiary question is presented in Chapter 8 ('Cost-effectiveness analysis').

4.1 Chapter overview

Descriptive statistics concerning the primary research question are presented across three key sections in this chapter. The first of these sections details the sample sizes of datasets for analysis, which is supported by information on sample size by geographical area. After this, levels of missingness are considered for the quantitative information from which datasets for analysis were formed. Lastly, this chapter presents descriptive summaries of two key measures used in subsequent inferential models: mathematics capability and exposure. The presentation of information on mathematics capability and exposure includes an investigation into each concept separately, as well as an exploration of the possible association between both concepts.

4.2 Samples for analysis

This chapter first presents information on the samples used in inferential models (presented in Chapters 5-7). As such, this section focuses on the cleaned datasets used for analysis: the 'Matched set', which includes primary and secondary cross-sectional data concerning Temeke and Ilala districts; the 'Household set', which gives cross-sectional data on Tanzania as a whole; and, the 'Panel set', which provides longitudinal data for 45 Tanzanian districts. Additionally, discussion of the 'Household set' features investigation into district population estimates, which is used to consider the survey structure of this cross-sectional data.

Ilala and Temeke cross-section data

The data used for analysis concerning Ilala and Temeke (the 'Matched set') had a total sample of 793. This data was ultimately used in a model with fixed effects at the enumeration area (EA) level.

Presentation of the Ilala and Temeke sample is therefore divided by EA (see Figure 4.1). There are 59 EAs in the sample, each of which contained between 2 and 33 children.[89]
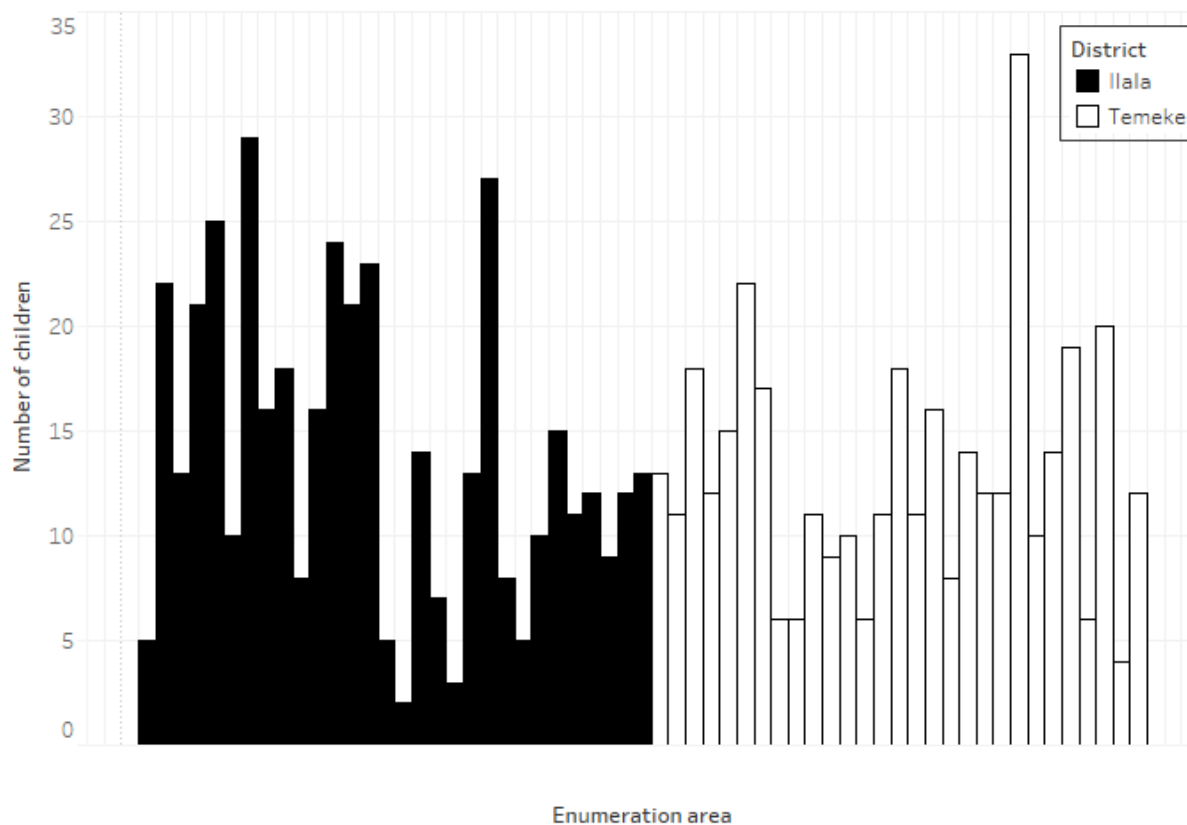
Figure 4. 1: Sample by enumeration area in Ilala and Temeke

Tanzania-wide cross-section data

The dataset for analysis concerning Tanzania as a whole ('the 'Household set') contained 39,923 children. This dataset was comprised of 13,752 households (the level at which fixed effects were employed in analysis, see Section 6.2) that were located in 1,670 EAs across 56 districts. The number of children per district is depicted below (Figure 4.2).

---

[89] It might be noted that data cleaning to create the 'Matched set' involved removing one child located in an EA without any other retained children. (The retention of this child would not have benefited the EA fixed-effects model employed.)
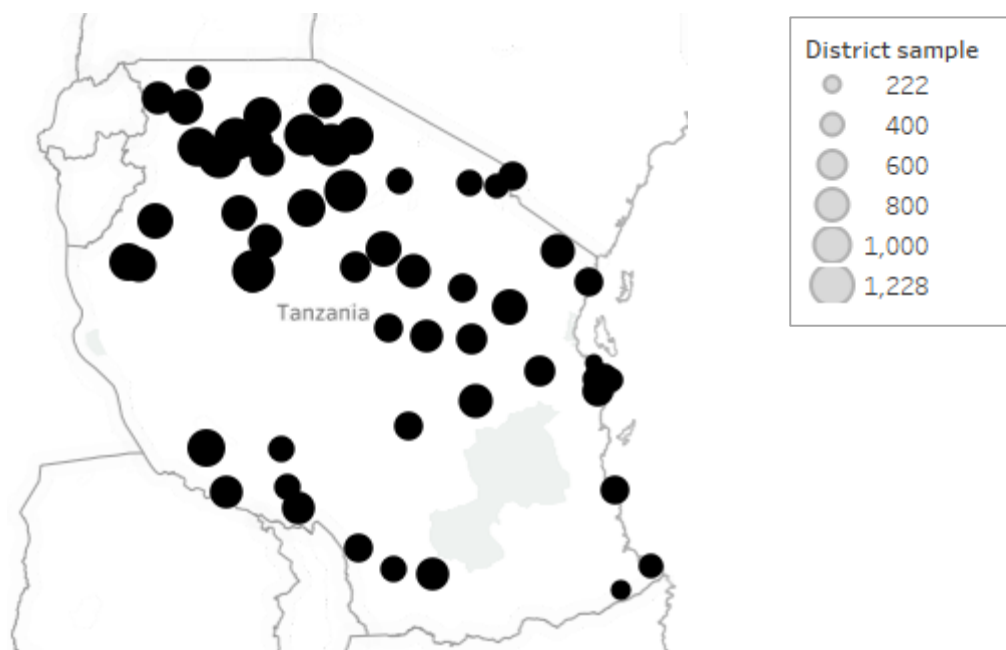
Figure 4. 2: Sample by district

This section now provides a graphic in which estimated district populations (in 2012) are presented for all districts featuring in Figure 4.2 (using population estimates from the Uwezo 2017 survey). Whist the estimated number of households varied markedly by district (from 25,179 to 275,580) (Figure 4.3), child samples remained comparatively constant (between 222 and 1,228) (Figure 4.2).
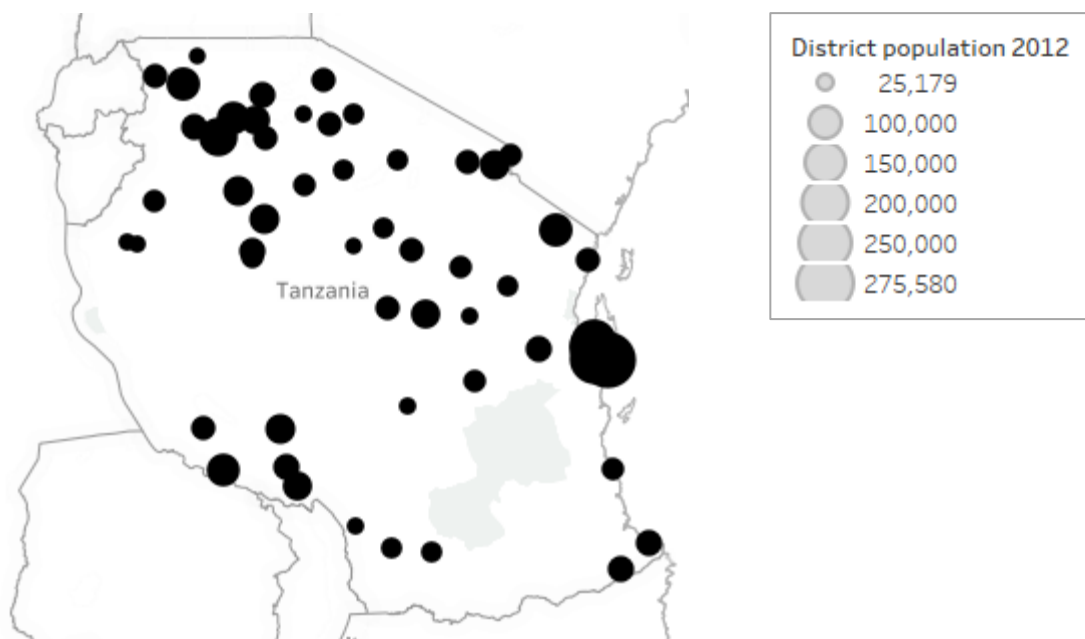
Figure 4. 3: Estimated district population

The distinction between Figures 4.2 and 4.3 suggested the applicability of sample weights when providing descriptive statistics for Tanzania as a whole.[90] Analysis using sample weights should more accurately reflect the likelihood of a household being selected within a particular EA, the probability of that EA being selected within a particular district and the probability of that district being chosen within Tanzania (Uwezo 2013).[91] Given that the weighting assigned to a household was influenced by the likelihood of a household being selected from an EA and the probability of that EA being selected, more sparsely populated districts were overrepresented in non-weighted calculations. The overrepresentation of more sparsely populated districts in non-weighted analysis likely contributed to lower unweighted national means for mathematics attainment and reported Ubongo Kids viewership. This was because districts with higher populations generally possessed higher levels of viewership and mathematics attainment (see Sections 4.4.2 and 4.4.3).

Tanzania-wide panel data

The 'Panel set' used for analysis contained information from 45 districts captured at multiple time points. District measures were mean district aggregates from Uwezo surveys, from which children

---

[90] Cross-sectional analysis involving the dataset for Tanzania as a whole (presented in later chapters) did not apply sample weights, given that weights cannot be included in fixed effects models (created using the Felm package in R). However, trialling was conducted on alternate models which suggested that sample weights had no substantive effect on coefficient estimates.

[91] It should be recognised that the sample weights provided by Uwezo are only strictly applicable to analysis involving the entire Uwezo 2017 sample (n = 48,530). They are, however, applied to the 'Household set' to provide weighted means, given that the 'Household set' retains the majority of the Uwezo 2017 sample (n = 39,923).

located outside the 45 districts of interest were removed. The Uwezo surveys used for aggregation had total samples ranging from 32,377 to 41,071 (in 2015 and 2011, respectively, see Figure 4.4). (Given that the 'Panel set' only featured aggregated district averages, the dataset used for analysis possessed a sample of 45 at each time point.)
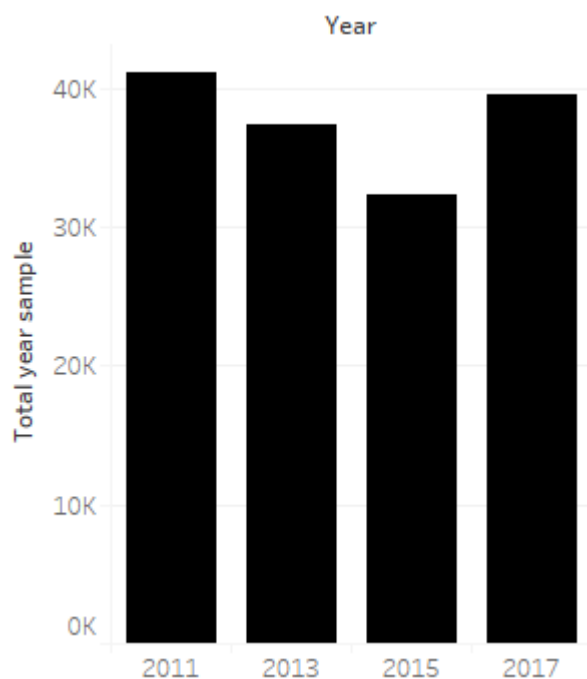
Figure 4. 4: Uwezo sample by year for the 45 districts common between each year

4.3 Levels of missingness

Before presenting descriptive information on key measures, this chapter considers the levels of missingness in quantitative data. This endeavour might initially appear unintuitive, given that the datasets used for analysis (describe above, see Section 4.2) possessed no missing values. However, these datasets for analysis were the product of a data preparation process that involved removing incomplete cases (for the 'Matched set' and 'Household set') or calculating district mean values (which ignored missing values, for the 'Panel set') (see Section 3.3.1.4.3). Investigation into missingness was therefore conducted using the 'original datasets' from which datasets for analysis were created. Examining missingness in the 'original datasets' provides two key benefits. Firstly, it gives some indication of the level of manipulation that was required to create the datasets for analysis. Secondly, exploring missingness justifies the exclusion of certain variables from the datasets for analysis (and, by extension, the inferential models used to examine the primary research question).

4.3.1   Cross-section data concerning Ilala and Temeke

This subsection provides information on missingness relevant to examination of the 'Matched set' – a cross-sectional dataset containing information from Ilala and Temeke district. To explore missingness relevant to the 'Matched set', descriptive analysis was conducted on secondary data captured by Uwezo in 2017 concerning Ilala and Temeke only (which was later cleaned and merged with primary data captured using the same sample frame to create the 'Matched set': Section 3.3.1.1.1). Focusing on Uwezo data from 2017 enabled this investigation to provide information on key variables that were omitted from inferential analysis (Chapter 5 – 'Matched set results'). Uwezo 2017 data also permitted exploration of the majority of variables used in the model presented in Chapter 5 ('Matched set results'), as well as information on the district and enumeration area (EA) in which each child was located (which were ultimately used in the application of model fixed effects and clustered standard errors). It should be noted, however, that two measures used subsequently in inferential analysis are not included in this missingness investigation. These measures are 'Ubongo Kids exposure' and 'exposure to other media', which were obtained through primary data collection (and therefore introduced into the 'Matched set' alongside Uwezo 2017 data only after merging Uwezo and primary information).

Table 4. 1: Number and percentage of missing values by variable – Ilala and Temeke

Variable	Number of missing	Percentage of missing
School type	216	17.50
Preschool	200	16.20
Any maths	53	4.30
Mother's education	37	3.00
Asset wealth score	15	1.22
Age	0	0
Enrolment	0	0
Sex	0	0
District	0	0
EA	0	0

This table presents the number of missing values and percentage of missing values for selected variables, in Uwezo 2017 data concerning Ilala and Temeke. The Uwezo data for 2017 collected from Ilala and Temeke contains 1233 children.
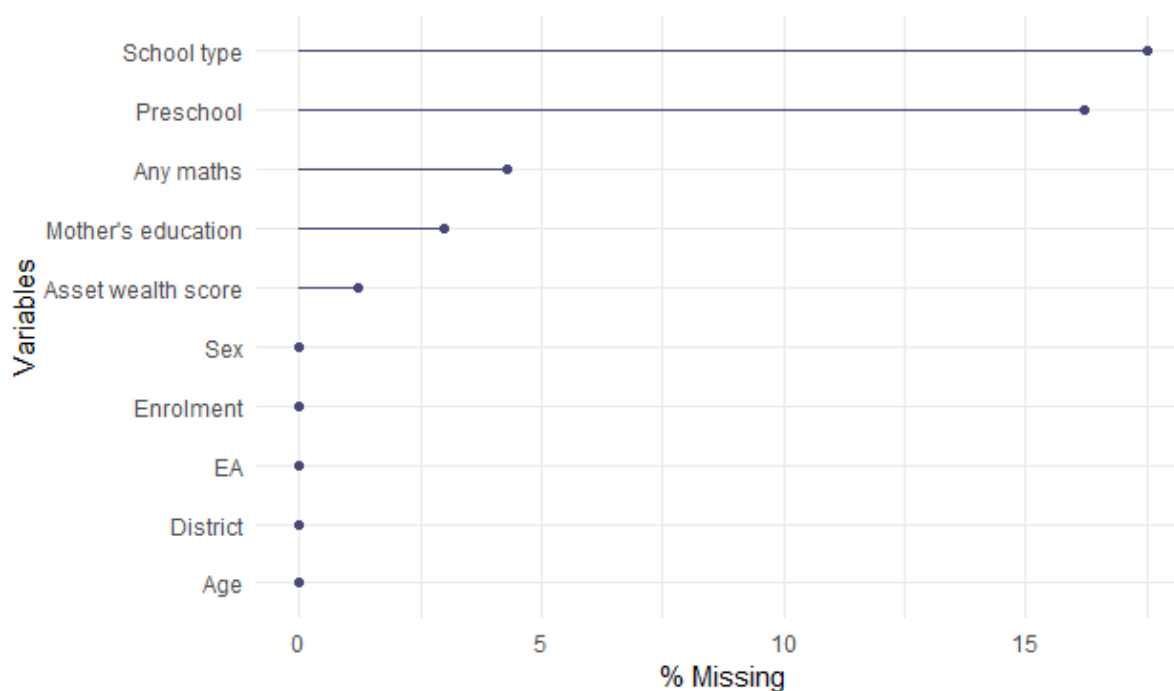
Figure 4. 5: Percentage missing by variable – Ilala and Temeke

This figure provides a visual representation of the percentage of missing values for variables in the Uwezo 2017 data concerning Ilala and Temeke.

The above graphic (Figure 4.5) shows the high levels of missingness amongst two variables that are frequently associated with learning outcomes: school type and preschool attendance. The type of school that a child attends has commonly been identified to be related to outcomes (e.g., Alcott & Rose, 2016). Similarly, children who attended preschool have been estimated to score approximately 0.10 standard deviations higher in previous Uwezo assessments (Bietenbeck, Ericsson & Wamalwa, 2017).[92] The relevance of school type and preschool attendance to learning outcomes warranted their inclusion in the dataset for analysis of Ilala and Temeke. However, the level of missingness for both variables precluded their otherwise justified retention. Children located in Ilala and Temeke from the Uwezo 2017 dataset did not provide information on preschool attendance and school type in 16.2% and 17.5% of cases, respectively. These variables were omitted from the dataset for analysis instead of imputing missing values or deleting the cases in which there was no information. This decision reflected the fact that information was not available for a large number of children and that the 'preschool' and 'school type' variables contained information that conflicted with other complete measures (which was considered using the Uwezo dataset for Tanzania: Section 4.3.2).

---

[92] As noted in Chapter 3 ('Methodology'), other variables that have frequently been found to be related to learning outcomes are not available whatsoever in Uwezo 2017 data (e.g., private tuition: Alcott & Rose, 2017).

Further exploration into missingness amongst the Uwezo 2017 data for Ilala and Temeke provided additional support for the decision to omit 'preschool attendance' and 'school type' from the dataset for analysis, as opposed to deleting cases where no information was available. There was a large amount of missing 'preschool' and 'school type' information where data was not missing for any other (retained) variable. The top three instances of variable missingness amongst children in the Uwezo 2017 sample for Ilala and Temeke featured 'school type' on its own (127 instances), 'preschool' on its own (107 instances) and 'school type' and 'preschool' together (59 instances). This is shown in the image below (Figure 4.6). Deleting cases for which there was no 'preschool' or 'school type' information would have substantially reduced the sample size of the dataset for analysis.

Figure 4. 6: Missingness dependence – Ilala and Temeke

The above figure shows the intersection between missing variables in cases from the Uwezo 2017 sample for Ilala and Temeke.

Aside from measures of 'preschool attendance' and 'school type', there was only a small amount of missing data for other variables in the Uwezo data for Ilala and Temeke. This did not exceed 5 percent. Imputation was not attempted for variables concerning asset wealth, mother's education or

mathematics. Instead, cases in which any of these variables were missing were deleted from the dataset for analysis.[93] This decision was justified by the fact that asset wealth, mother's education and mathematics information was available in the vast majority of cases, which meant that deleting incomplete cases was unlikely to have had a substantial effect on subsequent inferential results. Further, choosing not to impute promoted the replicability of findings: Uwezo will ultimately make the Uwezo 2017 dataset publicly available with all non-imputed measures used in quantitative analysis.[94]

4.3.2   Tanzania-wide cross-section data

This chapter now presents information on missingness relevant to examination of the 'Household set' – a cross-sectional dataset concerning Tanzania as a whole. This information is provided using Uwezo data for 2017, from which the 'Household set' was created (by removing children who had missing variables that were ultimately used in regression, then removing children with no assessed siblings, see Section 3.3.1.1.2). As with examination of Uwezo data for 2017 concerning Ilala and Temeke only (Section 4.3.1), using Uwezo 2017 data for Tanzania permitted investigation into variables that were omitted from the dataset for analysis. It might be noted that the investigation into variable missingness did not consider mother's education or asset wealth (as these were not used in the 'Household set' model).[95]

---

[93] Cases in which some but not all mathematics information is missing are retained (as the creation of capability estimates using IRT does not require data for all test items from each individual).

[94] Uwezo might impute mathematics results for all children whilst retaining non-imputed mathematics results in the future. Any later imputation conducted by Uwezo will not extend to other variables.

[95] These variables were not employed in inferential analysis of the 'Household set', as they did not vary within households. The model for analysis of the 'Household set' used household-level fixed-effects, so did not require their inclusion (Chapter 6 – 'Household set results').

Table 4. 2: Number and percentage of missing values by variable - Tanzania

Variables	Number of missing	Percentage of missing
Preschool	11459	23.61
School type	11243	23.17
Any maths	1528	3.15
Kiswahili	1208	2.49
Sex	0	0
Household	0	0
Exposure	0	0
Enrolment	0	0
EA	0	0
District	0	0
Age	0	0

This table presents the number of missing values and percentage of missing values for selected variables, in Uwezo 2017 data (for Tanzania as a whole). The Uwezo data for 2017 contains 48,530 children.
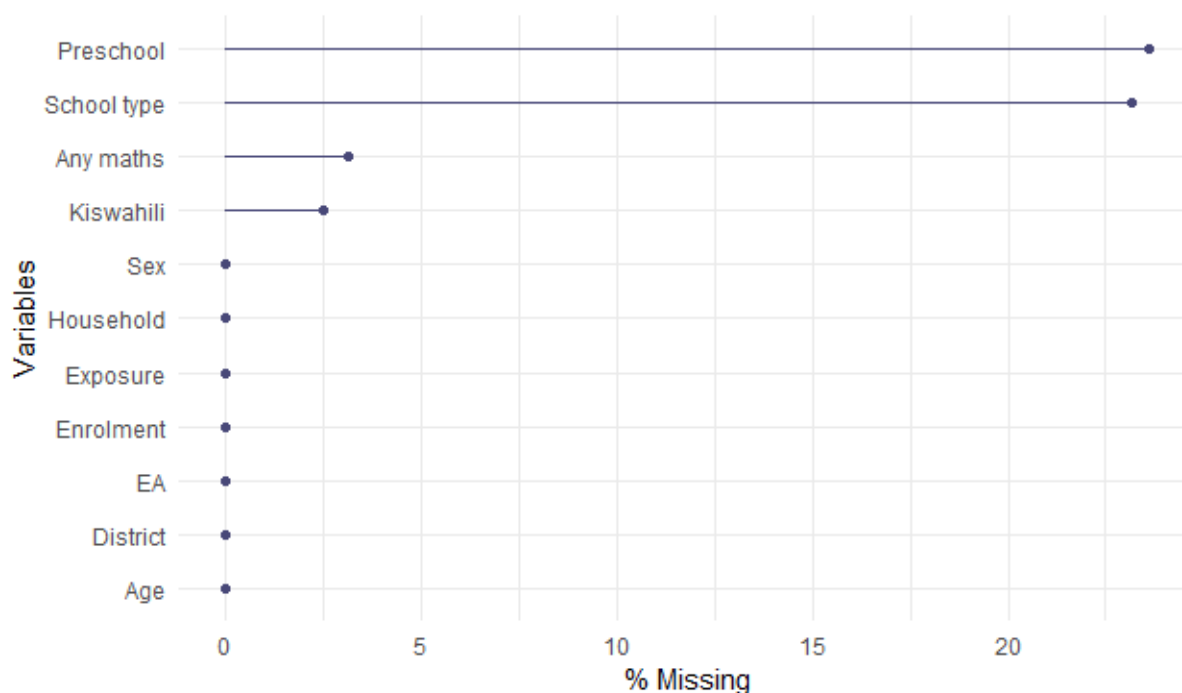
Figure 4. 7: Percentage missing by variable – Tanzania

This figure provides a visual representation of the percentage of missing values for variables in the Uwezo 2017 data concerning Tanzania as a whole.

The above table and image (Table 4.2 and Figure 4.7) depict a similar pattern of missingness to the tables and images produced to investigate data for Temeke and Ilala districts (Section 4.3.1). Once more, information concerning preschool attendance and school type was not available for a large percentage of children (23.61 percent and 23.17 percent, respectively). Subsequent inferential analysis of the 'Household set' therefore omitted these variables, despite their relevance to learning outcomes in East Africa (see Section 4.3.1). Examination of the intersection between missing variables further justified their exclusion from the datasets for analysis. As with investigation concerning Uwezo 2017 data for Ilala and Temeke, there were a large number of cases in which 'preschool' or 'school type' information was missing where information was present for all other variables (see Figure 4.8).
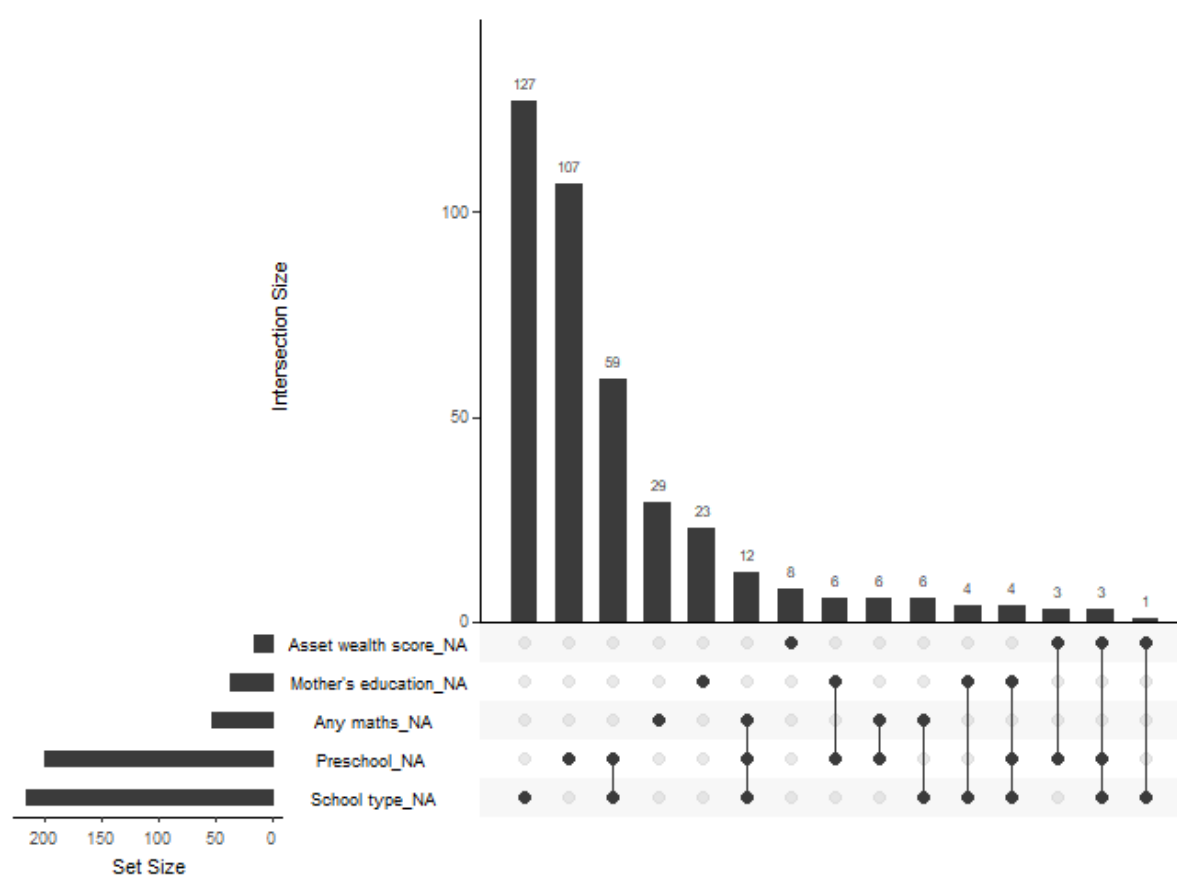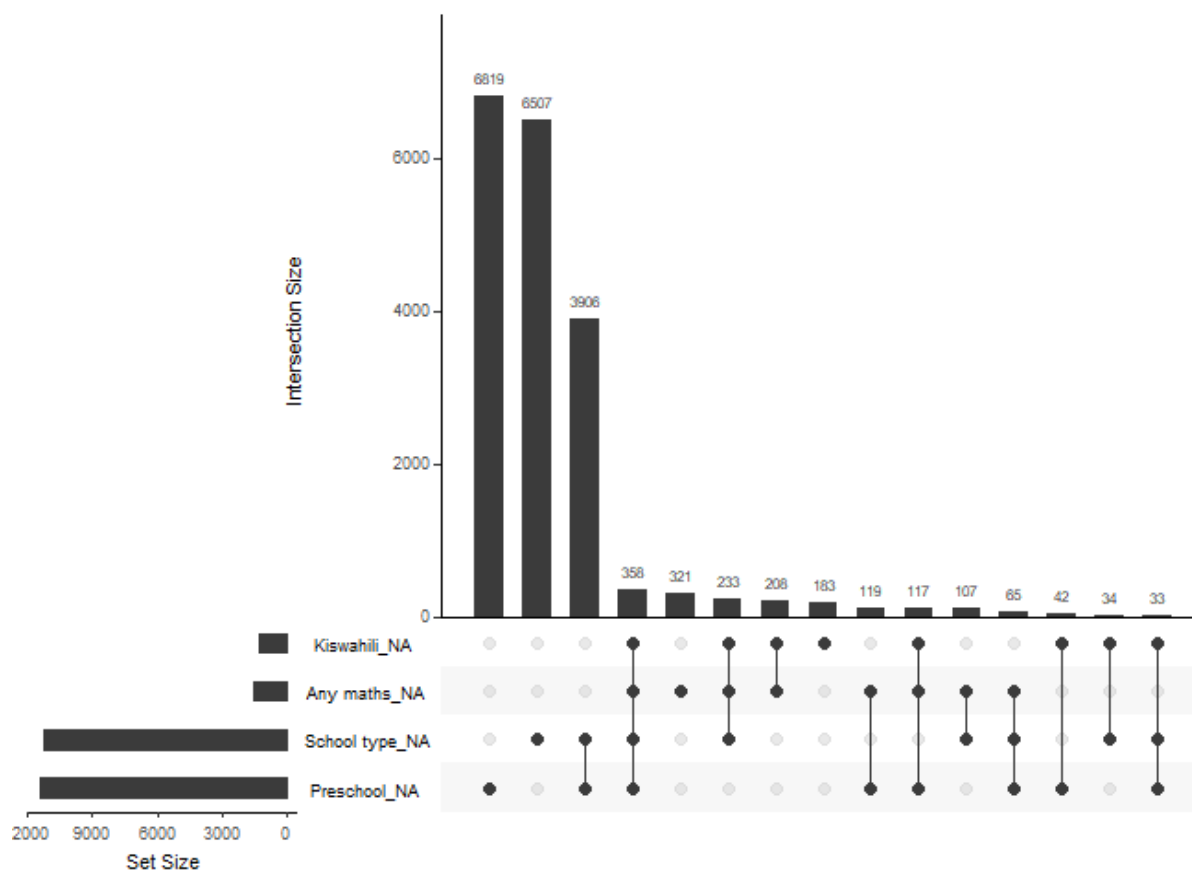
Figure 4. 8: Missingness dependence – Tanzania

The above figure shows the intersection between missing variables in cases from the Uwezo 2017 sample for Tanzania as a whole.

Additional investigation into the measures of 'preschool attendance' and 'type of school attended' further suggested that they contained data that conflicted with variables for which more complete information was available. For example, there were multiple children in the Uwezo 2017 data for Tanzania who were not reported to be enrolled in school yet simultaneously reported as attending a private or government school.[96] This indicates that an accurate imputation of values for 'preschool attendance' and 'type of school' could not have been conducted for any dataset drawing on information from Uwezo 2017.[97] Indeed, the conflict between school type and school enrolment

---

[96] School enrolment information was considered to be relatively trustworthy, because a) it had no missing information, and b) it had been described as such by Uwezo staff during informal conversations.

[97] This contention is supported by an investigation into the mean mathematics scores of each 'school type group' (treating missing values as a distinct group). Even in cases where all children considered were reported as currently enrolled in school, the mean mathematics score for the missing group was far lower than that for either the government or private school group. It might also be noted that some inaccuracy in this variable could have stemmed from the Uwezo 2017 survey design. Prior Uwezo data collections for Tanzania included another choice for school-type, "other". However, in the 2017 survey, administrators could only select "private" or "government". This likely caused some inaccurate reporting and some missing values for children who attended a faith-school or community school.

suggested that including the type of school that a child attended in analysis could have introduced bias to subsequent quantitative analysis.

As with Uwezo 2017 data concerning Ilala and Temeke, there was only a small amount of missing data for variables aside from 'preschool attendance' and 'school type'. This did not exceed 5 percent for any of these variables. Once more, variables were not imputed. Cases in which Kiswahili results or all mathematics test items were missing were instead deleted. This decision was supported by the fact that the replicability of analysis could have been enhanced by using non-imputed values.

4.3.3   Tanzania-wide panel data

The final investigation into missingness concerns the 'Panel set' – a dataset based on Uwezo survey information from multiple years. The 'Panel' dataset for analysis again contained no missing values. However, the dataset was formed of district aggregated means from Uwezo data in 2011, 2013, 2015 and 2017. This investigation focuses on the variables ultimately used in analysis of the 'Panel set'. These variables did not include preschool attendance and private tutoring, following the identification that there were high levels of missingness for both variables in Uwezo data for 2017 (which was used in creating the 'Panel set', see Section 4.3.3). Prior to aggregation, Uwezo information generally possessed no missing values for the variables subsequently employed in analysis.[98] The one variable for which there was missing data prior to district-level aggregation was mother's level of education. The presence of missing data concerning mother's education level prior to district aggregation might have ultimately reduced precision in the coefficient estimates produced through multilevel analysis (for this variable and, by corollary, for other variables: Chapter 7 – 'Panel set results'). The percentage of missing data by year is shown in the table, below (Table 4.3).

Table 4. 3: Percentage of missing values for mother's education – Tanzania-wide panel data

Year	Percentage of missing
2011	01.72
2013	11.66
2015	14.29
2017	07.39

---

[98] It should be noted that children with no mathematics information were removed from the Uwezo 2017 dataset before creating aggregates (reducing the initial 2017 sample by 3.08%). Mathematics information was available for all children prior to 2017, as values were imputed by Uwezo. Imputed values were used for pre-2017 Uwezo data when creating aggregates. This is because the proportion of children with missing mathematics information was relatively large in earlier Uwezo surveys. (Uwezo is yet to impute missing mathematics information for 2017.)

4.4 Information on variables

This chapter now considers the measures of mathematics capability and exposure. This information is presented in three subsections concerning mathematics (Section 4.4.1), exposure (Section 4.4.2) and the association between both measures (4.4.3). Each of these subsections focuses only on the datasets used for analysis (described in Section 4.2).

4.4.1   Mathematics

Ilala and Temeke cross-section data

Mathematics scores for the Ilala and Temeke dataset used in analysis had a mean of 0.019, a standard deviation of 0.806 and a variance of 0.650. The mean mathematics score is presented below by enumeration area (EA) (Figure 4.9). To facilitate understanding, it should be noted that mathematics capability scores were based on IRT estimations amongst individuals (thereby producing a mean of approximately zero). Figure 4.9 highlights EA variation in mathematics capability – the dependent variable used in subsequent inferential analysis of the data. The presence of such variation shows the importance of controlling for children's EA location in quantitative models (see Section 5.2).
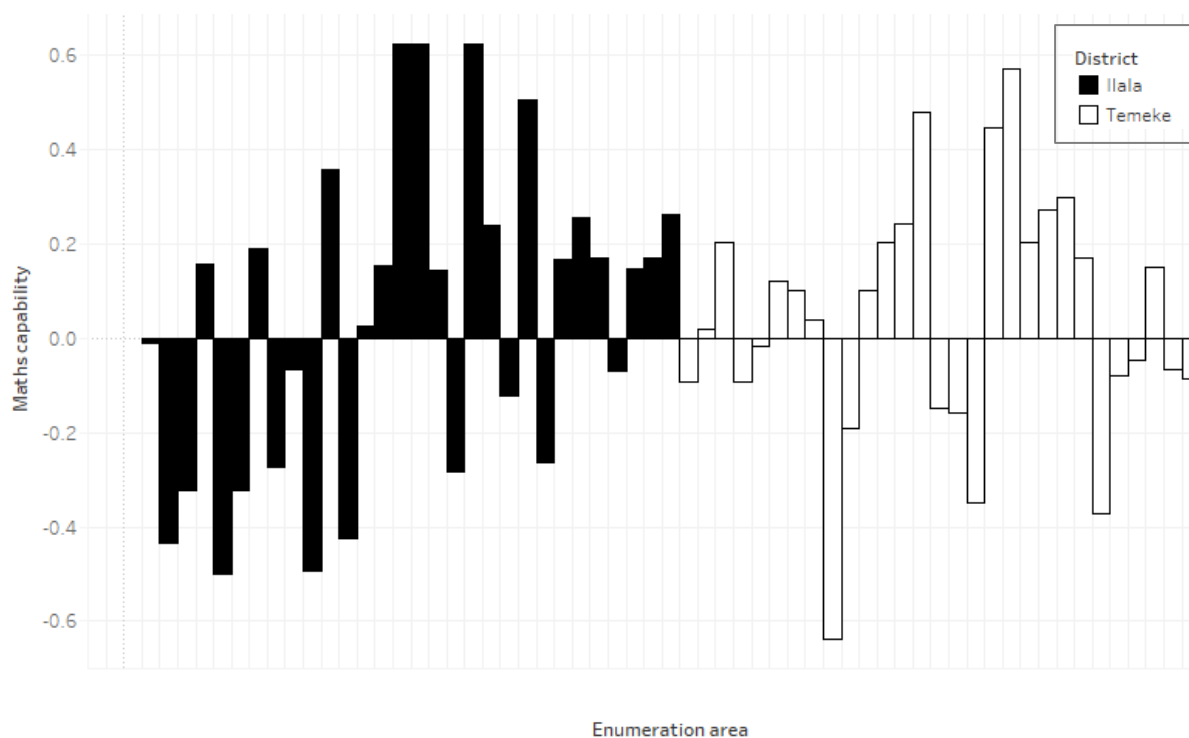
Figure 4. 9: Mathematics capability score and percentage of children who could complete all the main Uwezo mathematics tasks by EA

Tanzania-wide cross-section data

The mean mathematics score for the 'Household set' – used for cross-sectional analysis concerning Tanzania as a whole – had a sample mean of -0.007 and a weighted mean of 0.028. As noted previously, the weighted mean was higher given that districts with greater populations (which typically had higher sample weights) generally performed better in mathematics (Section 4.2). Mathematics scores had a standard deviation of 0.841 and a variance of 0.707. Non-weighted mathematics scores are presented by district in the image below (Figure 4.10). This graphic again highlights variance in mathematics capability by location, which underlines the importance of applying location-based fixed effects in later analysis.
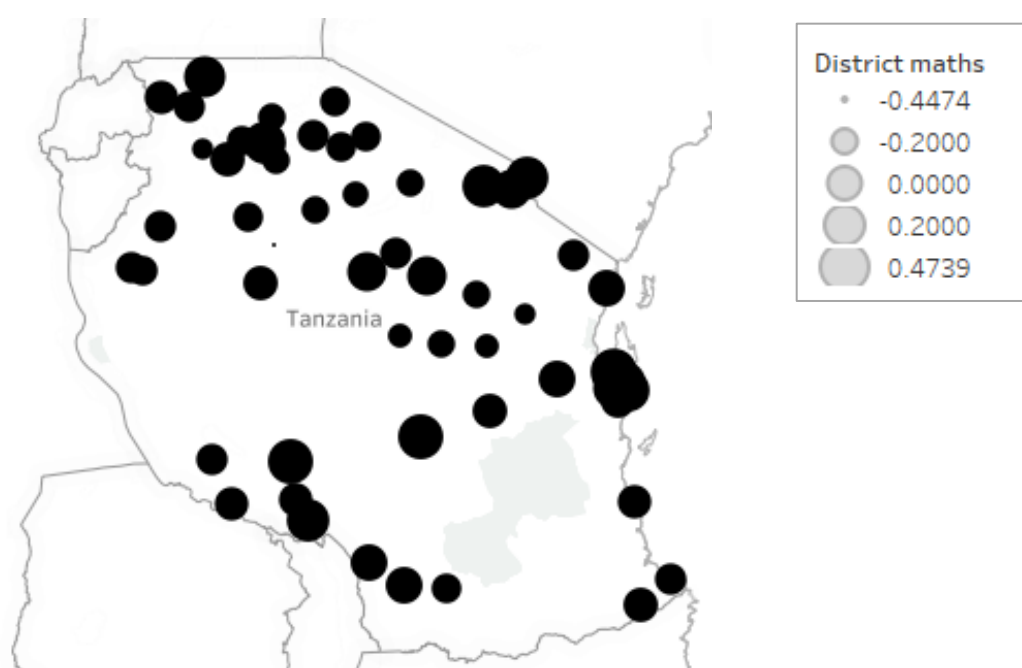
Figure 4. 10: Mathematics capability by district

Tanzania-wide panel data

The 'Panel set' used for longitudinal analysis featured mathematics capability rankings. As such, each year of data was made up of 45 sequential integers (1-45). Rankings at each time point were based on mean district mathematics capability. The standard deviation, variance and mean for district capability scores in each year are presented below (Table 4.4). This is followed by a violin plot by year for district mathematics capability (Figure 4.11). Table 4.4 shows a contraction in standard deviation and variance between 2013 and 2017 (the two years used in subsequent longitudinal analysis). The violin

plot suggests that this contraction from 2013 to 2017 particularly affects districts with relatively high capability scores (Figure 4.11). This identification supports the decision to examine change in district capability ranking as opposed to district capability means (Section 7.2).

Table 4. 4: Mean, standard deviation and variance for mathematics capability

Year	Mean	Standard deviation	Variance
2011	0.01	0.22	0.05
2013	0.01	0.27	0.07
2015	0.04	0.25	0.06
2017	0.02	0.21	0.04
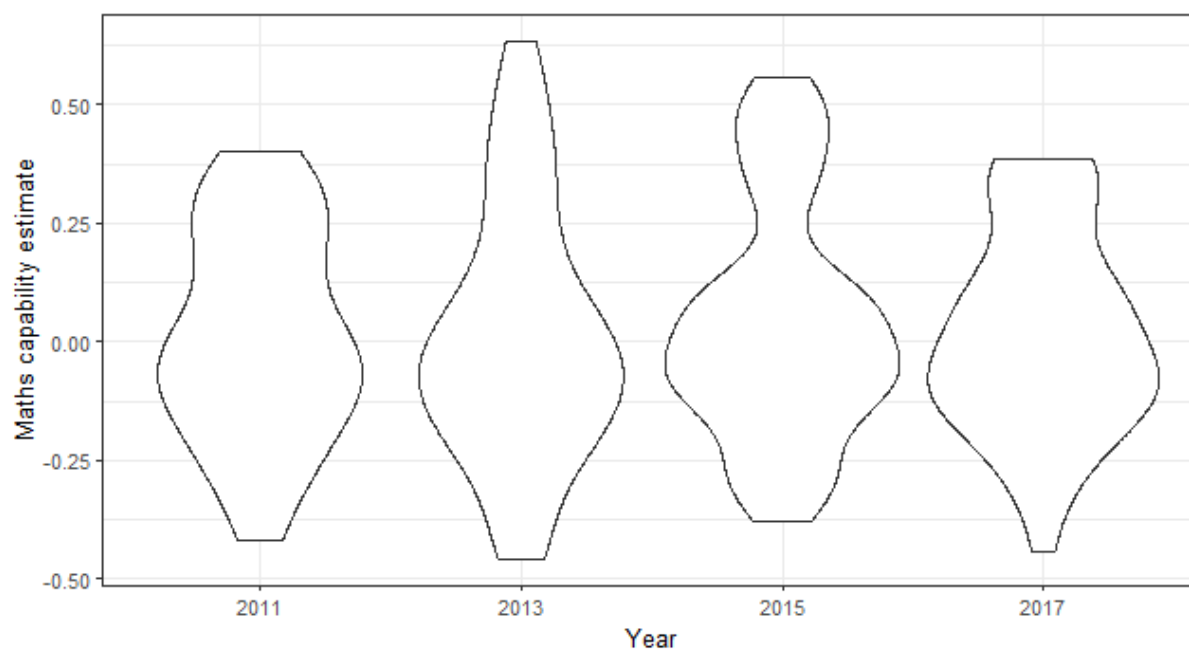
Figure 4. 11: Mathematics capability violin plots

4.4.2   Exposure

Ilala and Temeke cross-section data

The IRT scores for exposure in the Ilala and Temeke dataset for analysis (the 'Matched set') had a mean of 0.00, a variance of 0.72 and a standard deviation of 0.85. Mean exposure estimates for each EA are plotted below (Figure 4.12).
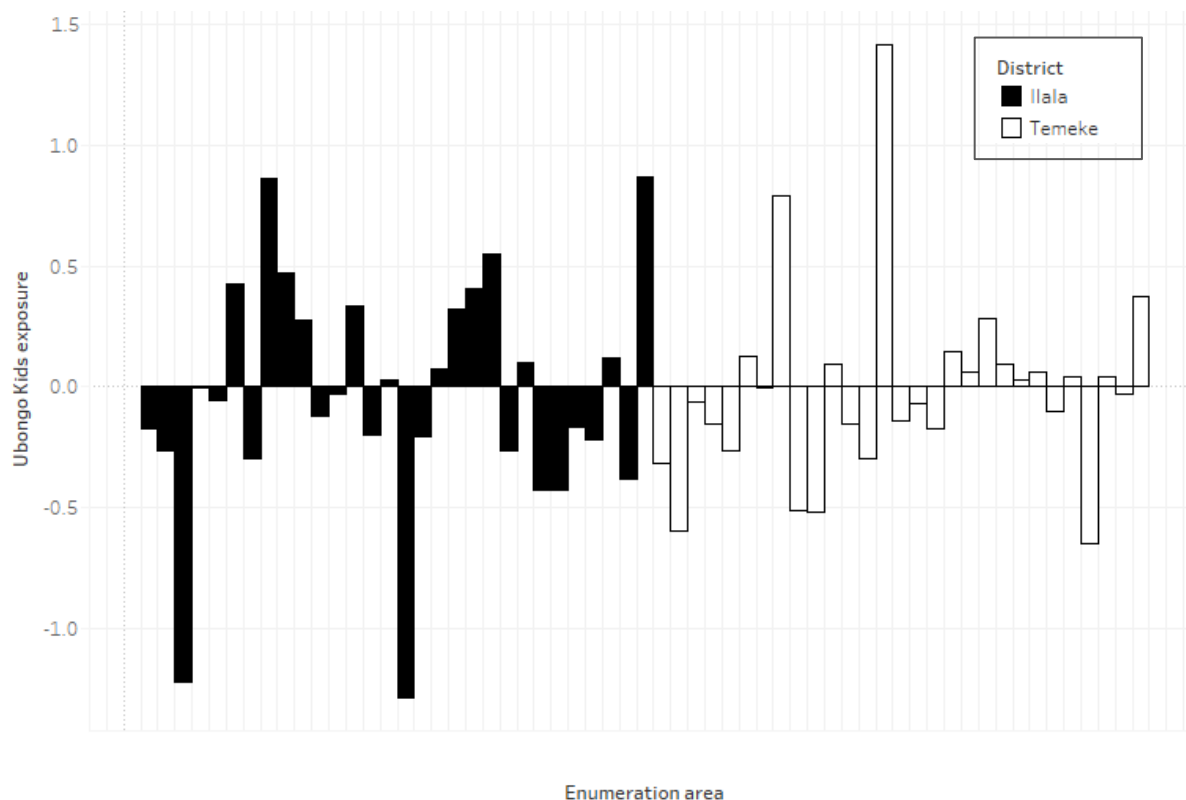
Figure 4. 12: Ubongo Kids exposure by enumeration area

In addition to exposure estimates created through IRT, the 'Matched set' included another child-level measure of exposure: self-reported viewership. While only IRT-created exposure scores were used in inferential analysis of the 'Matched set', the presence of self-reported viewership information presented an opportunity for comparison between the two measures. This comparison is beneficial given that self-reported viewership was the only measure of exposure available for analysis of the Tanzania-wide cross-section dataset and panel dataset (see Section 6.1 and 7.1). Should there be an association between self-reported viewership and IRT scores based on character recognition, this would support the use of a more basic binary self-reported measure in the analysis presented in Chapters 6 and 7 (see Section 3.3.1.3.2 for further discussion of exposure measures).[99]

Comparing enumeration area (EA) means for both exposure measures suggested an association (Figure 4.13). This was further supported by exploratory investigation into the correlation between self-reported viewership and exposure IRT scores amongst individuals (Figure 4.14). The presence of an association between self-reported viewership and exposure scores amongst individuals corresponded with findings from a linear regression between the two variables. Both were significantly related at $p=0.001$ with a coefficient of 0.39091. This suggested that the binary measure

---

[99] This is because character-based measures of exposure are favoured in the educational media literature (Section 3.3.1.3.2).

of viewership used in analysis of other datasets (see Sections 6.2 and 7.2) provided worthwhile information on exposure.
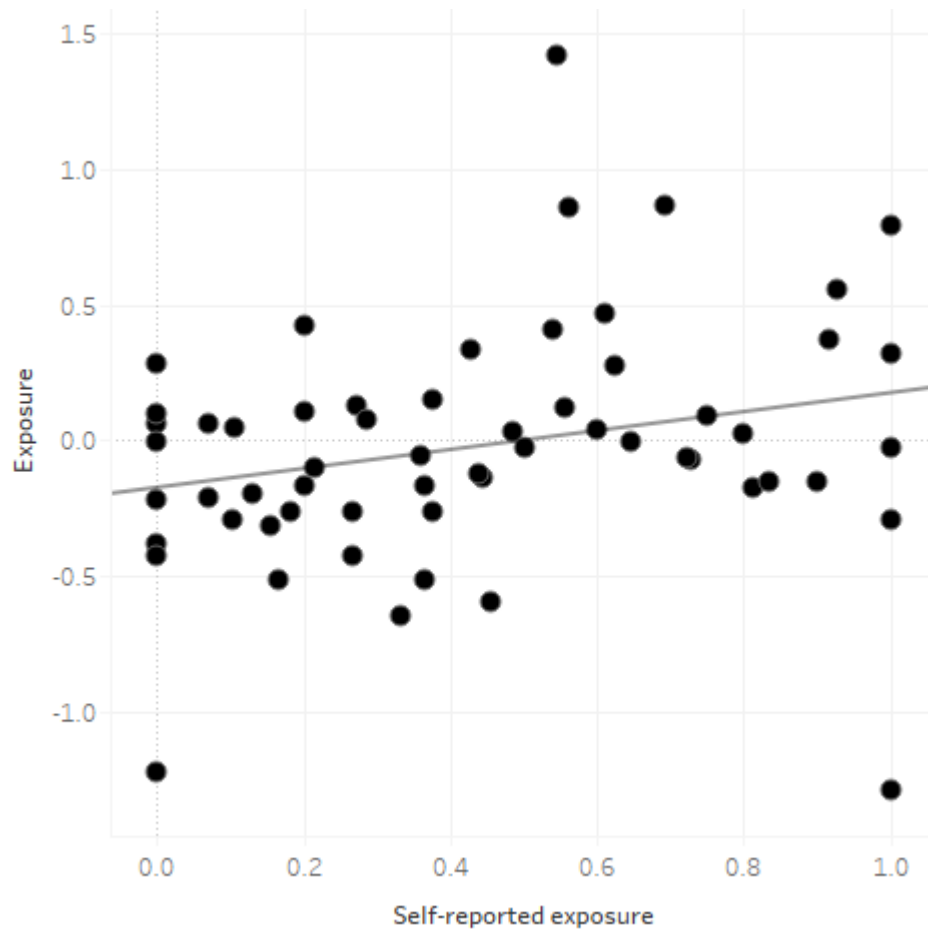
Figure 4. 13: Mean EA self-reported exposure and exposure IRT scores
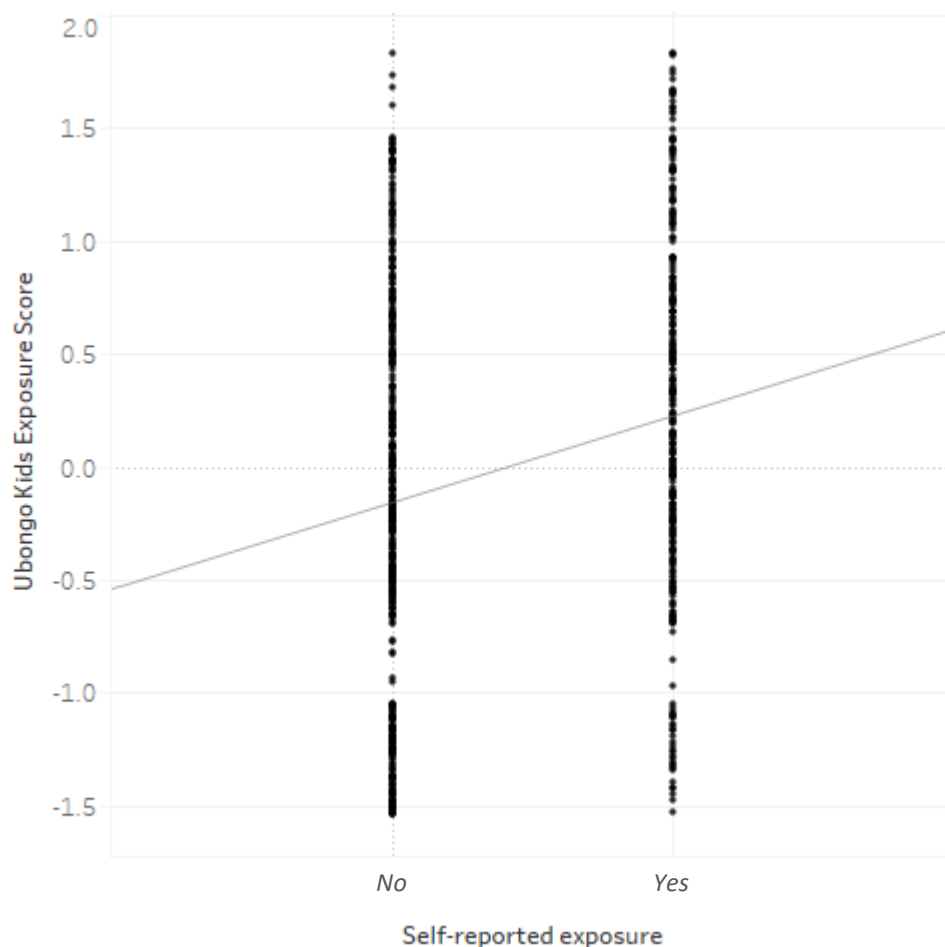
Figure 4. 14: Self-reported exposure and exposure IRT scores amongst individuals

This figure features exposure scores, to which a small amount of random noise (or 'jittering') was added. The addition of jittering enables this graphic to provide a clearer indication of IRT score distribution amongst individuals.

Although the above image (Figure 4.14) shows a correlation between self-reported viewership and exposure scores created from character recognition results, it also highlights a weakness in the binary self-reported measure. This is that numerous children who reported that they had not viewed Ubongo Kids in the past week often possessed IRT scores (based on character recognition results) that suggested some degree of previous exposure. This shows that the binary measure might provide a coarse measure of establishing who had been exposed and lead to an underestimation of who has benefited. This is of relevance to the provision of descriptive information on exposure regarding the 'Household set' and 'Panel set' (below, this section), as well as subsequent inferential analysis (see especially Section 6.3.2.1).

Tanzania-wide cross-section data

The mean level of self-reported viewership for the Tanzania dataset used in cross-section analysis was 0.1235, with a weighted mean of 0.1546. The standard deviation of reported exposure was 0.329 and

its variance was 0.108. Mean district self-reported exposure is shown by the image below (Figure 4.15). While variation in reported exposure by district is greater than that for mathematics capability (Figure 4.10), both maps feature a similar pattern. Reported exposure, like mean mathematics capability, was typically higher in districts with larger populations.
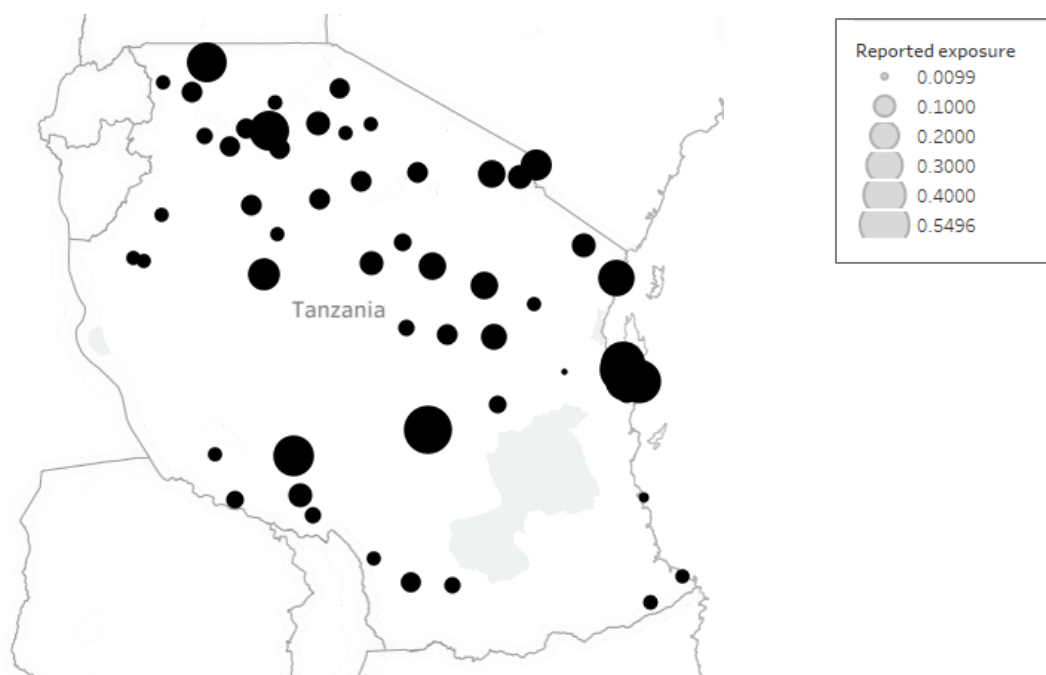
Figure 4. 15: Non-weighted exposure by district

Tanzania-wide panel data

Lastly, exposure is considered in the longitudinal dataset for analysis ('the Panel set'). The district aggregated viewership means contained in the 'Panel set' are presented below (Figure 4.16). The below graphic is a modified version of the image presented in Chapter 3, Figure 3.20.
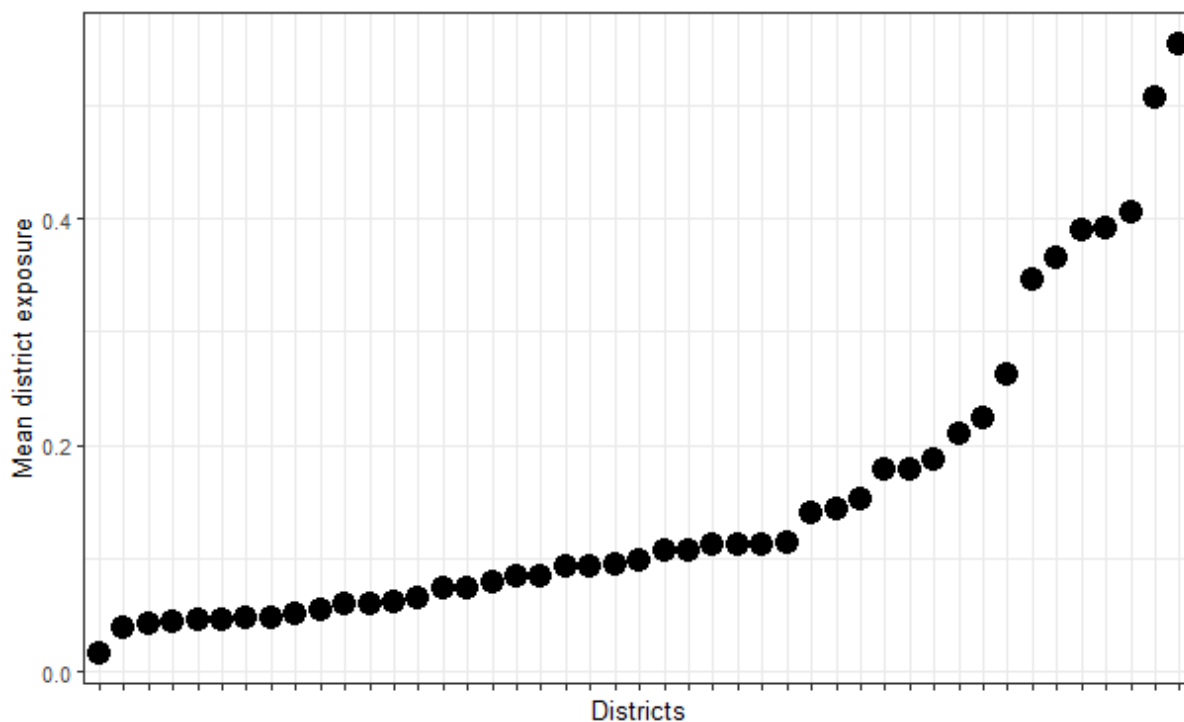
Figure 4. 16: Aggregated mean district recent viewership

4.4.3   Exposure and mathematics

The final piece of descriptive investigation presented concerns the correlations between mathematics scores and exposure estimates. This is explored using three graphics. These graphics each relate to a separate dataset used for analysis (in Chapters 5-7: 'Matched set results', 'Household set results' and 'Panel set results'). Additionally, regression-based exploration of the correlation between mathematics capability and exposure is presented for both cross-sectional datasets (before further statistical analysis involving controls in Chapters 5 and 6).

The first image provided concerns the dataset for Ilala and Temeke. This graphic depicts the mean mathematics score and mean Ubongo Kids IRT exposure estimate for each enumeration area (EA) (Figure 4.17). This figure suggests there to be some association between exposure to Ubongo Kids and mathematics capability. This was supported by statistical exploration into the correlation between mathematics and exposure. A linear regression including only Ubongo Kids exposure and mathematics capability showed that both measures were significantly related at p=0.001 with a coefficient of 0.207.
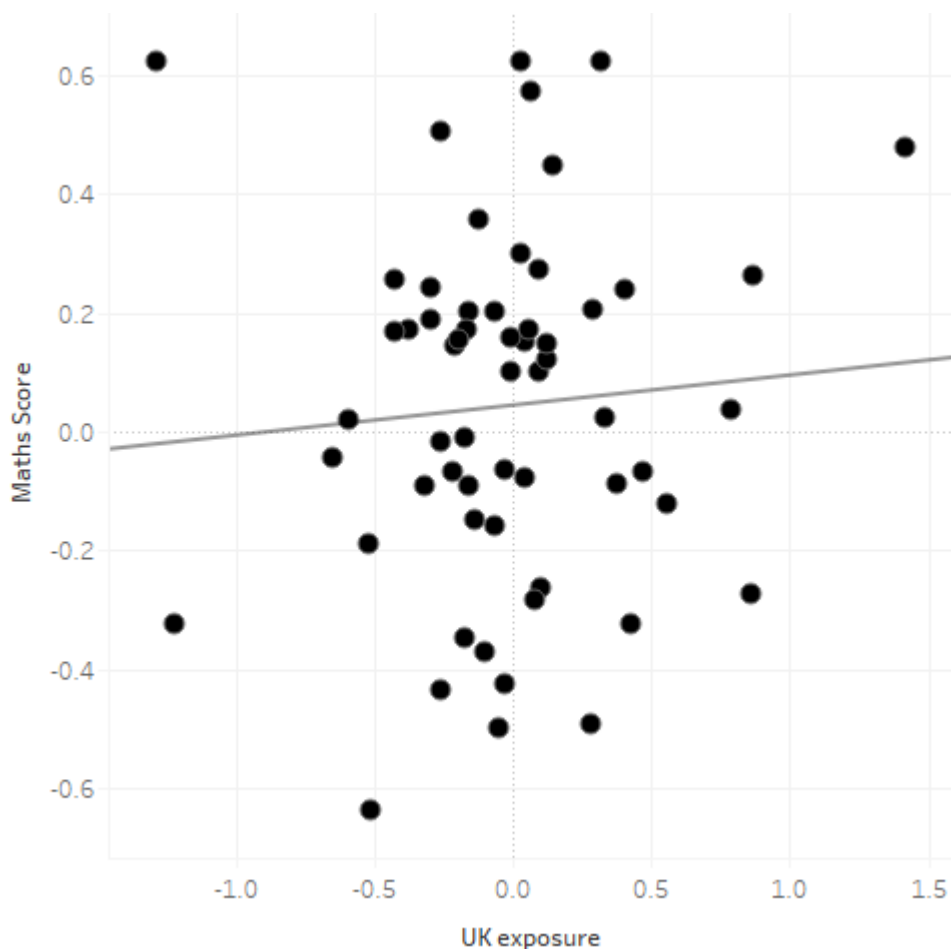
Figure 4. 17: Mean EA mathematics capability scores and exposure estimates in Ilala and Temeke

The next dataset considered is the cross-sectional dataset for Tanzania as a whole, the 'Household set'. Exploratory analysis showed that there was a correlation between individuals' mathematics capability and self-reported viewership (with both measures significantly related at p=0.001 with a coefficient of 0.363). This finding corresponds with an image produced to present mean mathematics capability scores and levels of self-reported viewership by district (Figure 4.18). This figure suggests a strong association between district mathematics capability and reported exposure. However, given the previous identification that district exposure followed a similar pattern to mathematics capability (both of which were generally higher in districts with larger populations) this trend is unsurprising. Indeed, this warranted the application of location-based fixed effects (alongside additional control variables) in later cross-sectional analysis of data for Tanzania (see Section 6.2).
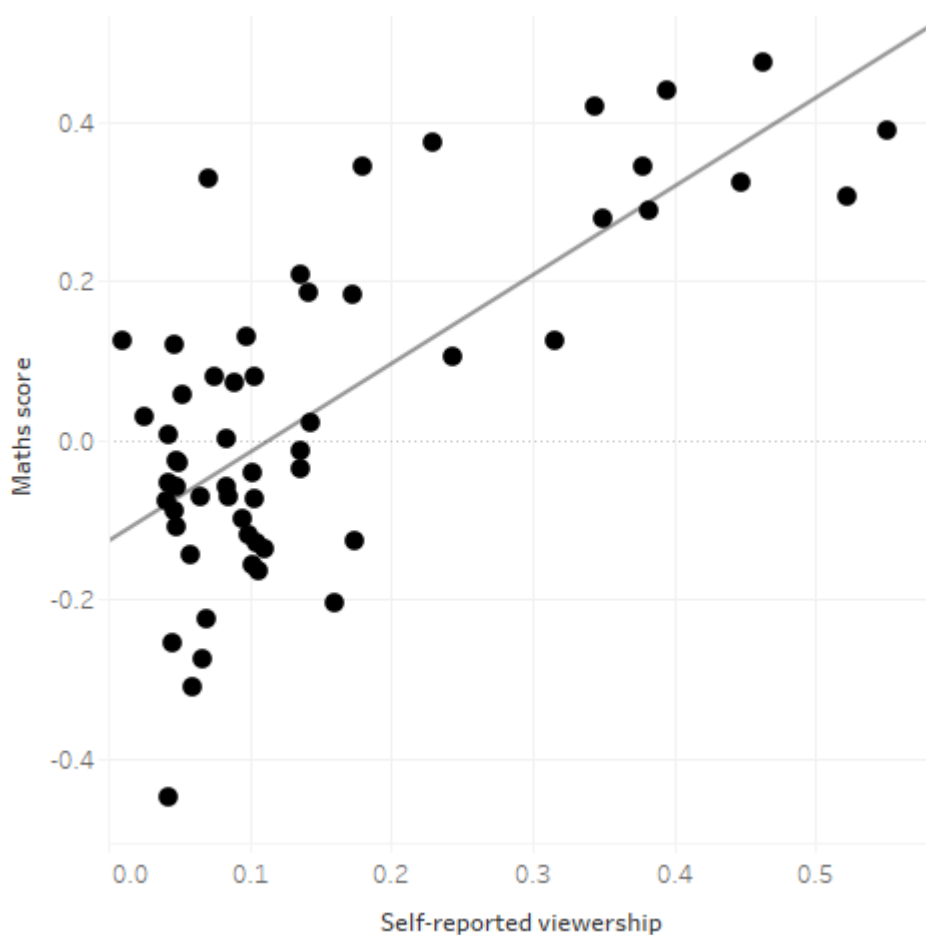
Figure 4. 18: Mean district mathematics capability scores and self-reported viewership

Lastly, violin plots are used to depict trends in capability ranking amongst exposure tertiles (Figure 4.19). Exposure tertiles were established from district mean measures of reported viewership from 2017 (see Section 3.3.1.4.3). In contrast with the relatively clear trend between exposure and mathematics capability shown in Figure 4.18 and (to a lesser extent) 4.17, changes in tertile ranking after the availability of Ubongo Kids (January 2014) provide only limited evidence of an association. This is because districts which possessed higher levels of exposure in 2017 already ranked higher than other districts in pre-exposure time points. This further emphasises the importance of applying extensive controls to subsequent cross-sectional models.
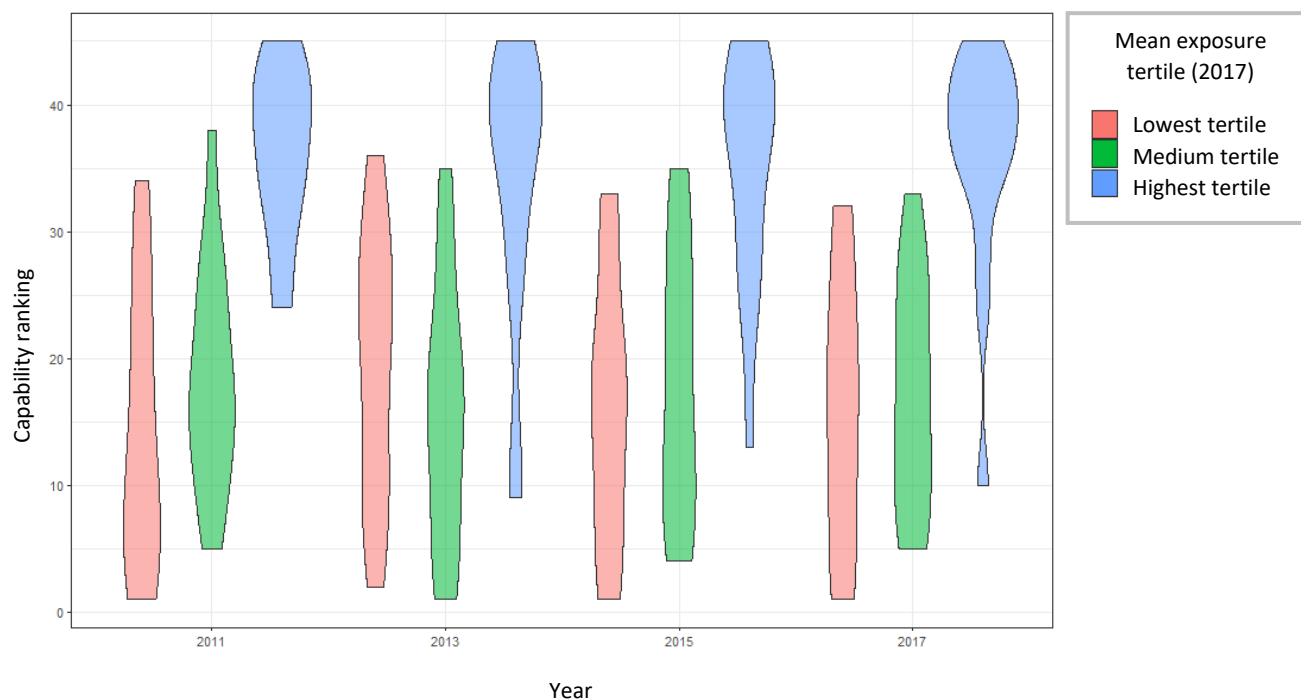
Figure 4. 19: Exposure tertile violin plot

4.5 Chapter Summary

This chapter has presented descriptive statistics relevant to the three models ultimately employed in inferential analysis (Chapters 5-7). Discussion began by presenting the sample sizes of the datasets used for analysis: the 'Matched set' concerning Ilala and Temeke (793 children); the 'Household set' concerning Tanzania (39,923 children); and, the 'Panel set' (which is formed of 45 districts measured at multiple timepoints). Examination of the 'Household set' included the presentation of sample sizes by district. District sample sizes were then compared against the estimated number of households per district (in 2012, Section 4.2). This comparison showed the importance of presenting weighted means for key measures in subsequent descriptive analysis (Sections 4.4.1 and 4.4.2).

This chapter then considered the level of missingness amongst secondary data used to create the datasets for analysis. Investigation into missingness showed the need to omit certain variables from the datasets for analysis. This was the case for 'preschool attendance' and 'school type', as there were large amounts of data missing and the information that was available conflicted with that provided by more trusted measures. Examination of the Uwezo data from 2011, 2013, 2015 and 2017 that was used to create aggregated district means (used in the 'Panel set'), showed that there was some missingness for information on 'mother's level of education' only. This was identified to demonstrate that district means concerning this measure could be prone to some inaccuracy, which might impact later inferential analysis.

Discussion then focused on key variables that featured in all datasets for analysis, including mathematics capability. The presentation of mathematics information from cross-sectional datasets highlighted the variation in capability by location. This underlines the importance of including location-based fixed effects in subsequent quantitative models (see Sections 5.2 and 6.2). Additionally, considering district mean capability scores in the 'Panel set' supported the decision to examine change in ranking (as opposed to raw means) in later analysis. This was partly because variance contracted substantially from 2013 to 2017 (the two years used in the longitudinal model: Section 4.4.1).

The final section in this chapter provided descriptive information on the relationship between mathematics capability and exposure to Ubongo Kids. Districts placed in high or medium exposure tertiles had marginally higher mathematics rankings in post-exposure time points than in 2013 (just before the launch of Ubongo Kids). Additionally, cross-sectional data provided an initial indication of an association. This was particularly true when considering district mean exposure against mathematics capability across Tanzania (using the 'Household set'). It was argued, however, that such a pattern was to be expected within the 'Household set' (Section 4.4.3): exposure, like mathematics capability, followed similar patterns to district population. As such, descriptive analysis of cross-sectional data suggested that Ubongo Kids exposure could influence mathematics capability, yet simultaneously reinforced the need to control for the effects of child location in subsequent inferential analysis.

5   Chapter 5 – Matched set findings

5.1   Chapter Introduction

This chapter presents analysis of a quantitative dataset concerning Ilala and Temeke. Analysis was conducted using the first of three models in this thesis that address the primary research question, which concerns the association between naturally occurring television exposure and mathematics capability. Findings were interpreted with regards to the material taught by Ubongo Kids (Section 5.3.1) and a logic-based consideration of exposure measurement (Section 5.3.2). In a subsequent chapter, results are also considered in light of qualitative data (Section 8.4).

Quantitative data were drawn from the dataset containing children from Ilala and Temeke, namely, the 'Matched set'. This dataset comprised 811 children in Ilala and Temeke who were surveyed by Uwezo in 2017 (to capture, amongst other things, child assessment and household socio-economic information), then assessed immediately afterwards in the primary data collection (to obtain additional exposure information: Section 3.3.1.1.1). In contrast with the primary research question, the characteristics of this quantitative dataset mean that findings do not concern an entire developing country. However, analysis still provides worthwhile indicative evidence on this research question, as results are representative of two large districts in Tanzania.[100]

5.2   Model findings

Quantitative analysis of the 'Matched set' used data obtained at one time point to examine the relationship between exposure to Ubongo Kids and mathematics capability. As such, a cross-sectional regression was employed that featured mathematics capability as the dependent variable and an estimate of latent exposure to Ubongo Kids as an independent variable. To help control for the possibility that the Ubongo Kids exposure variable merely reflected greater (total) media exposure or character recall ability, a latent measure of 'other media exposure' was also employed as an independent (control) variable. Additional control variables were included, to account for other factors commonly associated with child learning outcomes: age, sex, current school attendance, household asset wealth and mother's school attendance. Furthermore, the model featured enumeration area (EA) fixed effects – applied through EA dummy variables – in addition to clustered standard errors at the EA and district levels (see Section 3.3.1.4.1).[101] The coefficients for all control and treatment variables on mathematics capability are presented below (see Table 5.1 and Figure 5.1).

---

[100] The sample strategy employed by Uwezo gave a representative sample at the district-level (and, by extension, for Ilala and Temeke districts). The 'Matched set' is considered to retain this characteristic, as children who were and were not matched between Uwezo and primary datasets were found not to differ on key characteristics (see 'Chapter 3 - Methodology').

[101] EA dummies were significantly related to mathematics capability in 2 (of 59) cases.

Table 5. 1: Matched set model coefficient estimates

Independent variable	Estimate	Std. error	t value	p value	
Ubongo Kids exposure	0.093477	0.013096	7.1377	0.000	***
other media exposure	0.163344	0.07546	2.1646	0.031	*
age	0.134786	0.009252	14.5686	0.000	***
sex (female)	0.010776	0.032583	0.3307	0.741	
asset wealth	0.141355	0.288704	0.4896	0.625	
mother's school attendance	0.253188	0.037516	6.7488	0.000	***
school enrolment	0.28493	0.056026	5.0857	0.000	***

The above table presents the coefficient estimate for each independent variable, along with its standard error, t value and p value. On the right-hand side of the table, symbols are provided to indicate the level at which each coefficient estimate is significant. These are: '***' for 0.001, '**' for 0.01, '*' for 0.05, '.' for 0.1, and ' ' for 1.
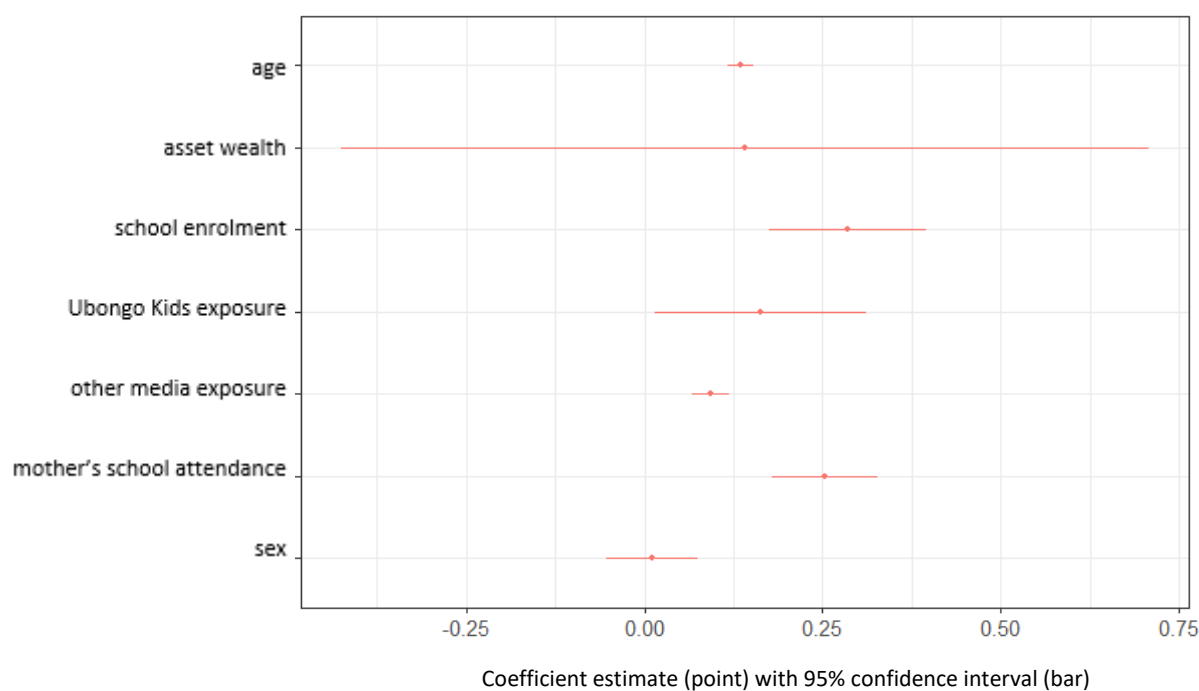
Figure 5. 1: Plot of Matched set model coefficient estimates

5.2.1   Control variable coefficients

All control variables included in the model were found to be positively associated with mathematics capability. That is, increases in the continuous measures of household wealth, age and exposure to other media led to increases in the dependent variable. Similarly, coefficients for the binary variables included in the 'Matched set' model were exclusively positive: higher mathematics capability was

associated with a child being female, being enrolled in school at the time of assessment and having a mother who previously attended school. The positive association between each control variable and mathematics capability was significant (at the $p = 0.05$ level), except for the binary measure of gender and continuous measure of household wealth. This is shown by the p-values and confidence intervals relating to coefficient values (see Table 5.1 and Figure 5.1).

Coefficient estimates were generally considered logical, given their correspondence with the identified literature. The significant coefficient estimate for general media exposure corresponded with the findings of Borzekowski and Macha (2010), where a score reflecting the number of media characters recognised was shown to correlate with mathematics outcomes (see Section 3.3.1.3.3). It is likely that the positive coefficient associated with 'other media exposure' reflected the effects of "current media exposure levels" as well as "a child's cognitive ability to recall [media figures]" (which were cited as reasons for including this control by Borzekowksi and Macha, 2010, p. 301). Indeed, the role of 'other media exposure' in controlling for cognitive recall ability (which is itself likely related to mathematics capability) was considered pivotal to creating a regression model that provided worthwhile information on the effects of Ubongo Kids exposure (Section 3.3.1.3.3).

Other coefficient results also mirrored the findings of previous research. Analysis of Uwezo data has suggested school attendance, age and mother's education to be positively related to learning outcomes (see Alcott and Rose, 2015; Jones, 2017; and, Alcott and Rose, 2016, respectively). Even the small and non-significant positive coefficient for gender (female) appears to be supported by prior research that has involved Uwezo data. Jones and Schipper (2015, p. 23), for example, found that "regression estimates consistently indicate that girls slightly outperform boys on the Uwezo tests".[102]

However, the lack of an apparent relationship between a child's household asset score and mathematics capability does not appear to correspond with the identified literature. This conflicts with previous research involving Uwezo data, where wealth has been identified as a key determinant of learning outcomes (see, for example, Mugo et al., 2015). However, it is proposed that the lack of significance for this independent variable was most likely influenced by sample size and, crucially, the application of EA-level fixed effects (within which there was limited variation in household wealth). The non-significant - albeit positive - relationship between household wealth and mathematics capability does not, therefore, warrant further discussion.

5.2.2 Treatment variable coefficient

The finding of most relevance to the primary research question was that Ubongo Kids exposure was positively related to mathematics capability, after controlling for 'exposure to other media' and all

---

[102] It might be noted, however, that the effects of gender could differ when considering examining contexts or wealth groups (see, for example, Alcott & Rose, 2017).

other variables typically associated with learning outcomes (that were adequately captured in the dataset for analysis). As for most of the control variables, the coefficient result for the treatment variable was considered rational. This result upheld an initial hypothesis, formulated after considering the relevant literature (in which educational television was frequently identified to have a positive effect on learning amongst viewers over 30 months: see Section 2.2.1). The result was also comparable to that obtained from the Tanzania-wide cross-section model (Section 6.2). To show the association between exposure to Ubongo Kids and mathematics capability, the cross-section model was used to predict values for mathematics capability against latent exposure (plotted in Figure 5.2, using the plot_model function in the sjPlot package).
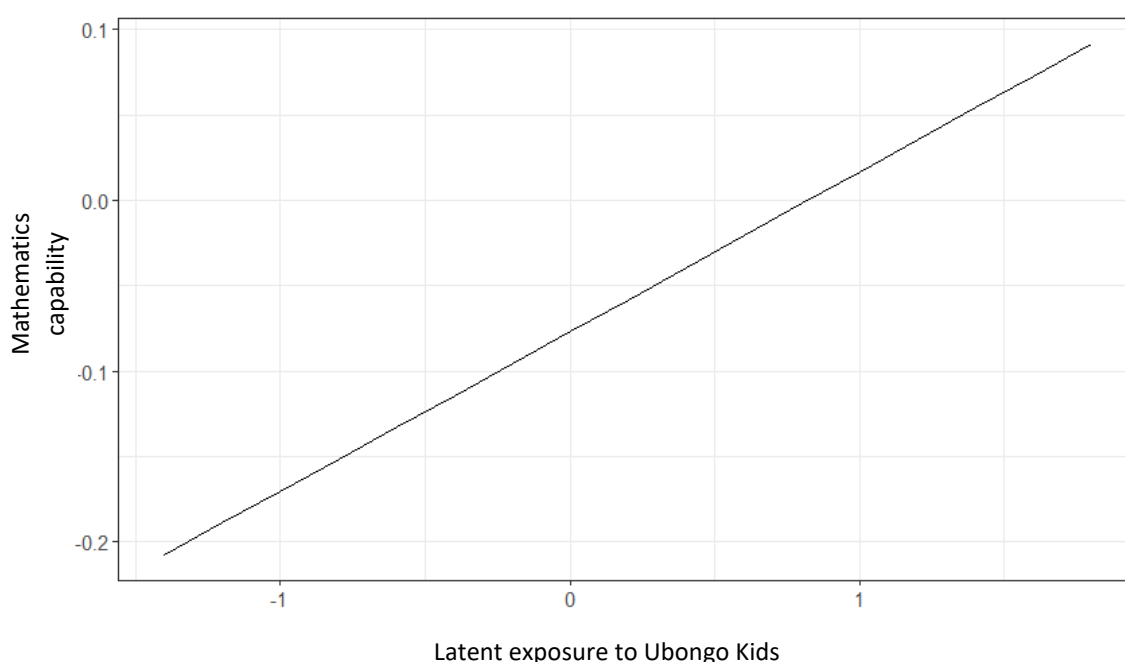
Figure 5. 2: Predicted capability values given latent exposure to Ubongo Kids

The identification of a positive relationship between exposure to Ubongo Kids and mathematics capability should, however, be treated cautiously. This model used cross-sectional information and therefore merely provided a snapshot of children in Ilala and Temeke (Cohen, Manion & Morrison, 2011). Using data from one time point meant that time invariant unobserved characteristics were not controlled for. Further, the cross-section model only employed controls for observed characteristics for which measures were available. This excluded private tutoring, which has been identified to have a positive effect on learning in different educational contexts (Alcott & Rose, 2017). In addition, some information on child characteristics related to learning outcomes that was available in the 'Matched set' was of insufficient quality to include in regression models (namely, the type of primary or secondary school a child attended and whether or not they had attended preschool: Section 4.3.1).

Furthermore, the consideration of measures used in the quantitative model suggests that the identified exposure coefficient might have been biased (see Section 5.3).

5.3   Consideration of measures

This section provides a consideration of two key concepts included in the cross-section model presented in this chapter. These concepts were exposure to Ubongo Kids and mathematics capability. A reflection on these concepts furthered understanding of regression findings by highlighting limitations in the accuracy and completeness with which they were measured. Mathematics capability is explored by considering Ubongo Kids' coverage of non-assessed topics (Section 5.3.1). Exposure is explored by considering potential shortcomings of the character recognition-based exposure measure (Section 5.3.2).

5.3.1   Mathematics capability

Non-assessed mathematics learning

The 2017 Uwezo assessment from which capability scores were derived included mathematics test items covering number recognition, counting, number patterns, addition and subtraction only. Investigation into the learning topics addressed in Ubongo Kids episodes showed that numerous episodes focused on competencies assessed by Uwezo (Appendix 2). This identification was supported by examination of an episode selected for content examination, 'Kibena and the Math Rats' (see Section 1.2.2.1.1). Kibena and the Math Rats covers the assessed topics of counting, addition and subtraction, as well as likely promoting number recognition (as certain numbers are spoken aloud as they are presented on screen). However, it was also recognised in Section 1.2.2.1 that episodes have addressed topics outside the scope of Uwezo assessment. Should it be accepted that Ubongo Kids promoted learning in mathematics domains that were not reflected in the model's dependent variable, the estimated association between Ubongo Kids and mathematics capability must be considered an underestimate.[103]

Non-assessed non-mathematics learning

While the dependent variable employed in regression models was based solely on mathematics items, Ubongo Kids might also have promoted learning in English and science. This could be the case, as science is the key focus in a small number of episodes and Ubongo Kids material is broadcast in both Kiswahili and English (Section 1.2.2.1). Non-mathematics learning based on these topics was not represented in the dependent variable for two key reasons. First, although information on English was

---

[103] The identification that Ubongo Kids material taught topics that went beyond those assessed by Uwezo is supported by test and item information examination, which suggested Uwezo mathematics questions to provide maximal information for 'Matched set' children with low to average theta (capability) levels (Section 3.3.1.4.1).

captured by Uwezo, such information was omitted from the dependent variable because it was not the focus of the primary research question (nor a key academic subject targeted by Ubongo Kids).[104] Second, information on science could not have been reflected in the dependent variable as no relevant information was available in the dataset used for analysis. Yet should any non-mathematics learning have occurred; it must again be recognised that the coefficient for Ubongo Kids exposure did not reflect all potential programme effects.

5.3.2 Exposure

An IRT-derived exposure estimate was employed as the treatment variable in analysis of the dataset for Ilala and Temeke. This estimate drew on results from five character recognition items. While character recognition results have been considered to provide a strong measure of exposure (Section 3.3.1.3.2), examination of the primary research question might still be benefited by exploration of findings with recourse to logic-based arguments. This section begins by considering whether one line of reasoning could point to inaccuracy in the quantitative exposure measure employed, before investigating how another line could suggest the measure of exposure to be relatively coarse.

Possible inaccuracy in exposure measurement

Logic-based reasoning suggests it to be possible that there was inaccuracy in the measurement of exposure used in the 'Matched set' model. This reasoning is based on the idea that children who have been exposed to Ubongo Kids might have 'underperformed' when assessed during primary quantitative data collection or, at least, have failed to positively identify all Ubongo Kids characters that they have encountered. Should either of these outcomes have occurred, it could be contended that exposure was underestimated. An underestimated exposure variable would itself have downwards biased the identified association between exposure and mathematics capability.

The identified association could have been biased downwards in this instance, as the quantitative measure of exposure was not truly continuous (despite IRT scores appearing to follow a normal distribution and being treated as continuous: Section 3.3.1.4.1). This was shown by 146 children possessing the equal lowest exposure score. Underestimation of exposure could have led to children with zero or low levels of actual exposure all obtaining equally low exposure scores. This would have reduced the apparent association between Ubongo Kids exposure and mathematics capability: presuming that (low levels of) exposure promote mathematics learning, the child's mathematics

---

[104] Indeed, including learning outcomes that were not specifically targeted by Ubongo Kids in a more broadly formulated dependent variable (DV) could have increased the likelihood of bias. This is because there would be a greater possibility of attributing DV variation associated with unobserved variables correlated with Ubongo Kids exposure, to Ubongo Kids. Further, adding information to the DV on subjects not specifically targeted by Ubongo Kids could have added additional (downwards) bias, as the apparent effect may have been diluted.

capability score would have been increased by exposure that was not reflected in their quantitative exposure estimate.

Possible coarseness in exposure measurement

Other limitations regarding the character-based measure of exposure should be recognised. These predominantly concern the coarseness of this measure. A review of the literature highlighted a lack of detail by showing that a broad range of factors could alter the influence of educational television exposure. It was noted that co-viewing and co-discussion could increase the benefits of exposure (see Section 2.2.2). Similarly, the platforms used to access Ubongo Kids might modify the effect of the programme (ibid). However, none of these concepts were reflected in exposure scores derived solely from responses to character recognition items.

Should the exposure of children in the 'Matched set' have been non-uniform then random bias would likely have occurred. Child viewing would have been non-uniform if, for example, child viewers both had and had not typically watched the show with a caregiver. This would have exerted random bias on exposure, as there is no reason to believe that differing exposure levels among children relate to them spending differing proportions of time co-viewing. The coarseness of the character-recognition based measure could therefore be considered to have imposed a non-directional bias on exposure estimates and, by extension, the identified association between exposure on mathematics capability.

5.4   Chapter summary

This chapter has cautiously recognised that findings from the 'Matched set' regression model suggested Ubongo Kids exposure to be associated with mathematics capability in Temeke and Ilala. The reasons for this caution include that the cross-section model employed did not control for mathematics capability prior to the intervention, or for certain other measures associated with learning outcomes (such as private tutoring and type-of-school attended). What is more, a logic-based investigation into key variables used in the regression model suggested that they were prone to inaccuracy. Shortcomings in the measurement of exposure were found likely to exert random and directional (downwards) bias on the identified coefficient for exposure. Additionally, the identification that Ubongo Kids could have delivered learning benefits not captured by Uwezo assessment suggested that the mathematics capability variable did not reflect all the potential effects of exposure.

While acknowledging these limitations, it should be emphasised that Ubongo Kids was identified to be positively related to mathematics capability through a study design that – in line with my philosophical perspective – sought to maximise objectivity. The regression model employed led to the finding of a significant result for Ubongo Kids exposure, even with the application of numerous

important controls and EA-level dummies (see Section 5.2.2). This finding might be considered more striking in light of investigation into key variables. While this investigation highlighted potential measurement bias, reasoning generally suggested the coefficient for exposure to be biased downwards (i.e., mis-measurement was more likely to lead to non-rejection of the null hypothesis that Ubongo Kids exposure was not associated with mathematics capability, or, a type II error). This reasoning and overall model findings are also broadly supported by thematic analysis presented in Chapter 8. Thus, both 'Matched set' model findings (Section 5.2) and qualitative results (Section 8.5) generally correspond with the argument that educational television exposure benefits mathematics capability.

6 Chapter 6 – Household set findings

6.1 Chapter introduction

This chapter presents information from the second of three quantitative models (initially described in Section 3.3.1) used to address the primary research question, which considers the association between educational television and mathematics capability within a developing country context. It is contended that this chapter provides the strongest information on the primary research question. This is because the quantitative model was benefited by using a large national sample, the size of which permitted application of household-level fixed effects (Section 3.3.1.4.1). The use of household fixed effects controlled for a large proportion of unobserved differences between children (in relation to the EA fixed effects applied to the 'Matched set' model: Section 5.2). The interpretation of results was supported by logic-based reasoning focusing on specific variables employed in the fixed effects model.

As in the previous chapter, Ubongo Kids was used as a vehicle through which to examine educational television in its entirety (see Section 5.1). However, in contrast to Chapter 5 ('Matched set findings'), results are directly relevant to an entire developing country, Tanzania. This is because the data employed in quantitative analysis (the 'Household set', n = 39,717) were derived from the Uwezo 2017 dataset. Uwezo obtained this dataset in a manner that provided information representative at the (district and) national level for children aged 6-16.[105]

6.2 Model findings

To examine the relationship between mathematics capability and naturally occurring exposure to Ubongo Kids across Tanzania, quantitative analysis used a fixed effects regression model with numerous controls. A cross-section regression approach was employed, given that analysis concerned national data collected at one point in time. Because of the lack of prior information on mathematics capability for individuals in the 'Household set', it might have been possible that the effects of Ubongo Kids were conflated with general intelligence. This could have occurred if, for example, more intelligent children were more likely to view Ubongo Kids, due to the probable relationship between intelligence and mathematics capability. To reduce the likelihood of such conflation, a control was specified for an intelligence proxy that Ubongo Kids was considered unlikely to impact – Kiswahili attainment. In addition to a variable for Kiswahili attainment, the regression model controlled for other child-level variables that have previously been identified to influence educational outcomes.

---

[105] Given that my manipulations involved excluding children without assessed siblings, it could be argued that claims of results being nationally representative should be made cautiously. Further, obtaining representative findings from the Uwezo data should involve the application of sample weights. These were not applied as the weights provided by Uwezo are applicable only to analysis involving the entire sample (which would include children with no assessed siblings). This said, the application of sample weights during trialling suggested that these have no substantive effect on regression coefficients.

These were age, sex and current school attendance. Due to the employment of household-level fixed effects, asset wealth and mother's education were not included as control variables given that they did not differ amongst children in the same households (c.f., Section 5.2).

It should also be noted that the variables for certain concepts differed from those included in the model presented in Chapter 5 (Section 5.2). Estimates of mathematics capability were created from child responses to the same questions, yet children from the data used in this chapter received different IRT scores to those from the data used in Chapter 5 (even if they provided the same responses to all mathematics questions). This is because IRT calculations were influenced by the responses of all children within samples (which differed between Chapter 5 and Chapter 6). In this chapter the measure of exposure was different to, and arguably less rich than, the estimate used in analysis of the 'Matched set' data (featuring in Chapter 5). The only measure of exposure available for all children in the secondary dataset employed in this chapter was a binary report of recent viewership. There was also no measure of 'other media exposure', which was again obtained through primary data collection and hence only relevant to Chapter 5 analysis. Findings are presented below (see Table 6.1 and Figure 6.1).

Table 6. 1: Household set model coefficient estimates

Independent variable	Estimate	Std. error	t value	p value	
Ubongo Kids exposure (binary)	0.126	0.035	3.631	0.000	***
Kiswahili attainment	0.950	0.024	39.940	0.000	***
age	0.063	0.002	26.768	0.000	***
school enrolment	0.193	0.021	9.067	0.000	***
sex (female)	-0.002	0.007	-0.277	0.782	

The above table presents the coefficient estimate for each independent variable, along with its standard error, t value and p value. On the right-hand side of the table, symbols are provided to indicate the level at which each coefficient estimate is significant. These are: '***' for 0.001, '**' for 0.01, '*' for 0.05, '.' for 0.1, and ' ' for 1.
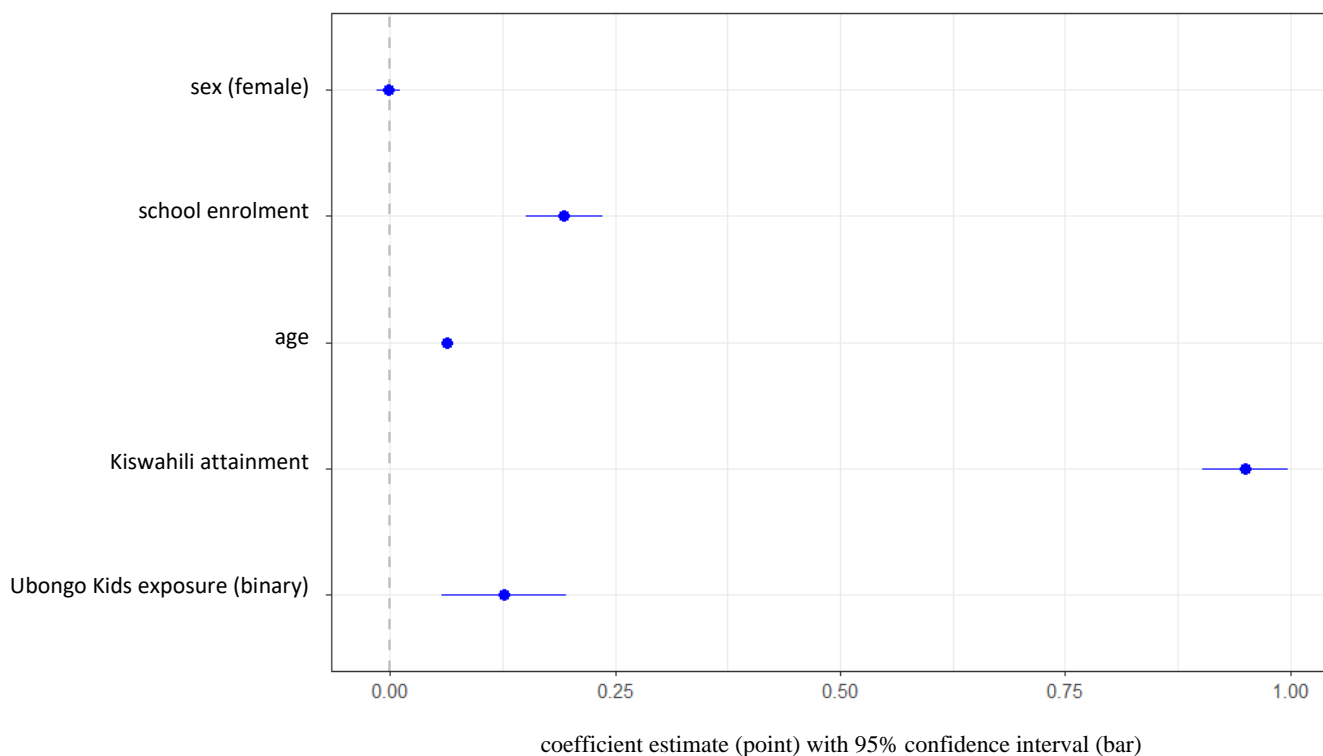
Figure 6. 1: Plot of Household set model coefficient estimates

6.2.1   Control variables

The identified control variable coefficients were broadly as anticipated. Increases in age and school enrolment were both positively and significantly related to mathematics capability. The coefficients for age and school enrolment corresponded with results from the 'Matched set' in Chapter 5 (see Section 5.2.1). These also mirror findings from the previous literature (see Jones, 2017, regarding age; and, Alcott and Rose, 2015, regarding school enrolment). Additionally, the proxy for general intelligence provided by Kiswahili attainment performed as expected. That is, general intelligence was found to be strongly related to mathematics capability. It could be noted that the Kiswahili attainment and mathematics capability variables were measured with error and, because both were derived from tests taken on the same day, those errors might have been correlated. The impact of this on the exposure coefficient is unknown. However, if the Kiswahili attainment coefficient was attenuated downwards due to measurement error, this would have led to upward bias in the exposure variable as the model would not have fully controlled for children's general intelligence.

In contrast with the coefficients for age, school enrolment and intelligence, the coefficient for sex might appear to conflict both with prior research in East Africa and other quantitative analysis in this thesis. Previous literature has suggested that girls typically outperform boys on Uwezo tests (Jones and Schipper, 2015). Additionally, the regression model applied to the dataset for Ilala and Temeke in

the previous chapter suggested (female) gender to have a positive – albeit non-significant – influence on mathematics capability (see Section 5.2.1). This said, the small negative coefficient for (female) gender on mathematics capability could be explained by the inclusion of the Kiswahili attainment variable. Investigation of the national dataset employed in this chapter showed that girls also strongly outperformed boys in Kiswahili. The inclusion of this variable as a control might therefore have reversed the effect of gender initially anticipated, since the gender effect worked through the Kiswahili variable rather than directly impacting mathematics attainment.

6.2.2   Treatment variable

While the regression model suggested that Ubongo Kids exposure was positively and significantly associated with mathematics capability, there are reasons to treat this finding with caution. Firstly, it was recognised that the quantitative model was not applied to a dataset which included child mathematics information from a pre-intervention timepoint. This increased the possibility that the apparent effect of Ubongo Kids was conflated with non-observed variables associated with mathematics capability (c.f., below), such as cognitive ability. Additionally, (limited) data availability meant that key child-level characteristics commonly associated with mathematics capability (such as private tutoring attendance) were not included in the model – a problem which also applied to analysis of the 'Matched set' (see Section 5.2).

This said, the model employed has many positive facets. The employment of household fixed effects likely gives the strongest means of obtaining inferential results from analysis of Uwezo data (with such an approach advocated by Jones, 2017; and Bietenbeck; Ericsson & Wamalwa, 2017). This fixed-effects approach controls for all observed and unobserved household-level differences between children. Further, key child-level characteristics were included as controls to address those differences that remained. These included a variable for Kiswahili attainment, which would have gone some way to control for children's varying motivation or non-mathematics capability (that otherwise might have been conflated with the potential effects of Ubongo Kids).[106] Lastly, the size of the Ubongo Kids coefficient (relative to other independent variables) meant that a significant finding could still have been identified even if the acknowledged model shortcomings tended to bias findings upwards. Table 6.1 suggests that, in this model, exposure to Ubongo Kids had approximately the same effect on mathematics capability as a child being two years older.[107]

---

[106] As recognised above (Section 6.2.1), however, error in the measurement of Kiswahili attainment could have diminished the extent to which this proxy for general intelligence provided an effective control.

[107] Still, when using results from the model in this chapter to approximate the standard deviation gain attributable to Ubongo Kids, the finding for this programme is small in relation to other interventions (see Section 9.6).

6.3 Consideration of measures

This section provides a consideration of two measures used in the model that featured in this chapter. These concepts were mathematics capability and Ubongo Kids exposure. As in Section 5.3, a reflection on the variables used in statistical analysis enhanced the understanding of findings. This reflection aided understanding by showing the variables employed to be limited in both depth and precision. Mathematics capability is considered by examining whether Ubongo Kids might have promoted learning in topics outside of the Uwezo assessment (Section 6.3.1). Exposure is considered by exploring whether the self-reported measure of exposure might have led to bias (Section 6.3.2).

6.3.1 Mathematics capability

The fixed-effects model employed in this chapter used capability scores derived from the Uwezo mathematics test (as in the model employed to investigate Ilala and Temeke data: Chapter 5). Following investigation in Section 5.3.1, it is considered probable that Ubongo Kids content imparted some knowledge on viewers that would not have been reflected in capability scores. Children could have furthered their knowledge of mathematics topics outside of the number recognition, counting, number patterns, addition and subtraction assessed by Uwezo (Section 5.3.1). Additionally, child viewers might have learnt science or English material (ibid).

Should learning in topics outside of the Uwezo assessment have taken place, application of the regression model in Section 6.2 would have underestimated the association between Ubongo Kids exposure and mathematics. This is because the regression model accounted for changes in capability, which was solely derived from Uwezo mathematics item responses. As such, quantitative analysis would not have accounted for any effect that Ubongo Kids had on learning in mathematics subjects outside the Uwezo assessment or any non-mathematics topics. The regression coefficient for Ubongo Kids might therefore fail to reflect all the possible effects that this programme had on learning.

6.3.2 Exposure

This chapter now provides logic-based arguments concerning exposure to Ubongo Kids. The exposure variable employed in this chapter was based on responses to a binary question on viewership, "did you watch Ubongo Kids in the last week?" (see Section 3.3.1.3.2). The responses recorded by Uwezo ("yes" or "no/do not know") clearly gave limited information. However, logic-based reasoning provides an opportunity to further consider this measure. First, reasoning is used to provide information on how the binary survey question might itself lead to inaccuracy in the measurement of exposure. Next, the implications of this inaccuracy on regression model findings regarding the coefficient for Ubongo Kids exposure are considered.

Reasoning primarily suggests that the exposure information captured by Uwezo was biased downwards. This is for two key reasons. First, children could have received some level of exposure to Ubongo Kids while not having watched in the past week. If this were to have occurred, children could

have faithfully responded negatively to the survey question despite having viewed the show on some previous occasion. Second, children might not have recognised the show when it was referred to by the Uwezo assessor. As a result, they could again have responded negatively when asked whether they had watched Ubongo Kids, even if the show had been watched in the past week.

Beyond the directional bias suggested by the reasoning above, quantitative exposure data was also susceptible to random bias. This was attributable to the lack of detail allowed by the binary measure. Neither of the binary response options gave detail on potential variation amongst viewers with regards to the way in which they watched or discussed the show. Similarly, differing viewership environments and media platforms were not reflected. Both these points suggest that the binary exposure proxy provided limited nuance.[108]

A lack of nuance and the likelihood that some exposure was missed would have acted to bias the exposure coefficient. Capturing information lacking in nuance from treated children was likely to have exerted *non-directional bias* on the coefficient for exposure. To demonstrate this point, the type of platform (e.g., television, tablet or phone) that the child used to watch Ubongo Kids is briefly considered. The use of differing platforms would have changed the nature of exposure and, most likely, had a non-zero effect on its educational benefits. However, there is no way of establishing which platform was used by each child who reported exposure. Indeed, while multi-platform use could influence learning (see Fisch, Lesh, Motoki, Crespo & Melfi, 2010), a search of the literature uncovered no evidence on the relative effectiveness of different educational media platforms in developing country contexts (Section 2.2.2). This shows that the binary exposure proxy's lack of detail was likely to have influenced the association between exposure and mathematics capability (identified in the regression model) in a random manner.

However, should the measure of exposure have been biased downwards, the association between Ubongo Kids exposure and mathematics capability would have tended to be an underestimate. Those who responded negatively to the question concerning Ubongo Kids viewership under Uwezo assessment could have received some exposure. This exposure could have occurred in the week leading up to the Uwezo survey or at any time before. The failure to recognise exposure would have led to an underestimation of the exposure coefficient, as children who had benefited from Ubongo Kids viewing would have been treated as not exposed.

---

[108] Pertinent information concerning this limited nuance was provided in Section 5.3.2 (which remains relevant here, despite the Ilala and Temeke model using a different measure of exposure).

6.4   Chapter summary

This chapter has provided further support for the notion that exposure to Ubongo Kids is positively related to mathematics capability. This was initially demonstrated through regression model findings, which showed a significant association. It was recognised, however, that these findings should be treated cautiously. This caution was recommended, as the regression model did not control for lagged mathematics capability, school-type or tutoring.

This being said, analysis of the data used in this chapter was conducted through what is likely the strongest quantitative model employed in this thesis. A large amount of unobserved heterogeneity was accounted for, through the application of household-level fixed effects. Further, the model controlled for key child characteristics for which there was within-household variance. Controls included a proxy for non-mathematics intelligence, which would have acted to limit the bias imposed by the possibility that children with higher levels of motivation or overall capability were more likely to watch Ubongo Kids. What is more, logical reasoning suggested that the statistical model could have underestimated the coefficient for Ubongo Kids. It was highlighted that the dependent variable could have led to underestimation of the total effects of exposure, as this did not reflect potential learning gains in non-assessed mathematics, English or science topics. Additionally, it was posited that the measurement of exposure could have been biased by missed cases of viewership. Should either of these possibilities have occurred, the identified association between mathematics capability and exposure should be considered an underestimate.

7 Chapter 7 – Panel data findings

7.1 Chapter introduction

The model presented in this chapter is again centred on examining the primary research question, which concerns the relationship between educational television and mathematics capability in a developing country context. This model is the third (and final) model presented that concerns the primary research question. Using multiple models meant that the main research question was investigated as fully as possible. Indeed, the model presented in this chapter was particularly helpful in this regard as it differed markedly from those in Chapter 5 and Chapter 6. In the previous two chapters, models were employed that used data from one post-intervention time point. This chapter features a multilevel longitudinal model that drew on Uwezo survey data from pre- and post-intervention periods.

The use of longitudinal data means that this design is, in some respects, superior to those featuring in previous chapters, which relied on cross-sectional data. In presenting these cross-section designs, it was acknowledged that the apparent positive effect of watching Ubongo Kids could be attributable to unobserved factors that a) were correlated with watching the cartoon, and b) increased mathematics capability. One example posited was the (unobserved and fixed) educational motivation of children, that might have both been correlated with exposure and related to mathematics outcomes. The advantage of a panel data approach is that it supports models which can allow for such differences between study members. This is a key justification for including a longitudinal approach in this thesis, which provides information on whether changes over time in district capability ranking are related to differing district levels of Ubongo Kids exposure.

However, it should be noted that the characteristics of the longitudinal dataset used for analysis in this chapter – the 'Panel set' – impinges on the ability of the model to uncover a causal link between exposure and outcomes. The 'Panel set' was derived from information collected by Uwezo in 2017 and 2013. (Descriptive analysis also presented in this chapter additionally incorporated data derived from the Uwezo 2015 and 2011 surveys.) Given that children were anonymised in Uwezo surveys prior to 2017, it was not possible to follow individual children through different time points. As such, change over time was considered using aggregated district averages.[109] There were 45 districts in common between the 2017 and 2013 surveys, so the 'Panel set' had a sample which only included these districts. (The same 45 districts were also common across all time points when including the Uwezo 2015 and 2011 surveys for the purposes of descriptive analysis presented in this chapter.)

---

[109] District averages were considered to provide accurate measures at each time point, as Uwezo captured representative district samples in all years.

Beyond the limited sample size, a comparison of change over time was hindered by limitations in the measures used for modelling (Section 7.3).

7.2 Model findings

Notwithstanding data limitations, the quantitative model was specified in a manner to give the most accurate estimation of treatment effect possible. District mathematics capability ranking was used as the dependent variable that changes over time. This was regressed on an interaction effect and multiple main effects. The main effects included district mean age, school enrolment, gender, school attendance and mother's school attendance (see Section 3.3.1.3) in addition to time.[110] These measures changed across time points, due to changes in district composition across Uwezo surveys. Controlling for these time variant measures meant that the model was less affected by fluctuations in district characteristics. Additionally, the model included a fixed (main) effect for district mean level of Ubongo Kids exposure (measured in 2017). This variable played a key role in controlling for the baseline differences between districts, with these differences shown through the descriptive tertile-exposure-group plot presented below (see Figure 7.1). Lastly, information on the influence of exposure was provided through an interaction effect between district self-reported viewership (measured in 2017) and time (with 2013 specified as time point 0 and 2017 as time point 1).

All variables detailed above were included in a multilevel design featuring random group (i.e., district) intercepts.[111] While a multilevel design produced the same coefficient results as a difference in difference (DiD) analysis of the 'Panel set', the former accounted for clustering by design (so did not require the specification of clustered standard errors). Indeed, multilevel approaches give a more accurate estimation of errors values than DiD approaches with clustered standard errors (Cheah, 2009). Findings are presented below (see Table 7.1).

---

[110] Fixed effects did not include a measure of asset wealth. This was because there was no consistent measure of this concept between 2017 and 2013 data.

[111] A random slopes model could not be employed, as this was not supported by the data.

Table 7. 1: Panel set model – independent variable coefficients

Independent variable	Estimate	Std. error	t value	p value	
mean Ubongo Kids exposure	34.997	8.814	3.971	0.000161	***
time	5.276	3.514	1.501	0.137602	
mean Ubongo Kids exposure*time	1.000	10.376	0.096	0.923654	
mean school enrolment	25.561	14.382	1.777	0.079222	.
mean sex (female)	54.848	33.627	1.631	0.106765	
mean mother's school attendance	31.335	8.362	3.747	0.000341	***
mean age	14.229	4.637	3.069	0.002915	**

The above table presents the coefficient estimate for each independent variable, along with its standard error, t value and p value. On the right-hand side of the table, symbols are provided to indicate the level at which each coefficient estimate is significant. These are: '***' for 0.001, '**' for 0.01, '*' for 0.05, '.' for 0.1, and ' ' for 1.

7.2.1 Main effects

The identified main effects broadly corresponded with previous research. District measures with higher mean ages ranked higher in terms of mathematics capability (see, for example, Jones, 2017). The same was true for district measures in which a higher proportion of children's mothers had attended school (Alcott & Rose, 2016). The effects of district mean child gender (female) and school enrolment were also positive. Here, the directions of coefficients were in line with previous research (see Jones & Schipper, 2015; and, Alcott & Rose, 2016, concerning gender and enrolment, respectively). However, the effects of these variables on capability ranking were not significant. These findings were likely influenced by sample size – an issue which might also have affected results concerning the interaction between time and exposure (see below). In this model, the main effect for time also exerted a non-zero (albeit insignificant) effect on district capability rank. This accounted for the differing effects of other (time variant) independent variables in 2013 and 2017.

Lastly, the main effect for mean district exposure (measured in 2017) was positive and significant. Given that this finding occurred in a model that included an interaction effect between exposure and time, this suggests that the measurement of district exposure in 2017 encompassed other unobserved variables associated with capability rank. To demonstrate this point, the differences in capability ranking between districts when grouped into exposure tertiles are presented (see Figure 7.1, which was originally presented in Chapter 4, Figure 4.19).[112] This plot highlights how controlling for exposure dosage at the pre-intervention time point played an important role in addressing potential

---

[112] It was also recognised in Section 3.3.1.4.3 that district rankings prior to the launch of Ubongo Kids might suggest that an assumption of longitudinal analysis is violated. That is, while intervention allocation should be unrelated to baseline dependent variable results, districts in the high exposure tertile typically have higher mathematics capability rankings.

omitted variable bias. Indeed, this finding warranted the inclusion of more extensive controls in cross-section models (where data from previous time points could not be included).[113]
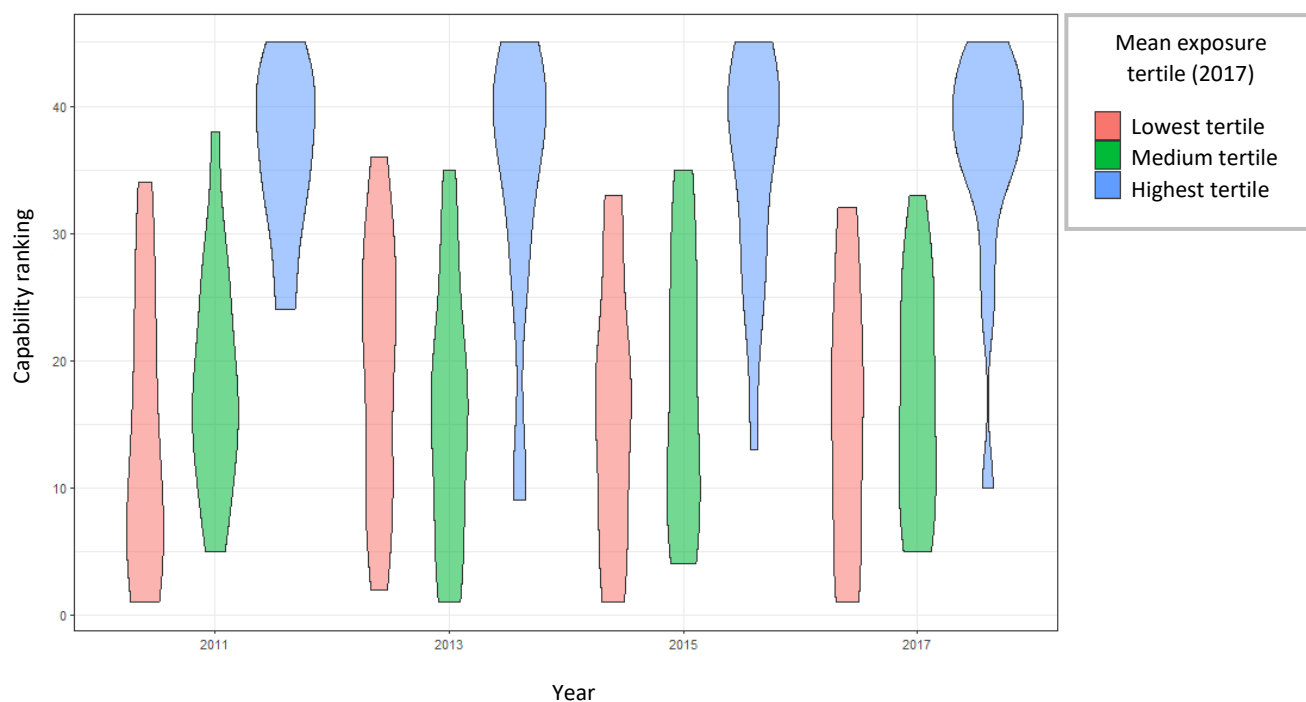
Figure 7. 1: Exposure tertile violin plot

7.2.2   Interaction effects

The interaction between time and exposure gave information on the effect of treatment. Specifically, the coefficient for this interaction showed the change in district mathematics capability ranking between 2013 and 2017 attributable to the percentage of children in each district who reported viewing Ubongo Kids (in the week before Uwezo's 2017 survey). This apparent effect of Ubongo Kids exposure on mathematics capability was positive. To show this effect, the multilevel model was used to predict capability rankings (using the predict function in R). District predicted ranks at time points 0 (2013) and 1 (2017) are plotted on a graph featuring lines that are coloured according to mean district exposure (using the R's ggplot function: see Figure 7.2).

---

[113] In Chapter 6 ('Household set findings'), for example, a proxy for general intelligence was included to control for the possibility that reported Ubongo Kids viewing was associated with unobserved variables associated with mathematics capability (such as child motivation: see Section 6.2).
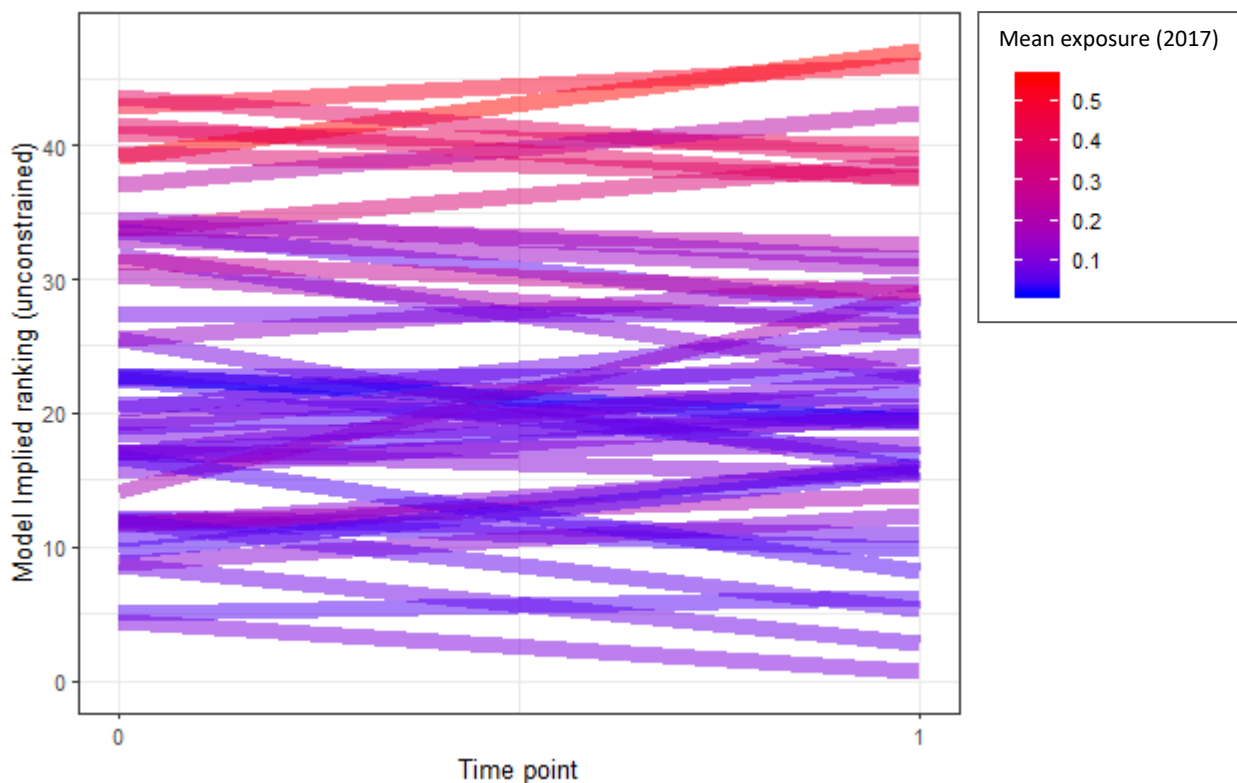
Figure 7. 2: Predicted district ranking

The identification of a positive coefficient appears to mirror descriptive information on the changes in rank of exposure tertiles between 2013 and 2017 (see Figure 7.1). It also complies with findings from the literature on television in developing contexts (Borekowski & Henry, 2011; see further Section 2.2.1). Additionally, the coefficient direction corresponds with findings from the cross-section models considered in Chapter 5 and Chapter 6. Yet, unlike the models discussed in the previous two chapters, the multilevel model considered here did not suggest educational television exposure to promote mathematics capability. The lack of a significant result might, in itself, suggest that findings from the cross-sectional models in this study should be queried.

However, the effect identified through this multilevel approach must be considered in light of various factors. Firstly, models concerning change over time give less variation than cross-section models and hence are more likely to produce insignificant results. Secondly, the limitations of the panel dataset used in this chapter could have reduced the chance of obtaining a significant finding still further. A key limitation that might have caused an insignificant result was the use of a small sample size (n = 45). A Third important factor is provided by limitations in the measures used (Section 7.3).

7.3   Consideration of measures

This section gives information on the exposure and mathematics variables employed in the multi-level model. As in Sections 5.3 and 6.3, exploring these two measures allows model findings to be better understood. In this instance, the non-significant interaction effect between Ubongo Kids exposure and mathematics rank could partially be explained by limitations in the measurement of both variables. The rankings-based mathematics variable is considered in Section 7.3.1, while the district exposure measure is considered in Section 7.3.2.

7.3.1   Mathematics rank

Mathematics ranking in 2013 and 2017 was based on Uwezo mathematics testing. As such, the multi-level model remains susceptible to shortcomings in the measurement of mathematics in 2017 (recognised in Section 5.3.1 and 6.3.1). That is, sampled children in selected districts who had been exposed to Ubongo Kids might have increased their knowledge of topics outside of those assessed by Uwezo (in 2017: Section 5.3.1). This point applies similarly to 2013 data, where mathematics testing was confined to number recognition, counting, number patterns, addition, subtraction and multiplication (with the latter not included in Uwezo 2017 testing: Section 3.3.1.4.3). As numerous topics taught by Ubongo Kids were not assessed in both 2013 and (particularly) 2017, change in district mathematics capability due to viewing will have been underestimated.

What is more, the fact that the Uwezo 2017 test provided even less information for test takers with higher capability levels than the Uwezo 2013 test promoted an increase in ceiling effects over time (Section 3.3.1.4.3). This feature of the available mathematics information led to the use of district capability rankings in statistical analysis. Retaining 'raw' district means would otherwise have biased analysis, as increasing ceiling effects caused a fall in mean district capability amongst higher performing districts (from 2013 to 2017: Figure 3.2.1). However, employing a rank-based measure acted to reduce precision. This is because changes in mean district capability over time would not influence results should they have no effect on district rank (Section 3.3.1.4.3). The limited precision of the mathematics dependent variable would have reduced the chances of identifying a significant result from the multilevel model.

7.3.2   Exposure

Logic-based reasoning was also used to inform the interpretation of exposure. The proxy for exposure featuring in the multilevel model was based on child reports of recent viewership captured by Uwezo. Specifically, the mean of responses from all children in each selected district was employed to create a continuous district level measure. This measure was incorporated directly into the multilevel approach (see further Figure 4.16, which plots the district exposure values used in analysis). Section 7.3.2 examines this exposure variable in various manners. First, there is consideration of the variation

in exposure and whether this might have been underestimated. Second, the implications of this potential variation and underestimation are discussed.

The self-reported viewership measure on which the district exposure proxy was based provided coarse information. Following Section 6.3.2, it is acknowledged that those who reported exposure in different districts might therefore have been exposed in quite different manners. For example, a report of recent viewership provided no data on co-viewing or episode discussion during or after viewership. What is more, the level of exposure by district was likely inconsistent over time. Changes in exposure intensity over time would also not have been reflected in the exposure measure, as this was based solely on information captured in 2017 (Section 3.3.1.4.3).

Additionally, it could have been the case that exposure was affected by directional bias. The reasoning presented in Chapter 6 ('Household set results') suggested it to be possible that the employment of a binary measure of viewing would have led to exposure being underestimated (Section 6.3.2.3). It was argued that this could have occurred as a) some viewers might not have understood their assessor's reference to Ubongo Kids, and b) those who had not watched in the past week could have received exposure before this. Should the first of these two possibilities have occurred, a child would (inaccurately) have not reported viewing the programme during Uwezo assessment. Alternately, if a child had watched Ubongo Kids on numerous occasions but not in the week prior to Uwezo testing, they could have (accurately) responded negatively to the question on viewership.

The potential biases identified above would have had differing implications for the identified interaction effect. Random variations in the nature of exposure – only amongst those considered to have been exposed – would have led to larger standard errors, thereby decreasing the likelihood of obtaining a significant result. Variations in the amount of district exposure over time could have produced attenuation bias leading to a lower t value, again reducing the chance of a significant finding. Conversely, systematic directional (downwards) bias in the measurement of exposure would have biased the interaction coefficient upwards but simultaneously raised standard errors, in a way that would not have affected the likelihood of identifying a significant result. While various biases would have been present, it was probable that their overall impact was to reduce the possibility of a significant interaction effect between exposure and mathematics capability.

7.4   Chapter summary

This chapter has presented a model that differed from those in the previous two chapters. The model drew on data from pre- and post-intervention time points to examine the effects of district-level Ubongo Kids exposure on district mathematics capability ranking. This chapter helped to build up information on the primary research question, by using a model that took account of district mathematics capability before the Ubongo Kids programme was launched. In this respect, the model

might be considered superior to the cross-section approaches presented in Chapter 5 and Chapter 6. The failure to identify a significant result through this model could therefore call into question the veracity of the significant findings in previous chapters.

However, other features of the multilevel model suggested it to possess limitations. It was recognised that the dependent variable failed to account for all learning potentially attributable to Ubongo Kids and was imprecise. Additionally, the treatment variable was considered likely to be downwards biased, did not account for variations in the nature of exposure and failed to reflect probable changes in treatment intensity over time. While the implications of these shortcomings differed, it was generally acknowledged that they reduced the chances of significant results. Perhaps most importantly, these measures were used in a model with a small sample size (n = 45). Indeed, the likely implications of this sample size (and coarse dependent variable) appear to have been substantial, as the finding for the main effect of school enrolment was also not significant (which is particularly at odds with existing literature: Alcott & Rose, 2015). Thus, the lack of a significant interaction effect does not necessarily disprove cross-sectional results. Indeed, the positive (albeit insignificant) coefficient on the interaction between Ubongo Kids exposure and time could even provide some tentative support for findings from other models (including the strongest quantitative approach: Section 6.2).

8. Chapter 8 – Qualitative data findings

8.1 Introduction

Chapter 8 presents information from the thematic analysis of interview data from seven children and four caregivers. This chapter begins by providing information on the role of qualitative data in this thesis (Section 8.2) as well as its strengths and weaknesses (Section 8.2.1). After this, the codes and themes generated through the thematic analysis are detailed (Section 8.3). The implications of identified themes and codes for quantitative findings are then considered (Section 8.4). In doing so, qualitative analysis is presented that supports the positive findings identified in Chapter 5, 6 and 7. Further, interviewee responses are interpreted to correspond with prior logic-based reasoning that suggested (amongst other things) that quantitative exposure measures have typically been downwards biased. This chapter concludes with a summary of the findings of qualitative analysis (Section 8.5).

8.2 The role of qualitative investigation

Qualitative investigation comprised a minor portion of this research into educational television. The qualitative data analysis presented in Chapter 8 is primarily intended to support the interpretation of quantitative results (in Chapters 5, 6 and 7). The larger-scale quantitative data are complemented and illuminated by thematic analysis of the smaller interview sample, as qualitative findings a) provide information pertinent to overall quantitative findings and b) aid the interpretation of measures employed in statistical models. The qualitative themes uncovered in this chapter also inform the cost-effectiveness analysis provided subsequently (Chapter 9).

8.2.1 Strengths and weaknesses of qualitative investigation

Interview data were obtained from four families drawn from the backup dataset collected from Dar es Salaam (described in Section 3.4.1). Within these four families, all children aged 6-16 years and one caregiver were interviewed. This gave a total qualitative sample of eleven interviewees. The characteristics of this qualitative dataset mean that findings may of course not generalise across an entire low-income country. This contrasts with the primary research question, which concerns Tanzania as a whole.

It has also been recognised that the qualitative sample did not include households featuring in any quantitative dataset employed in Chapter 5, 6 or 7 (Section 3.4.3). As such, it could not be established quantitatively whether a specific interviewee had or had not benefited from viewing. Quantitative data that could have aided the identification of exaggerated or fabricated interview responses were unavailable. Indeed, the influence of Mussa, Titus and myself as interviewers might have increased the likelihood of exaggerated or fabricated responses, should interviewers have believed we desired particular responses (Section 3.4.3).

Having acknowledged these limitations, interviews still provided data that undoubtedly benefited the interpretation of quantitative results. Interviewees possessed similar characteristics to members of all quantitative datasets, despite qualitative and quantitative samples otherwise being separate. Additionally, qualitative data collection and analysis were conducted in a way that limited bias (Section 3.4.3). Perspectives that might have been perceived to conflict with the assumed 'research desires' of Titus, Mussa or myself were sought and highlighted (e.g. Section 8.3.2). Further, emphasis is placed on excerpts that were "substantiated by concrete examples" (Hennessy, Hassler & Hofmann, 2015, p. 14). This is because both storyline and co-viewing examples likely increased the credibility of evidence concerning learning and reported exposure (Section 8.4.2.1).

8.3 Themes and codes

Themes and codes were identified through a recursive six-phase process (described in Section 3.4.3). To facilitate the presentation of findings, the final themes and upper level codes applied to the data are depicted below with themes in red and codes in blue (Figure 8.1, originally presented in Section 3.4.3).
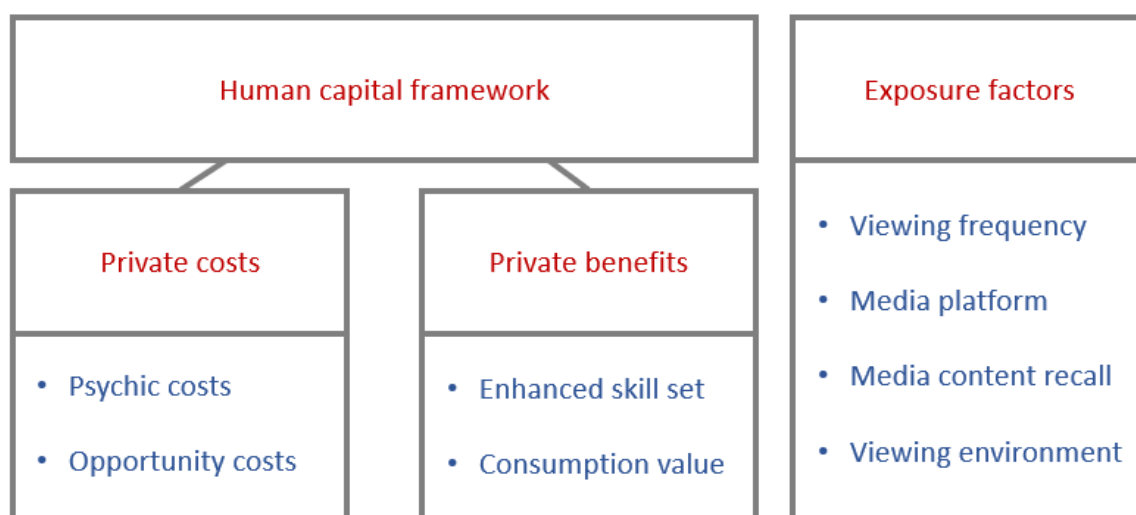
Figure 8. 1: Theme and code structure

8.3.1 Private costs

Thematic analysis identified the possible presence of private viewing costs. These centred on psychic costs and opportunity costs. Analysis indicated that both these costs were likely to be insubstantial. For psychic costs, only limited evidence was found that children suffered stress or hardship while engaging with the educational content on Ubongo Kids. Further, interviewee responses that detailed psychic costs focused on challenges with learning educational concepts in themselves as opposed to

any difficulty associated with the way they were delivered. In fact, both children and caregivers portrayed television as a comparably stress-free form of learning (see Excerpt 8.1).

Excerpt 8.1: Caregiver suggestion that cartoon learning imposes a relatively low psychic cost

Caregiver (father), 46 years, Pugu

_____

Caregiver: he's learning, so let's say Ubongo kids teaches him more than when he's in the class

…

Titus: How do you think it teaches him more than in the class?

Caregiver: Because that teaches

Titus: Mhh...h

Caregiver: In calm way

Titus: Mhh...h

Caregiver: In a gentle way

Titus: Mhh...h

Caregiver: Without forcing him

Titus: Mhh...h

Caregiver: Or doing things in a hurry

Additionally, qualitative analysis suggested that the opportunity cost of watching Ubongo Kids was often negligible. Time spent watching Ubongo Kids was frequently at the expense of playing games (Excerpt 8.2) or viewing different television programmes (Excerpt 8.3).[114]

Excerpt 8.2 Caregiver statement on what their children would be doing if they were not watching Ubongo Kids

---

[114] While Excerpt 8.3 features reference to Dora the Explorer, most alternate programmes available are non-educational.

Caregiver (father), 32 years, Pugu

_____

Joe: Aah, what could your kids do if they were not watching Ubongo Kids?

Caregiver: If they could not it could just be playing games only

Joe: Mhh

Caregiver: Only games

Musa: What kind of game, maybe, do you know that they could play?

Caregiver: Games like that – sometimes, you see, they go to a football match

Excerpt 8.3: Child statement on what they would be doing if they were not watching Ubongo Kids

Child, 11 years, Pugu

_____

Joe: Okay, [child's name], if you were not watching Ubongo Kids what would you do?

Child: If I did not watch Ubongo Kids

Musa: Mhh!

Child: I could watch other channels

Joe: Mh mm, yes

Child: I could watch other things

Musa: Like what?

Child: Other cartoons

Musa: For example, which cartoon?

Child: For example, I can search and see which cartoon I find, if I find Dora I can watch, I find any I can watch

Musa: Mhh!

Child: If I find nothing I go outside

8.3.2 Private benefits

Consumption value

While the private costs incurred through viewing were typically portrayed as limited, thematic analysis uncovered numerous indications that Ubongo Kids delivered meaningful private benefits. Firstly, interviewee responses often suggested Ubongo Kids viewing to provide a non-zero consumption value. That is, children often reported that the show entertained them (as well as educating them, see Excerpt 8.4 and 8.5). These child responses corresponded with information from adult interviewees concerning their co-viewing experiences (Excerpt 8.6) and Ubongo Kids' theory of change (which features entertainment as a key component: Section 1.2.2).[115]

Excerpt 8.4: Child report of enjoyment from watching Ubongo Kids (1)

Child, 10 years, Somangira

_____

Musa: What exactly makes you happy [or, enjoy, watching Ubongo Kids]?

Child: The way they dance

Musa: Mm! What else apart from dancing?

Child: Mh! Like when they play football in the water

Musa: Mh! What makes you enjoy seeing football in the water?

Child: They do calculations until they get the football!

Musa: When they do the calculation until they got the ball, what makes you happy… when they are doing those calculations?

Child: On maths?

Musa: Yes, the time they were doing calculations

Child: They make me happy doing maths

---

[115] The idea that Ubongo Kids provides consumption value to viewers also broadly corresponds with the qualitative data presented earlier in this section, where interviewee responses suggested that Ubongo Kids could be a substitute for other leisure activities (Excerpt 8.2 and 8.3).

Excerpt 8.5: Child report of enjoyment from watching Ubongo Kids (2)

Child, 11 years, Pugu

_____

Joe: What do you enjoy about Ubongo Kids?

Child: Mhh! what do I enjoy about Ubongo Kids?

Joe: Yes

Musa: Mh

Child: I enjoy how is teaching me

Musa: Mh! Apart from teaching what else makes you enjoy it when you watch Ubongo Kids?

Child: When I saw them playing! What! Doing body exercise I just like to watch

Excerpt 8.6: Adult report of enjoyment from watching Ubongo Kids

Caregiver (father), 46 years, Pugu

_____

Joe: Why, aah, do you think it's entertainment?

Caregiver: It's entertainment because there are animals

…

Musa: Apart from animals why do you think that is entertaining, on your side

Caregiver: Those animals were arguing one after another aah … ah! Eeh, and they make some jokes and make me enjoy.

Enhanced skill set

Qualitative data featured multiple reports concerning the enhancement of viewers' skill sets, although the strength of a thematic claim from this material should not be overstated. Interviews possessed various weaknesses (identified above: Section 8.2.1). Amongst these, it should be reiterated that there

could have been some response inaccuracy (partially attributable to the sense contributed by Mussa, Titus or myself). Additionally, negative statements relating to the 'Enhanced skill set' code were identified. One child stated that they did not learn from the show.[116] Further, information from one caregiver suggested that learning material was not always targeted appropriately for viewers (Excerpt 8.7).

Excerpt 8.7: Suggestion that some learning content was already known

Caregiver (father), 46 years, Pugu

_____

Mussa: Do you discuss anything about the program, or after the programme has finished?

Caregiver: Mhh

Mussa: Either you speaking to them, or discussing together?

Caregiver: Aah. Err, when we discuss you find many times [Child A] is the top level

Mussa: Mhh

Caregiver: So [Child A] will be found contributing and telling others that "I know that", but others learn, and Daddy (referring to himself) continues with other stuff

Still, the significance of acknowledged interview weaknesses and negative reports concerning this theme are limited. Negative reports appeared only very occasionally. Additionally, the likelihood of interview inaccuracy was restricted in multiple instances. This is because interviewees frequently provided learning claims that were embedded within storyline narrative. Responses featuring narrative were less prone to fabrication (see also Section 8.4, for information on a further positive implication of interviewee responses featuring both educational and narrative content). Positive excerpts relating to the code of 'enhanced skill set' are now provided. These centre on the topics of division (Excerpt 8.8), fractions and decimals (Excerpt 8.9) and the production of sound (8.10).

---

[116] When the child was encouraged to say why they did not learn from the show, they responded that learning did not occur because it was not their intention: "I didn't sit to learn".

Excerpt 8.8: Child recollection of division content

Child, 11 years, Pugu

_____

Child: Kiduchu and Kibena came to help Ngedere and they were helped to get a ride

Mussa: How did Kiduchu and Kibena help Ngedere?

Child: To find the number when they switched on the radio to get a song

Mussa: How did they help him to get those number on radio?

Child: Oh! [They] helped by dividing numbers. And subtraction.

Mussa: How did using division help the monkey tune his radio?

Child: Naa! They saw

Mussa: Mh?

Child: ... Using 5 divided by 10 to help him

Excerpt 8.9: Child learning claim regarding fractions and decimals

Child, 11 years, Pugu

_____

Child: Mh! one day I watched Ubongo Kids I understood before the teacher teaching, and my sister insisted …

Mussa: Mh!

Child: It taught me

Mussa: What did you understand on Ubongo Kids before you were taught?

Child: To change fractions to decimals

Mussa: To change fractions to decimals?

Child: Mh!

Mussa: Do you remember who was teaching that when you were watching?

Child: Mama Ndege

Mussa: To change fractions to percentages, and what was she doing when teaching that - to change a fraction to a percentage?

Child: She was teaching good, but she was saying if you were given five over three you take five times a hundred because it is a fraction, it is a percentage. When you times to a hundred you get three hundred so a hundred you divide to the number below that is five, three hundred divide by five you will get … when you get (your) answer you put your answer in a percentage.

Excerpt 8.10: Child recollection of a science concept

Child, 11 years, Pugu

_____

Mussa: And do you remember which instruments they were creating?

Child: It is a bottle, they put beans inside it

Mussa: Mh!

Child: And the bucket they started to beat it

Mussa: Aha!

Mussa: So after putting beans inside the bottle how did they use it as an instrument?

Child: They are shaking like this

Mussa: When they do like that what happened?

Child: It produces the sound

8.3.3   Exposure factors

The theme labelled as 'Exposure factors' was formed of qualitative information pertaining to viewing frequency, viewing environment, media platforms and media content recall. Data within each of these codes showed the notion of exposure to be more complex than was accounted for in any quantitative measure. Information coded under 'media platforms' and 'viewing environment' has relatively consistent implications for the interpretation of all quantitative models, yet the effects of 'viewing frequency' and 'media content recall' might differ by statistical approach (Section 8.4). To elucidate this theme, two excerpts labelled as 'media content recall' and 'viewing frequency' are provided below.

Excerpt 8.11: Viewing frequency data, suggesting that exposure 'a long time' ago could still influence current mathematics ability

Child, 14 years, Pugu

_____

Joe: Okay, aa! When was the last time you watched Ubongo Kids?

Child: It's a long time

…

Joe: Ah!, can you tell me what happened in Ubongo Kids when you watched?

Child: I watched it … I understood there by watching it

Mussa: Mh!

Child: It is not necessary to use a whistle! to use… any audio sound is an action of collision between two things at the same time, so there I learn others they went they use their mouth and sound came out, others, aah! They went there with their things - like I can't explain.

Excerpt 8.12: Media content recall data, featuring non-prompted name recollection of multiple Ubongo Kids characters by a child

Child, 13 years, Somangira

_____

This section follows the child having responded to a question on which shows they watched. The child's answers included "Kibena's programme". Kibena is a key character on Ubongo Kids. Both Kibena and Mzee Kigo feature in the character recognition test.

Joe: Okay! Aaa ... when was the last time you watched Kibena's Program?

Child: I watched it, a day like...it was Sa... today, Saturday! It was Saturday.

Mussa: Mh!

Child: Sunday I came here

Joe: OK. Aah! (Do) you watch Ubongo Kids?

Child: Uh uh (meaning no), I do not watch

Joe: Aah, okay can you tell me What happened to Kibena - the last time you watched?

Child: Aah! There Kibena ..a, she was transporting mice/rats

Joe: Mh!

Child: And going to Mzee Kigo's farm to tell him that she transported the mice

Mussa: Mhh!

Child: He was grateful, the mice were counted. One of them fell in water.

Joe: Okay!

Mussa: Why was Kibena moving the mice?

Child: They were eating crops

Mussa: Whose crops?

Child: The old man! I don't remember the name. …

8.4   Implications of codes and themes for primary research question findings

8.4.1   Mathematics learning findings

Qualitative data provides support for positive quantitative results that is consistent yet not unanimous. Exceptions are provided by the single child response claiming that they did not learn from watching Ubongo Kids (Section 8.3.2) and a caregiver report indicating that mathematics content is insufficiently targeted to facilitate learning in all instances (Excerpt 8.7). These outlier examples could potentially be invoked in seeking to explain the single model that produced a non-significant (albeit positive) exposure coefficient (Section 7.2). However, to do so would entail applying an argument based on scant evidence. Indeed, it is probable that non-significant results were chiefly attributable to model weaknesses (Section 7.4).

Moreover, the vast majority of coded information concerning the 'enhanced skill set' of viewers indicated that Ubongo Kids exposure led to mathematics learning. Material under the theme of 'Private benefits' in Excerpt 8.8 featured a recollection of subtraction (and division) within a storyline example. Should the child's exposure to this educational cartoon content have furthered their understanding of subtraction, this could potentially have been reflected in their mathematics capability

score (which was influenced by the 'main mathematics item' that included subtraction assessment: Appendix 2). As noted above, the fact that reference to educational content was embedded within recollection of programme narrative likely reduced the chance that this learning claim was falsified.

Further, the implications of educational content and storyline being embedded can also be explored by returning to the transfer of learning theory. This theory was initially invoked to explain *why* educational television might exert a positive effect on mathematics capability (see Section 1.3.2). According to the transfer of learning theory, the first key stage of transfer involves "initial learning or comprehension" (Fisch, Kirkorian & Anderson, 2005, p. 379). This stage has itself been examined using the capacity model (see, for example, Piotrowski, 2014). The capacity model perceives demands on a viewer's limited working memory resources (or, capacity) to stem from three elements: processing of narrative, processing of educational content and "the distance between the educational and narrative content in the program" (Piotrowski, 2014, p. 313). In cases where "educational content is woven tightly into the narrative", the processing of both educational and narrative concepts becomes "complementary rather than competitive, and comprehension is (therefore) likely to be strengthened" (Fisch, 2004, pp. 144-145). The recollection of educational content embedded in a storyline suggested that educational and narrative content were closely related.[117] As such, it could be contended that the format of Ubongo Kids *should* have been facilitative of enhancing mathematics capability.

Interpretation of the information coded under the same theme ('Private benefits') that features in Excerpt 8.9 is more complex. Excerpt 8.9 also features reference to a mathematics concept covered in Ubongo Kids. As such, this might support the notion that Ubongo Kids exposure promotes mathematics learning. However, this excerpt only references mathematics material falling outside the 2017 Uwezo assessment from which mathematics capability estimates were derived. The contents of this excerpt correspond with the reasoning in Section 5.3.1, where it was noted that Ubongo Kids episodes covered material which did not feature in the Uwezo test. Should it be accepted that Ubongo Kids promoted learning in mathematics domains that were not reflected in the model's dependent variable, not all learning attributable to viewing would be reflected in quantitative results (a possibility originally presented in Section 5.3.1).

Qualitative data provides additional support for the argument that quantitative models fail to reflect the total benefit of viewing. Interview responses concerning the benefits of watching Ubongo Kids frequently featured reports or suggestions of learning outside of the mathematics domain. Non-mathematics learning was not represented in the dependent variable used in any model. This decision was made for two key reasons. First, although information on English and Kiswahili was captured by

---

[117] This reflects the proximity of educational content and narrative identified when examining programme scripts (see Section 1.3.2).

Uwezo, such information was omitted from the dependent variable because neither was the focus of the primary research question (nor a key academic subject targeted by Ubongo Kids).[118] Second, information on other educational topics that were targeted by Ubongo Kids (e.g., science) was not reflected in the dependent variable, as no relevant information was available in the data used for analysis. Should non-mathematics learning have occurred (as suggested by Excerpt 8.10 and prior reasoning in Section 6.3.1), it must also be recognised that quantitative findings do not account for all benefits delivered by Ubongo Kids.

Further, the argument that Ubongo Kids provides some benefits that would not have been reflected in quantitative findings is strengthened by considering information coded as concerning 'consumption value'. Interviewee responses frequently suggested Ubongo Kids viewing to possess a non-zero consumption value. That is, children often reported that the show entertained them (as well as educating them, see Excerpt 8.4 and 8.5). These child responses corresponded with information from adult interviewees concerning their co-viewing experiences (Excerpt 8.6) and the possibility that Ubongo Kids could be a substitute for other leisure activities (suggested by Excerpt 8.2 and 8.3).

It is recognised that quantitative exploration of the primary research question might also have failed to reflect various private costs. However, thematic analysis suggested these to be minimal. Ubongo Kids might have been viewed at the expense of productive work, meaning that there could have been opportunity costs of viewing. Yet, reports typically suggested the programme to have been a substitute for leisure (as recognised in the previous paragraph). Further, qualitative material comprising the theme of 'Private costs' suggested it to be possible that viewers incurred psychic costs but that these were likely insubstantial (Section 8.3.1). Indeed, the adverse psychic effects of learning from educational television have been portrayed as small in relation to learning in school (Excerpt 8.1).

8.4.2   Exposure findings

Material coded under the theme of 'Exposure factors' has differing implications for each of the quantitative models created to examine the primary research question. The implications of qualitative information differed, as models used varying measures of exposure. Quantitative analysis of the data from Ilala and Temeke (the 'Matched set') featured an exposure estimate created from an item response theory (IRT) model applied to five character recognition items. The model concerning Tanzania (the 'Household set') drew on individual responses to a question on recent viewership. Additionally, the 'Panel set' model used the same individual responses but averaged by district. To

---

[118] Indeed, including learning outcomes that were not specifically targeted by Ubongo Kids in a more broadly formulated dependent variable (DV) could have increased the likelihood of bias. This is because there would be a greater possibility of attributing DV variation associated with unobserved variables correlated with Ubongo Kids exposure, to Ubongo Kids. Further, adding information to the DV on subjects not specifically targeted by Ubongo Kids could have added additional (downwards) bias, as the apparent effect may have been diluted.

explore the implications of qualitative exposure information, this section contains three sub-sections concerning each model individually (Section 8.4.2.2, 8.4.2.3 and 8.4.2.4). These are preceded by a subsection concerning the coarseness amongst all quantitative exposure measures highlighted by qualitative material relevant to the 'Exposure factors' theme (Section 8.4.2.1).

8.4.2.1   Implications for all models

Qualitative information highlighted limitations regarding the measures of exposure employed in all quantitative models. These limitations predominantly concerned the lack of detail provided by exposure measures, which were initially acknowledged (without qualitative support) in Sections 5.3.2, 6.3.2 and 7.3.2). Interviews highlighted this lack of detail by providing information on more nuanced concepts. These are categorised under three subheadings:

- Longer instances of Ubongo Kids content recall
- Co-viewing and discussion of Ubongo Kids
- Viewing environment and platform usage

Longer instances of Ubongo Kids content recall

Interviews provided longer instances of programme recall, that were not identified through any quantitative measure. These included child recollection of (1) storylines, (2) educational material and (3) both storylines and educational material simultaneously. This subsection focuses on (1) storyline recollection only. Responses that featured programme narrative only appeared to provide exposure information that went beyond what was obtained through quantitative exposure measures based on both character recognition and self-reported viewership. Instances where (2) educational and narrative content were presented together gave valuable exposure information as well, but also (more importantly) provided an indication that viewing *should* have led to learning (in accordance with Fisch's theory of learning transfer, 2004). This material was considered above with regards to the implications of thematic analysis on overall findings (Section 8.4.1). Conversely, (3) interviewees' apparent recollection of educational material alone might be relatively uninformative due to the possibility of fabrication (when claims or reports were not presented in the context of a programme storyline).

Storyline recollection provides detail on exposure that would not have been reflected in the self-reported viewership and character recognition measures. Regarding character recognition, children who identified the same number of characters during 'backup' assessment gave storyline information with different levels of detail.[119] Similarly, the self-reported viewership measure that featured in Chapter 6 and 7 gave no information on the extent to which a child could recall Ubongo Kids

---

[119] IRT-derived exposure estimates for children recognising the same number of characters could still have varied, if they recognised different Ubongo Kids characters (Section 3.3.1.3.2).

narrative. Indeed, interviews suggested it to be probable that numerous children who had not reported recent viewership would have been more capable of recalling programme narrative than other children who had reported recent viewership (Excerpt 8.11). The differing levels of storyline recall exhibited by interviewees suggested exposure to be a more nuanced concept than any quantitative exposure measure accounted for.

Co-viewing and discussion of Ubongo Kids

Separate qualitative information provided further evidence of coarseness in the measure of exposure used in quantitative analyses. Amongst this information was interviewee reference to Ubongo Kids co-viewing and discussion of stories or taught concepts. Co-viewing and co-discussion could undoubtedly alter the nature of exposure (Section 2.2.2). However, neither are reflected in exposure measures based on character recognition or child-reported viewership. Because interview responses suggest co-viewing and co-discussion to have occurred (see Excerpt 8.7), all exposure measures employed were likely to have lacked precision.

Viewing environment and platform usage

Quantitative exposure measures also failed to account for complexity with regards to viewing environment and platform usage. This again highlighted how limited nuance is provided by all exposure measures used in statistical analysis. Child interviewees typically accessed Ubongo Kids "at [their] home place" and watched the show using television (because they "get picture and sound"). However, there was some variation.

Variation in viewing environment and platform use was evidenced using interview responses from two separate children. One child reportedly used the family television and their parents' smartphone to access Ubongo Kids material. Another child only watched the show at her "grandma's place" in a nearby village. The existence of this variation amongst the interview sample suggested that some children in the quantitative sample would have received exposure to Ubongo Kids in differing viewing environments and using various media platforms. This viewing environment and platform information is therefore an important omission from quantitative exposure estimates, which undoubtedly has the potential to alter the effect of viewing.

8.4.2.2   *Implications for the model applied to the Ilala and Temeke dataset*

Following initial exploration of the character recognition-based exposure measure in Section 5.3.2, qualitative findings are employed here to further understanding. Qualitative data were well-suited to this task, as it was possible to compare qualitative and quantitative data from the same participants. Members of the backup quantitative dataset - from whom interviewees were drawn - all participated in a survey that captured numeric exposure information. Therefore, while information from interviews came from a different source to the dataset to which the regression model was applied, it was at least

from a group for whom corresponding exposure information was captured. This section now considers whether interview findings point to inaccuracy in the quantitative exposure measure employed.

Two interviewees provided responses which appeared inconsistent with their quantitative results (from the backup dataset). In these cases, children's answers included references to characters that they did not recognise during survey assessment.[120] In Excerpt 8.12, for example, the child spontaneously referred to multiple Ubongo Kids characters that they did not positively identify during the 'backup' quantitative survey. Reference to characters not recognised during survey assessment might suggest that a character recognition-based measure underestimated exposure to Ubongo Kids. If exposure was to have been underestimated, then its association with mathematics capability would have been biased downwards (as explained in Section 5.3.2).

It should be noted, however, that while exposure information from child interviews and surveys might be considered broadly comparable, data obtained from both sources were necessarily distinct. Interview data were captured through semi-structured questioning and probing from Mussa, Titus and myself, and, most importantly, merely detailed recollection of a character name (as opposed to positive identification of a character from a picture). Neither of these characteristics applied to data captured in the backup survey. Further, it was possible that children viewed Ubongo Kids in the period between survey assessment and interview, which were approximately two weeks apart. For this reason, interview data did not necessarily suggest that the quantitative measure possessed weak external reliability. There is no clear evidence to suggest that test-retest results for individual children would have differed (in the absence of additional exposure between tests), nor that different quantitative exposure results would have been captured by different assessors. Interviews only supported the idea (originally presented in Section 5.3.2) that an underestimation of exposure that would downwards bias the exposure coefficient was likely.

Overall consideration of exposure regarding the Ilala and Temeke dataset

Consideration of the qualitative findings on exposure suggests that the exposure measure used in Chapter 5 was susceptible to both random and directional bias. Both these findings correspond with the logic-based reasoning initially employed in Section 5.3. Interviewee responses detailing storyline recall, the specifics of platform usage, viewing environment, co-viewing and discussions of programme content all indicated that quantitative estimates of exposure were coarse. The comparatively rich data from interviewee responses highlighted that exposure estimates based only on

---

[120] No counterexamples are available in this instance. This is because all character recollection during interviews occurred spontaneously. That is, children were not, for example, asked to state the name of every Ubongo Kids character they knew. Given this format, it is possible that a child did not happen to refer to all Ubongo Kids characters whose name they could recall during interview discussion.

character recognition results gave information lacking in depth. This might be exemplified by information suggesting child-parent co-viewing of programme content (see Excerpt 8.7). Such co-viewing would undoubtedly have altered the nature of exposure and likely impacted on any resultant learning. Yet, there was no evidence to suggest that children's levels of exposure related to them spending different proportions of viewing time while co-viewing. It was therefore assumed that the limited detail provided by the character-recognition based measure imposed no directional bias on exposure estimates (and by extension the association between exposure and mathematics capability).

In contrast, comparing child character recognition under survey assessment against the characters they mentioned spontaneously during interviews suggested quantitative measurement to underestimate exposure. Excerpt 8.12 provided clear support for this argument. Here, a child who had failed to positively identify any Ubongo Kids characters from images during survey assessment, mentioned multiple characters without their name being provided by my research assistant or myself. From quantitative information alone, this child's exposure estimate would have corresponded exactly with that of a child who had never interacted with Ubongo Kids learning material in any fashion. However, qualitative information suggested that such correspondence would have been misleading, as this child strongly appeared to have had some degree of exposure. Therefore, the exposure estimate could be considered to provide an underestimate, which would cause the apparent relationship between exposure and mathematics capability to be biased downwards: presuming that (low levels of) exposure promoted mathematics learning, the child's capability score would have been increased by viewing that was not reflected in their quantitative exposure estimate.

8.4.2.3 *Implications for the model applied to the dataset for Tanzania*

The exposure variable employed in Chapter 6 was based on a binary viewership question, "did you watch Ubongo Kids in the last week?" (see Section 3.3.1.3.2). The logic-based reasoning employed in Section 6.3.2 suggested it to be possible that this exposure measure was susceptible to bias. Here, interviews are first used to further consider how the binary survey question might lead to inaccuracy in the measurement of exposure. Next, the implications of any identified inaccuracy on regression model findings are considered.

Interview dialogue suggested directional bias in the exposure information captured by Uwezo. In accordance with prior logic-based reasoning (Section 6.3.2), responses primarily supported the idea that the exposure information captured by Uwezo was biased downwards. Qualitative analysis suggested that this could have occurred for two key reasons. First, many children would have received some level of exposure to Ubongo Kids while not having watched in the past week. For this reason, children could have faithfully responded negatively to the survey question despite having viewed the show on some previous occasion (Excerpt 8.11). Second, children might not have referred to the show by its given name. As a result, they would again have responded negatively when asked whether they

had watched Ubongo Kids, even if the show had been watched in the past week. This point was demonstrated by Excerpt 8.12, in which a negative response on viewership was provided concerning Ubongo Kids but a positive response was made regarding "Kibena's programme" (with Kibena being a key Ubongo Kids character).

However, interviews suggested a new form of directional bias not considered in previous reasoning (Section 6.3.2). This is that children might have responded positively (and inaccurately) to the question on recent viewership during Uwezo assessment. It is posited that such inaccuracy could have arisen due to a child's desire to please a Uwezo assessor (or, potentially, provide the 'correct' answer). Interview data indicated that this possible cause of response inaccuracy should be considered, as one caregiver response suggested that a research assistant and I were understood to be associated with Ubongo Kids. While this point remains important, it should be noted that children assessed by Uwezo might be less inclined to provide 'positive' information than children in the qualitative sample. Uwezo assessors had no connection to Ubongo Kids and were unlikely to be perceived as such.

Overall consideration of exposure regarding the Household dataset

Section 6.3.2 explained how different biases identified through logic-based reasoning might influence the exposure coefficient. There, it was argued that a lack of nuance in the quantitative exposure measure (also identified through qualitative analysis: Section 8.4.2.1) would exert non-directional bias on the relationship between exposure and mathematics capability (Section 6.3.2). Additionally, it was noted that downwards bias on the exposure variable (which was again evidenced by interview data: Excerpt 8.11) would have led to the association between Ubongo Kids exposure and mathematics capability being underestimated (Section 6.3.2). However, interview analysis also highlighted the (limited) possibility that the measurement of exposure could have been biased upwards by Uwezo survey respondents falsely reporting viewership to please assessors. This would also have led to an underestimation of the coefficient for exposure, because some children treated as exposed in the model in Chapter 6 might not have benefited from viewing Ubongo Kids.

8.4.2.4 *Implications for the model applied to the Panel dataset*

The implications of thematic analysis findings closely mirror those explored in Chapter 7 (Section 7.3.2). This is because two of the potential biases to exposure identified through logic-based reasoning were also suggested by interview data. Specifically, thematic analysis suggested that the self-reported viewership measure failed to account for variations in the nature of exposure (Section 8.4.2.1) and underrepresented exposure in the majority of cases (Section 8.4.2.3). In Section 7.3.2, it was explained that variations in the nature of exposure were not expected to have biased the identified interaction effect between exposure and time in any particular direction. This variation would,

however, have increased the standard error of the interaction effect thereby decreasing the likelihood that this effect would be significant. Further, it was noted that systematic downwards bias in district exposure would have increased both the standard errors and coefficient for the interaction effect (Section 7.3.2).[121]

8.5   Summary

This chapter has provided information on the thematic analysis conducted on interview data captured from four families in Dar es Salaam. While employment of thematic analysis highlights shortcomings in quantitative models (presented in Chapter 5, 6 and 7), qualitative analysis was itself recognised to be limited. The 'sense' contributed by Mussa, Titus or I could have contributed to inaccuracy amongst responses. Further, quantitative and qualitative data were not typically available from the same children, which restricted opportunities to identify inaccuracy. However, qualitative analysis undoubtedly remained worthwhile. Bias was limited wherever possible by asking interviewees to elaborate on responses and provide concrete examples, as well as highlighting conflicting opinions. Additionally, findings were directly relevant to the interpretation of quantitative models.

Through thematic analysis, themes were identified concerning private costs, private benefits and factors relating to Ubongo Kids exposure. Material relevant to the theme of private benefits acted to support the positive results identified across all quantitative results chapters. This material suggested that children's exposure to Ubongo Kids material furthered their understanding of mathematics content that was included in the tests that produced capability scores. Indeed, the fact that learning claims were embedded in episode narratives was also considered to give an indication of *why* Ubongo Kids supported learning (by invoking the 'Transfer of learning' theory). Further, thematic analysis supported logic-based reasoning (e.g. Section 5.3.1) which showed it to be probable that all quantitative findings failed to reflect the total benefits of viewing Ubongo Kids. Thematic analysis evidenced this through material coded under both 'enhanced skill set' and 'consumption value', which suggested that learning gains and viewer enjoyment were likely to have been obtained (that would not have been reflected in quantitative models).

Lastly, the theme named 'Exposure factors' benefited the interpretation of all models employed in each of the previous three chapters. Information under this theme substantiated prior logic-based reasoning by highlighting coarseness in the exposure measures used in all quantitative models. Further, the specific manners in which each exposure measure might have been biased was considered. Overall, qualitative exposure information was found to suggest that models featured a downwards biased exposure measure. This would have led to an underestimation of the association

---

[121] Thematic analysis neither proved nor disproved the reasoning-based idea that exposure intensity would have varied over time, which was identified in Section 7.3.2 to reduce the t value of the interaction effect between exposure and time.

between exposure and mathematics capability in Chapter 5 and 6. Conversely, qualitative exposure data suggests that the apparent effect identified through the 'Panel set' model (Chapter 7) could have been inflated, but that standard errors would also have been increased.

9 Chapter 9 – Cost-effectiveness analysis

9.1 Introduction

This chapter presents findings on the subsidiary research question, which concerns the cost effectiveness of educational television. Exploration of the subsidiary research question required the employment of cost-effectiveness analysis (CEA). CEA permits determination of "the least cost approach to meeting such educational objectives as… raising [attainment]" or the relative gains in attainment achieved by differing interventions for a given cost (Levin, 1988, p. 52), with this chapter focusing on the latter. Accordingly, cost-effectiveness estimates for the television programme, Ubongo Kids, were ultimately compared against results for alternate projects in similar contexts.

This chapter begins by considering what information can be provided by CEA and re-familiarising the reader with the CEA approach used (following initial discussion in Section 3.3.2). After this, the justifications for choosing various figures to populate the J-PAL equation are discussed. Following CEA estimation, results concerning Ubongo Kids are compared against other programmes pursuing similar objectives. This comparison of CEA results suggested that Ubongo's educational television intervention had been highly cost effective, due predominantly to its large scale. The extent to which this television programme was found to be cost effective demands the attention of educational policy makers seeking a more efficient allocation of resources.

9.2 Cost-effectiveness analysis approach

CEA provides a valuable means of examining relative cost effectiveness. CEA enables this through the comparison of cost-effectiveness results for different projects assessed under the same methodology. Such comparisons are, of course, not without their limitations. For example, cost-effectiveness comparisons do not account for differences in the location of programmes' beneficiaries (see Section 2.3.1.1). Further, there are many topics that CEA does not address. CEA alone cannot show whether a programme, in itself, is a worthwhile investment. Such investigation is the domain of cost-benefit analysis (CBA). This is because CBA involves "a direct comparison of the monetary cost and benefit of a particular intervention where cost and benefit consequences can both be measured in monetary terms" (Levin & Belfield, 2015, p. 402). Conversely, the benefits that influence CEA calculations are limited to selected outcomes (which, in this instance, solely concern mathematics capability).[122]

Indeed, it must be recognised that all forms of cost analysis are restricted in scope. There are numerous possible eventualities from child engagement with interventions that are highly challenging

---

[122] While CBA could provide additional information, this is not employed in this thesis because the creation of accurate results is obstructed by (lack of) data availability (Section 2.3).

to quantify. The opportunity cost of watching educational television, for example, was assumed to be zero (Section 9.5). This assumption was made on the basis that educational television had often been viewed instead of leisure activities such as watching other programmes or playing games (Section 8.3.1). However, it is undoubtedly conceivable that television could have been viewed at the expense of other activities, such as those based on family interaction. The true cost of losing family time is extremely difficult to measure and, therefore, not commonly reflected in any form of cost analysis.[123] Another potential cost of viewing media that was suggested by caregiver discussion of non-Ubongo Kids programmes is that children might watch inappropriate content. Like the cost of losing family time, the costs of watching inappropriate content inadvertently included in an educational programme would not influence cost-effectiveness results.[124]

The CEA calculations made in this chapter followed the J-PAL estimation method. Use of the J-PAL approach ensured that the numerous assumptions underlying CEA calculations were the product of some degree of deliberation.[125] Such assumptions included the discount rate selected by J-PAL, which followed consideration by Dhaliwal, Duflo, Glennerster and Tuloch (2012) (see Section 9.5). Additionally, employment of J-PAL's method enabled comparison between cost-effectiveness results regarding Ubongo Kids and numerous other interventions for which CEA estimates have previously been produced. This is because the J-PAL method has been used for multiple CEA estimations, concerning interventions ranging from computer assisted learning to conditional cash transfers (see Banerjee, Cole, Duflo & Linden, 2007; and, Baird, McIntosh & Ozler, 2011, respectively). Using a different CEA methodology would have precluded the possibility of comparison with these projects (as CEA estimations created through differing methods cannot be compared, Levin & McEwan, 2001).

The J-PAL method requires information on the cost, influence and number of beneficiaries supported by a programme to give a finding in the form of change in standard deviations per $100 spent. Ultimately, a CEA result is produced using the following equation (initially presented in Section 3.3.2):

Change in standard deviations per $100 spent = 100/(Cost/Influence*Beneficiaries)

In this equation, Cost refers to the total programme cost, Influence refers to the standard deviation gain amongst an individual that is attributable to the programme and Beneficiaries refers to the total number of children who have received the programme. Populating this equation ostensibly requires

---

[123] This is in spite of family interaction likely contributing to children's human capital formation (according to the model of family relations: Becker, 1981).

[124] Indeed, the very notion that private costs could be incurred by watching inappropriate media content fell outside the preconceived human capital framework used in this study (see Section 3.4.3).

[125] This being said, certain assumptions are subsequently altered to conduct a brief sensitivity analysis (Section 9.7.1).

only one set of estimations. However, the production of one CEA result relevant to Ubongo Kids' ongoing functioning and another result relevant to activities since inception (Section 9.7) necessitates the employment of two sets of estimations (Section 9.3). These estimations reflect different programme duration assumptions (2013-2017 and 2017 only), as well as differing cost and influence estimates.

In producing both CEA estimations, fidelity to the J-PAL method was maintained. This promoted the validity of comparisons with CEA results for other projects (see Section 9.7.1). Tailoring the J-PAL CEA formula might have allowed it to, for example, reflect the consumption value of viewing Ubongo Kids (suggested by thematic analysis: Section 8.3.2). Yet, such alterations have not typically featured in other CEA estimations against which Ubongo Kids was compared. Indeed, identified CEA estimations that were based on modified versions of the J-PAL approach are potentially susceptible to criticism. For example, investigation into Camfed's Tanzania activities used a CEA calculation that accounted for the reduction in school dropout attributable to the programme (Sabates, Rose, Delprato, & Alcott, 2018). However, this estimation simultaneously excluded the money spent by Camfed on school fees from programme costs. Given that funding school fees likely reduced the chance of dropout, an argument could be made that this adaptation of the J-PAL approach overestimated Camfed's cost-effectiveness.[126]

9.3   Programme duration

Two programme duration estimates were selected to permit investigation into Ubongo Kids' cost effectiveness. The time period used to examine Ubongo Kids cost effectiveness since the outset of the programme was 2013 to 2017. This time period spanned the inception of Ubongo Kids (July 2013), through its first broadcast (January 2014), up to the Uwezo survey from which influence findings were derived (December 2017). Additionally, this estimate ensured that the programme duration period covered all time points at which a child could have been exposed to Ubongo Kids. Using this period is, however, considered likely to produce a highly conservative CEA estimate.

This contention relates to limitations in gauging the number of beneficiaries. The number of beneficiaries was established by calculating the proportion of children who reported recent viewership during the 2017 Uwezo assessment (Section 9.4). Assuming the intervention to have taken place between 2013 and 2017 led to a CEA calculation that treated self-reported viewers (in 2017) as the only Ubongo Kids beneficiaries since the programme's beginning. Child-level viewership data was

---

[126] It should be noted that the paper by Sabates, Rose, Delprato and Alcott (2018) also presents a CEA finding that is not modified to account for increased school retention, although this is not ultimately used in the CEA comparisons made. School fees were omitted from all calculations in the policy brief, on the basis that these were abolished in Tanzania after data collection took place.

available only for December 2017, so viewership amongst children at earlier points could not have been established. The programme duration estimate relevant to Ubongo Kids' ongoing operations was considered less likely to lead to downwards biased cost-effectiveness findings.

This estimate assumed the intervention to have lasted one year (or, to have occurred in 2017 only). Such an assumption was likely susceptible to both upwards and downwards bias (see Section 3.3.2). Like the estimate for Ubongo Kids since its inception, it is probable that the exposure of numerous children in 2017 was not reflected in Uwezo survey responses. However, choosing 2017 as the start point of the intervention also introduced the potential of bias in the other direction. That is, there could have been some children (who reported Ubongo Kids viewership) that benefited from watching the programme prior to 2017. The 2017 duration estimate might therefore have provided a compromise between biases in both directions.

9.4   Number of beneficiaries

This chapter now discusses the number of Ubongo Kids beneficiaries. An estimate for this value was produced using Uwezo (2017) data and United Nations (UN) populations estimates (https://population.un.org/wpp/Download/Standard/Interpolated/). Uwezo data provided the only means of gauging the proportion of children (6 to 16 years) exposed to Ubongo Kids from nationally representative child-level data. This was established by calculating the weighted mean of responses in the Uwezo dataset to a question regarding recent viewership, "Did you watch Ubongo Kids in the past week?". This calculation suggested that 16.68 percent of children had watched the show.[127]

To establish the number of child viewers, this percentage was multiplied by the total number of children aged 6 to 16 in Tanzania (as of 2017). The total number of children in Tanzania was calculated using UN population estimates. UN estimates for Tanzania were created using censuses and other official reports, with adjustments for underenumeration (https://population.un.org/wpp/DataSources/834). These estimates could be broken down by country, year and age, which permitted approximation of the total Tanzanian population aged 6 to 16 years for 2017. This source suggested that there were 15,842,916 children between 6 and 16 years. Multiplying this number by the percentage of children of the same age who view Ubongo Kids gives 2,643,250 children.

---

[127] It could be noted that this figure differs slightly to the estimate presented in the chapter providing descriptive statistics. There, the weighted mean of children in the cross-sectional dataset concerning Tanzania as a whole that reported recent exposure was identified to be 15.46 percent (see Section 4.4.2). The estimate in Section 4.4.2 differs because it was based on a dataset that excluded various children, including those with no assessed siblings.

Before proceeding, it should be reiterated that this figure is likely an underestimate of the number of programme beneficiaries. In this thesis, it has been identified that estimates of Ubongo Kids exposure based on self-reported viewership are likely to be biased downwards. It was recognised that members of the dataset concerning Ilala and Temeke children who did not report recent viewership were frequently able to recognise multiple Ubongo Kids characters (thereby suggesting exposure at some previous point: Section 4.4.2). Further, members of the qualitative sample were able to recall detailed information from Ubongo Kids shows despite not having watched the show in the recent past (Section 8.3.3). Both these points suggested that viewership would be underestimated by a binary exposure proxy based on behaviour in the previous week.

What is more, the same estimation of total viewers was used in both CEA calculations (see Section 9.7). That is, calculations featured the same number of beneficiaries regardless of whether the intervention was assumed to have occurred in 2013-2017 or 2017 only. The lack of child level data precluded investigation into viewership prior to 2017. This meant that CEA calculations based on the programme duration estimate for Ubongo Kids from its outset (2013-2017) were likely to be particularly downwards biased. It is probable that many children who received exposure between the first broadcast of Ubongo Kids and the 2017 Uwezo survey (that were included in the Uwezo sample) did not report viewership during the 2017 survey. This bias would apply to a lesser extent when assuming the intervention to have occurred in 2017 only (see Section 9.3).

9.5   Costs

In accordance with the employment of two programme duration periods, Ubongo Kids costs were estimated both for 2017 alone and 2013-2017. These estimates were derived from costs submitted by Ubongo in accordance with the J-PAL basic costing template (available at: https://www.povertyactionlab.org/research-resources/cost-effectiveness). This template facilitated the establishment of total programme costs, by providing a framework for obtaining cost data in the following categories:

- Programme administration and staff costs
- Targeting costs
- Staff training
- Implementation and programme material costs
- Monitoring costs
- Participant training
- User costs
- Averted costs

The figures submitted for multiple cost categories comprising the J-PAL template had a value of zero. These categories were 'Participant training', 'User costs' and 'Averted costs'. 'Averted costs' were zero because the intervention was unlikely to have prevented any costs being incurred by beneficiaries or other education providers. Additionally, 'Participant training' was zero because there were no specific trainings required for the viewing of Ubongo Kids during the intervention period. 'User costs' were also reported as naught. The justification for 'User costs' being equal to zero was less clear-cut than for 'Participant training' and 'Averted costs'. This is because the 'User' devoted time resources to watching Ubongo Kids and could only have accessed the show using a media platform, which likely cost a non-zero amount.[128] For this reason, the establishment of this figure was discussed by Ubongo and I.

While Ubongo and I acknowledged that there could have been some positive 'User costs', it was decided that an estimate of zero was most appropriate. Media devices such as televisions were indeed likely to have cost a non-zero amount, yet it would have been farfetched to assume that the cost of such devices was incurred solely for the purposes of watching Ubongo Kids. Additionally, thematic analysis suggested that the opportunity cost of watching Ubongo Kids was often negligible (Section 8.3.1). Time spent watching Ubongo Kids was frequently at the expense of playing games (Excerpt 8.2) or viewing different television programmes (Excerpt 8.3).

Ubongo provided non-zero estimates for all categories other than 'User costs', 'Averted costs' and 'Participant training'. These estimates reflected costs such as office rent (a component of 'Administration and staff costs'), the fees spent on external voice actors (an 'Implementation and material' cost), the cost of staff courses (a 'Staff training' cost), the subscription fee paid to an SMS-based viewership survey (a 'Monitoring' cost) and the fee paid to broadcasters for Ubongo Kids to be aired in 2014 (a 'Targeting cost'). From financial records, Ubongo could immediately provide 'Targeting' and 'Monitoring' costs relevant to Tanzania alone. However, all other costs could not be disaggregated from Ubongo's activities in other countries. That is, 'Administration and staff costs', 'Implementation and material costs' and 'Staff training' costs were initially relevant to Tanzania and all other countries in which Ubongo Kids aired before 2018 (Kenya, Rwanda and Uganda). To estimate costs relevant to Tanzania alone, 'Administration and staff costs', 'Implementation and material costs' and 'Staff training' costs were multiplied by the approximate percentage of Ubongo

---

[128] Other forms of cost analysis could also account for user psychic costs. However, these are not included in the J-PAL framework. Indeed, such costs were likely to have been minimal: a caregiver reported that television taught in a "gentle" or "calm way" relative to classroom learning (Excerpt 8.2).

Kids viewers that were based in Tanzania. The percentage of viewers in Tanzania is presented in the table below (Table 9.1).[129]

Table 9. 1: Percentage of viewers based in Tanzania

Year	2013	2014	2015	2016	2017
Percentage	100%	100%	53%	42%	23%

This table features the approximate percentage of viewers in Tanzania in different calendar years. The percentage for 2013, the year prior to Ubongo Kids' first broadcast in 2014, is set at 100 percent. This percentage is chosen because the initial development of Ubongo Kids was geared towards broadcasting in Tanzania alone.

The above table shows that the percentage of Ubongo Kids viewers that were based in Tanzania fell after 2014. This is because Ubongo Kids was initially broadcast in Tanzania alone, but became available in Kenya, Rwanda and Uganda from 2015 onwards. The reduction in the percentage of Tanzania-based viewers contributed to falling cost estimates from 2015. Cost estimates are depicted below in Tanzanian Shillings (TZS) (Figure 9.1). It might also be noted that the presence of targeting costs in 2014 only produced a spike in total costs for that year. After Ubongo Kids' first year of broadcast (2014), free airtime was negotiated with television stations.[130]

---

[129] The proportion of Ubongo Kids viewers based in Tanzania was established using historic monitoring data initially obtained from GeoPoll. GeoPoll is an organisation that provides research services that include text message surveys, from which viewership proportions were estimated.

[130] No targeting costs other than the fee paid to broadcasters in 2014 were incurred by Ubongo in targeting or raising awareness amongst potential beneficiaries (in 2014 or any other year).
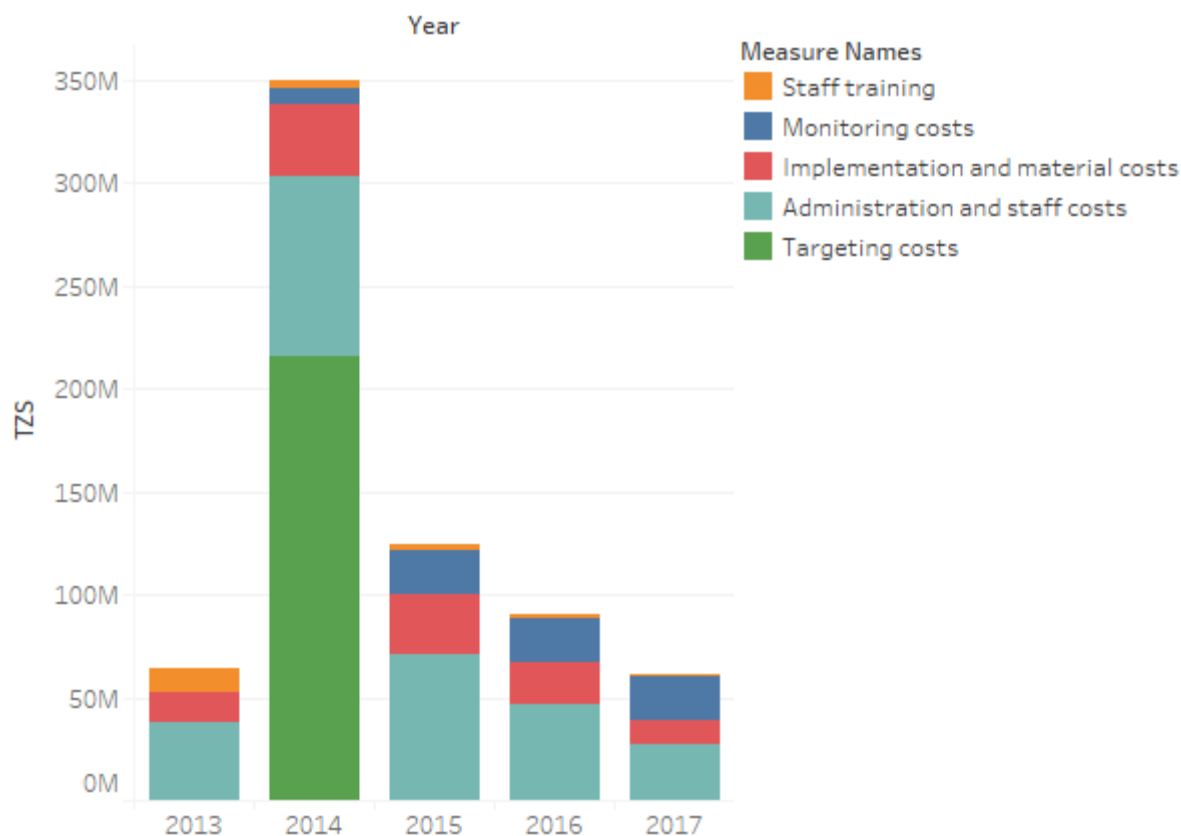
Figure 9. 1: Cost estimates by year in Tanzanian Shillings

This figure features costs at the time incurred. That is, it does not display costs that are benchmarked for inflation.

The amounts in Shillings depicted above were converted into their present value at a 'year of analysis' (common amongst all compared programmes). When assuming the intervention to have taken place in 2017 only, this merely involved converting the sum of all categories in 2017 (TZS) into 2011 USD.[131] This gave a total intervention cost of $25,481.12 and a cost per child of under $0.01 (using the number of beneficiaries estimated in Section 9.3). When assuming the intervention to have taken place from 2013-2017, the calculation of present value in 2011 became slightly more involved. The stages required to produce the cost estimate for analysis are detailed below (Table 9.2).

---

[131] The selected 'year of analysis' must be consistent for all programmes compared (Dhaliwal, Duflo, Glennerster & Tuloch, 2012). This was therefore set at 2011, to permit direct comparison with other CEA estimates based on the J-PAL approach. This year differs to Ubongo Kids' base year (whether considering the programme's ongoing operations or activities since inception). As such, the costs ultimately used to create Ubongo Kids CEA estimates are deflated back to 2011 dollars.

Table 9. 2: Conversion of TZS estimates to 2011 USD present value estimates

Row	Cost	2013	2014	2015	2016	2017	Total
1	total cost/year, TZS	63,823,527.05	349,537,202.92	124,008,964.40	90,156,995.05	61,491,195.08	689,017,884.50
2	Total cost/year, 2013 USD	39,144.23	208,091.26	55,922.61	40,230.09	26,358.10	369,746.29
3	PV of cost/year, 2013 USD	39,144.23	189,173.88	46,217.03	30,225.46	18,002.93	322,763.54
4	PV of cost/year, 2011 USD	37,841.84	182,879.74	44,679.31	29,219.81	17,403.95	312,024.64

The first row of this table gives the total of all cost categories by year. Row 2 features estimates that were converted to USD at programme base year values (2013). In the third row, the present value of costs in 2013 USD are presented. The present value of costs was calculated using a discount rate of 10 percent. Discount rates adjust for the choice faced by project funders between incurring costs in a current year, or deferring costs to invest until spending in a later year (J-PAL: https://www.povertyactionlab.org/research-resources/cost-effectiveness). The specific discount rate selected followed J-PAL analysis (with the justification for J-PAL's selection of this discount rate discussed by Dhaliwal, Duflo, Glennerster & Tuloch, 2012). The fourth row gives the present value of costs from Row 3 converted into USD at the year of analysis (2011). The total of costs in this row for all years (2013-2017) gives the total cost of Ubongo Kids from its outset ($312,024.64, in blue font on Row 4). Dividing this amount by the total number of viewers suggested a per child cost of just under $0.12. The cost per child for both duration estimates are presented below, alongside other educational interventions measured in accordance with the J-PAL approach.
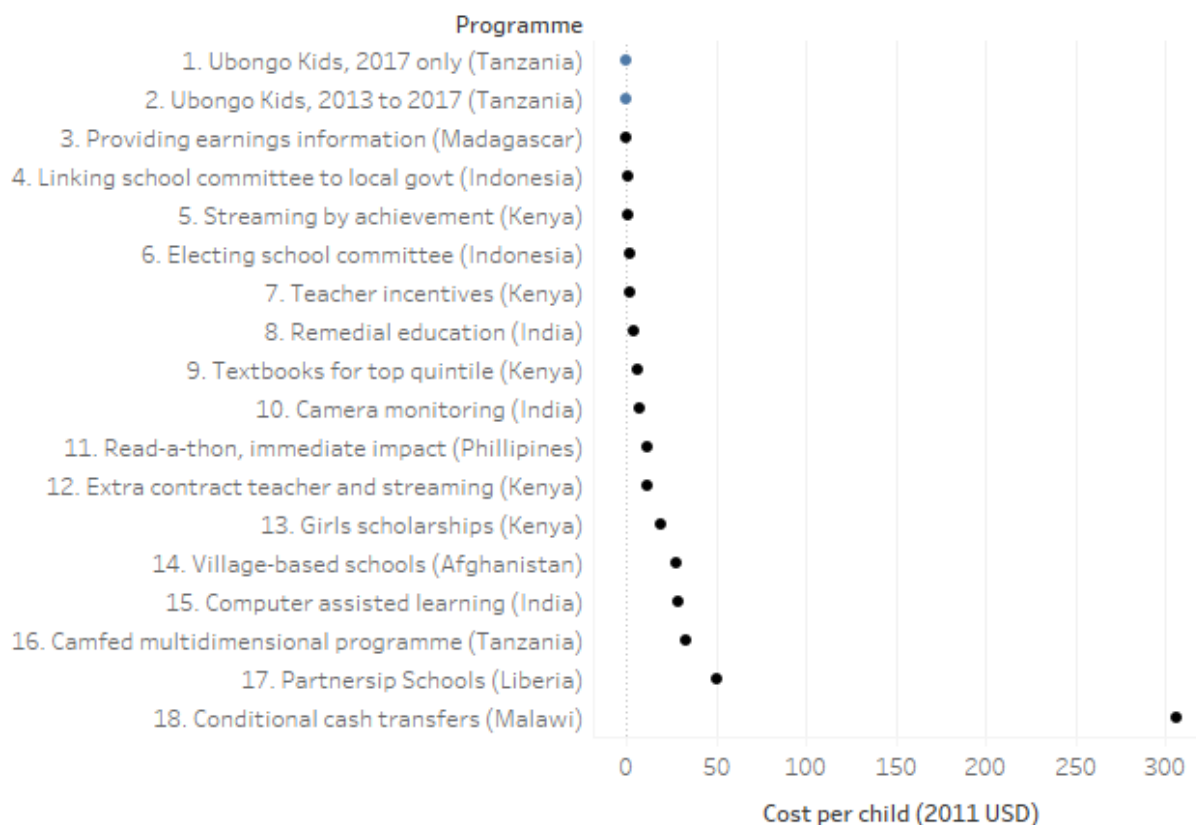
Figure 9. 2: Total programme costs

The above figure shows cost per beneficiary (in 2011 USD) amongst all interventions for which positive (and significant) J-PAL-based CEA findings are available. Costs cover all programme costs recognised by J-PAL (https://www.povertyactionlab.org/research-resources/cost-effectiveness) with two exceptions. The figure for the 'Partnership Schools' project in Liberia [estimate 17] was based on "the lowest possible cost associated with the program" (Romero, Sandefur & Sandholtz, 2017, p. 49). Additionally, the estimate for Camfed in Tanzania [16] omitted school and examination fee costs (which were removed in Tanzania following data collection, Sabates, Rose, Delprato & Alcott, 2018). Estimates [3] to [15] and [18] were taken from J-PAL, who compiled information from the following studies (each of which features in the above graphic): Baird, McIntosh & Ozler (2011) [18]; Kremer, Miguel & Thornton (2009) [13]; Burde & Linden (2013) [14]; Nguyen (2008) [3]; Glewwe, Kremer & Moulin (2009) [9]; Abeberese, Kumler, & Linden (2014) [11]; Banerjee, Cole, Duflo & Linden (2007) [8] and [15]; Kremer, Duflo & Dupas (2011) [5] and [12]; Glewwe, Ilias & Kremer (2010) [7]; Duflo, Hanna & Ryan (2012) [10]; and, Pradhan, Suryadarma, Beatty, Wong, Alisjahbana, Gaduh & Artha (2011) [4] and [6].

Figure 9.2 compares only educational programmes for which J-PAL CEA estimates are available, thereby excluding television-based interventions other than Ubongo Kids. It was, however, possible to situate cost estimates regarding Ubongo Kids alongside those available for a small number of other television programmes (that were not calculated under the J-PAL approach). Amongst the previous television studies identified (that were publicly available), reported per-child cost estimates differed

starkly to those for Ubongo Kids (Section 2.3.1.3). These estimates ranged from $17.81 (from a 2002 publication focusing on Brazil, by Wolff, de Moura Castro, Navarro & Garcia) to $1305.73 (for the costliest year of a longstanding Mexican television project: Perraton, 2005).[132] The substantially lower per-child costs for Ubongo Kids reflect differences by intervention in the number of beneficiaries affected. This was likely attributable to intervention format in multiple instances (with, for example, the Mexican television project being classroom-based: Perraton, 2005). Further, changes in the cost and availability of technologies has undoubtedly led to reductions in per-child costs of television-based interventions over time.[133] This point highlights a shortcoming of CEA, which is that results can quickly become obsolete in "areas where change may occur rapidly" (Levin, 1988, p. 56). Indeed, it is conceivable that near-term changes in technology cost and availability will bring about lower costs-per-viewer figures than were estimated in this thesis.

9.6 Influence

This section details the creation of two influence estimates. These estimates were both derived from the most accurate model employed to answer the primary research question, which concerns the association between Ubongo Kids exposure and mathematics capability. The model used to answer this question with greatest precision was the household-level fixed effects approach applied to the cross-sectional dataset regarding Tanzania (see Section 6.2). This model was considered to give the most accurate information on the effects of Ubongo Kids for multiple reasons. These include that the fixed effects approach controlled for observed and unobserved household-level differences between children. In addition, remaining child-level differences were also controlled for using multiple independent variables (including a proxy for non-mathematics attainment). It should also be recognised that this model was applied to a large-scale dataset representative of Tanzania as a whole (n = 39,717).

When assuming the intervention to have taken place in 2017 only, the creation of an influence estimate for use in CEA calculations was straightforward. This merely required multiplying the model coefficient for Ubongo Kids exposure by the standard deviation of the mathematics capability variable (in the dataset for cross-sectional analysis concerning Tanzania). The coefficient for Ubongo Kids (0.126) was given alongside those for all other independent variables comprising the regression model presented in Section 6.2. The standard deviation of mathematics capability (0.841) was calculated

---

[132] These figures were inflated to 2011 USD. However, these cost-per-child estimates likely remain incompatible with figures produced under the J-PAL approach (due, for example, to the exclusion of certain costs captured under the J-PAL methodology). For this reason, any cost-per-child comparison with the Ubongo Kids intervention should be considered highly tentative.

[133] In fact, widespread television ownership in Tanzania meant both that Ubongo Kids reached a large number of children and that the cost of this technology could be excluded from cost estimations (given that televisions would likely have been purchased anyway, for purposes other than watching Ubongo Kids).

independently from the dataset for analysis. The multiplication of these figures provided the influence per individual (0.106 standard deviation gains, see Figure 8.3).[134] Lastly, the total influence over all beneficiaries was calculated by multiplying the number of beneficiaries (or, viewers, established in Section 9.4) by the influence per individual. This gave a total influence of 280,862.294. These figures are presented alongside upper and lower bound estimates in Table 9.3.

When assuming that the intervention took place from 2013 to 2017, the calculation of influence differed. This estimation required discounting to be taken into account. In the same manner as for costs, programme benefits must be "discounted back to the base year of the program so inflation is compounded over the correct number of years" (Dhaliwal, Duflo, Glennerster & Tuloch, 2012, p. 37).[135] This enables calculations to give a finding for the present value (in 2013) of later learning gains. Application of a discount rate acts to reduce the present value of programme benefits (or costs) by an amount that increases the longer that a benefit (or cost) is incurred after the programme base year. Calculations used the same discount rate applied to programme costs, 10 percent. As such, translating standard deviation gains established in 2017 to the programme base year of 2013 required the influence per individual (0.106) to be divided by $(1+0.1)^4$.[136] This suggested there to be an influence per individual of 0.073 (see Figure 9.3). Multiplying this number by the total number of viewers (from Section 9.4) gave a total influence of 191,832.726. These figures are presented alongside estimates for an assumed programme time period of 2017 only (in Table 9.3, below).

Table 9. 3: Influence estimates

Duration assumption	Influence per individual			PV of influence per individual			Total influence over all beneficiaries		
	point estimate	*upper bound*	*lower bound*	*point estimate*	*upper bound*	*lower bound*	*point estimate*	*upper bound*	*lower bound*
2013-2017	0.106	0.163	0.049	0.073	0.112	0.033	191832.726	295165.422	88500.030
2017	0.106	0.163	0.049	0.106	0.163	0.049	280862.294	432151.695	129572.894

This table features the influence, present value (PV) of influence and total influence for both programme duration estimates. Upper and lower bound estimates are at the 90% confidence interval.

---

[134] All influence estimates are presented in the format of standard deviation gains. The use of this terminology reflects that used in prior CEA literature (e.g., Dhaliwal, Duflo, Glennerster & Tuloch, 2012) and is not intended to imply causality.

[135] When assuming the programme to have taken place in 2017 only, no discount rate needs to be applied.

[136] Note that this formula treats influence amongst beneficiaries as something that only occurs in 2017. This is a product of data availability. The data cannot, for example, allow the identification of influence amongst different individuals in years prior to 2017.
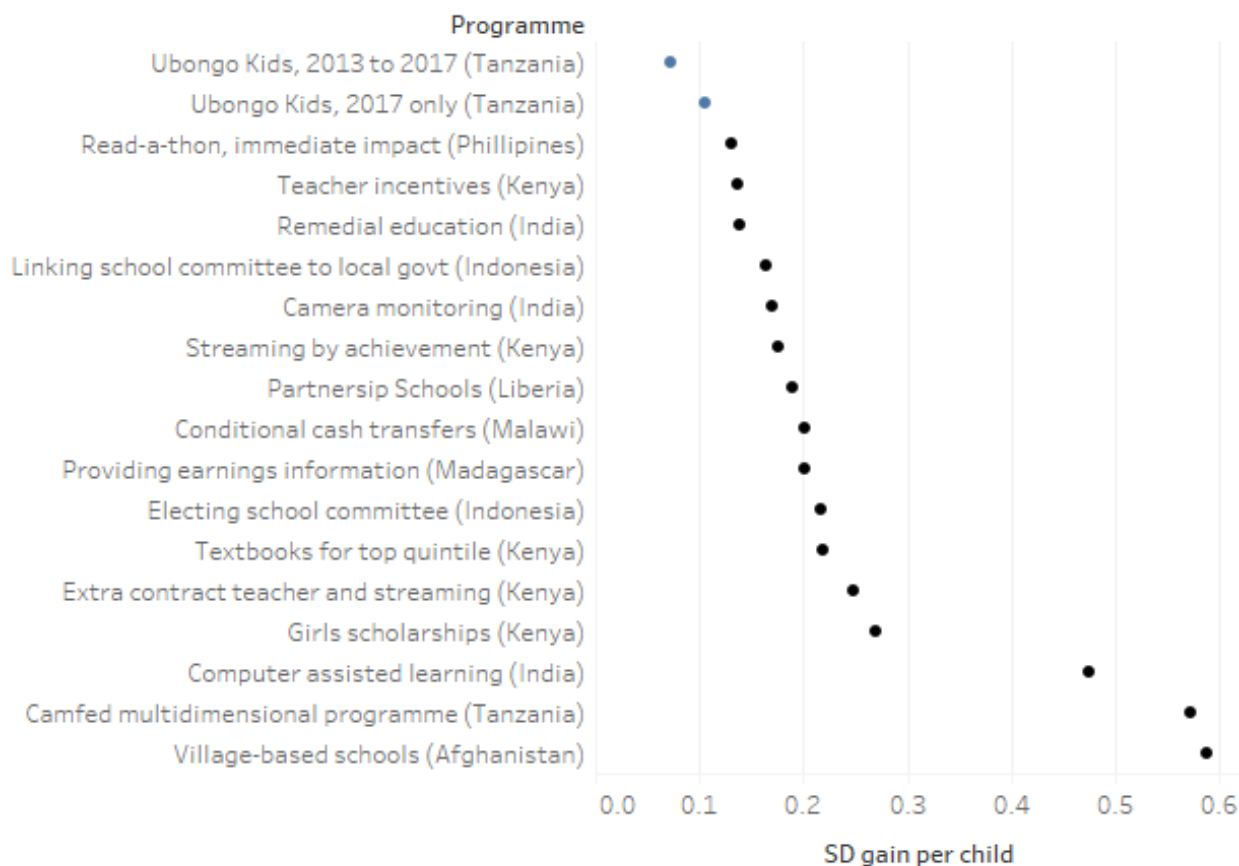
Figure 9. 3: Standard deviation gain per child

The above figure shows standard deviation learning gains per child for all programmes that ultimately feature in cost-effectiveness comparisons (Figure 9.4). References for the programmes in the above graphic follow those in Figure 9.2. Programmes are again ordered according to their relative cost-effectiveness performance (Figure 9.4).

Before proceeding to CEA calculations, it should be recalled that the coefficient value from which the influence estimates in Table 9.3 were derived could have been downwards biased. This was demonstrated using both logic-based reasoning and thematic analysis, which suggested it to be likely that numerous children who did not report viewership during Uwezo assessment in 2017 had received some level of exposure (Section 6.3.2; Section 8.3.3). This would have acted to reduce the identified coefficient for Ubongo Kids, as children who had benefited from watching the show would not have been treated as being exposed in the model. Additionally, reasoning and thematic analysis suggested that Ubongo Kids exposure could promote learning outside of assessed mathematics topics and provide consumption value (Section 6.3.1; Section 8.3.2). Again, this indicates that the total effects of Ubongo Kids exposure were underestimated in quantitative results, as these results were not influenced by learning in other academic subjects.

9.7   Cost-effectiveness comparison and analysis

This section provides cost-effectiveness estimates for the Ubongo Kids intervention, after which these estimates are compared against those for differing interventions. To calculate cost effectiveness, the cost of the intervention (established in Section 9.5) was divided by the estimated 'total influence over all beneficiaries' (from Table 9.3). This calculation was performed using point estimates for influence in addition to upper and lower bound findings. The figures relevant to each of the two programme duration periods were used to provide two separate CEA results. These findings are detailed in Table 9.4.

Table 9. 4: CEA calculations

	$ cost per additional SD			Additional SD/$100		
Duration assumption	*point estimate*	*upper bound*	*lower bound*	*point estimate*	*upper bound*	*lower bound*
2013-2017	1.63	1.06	3.53	61.48	94.60	28.36
2017	0.09	0.06	0.20	1102.24	1695.97	508.51

The above table plots the cost per additional standard deviation (in 2011 USD) and additional standard deviations per $100 spent (in 2011 USD).

9.7.1   Comparison of estimates

Using the measures of additional standard deviations per $100 USD spent (Table 9.4), estimates concerning Ubongo Kids were plotted against those for other interventions. The method of CEA employed in this chapter permitted direct comparison against numerous differing programmes. Point estimates and upper and lower bounds at a 90 percent confidence interval for additional standard deviations per $100 are plotted below (Figure 9.4).
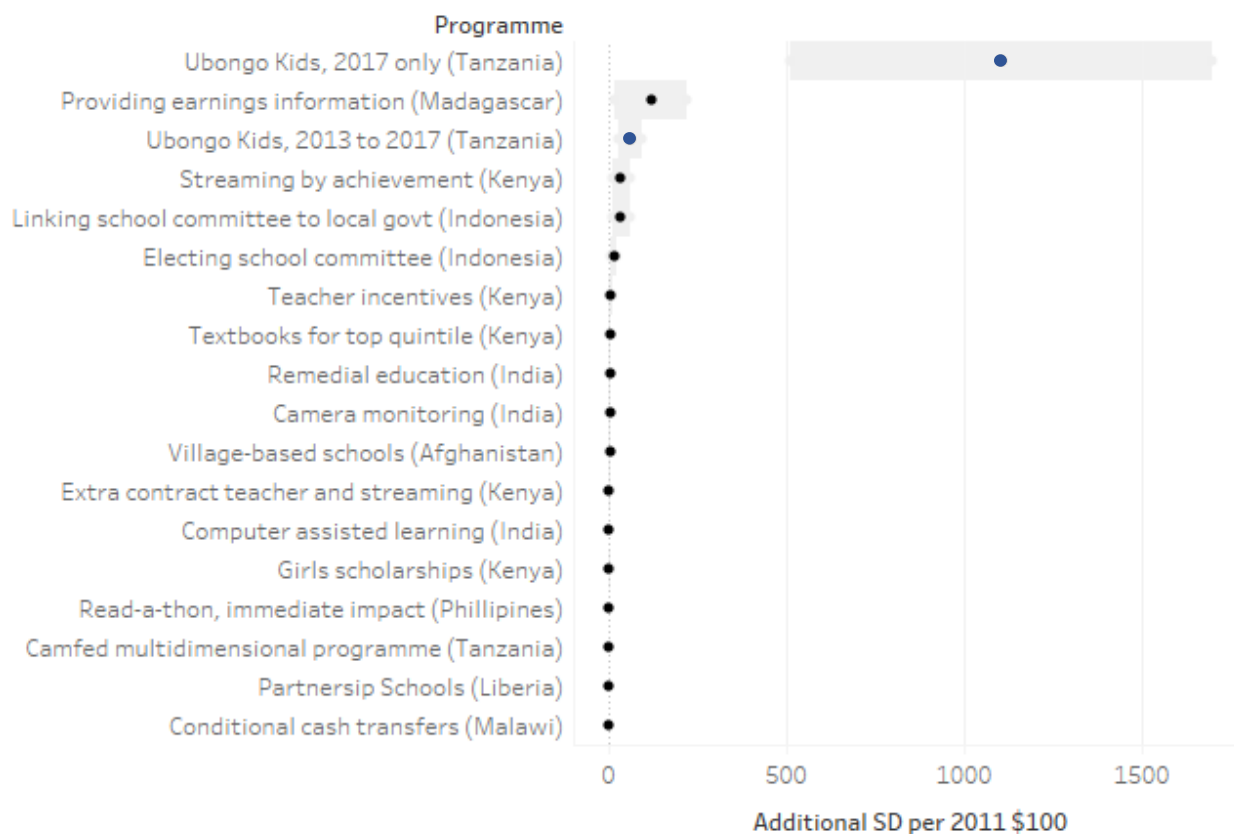
Figure 9. 4: Cost-effectiveness comparisons

The above figure shows learning gains per $100 (2011) USD from all available CEA calculations that used the J-PAL approach (where cost-effectiveness estimates were positive and based on significant influence findings). The Camfed finding was produced using a modified version of the J-PAL approach, which allowed estimations to reflect the reduction in dropout rate attributable to the programme. References for the programmes in the above graphic follow those in Figure 9.2.

This section now concludes by discussing how the Ubongo Kids intervention compared to other interventions in developing country contexts. Whether considering the cost effectiveness of Ubongo Kids relevant to its ongoing operations or activities since inception, CEA results compared favourably. The CEA estimate for 2017 only suggested that Ubongo Kids delivered greater child learning benefits per dollar spent than any other programme for which CEA results were available. The 2013-2017 CEA estimate was topped only by a Madagascar-based intervention, where statistics demonstrating the returns to education were provided to those attending treatment schools (Nguyen, 2008).

Relative cost effectiveness can be attained through both a comparatively low cost or high influence per beneficiary. Improving resource allocation - a key concern of economics of education approaches including human capital theory - would be accomplished by investing a greater proportion of

educational resources in cost-effective approaches (regardless of whether cost effectiveness was obtained through a low cost or high influence). In this instance, the strong cost effectiveness of Ubongo Kids relative to alternate interventions is largely attributable to its cost as opposed to influence per viewer. The standard deviation gain amongst individual children is small relative to other interventions assessed under the J-PAL methodology (see Figure 9.3). However, the extremely low marginal cost per beneficiary clearly outweighs this influence result (when establishing cost effectiveness).[137] The reach of Ubongo Kids is reflective of its appeal (Section 8.3.2) coupled with its wide availability. Indeed, the scale of the Ubongo Kids intervention permits cost-effectiveness results typically out of reach for programmes not delivered through a form of mass media.[138]

The scale achieved by Ubongo Kids highlights a limitation of the CEA comparison made. That is, the comparison made in this chapter did not account for the fact that certain forms of intervention could not feasibly produce CEA results that approached those obtained by television-based initiatives. Such interventions might include school-based projects, where per-child costs are inevitably higher. Further, the projects compared differed in terms of context, population age and education level (as recognised by Sabates, Rose, Delprato & Alcott, 2018). What is more, the measure of Ubongo Kids influence used in CEA calculations was based on a cross-sectional model. The specific features of this model meant that it likely gave worthwhile findings (and might have even been biased downwards: Section 9.6). Yet, cross-sectional approaches are typically less precise than the longitudinal models against which results were compared (and particularly those in which the treatment was randomised, such as that used by Nguyen, 2008).

Lastly, it must be recognised that the results of cost-effectiveness comparisons were influenced by the choice of CEA approach. Employing the J-PAL methodology necessitated adherence to various assumptions that influenced the 'additional standard deviation per $100' figures reached. Key amongst these was the discount rate used to convert both the cost and influence of Ubongo Kids to its 'base year', which affected Ubongo Kids' apparent cost effectiveness when examining the intervention since its inception.[139] To gauge the change in results that could have arisen from adopting different CEA approaches, sensitivity to differing discount levels was investigated. This was done by assuming discount rates of 0.12 and 0.08 (instead of the 0.1 figure used by J-PAL). Increasing the discount rate to 0.12 reduced the standard gain per $100 spent to 58.63 (from 61.48, when using a

---

[137] Indeed, the pattern of cost per beneficiary (Figure 9.2) gives a clear indication of ultimate cost-effectiveness positions (Figure 9.4).

[138] The use of technology in itself does not, of course, guarantee strong cost-effectiveness results. This is indicated by the high cost-per-child estimates for dated television interventions (recognised in Section 9.5).

[139] Discount rate had no effect on CEA calculations regarding Ubongo Kids' ongoing activities. This is because the period of concern for ongoing activities was 2017 only (or, all times up to twelve months prior to the Uwezo data collection at the end of 2017).

discount rate of 0.1).[140] Conversely, using a discount rate of 0.08 increased the apparent standard deviation gain per \$100 spent to 64.50. These figures show that some variation arises from using different discount rates. However, the relative consistency of Ubongo Kids' results suggests that strong cost-effectiveness performance would have been identified under alternate CEA approaches.

9.8   Chapter summary

This chapter began by discussing the concept of cost effectiveness and the specific estimation method through which CEA results would be generated. The following section detailed how the adoption of two programme duration estimates supported investigation into Ubongo Kids' ongoing cost effectiveness and cost effectiveness since inception. This necessitated the employment of different costs and influence values, that were ultimately used to produce two CEA findings. Limitations in comparing the cost effectiveness of Ubongo Kids with alternate programmes were recognised. These included that Ubongo Kids differed in form, target population and even method of evaluation from other interventions.

In spite of these qualifications, it must be noted that Ubongo Kids performed strongly relative to alternative projects. Indeed, this performance was particularly impressive given that cost-effectiveness results for Ubongo Kids were likely underestimated. Underestimation was probable, because the estimate of influence used in calculations could have been biased downwards (Section 9.6) and the number of programme viewers under-counted (especially when considering Ubongo Kids activities since inception: Section 9.4). The relative cost effectiveness of Ubongo Kids suggests that if a greater proportion of educational resources were allocated to television-based interventions, learning outcomes amongst children in developing contexts would be improved. This finding should help inform policy makers pursuing this outcome, especially given that no recent cost information concerning educational television is publicly available (Section 2.4).

---

[140] Increasing the discount rate results in a reduction in apparent cost effectiveness because reductions in the present value of influence outweigh reductions in the present value of costs (which were relatively small due to a large portion of programme cost being incurred soon after the base year).

10 Chapter 10 – Conclusion

10.1 Introduction

This thesis sought to address two research questions concerning educational television. The primary research question asked, 'What is the association between naturally occurring exposure to educational television and mathematics capability in a developing country context?'. The subsidiary research question asked, 'How cost-effective is educational television in a developing country context?'. The purpose of this chapter is to synthesise the findings from investigation into both research questions, detail the methodological, empirical and theoretical contributions of this thesis, draw out some implications for policy and consider potential future research.

This chapter is structured as follows:

- Section 10.2 provides a summary of findings, divided between the primary research question and subsidiary research question
- Section 10.3 details the empirical, methodological and theoretical contributions of this thesis
- Section 10.4 makes suggestions for future research
- Section 10.5 gives a final summary that concludes this chapter and thesis

The findings summarised in Section 10.2 suggest that normal exposure to educational television is positively related to mathematics capability and that educational television interventions are highly cost effective. In themselves, these findings are a significant contribution to the literature on educational media. This is in part because they act to triangulate similar findings from other studies concerning educational television in developing contexts that have typically concerned television exposure in controlled settings (Section 10.3). While research into normal educational television exposure could be progressed still further using additional approaches (Section 10.4), this thesis represents a substantial step forward in advancing knowledge on an important topic.

10.2 Summary of findings

Chapters 5 to 8 presented findings regarding the primary research question, while Chapter 9 reported findings concerning the subsidiary research question. In all chapters, a Tanzanian cartoon, Ubongo Kids, was used as the vehicle for investigation. Findings relevant to the primary research question were based on three quantitative models. These were 1) a cross-section approach applied to a dataset concerning Ilala and Temeke districts, 2) a cross-section design applied to a dataset for Tanzania as a whole, and 3) a multilevel model applied to a panel dataset comprising 45 districts across Tanzania. Information on these models is presented in Section 10.2.1. After this, findings concerning the subsidiary research question from Chapter 9 are detailed in Section 10.2.2.

<div style="border:1px solid black;">

Key findings

Exposure to educational television is positively associated with mathematics capability amongst Tanzanian children.

_____

Educational television provides a highly cost-effective means of enhancing mathematics capability.

</div>

10.2.1 Findings concerning the primary research question

Overview

Exposure to Ubongo Kids was identified to be positively associated with mathematics capability in all models. That is, the coefficient for Ubongo Kids exposure was a positive value when running a cross-sectional design regarding Ilala and Temeke (Section 5.2), a cross-section regression concerning Tanzania (Section 6.2) and a district-level longitudinal design (Section 7.2). However, while the coefficient value was consistently positive, significant findings were identified for cross-section approaches only. Still, the lack of a significant result from the longitudinal design was not considered to disprove cross-sectional findings. This is because the longitudinal design possessed various methodological shortcomings, which acted to counter its inherent advantages over cross-sectional approaches (as explained in the following subsection).

Panel data model

A multilevel longitudinal model was applied to a panel dataset featuring 45 districts measured in 2013 and 2017 (Chapter 7). It was recognised that using this model would help to strengthen investigation into the primary research question by addressing numerous limitations of the other (cross-section) designs in this thesis. Amongst the limitations addressed, it was considered particularly advantageous that longitudinal analysis would control for unobserved (and fixed) district-level characteristics that might be correlated with mathematics capability. However, the advantages of the longitudinal approach were countered by various shortcomings of the dataset used for multilevel analysis – the 'Panel set'.

The Panel set was produced using child-level information captured by Uwezo in 2013 and 2017. Due to child data being anonymised in 2013, it was not possible to create a dataset in which individual children were followed from a pre-exposure time point (2013) to a post-exposure time point (2017) (cf. Section 10.4). As such, the Panel set merely contained 45 district-level averages at each time point. In addition to this sample limitation, the Panel set also suffered from differences in the

measurement of mathematics outcomes across Uwezo surveys. Changes in the measurement of mathematics from 2013 to 2017 led to an increase in test ceiling effects. Ceiling effects were addressed by using district mathematics ranking in quantitative analysis. However, considering change in district ranking acted to further reduce precision (see Section 3.3.1.4.3).

The Panel set model suggested that district-level exposure to Ubongo Kids had a positive albeit non-significant effect on change in district capability ranking over time. That is, the interaction between Ubongo Kids exposure (measured in 2017 only) and time was not identified to be significant. While this result fails to disprove the null hypothesis that Ubongo Kids exposure had no positive effects on mathematics capability over time, it must be considered in light of the shortcomings recognised above. Indeed, the combination of an imprecise dependent variable and small sample size likely influenced the production of other non-significant coefficient results (see Section 7.2). These included the lack of a significant finding for the main effect of being enrolled in school – a characteristic that has consistently been identified to be related to learning outcomes (see, for example, Rose & Alcott, 2015). Given these limitations, more accurate information on the primary research question was likely uncovered through the models applied to cross-sectional datasets.

Ilala and Temeke model

Chapter 5 presented the cross-sectional design applied to the dataset for Ilala and Temeke. This dataset, the 'Matched set', was based on a sample of 811 children who were surveyed by Uwezo in 2017 and then participated in primary data collection immediately afterwards. Due to their presence in both datasets, children in the 'Matched set' possessed demographic and mathematics outcome information (captured during the Uwezo survey) as well as character recognition-based Ubongo Kids exposure information (obtained from primary data collection). This information permitted the application of a cross-section model with extensive controls. These helped to control for observable differences between children that were correlated with mathematics capability.

The cross-sectional model showed that Ubongo Kids exposure was positively related to mathematics capability in Ilala and Temeke districts. This finding was deemed relatively convincing, given that the model employed control variables for 'exposure to other media' (based on child recognition of characters that do not feature in Ubongo Kids) in addition to numerous other child characteristics typically associated with learning outcomes (excluding tutoring and type-of-school attended: see below). Further, logic-based reasoning even suggested it to be likely that the exposure coefficient was underestimated by the model. This was because children might have learnt information in topics that were not covered during Uwezo assessment, and exposure might have been missed amongst certain children.

This said, there were other potential limitations with the cross-sectional model which could reduce confidence in the findings concerning exposure. Children's mathematics capability prior to the

intervention was not controlled for. There were also no control measures for certain other important child characteristics, such as prior pre-school attendance or the type of school currently attended (due to the large number of missing values in the Uwezo data, outlined in Section 4.3.1). Lastly, given that the primary research question considers educational television in the context of an *entire* developing country, it should be recognised that findings concerning Ilala and Temeke might not apply directly to Tanzania as a whole. This is emphasised by the stark differences in district characteristics across Tanzania (shown in Section 4.4).

Tanzania model

The other cross-sectional approach used to examine the association between educational television and mathematics capability was applied to a national sample (Chapter 6). In using a comparable design to the model employed to analyse the Ilala and Temeke dataset, the Tanzania-wide cross-section model possessed several of the same limitations. These included that the model did not control for children's school type, participation in private tutoring, or pre-exposure mathematics capability. Because such measures were not controlled for, questions about the validity of inferring a causal relationship undoubtedly remain.

However, the model was specified in order to minimise bias created by omitted variables as much as possible. Child-level heterogeneity was accounted for using available controls, which included a measure of non-mathematics attainment. This would have acted to limit bias resulting from children with higher levels of overall 'intelligence' or motivation being more likely to watch Ubongo Kids. What is more, all within-household (unobserved and observed) differences between children were controlled for using household-level fixed effects. The employment of household-level fixed effects was permitted by the scale of the dataset used for analysis. Within this dataset, there were a large number of children in households with assessed siblings (of which 39,717 were in the final dataset for analysis). The dataset for analysis featured all districts sampled in the Uwezo (2017) survey, which was structured to be representative of Tanzania as a whole. As such, *the identification of a significant coefficient for Ubongo Kids through this model provides persuasive evidence of relevance to Tanzania in its entirety*.

By concerning Tanzania as a whole, this cross-section approach addresses the primary research question most directly. This is because the question considers the relationship between television exposure and mathematics capability in the context of an entire developing country, which both other designs covered to a lesser degree. The longitudinal model and cross-sectional model concerning Ilala and Temeke used information from 45 and 2 districts, respectively, while the nationally representative sample employed in the Tanzanian cross-sectional approach featured 56 districts. It might also be noted that the relative strength of the Tanzania-wide cross-section model led to it being selected as the basis of the influence result employed in cost-effectiveness calculations (see Section 9.2.2). The

findings from all quantitative models are now expanded upon using qualitative data and educational television theory.

Qualitative findings

The qualitative information captured from interviewees in four Dar es Salaam-based families typically supported positive quantitative findings. Children frequently claimed to have learnt mathematics material included in Uwezo tests (Section 8.4.1). While learning claims could have been inaccurate, numerous interviewees provided claims within the recollection of an episode storyline. It was recognised that such learning claims possessed a reduced chance of fabrication. Further, the reference to educational content within an episode narrative was understood to provide an indication as to *why* educational television exposure might lead to greater mathematics capability.

To make this argument, Section 8.4.1 referred to the first key stage of Fisch's transfer of learning theory (Fisch, 2004). This stage, which involves initial comprehension, has itself been considered using the capacity model (Piotrowski, 2014). According to this model, where educational and narrative contents are interwoven, the processing of both concepts becomes "complementary rather than competitive, and comprehension is likely to be strengthened" (Fisch, 2004, p. 145). Because simultaneous interviewee reference to educational and narrative content suggests that children perceived both concepts to be intertwined in Ubongo Kids episodes, qualitative data hint that the intervention *should* benefit mathematics learning.

Further, interview material relevant to the themes of 'Private benefits' (Section 8.3.2) and 'Exposure factors' (Section 8.3.3) supported the logic-based reasoning presented in quantitative results chapters that showed it to be probable that statistical findings were downwards biased. Interviewee responses detailing the 'Private benefits' of viewing included learning claims made with regards to topics outside of the Uwezo assessment. Should Ubongo Kids have promoted learning in topics that were not assessed by Uwezo, it would have been the case that statistical findings failed to reflect the total benefits of Ubongo Kids exposure. Additionally, information under the 'Exposure factors' theme suggested that the exposure measures used in each quantitative model were susceptible to both random and directional biases. Overall, the identification of these biases supported the idea that statistical approaches underestimated the relationship between exposure and mathematics capability (in Chapter 5 and 6) and reduced the likelihood of identifying significant results for the intervention (in Chapter 5, 6 and 7).

10.2.2 Findings concerning the subsidiary research question

The subsidiary research question concerns the cost effectiveness of educational television in a developing country context. Cost effectiveness was investigated both with regards to the ongoing operations of Ubongo Kids as well as all activities since inception. The examination of cost

effectiveness was conducted using the J-PAL method. Employing this method ensured that cost-effectiveness analysis (CEA) calculations possessed a degree of rigour (because the assumptions used by J-PAL were the product of thorough deliberation, by Dhaliwal, Duflo, Glennerster and Tuloch, 2012). Further, using J-PAL's approach facilitated comparison with numerous other educational programmes that had been assessed in the same manner.

Cost-effectiveness estimates were created through a formula that drew on Ubongo Kids' cost, influence, and number of viewers. The number of Ubongo Kids viewers was estimated from information on programme viewing provided by children in the Uwezo 2017 survey. The estimated total cost of the Ubongo Kids intervention was obtained by summating the costs provided by Ubongo for various programme categories. Costs includes items such as staff wages – included in the category of 'programme administration and staff costs' – and staff courses – which fell under 'staff training'. Influence was established using the coefficient for Ubongo Kids exposure from the Tanzania-wide cross-section model multiplied by the standard deviation of the dependent variable (in the dataset for analysis).

The cost effectiveness of Ubongo Kids – whether considering ongoing operations or activities since inception – compared very favourably against other interventions. Cost-effectiveness calculations concerning the ongoing operations of Ubongo Kids suggested it to be more cost effective than any other programme to have been assessed using the J-PAL methodology. Additionally, when estimating the cost effectiveness of Ubongo Kids' activities in Tanzania since its inception (2013), the result was only bettered by one other programme (where children were provided with statistics demonstrating the returns to education: Nguyen, 2008). These CEA results were striking, given that the cost effectiveness of Ubongo Kids was likely underestimated. This probable underestimation was due to CEA calculations using downwards biased measures of programme influence and number of beneficiaries (with the latter particularly affecting CEA estimates for Ubongo Kids' activities since inception: Section 9.8).

Ubongo Kids' strong CEA results were the product of low per-viewer costs as opposed to high per-viewer influence. Before calculating CEA results, initial influence and cost comparisons were made between Ubongo Kids and the interventions against which its cost effectiveness would ultimately be judged. The influence of Ubongo Kids' ongoing operations and activities since inception was exceeded by all other programmes assessed using the J-PAL methodology (that had a positive and significant influence finding).[141] The cost of the Ubongo Kids intervention per child was, however, extremely low compared to other programmes. The low cost-per-viewer therefore drastically

---

[141] Ubongo Kids might have been relatively likely to have claimed this unfortunate title. This is because the influence of the programme was established using a larger sample than was used for any other J-PAL assessed programme, meaning that a significant result could have been identified with a smaller coefficient.

outweighed the relatively small learning gains attributable to Ubongo Kids, when producing CEA estimates.

It might be noted, however, that the cost-per-beneficiary of Ubongo Kids highlights a shortcoming of comparing cost-effectiveness results. CEA comparisons do not account for the fact that certain interventions might differ drastically in form to the programmes against which they are considered. Educational programmes such as those involving cash-transfers (see, for example, Baird, McIntosh & Ozler, 2011) possess an almost zero chance of achieving cost-per-beneficiary results that approach those of the television-based programme under discussion. Beyond this, CEA comparisons did not account for the fact that Ubongo Kids targeted a different population and was even measured using a different method to alternate programmes.

Yet despite these shortcomings in the comparison of CEA results, the vast extent to which Ubongo Kids results bettered those from alternate interventions provides valuable information to policy makers. The availability of this information could act to promote the key economics of education concern of resource allocation (Dearden, Machin & Vignoles, 2011). That is, policy maker action upon CEA findings would promote a redirection of available educational resources towards television-based programmes that could act to enhance human capital development in low-income nations. Following the adopted theory (HCT), this human capital development should ultimately produce wage gains (Section 1.3.1).

10.3 Methodological, empirical and theoretical contributions

This subsection details the methodological, theoretical and empirical contributions of this thesis. These contributions are of relevance to differing groups. Methodological contributions are of particular importance to researchers investigating educational media. These contributions focus on the measurement of naturally occurring exposure to television, which proved a key consideration in each design employed. The theoretical guidance provided in this thesis is likely also most pertinent to researchers concerned with educational media. Theoretical considerations include reflection on the applicability of HCT to the study of educational television. Lastly, the empirical contributions of this thesis are of relevance to policy makers. These contributions centre on the effects and cost effectiveness of educational television programmes in developing country contexts. Methodological, empirical and theoretical contributions are addressed under separate sub-headings, below.

Theoretical contributions

This thesis in itself provides evidence of the applicability of HCT to research involving educational television. This theory has not typically been explicitly adopted by studies concerning educational television (notwithstanding dated acknowledgement of the role of classroom-based television in

human capital formation: Carnoy, 1975).[142] However, while this thesis has indeed applied HCT, it must be recognised that not all costs and benefits in the human capital-based framework created for this investigation (Section 1.3.1) were adequately measured. For example, investigation into the enhanced skillset obtained by watching Ubongo Kids was based on a test featuring very few mathematics concepts. Conversely, Ubongo Kids likely supported the learning of other academic subjects and alternate mathematics topics (see Section 8.4.1). Such shortcomings show that this specific application of HCT to the study of educational television was undoubtedly imperfect.

It is also possible to consider the extent to which the Ubongo Kids intervention is aligned with HCT. At the outset of this thesis, it was recognised that ideas within Ubongo Kids' theory of change (ToC) corresponded with the form of HCT employed (Section 1.3.1). Amongst the ties recognised between Ubongo Kids' ToC and human capital, it was noted that both supported an investment in skills amongst the young. That is, the ToC states Ubongo Kids' intention to target primary age children, while Heckman's (2000) interpretation of HCT posits that the human capital return is higher when investigating into younger people. Findings concerning the primary research question (Chapters 5, 6, 7 and 8) provided empirical support for alignment between Ubongo Kids and HCT. This is because the identification that Ubongo Kids was positively related to mathematics capability suggests that the intervention may lead to human capital development amongst young viewers.

This thesis additionally provides information on a theory specific to educational television, Fisch's (2004) transfer of learning theory. This chapter has previously recognised that this theory could be invoked to provide a possible explanation for why educational television might have a positive effect on mathematics capability (Section 10.2.1). It should also be recognised, however, that identifying Ubongo Kids to be associated with mathematics capability and have delivered educational content that was intertwined with episode narrative might itself be perceived to corroborate (the initial comprehension stage of) Fisch's transfer of learning theory. This is the case because empirical findings align with the theory's recognition that understanding of educational content amongst viewers is increased where the distance between narrative and educational material is small (Piotrowski, 2014).

Methodological contributions

This thesis offers valuable information on the measurement of naturally occurring exposure to educational television. Methodological advancement in this area might well be required, given that the majority of previous television-focused studies in developing contexts have concerned controlled

---

[142] It could be recognised that the methods employed in numerous investigations into the effect of educational television, which did not specifically refer to the notion of human capital, did provide information on skillset development that is clearly compatible with HCT approaches (Section 2.2.1).

exposure in artificial contexts (see Section 2.4).[143] Two separate exposure measures were employed in quantitative models used to investigate the effects of educational television. It was not initially intended that distinct exposure measures would be used across different designs. The exposure measure first submitted to Uwezo for inclusion in their national assessment (in 2017) corresponded precisely with the character recognition measure that features in the primary dataset (see Section 3.3.1.3.2). However, the limited space available in the Uwezo survey booklet led to strict format requirements. These requirements necessitated creating a simple binary measure of self-reported child viewership. This appeared only to be a disadvantage, as information from a self-reported measure is unlikely to be as accurate as that from a character recognition-based assessment. However, this resulted in two measures of exposure being available for children in the 'Matched set', which enabled comparison between both approaches.

Conducting this comparison provided an important validation of the binary measure of exposure. This is because the comparison suggested an association between exposure scores (derived from character recognition results) and self-reported exposure. Given that self-reported exposure was positively associated with a more reputable measure, self-reported exposure was considered to provide an acceptable exposure proxy for use in Tanzania-wide analysis. This finding could aid future research efforts concerning educational television: in instances where character-based recognition is not possible (due to, for example, survey space limitations), relevant information can be captured merely from self-reported viewership responses.

However, future research that follows this methodological guidance should also take note of the limitations of the binary exposure measure that were uncovered. The recognised limitations included that self-reported information provided a proxy for exposure that was most likely an underestimate. This was suggested by various pieces of information. The comparison of child character recognition results against viewership reports showed that numerous children who did not report viewing Ubongo Kids recently were able to recognise characters from the show (Section 4.4.2). Additionally, interviews with children in Dar es Salaam suggested that those who did not report recent viewing had received some exposure at earlier points (Section 8.4.2.3). This was shown to be the case, as numerous children, who reportedly had not watched in the past week, were capable of recalling in-depth information from earlier shows.

This thesis also contributes methodologically to the measurement of educational television exposure when this is assessed using character recognition items. This is because this thesis was the first to employ a latent exposure estimate that was derived from character recognition responses. This

---

[143] For example, studies concerning international variants of Sesame Street have examined the impact of delivering specific amounts of multimedia learning material to set groups during a 6-week study period (Borzekowski & Macha, 2010).

exposure estimate provides a more precise way of differentiating between test takers than merely summing the total number of characters recognised (as had featured in all previous research, see, for example, Borzekowksi & Macha, 2010). Additional precision is generated by creating IRT scores, which reflect varying item difficulty and discrimination.

What is more, this research has considered character recognition results against interview data. Character recognition is indeed a measure that typically provides "a reliable assessment of children's exposure" (Rimal, Figueroa & Storey, 2013, p. 605). However, it has been noted that further research into this approach is required (Cole & Lee, 2016). Findings from the comparison between character recognition and interview data suggested that using character recognition results as the sole measure of exposure could lead to bias. This was identified to be the case because children frequently made spontaneous references to characters during interviews, that they were not able to identify when presented with an image during survey assessment. In some instances, children were not able to recognise any Ubongo Kids characters from images but were capable of spontaneously recalling multiple character names during discussions (see Section 8.4.2.2). Such occurrences suggest that lower levels of exposure could be missed by character recognition-based tests.

Empirical contributions

This thesis also makes a meaningful empirical contribution to the literature on educational television, by addressing numerous research gaps identified in the structured literature search (presented in Chapter 2). Examination of the primary research question tackled the lack of research concerning mathematics-focused television programmes directed at school-age children (outside of America). Further, exploration into the association between naturally occurring exposure to television and mathematics capability provided a valuable addition to research in developing contexts that has largely been conducted in controlled environments. Conducting research only in controlled settings provides limited information on the day-to-day consumption of educational television. The nature of exposure could, for example, be altered by co-viewing and discussion between a child and their caregiver (see Section 2.2.2). Educational policy makers should therefore be informed by positive findings concerning normal television exposure, which act to triangulate existing studies in regulated environments.

Indeed, the cost-effectiveness findings produced in examining the subsidiary research question should certainly make policy makers take notice. Prior to this thesis, no CEA results had been produced for educational television interventions in developing country contexts. In addressing this gap, CEA results showed educational television to be a highly cost-effective intervention. This suggests that directing a greater portion of resources towards educational television gives a viable means of promoting human capital development and, ultimately, producing income gains (following HCT: Section 1.3.1).

10.4 Recommendations for further research

Identifying the methodology through which educational television research could be best advanced is challenging. Randomisation at the individual level is not considered the most suitable approach, as the delivery of exposure in a controlled environment is considered to differ so starkly to day-to-day viewing that other designs are preferred. Existing investigation into the effects of educational television in developing countries also features a relative wealth of RCTs in artificial settings (Section 2.2.4). Similarly, randomisation of a television intervention at a regional level within a country that had hitherto received no exposure is not encouraged due to implementation difficulties. The first difficulty would be identifying a country that had not received exposure to the educational television intervention under investigation. If this intervention were to be Ubongo Kids, options would be limited given that this programme has already been televised in 31 countries (on free-to-air and pay-to-view channels). After identifying the country, the next difficulty would involve negotiating with broadcasters to ensure that the show was aired only in randomly selected regions (of which there must be enough to permit a study with sufficient power).[144]

Instead, this thesis proposes a method that is both rigorous and feasible to implement using data that has already been collected: a longitudinal design concerning the effect Ubongo Kids. This design would address the identified shortcomings of the longitudinal model employed in this thesis, which was limited by being based on district-aggregated arithmetical means. That is, the model examined the effect of district mean viewership on change in district mathematics capability ranking over time. This led to coarseness in the measurement of key variables and produced a limited sample size. Both these problems would be tackled by creating a child-level longitudinal model. The creation of such a model was not attempted in this thesis due to the fact that the 2013 Uwezo dataset does not contain child identification information. It might, however, be possible to circumvent the lack of identification information in publicly available data through negotiation with Uwezo.

Should Uwezo grant temporary access to personal identification information for their 2013 dataset, a child-level longitudinal approach would become possible. While Uwezo does not grant public access to personal information to maintain participant confidentiality, it is certainly possible that my relationship with the organisation could provide the basis for access to this data at a future point. Indeed, negotiations to achieve such information from Uwezo's 2017 dataset were successful (with this information used to merge primary and Uwezo data for children in Temeke). If 2013 identification information were obtained, children could be matched between datasets using child age and location as well as caregiver telephone number data (which were drawn on to create the dataset for analysis of Ilala and Temeke, see Section 3.3.1.1.1). This process would also likely benefit from

---

[144] Notwithstanding the difficulties in carrying out such a design, it must also be recognised that this approach would still be susceptible to some bias. Children in one area might, for example, benefit from a region-specific school programme during the study.

employing text string matching techniques to identify children in both the 2013 and 2017 data (with these techniques being used to aid deduplication in the structured literature search: Section 2.2). This would result in a dataset featuring individual children who were assessed at both pre- (2013) and post-exposure (2017) timepoints, thereby creating a cohort aged 7-12 at baseline and 11-16 at endline.[145]

The analysis of data captured from children in 2013 and 2017 could also involve investigation into concepts beyond the relationship between educational television exposure and mathematics capability. Such research might, for example, draw on household television ownership (considered previously for 2017: Section 1.2.1). Doing so would enable future research to explore whether household television ownership acted to enhance or diminish the effects of exposure over time. It is theorised that television ownership would diminish the effects of exposure, as children with no household television access who had reported exposure would have co-viewed the show in a location outside of their home (with co-viewing and co-discussion likely to enhance the effects of viewing: Section 2.2.2). This said, it remains possible that children with household television access who reported Ubongo Kids exposure in 2017 could have been exposed more consistently in the lead up to this time point (due to ease of access). If this were the case, longitudinal models might suggest that household television access enhances the effect of exposure. This would, however, likely be a product of the coarse exposure variable available (that is limited in the information it provides on exposure intensity), as opposed to exposure itself being enhanced by household television ownership.

10.5 Final summary

This thesis set out to examine two key questions. These concerned the association between educational television exposure in everyday settings and mathematics capability, and the cost effectiveness of educational television interventions in developing country contexts. Investigation into normal television exposure in one such context, Tanzania, suggested it to be positively related to mathematics learning. Two of three models employed to examine this question showed that exposure to Ubongo Kids – employed as a vehicle through which to investigate educational television as a whole – was positively and significantly associated with mathematics capability. While the third model applied to the primary research question produced a non-significant (albeit positive) result, it was considered probable that this finding was attributable to shortcomings in the research design. Investigation of the subsidiary research question suggested educational television to be a highly cost-effective intervention, relative to other programmes that have been assessed using the J-PAL approach. It is hoped that these findings both capture the immediate attention of policy makers

---

[145] Uwezo's survey at both time points was based on random sampling. It is highly probable that the vast majority of children assessed in 2013 would not, therefore, have been re-assessed in 2017. However, the scale of Uwezo's national survey means that there is a good possibility that sufficient children could be matched between datasets to permit worthwhile child-level longitudinal research.

seeking to increase human capital development in low-income contexts and stimulate further research into educational television and other widely accessible educational technologies.

References

Abeberese, A. B., Kumler, T. J., & Linden, L. L. (2014). Improving reading skills by encouraging children to read in school: A randomized evaluation of the Sa Aklat Sisikat reading program in the Philippines. *Journal of Human Resources*, *49*(3), 611–633.

Acemoglu, D., & Autor, D. (2011). *Lectures in labor economics*. https://economics.mit.edu/files/4689

ACER: The Australian Council for Educational Research, Centre for Global Education Monitoring. (2015). *Report on the concurrent validity and inter-rater reliability studies of Uwezo*. http://www.educationinnovations.org/sites/default/files/Concurrent%20Validity%20and%20Inter-rater%20Reliability%20-%20Uwezo%20Kenya%20-%20June%202015.pdf

Alcott, B. (2016). *Cognitive Framing and Its Theoretical Implications for Tuition Fee Policies in England* (Working Paper 04/2016; Faculty of Education University of Cambridge Working Paper Series). University of Cambridge, Faculty of Education. https://foeworkingpapers.files.wordpress.com/2017/01/alcott-working-paper-submission-oct-2016.pdf

Alcott, B., & Rose, P. (2015). *How can education systems become equitable by 2030?* (DFID Think Pieces). DFID. http://uis.unesco.org/sites/default/files/documents/how-can-education-systems-become-equitable-by-2030-learning-and-equity_pauline-rose_benjamin-alcott_heart_2015-en.pdf

Alcott, B., & Rose, P. (2016). Does private schooling narrow wealth inequalities in learning outcomes? Evidence from East Africa. *Oxford Review of Education*, *42*(5), 495–510.

Alcott, B., & Rose, P. (2017). Learning in India's primary schools: How do disparities widen across the grades? *International Journal of Educational Development*, *56*, 42–51.

Anderson, D. R., Huston, A. C., Schmitt, K. L., Linebarger, D. L., & Wright, J. C. (2001). Early Childhood Television Viewing and Adolescent Behavior: The Recontact Study. *Monographs of the Society for Research in Child Development*, *66*(1), 1–147.

Anderson, D. R., Lavigne, H. J., & Hanson, K. G. (2013). The Educational Impact of Television. In A. N. Valdivia (Ed.), *The International Encyclopaedia of Media Studies*. Wiley-Blackwell.

Anderson, D. R., & Subrahmanyam, K. (2017). Digital Screen Media and Cognitive Development. *Pediatrics*, *140*, S57–S61.

Anfara, V. A., Brown, K. M., & Mangione, T. L. (2002). Qualitative Analysis on Stage: Making the Research Process More Public. *Educational Researcher*, *31*(7), 28–38.

Arksey, H., & Knight, P. T. (1999). *Interviewing for social scientists: An introductory resource with examples*. Sage.

Baird, S., McIntosh, C., & Özler, B. (2011). Cash or condition? Evidence from a cash transfer experiment. *The Quarterly Journal of Economics*, *126*(4), 1709–1753.

Baker, F. B. (2001). *The basics of item response theory* (2nd ed.). ERIC.

Ball, S., & Bogatz, G. A. (1970). *The First Year of Sesame Street: An Evaluation. Final Report, Volume III of V Volumes* (PR-70-15; ETS Program Report).
https://eric.ed.gov/?q=%22sesame+street%22&pg=15&id=ED047823

Ball, S., & Bogatz, G. A. (1973). *A Summary of the Major Findings from 'Reading with Television: An Evaluation of the Electric Company.'*
https://eric.ed.gov/?q=%22educational+television%22+AND+%22gain%22&pg=4&id=ED115679

Banerjee, A. V., Cole, S., Duflo, E., & Linden, L. (2007). Remedying education: Evidence from two randomized experiments in India. *The Quarterly Journal of Economics*, *122*(3), 1235–1264.

Barr, R., Lauricella, A., Zack, E., & Calvert, S. L. (2010). Infant and early childhood exposure to adult-directed and child-directed television programming: Relations with cognitive skills at age four. *Merrill-Palmer Quarterly (1982-)*, 21–48.

Batada, A., Banarjee, S., & Subramanian, M. (2016). Community engagement in India. In Charlotte F. Cole & J. H. Lee (Eds.), *The Sesame Effect: The Global Impact of the Longest Street in the World*. Routledge.

Beaujean, A. A. (2014). *Latent variable modeling using R: A step-by-step guide*. Routledge.

Becker, G. (1981). *A Treatise on the Family*. Harvard University Press.

Becker, G. S. (1964). *Human capital: A theoretical and empirical analysis, with special reference to education*. National Bureau of Economic Research.

BERA: British Educational Research Association. (2011). *Ethical guidelines for educational research* (2nd ed.). https://www.bera.ac.uk/researchers-resources/publications/ethical-guidelines-for-educational-research-2018

Beyens, I., Eggermont, S., & Nathanson, A. I. (2016). Understanding the relationship between mothers' attitudes toward television and children's television exposure: A longitudinal study of reciprocal patterns and the moderating role of maternal stress. *Media Psychology*, *19*(4), 638–665.

BGM Policy Innovations. (2011). *An evaluation of the educational impact of Galli Galli Sim Sim anganwadi intervention in Mumbai*.

Bietenbeck, J., Ericsson, S., & Wamalwa, F. M. (2017). *Preschool Attendance, School Progression, and Cognitive Skills in East Africa* (Working Paper IZA DP No. 11212; Discussion Paper Series, pp. 1–38). IZA Institute of Labor Economics.

Blimpo, M. P., & Evans, D. K. (2011). *School-Based Management and Educational Outcomes: Lessons from a Randomized Field Experiment*. World Bank. http://siteresources.worldbank.org/EDUCATION/Resources/Blimpo-Evans_WSD-2012-01-12.pdf

Bonus, J. A., & Mares, M.-L. (2015). Learned and Remembered But Rejected Preschoolers' Reality Judgments and Transfer From Sesame Street. *Communication Research*, *46*(3), 375–400.

Borzekowski, D. (2018). A quasi-experiment examining the impact of educational cartoons on Tanzanian children. *Journal of Applied Developmental Psychology*, *54*, 53–59.

Borzekowski, D. L., & Henry, H. K. (2011). The Impact of" Jalan Sesama" on the Educational and Healthy Development of Indonesian Preschool Children: An Experimental Study. *International Journal of Behavioral Development*, *35*(2), 169–179.

Borzekowski, D., Lando, A. L., Olsen, S. H., & Giffen, L. (2019). The Impact of an Educational Media Intervention to Support Children's Early Learning in Rwanda. *International Journal of Early Childhood*, 1–18.

Borzekowski, D., & Macha, J. E. (2010). The role of Kilimani Sesame in the healthy development of Tanzanian preschool children. *Journal of Applied Developmental Psychology*, *31*(4), 298–305.

Boulet, M.-M., Boudreault, S., & Guerette, L. (1998). Effects of a Television Distance Education Course in Computer Science. *British Journal of Educational Technology*, *29*(2), 101–111.

Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. *Review of Research in Education*, *24*, 61–100.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101.

Brenner, M. (2006). Handbook of Complementary Methods in Education Research. In J. Green, G. Camilli, & P. Elmore, *Handbook of Complementary Methods in Education Research* (pp. 357–370). American Educational Research Association.

Burde, D., & Linden, L. (2013). The Effects of Village-Based Schools: Evidence from a Randomized Controlled Trial in Afghanistan. *American Economic Journal: Applied Economics*, *5*, 27–40.

Bureau of Labor Statistics. (n.d.). *CPI Inflation Calculator*. Retrieved 20 October 2019, from http://www.bls.gov/data/inflation_calculator.htm

Carlson, S. A., Fulton, J. E., Lee, S. M., Maynard, L. M., Brown, D. R., Kohl, H. W., & Dietz, W. H. (2008). Physical Education and Academic Achievement in Elementary School: Data From the Early Childhood Longitudinal Study. *American Journal of Public Health*, *98*(4), 721–727.

Carnoy, M. (1975). The economic costs and returns to educational television. *Economic Development and Cultural Change*, *23*(2), 207–248.

Charmaz, K., & Bryant, A. (2011). Qualitative Data Analysis. In D. Silverman (Ed.), *Qualitative Research* (3rd ed.). SAGE Publications.

Cheah, B. C. (2009). Clustering standard errors or modeling multilevel data. *University of Columbia*, 2–4.

Chen, W. H., & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics*, *22*(3), 265–289.

Cohen, L., Manion, L., & Morrison, K. (2011). *Research methods in education*. Routledge.

Cole, C. F., & Lee, J. H. (2016). Measuring the difference. In C. F. Cole & J. H. Lee (Eds.), *The Sesame Effect: The Global Impact of the Longest Street in the World*. Routledge.

Cole, Charlotte F. (2016). The Global Sesame Effect. In Charlotte F. Cole & J. H. Lee (Eds.), *The Sesame Effect: The Global Impact of the Longest Street in the World*. Routledge.

Corder-Bolz, C. R. (1980). Mediation: The role of significant others. *Journal of Communication*, *30*(3), 106–118.

Cotton, J., & Baker, S. T. (2019). A data mining and item response mixture modeling method to retrospectively measure Diagnostic and Statistical Manual of Mental Disorders-5 attention deficit hyperactivity disorder in the 1970 British Cohort Study. *International Journal of Methods in Psychiatric Research*, *28*(1), 1–13.

Crawford, C., & Vignoles, A. (2014). *Heterogeneity in graduate earnings by socio-economic background* (W14/30). IFS Working Paper. https://www.ifs.org.uk/publications/7419

Crawley, A. M., Anderson, D. R., Wilder, A., Williams, M., & Santomero, A. (1999). Effects of repeated exposures to a single episode of the television program Blue's Clues on the viewing behaviors and comprehension of preschool children. *Journal of Educational Psychology*, *91*(4), 630–637.

Creswell, J. W. (2015). *A concise introduction to mixed methods research*. SAGE Publications.

Creswell, J. W., & Plano Clark, V. L. (2007). *Designing and conducting mixed methods research*. SAGE Publications.

Crotty, M. (1998). *The foundations of social research: Meaning and perspective in the research process*. SAGE Publications.

de Bruin, B., & Floridi, L. (2017). The Ethics of Cloud Computing. *Science and Engineering Ethics*, *23*(1), 21–39.

Dearden, L., Machin, S., & Vignoles, A. (2011). The contribution of the economics of education to education. *Disciplines of Education. Their Role in the Future of Education Research*, 85–102.

Delaney, J. A., & Kearney, T. D. (2015). The impact of guaranteed tuition policies on postsecondary tuition levels: A difference-in-difference approach. *Economics of Education Review*, *47*, 80–99.

Delprato, M., & Sabates, R. (2015). A multilevel analysis of late entry in Nigeria. *International Journal of Research & Method in Education*, *38*(1), 83–99.

Denzin, N., & Lincoln, Y. (1994). *Handbook of qualitative research*. Thousand Oaks.

Denzin, N., & Lincoln, Y. (2013). *The Landscape of Qualitative Research* (4th ed.). SAGE Publications.

Dhaliwal, I., Duflo, E., Glennerster, R., & Tulloch, C. (2012). *Comparative Cost-Effectiveness Analysis to Inform Policy in Developing Countries*. https://www.povertyactionlab.org/sites/default/files/publications/CEA%20in%20Education%202013.01.29_0.pdf

Docherty, S., & Sandelowski, M. (1999). Focus on qualitative methods: Interviewing children. *Research in Nursing & Health*, *22*(2), 177–185.

Duflo, E., Hanna, R., & Ryan, S. P. (2012). Incentives work: Getting teachers to come to school. *American Economic Review*, *102*(4), 1241–78.

Edejer, T. T.-T., Aikins, M., Black, R., Wolfson, L., Hutubessy, R., & Evans, D. B. (2005). Cost effectiveness analysis of strategies for child health in developing countries. *British Medical Journal*, *331*(7526), 1177.

Eide, P., & Allen, C. B. (2005). Recruiting transcultural qualitative research participants: A conceptual model. *International Journal of Qualitative Methods*, *4*(2), 44–56.

Eklund, M., Erlandsson, L.-K., & Hagell, P. (2012). Psychometric properties of a Swedish version of the Pearlin Mastery Scale in people with mental illness and healthy people. *Nordic Journal of Psychiatry*, *66*(6), 380–388.

Elks, P. (2016). *The impact of assessment results on education policy and practice in East Africa: DFID think piece*. DFID. http://www.heart-resources.org/wp-content/uploads/2016/04/Think-Piece-the-impact-of-assessment-results-on-education-policy-and-practice-in-East-Africa.pdf

Evans, D. K., & Popova, A. (2016). Cost-Effectiveness Analysis in Development: Accounting for Local Costs and Noisy Impacts. *World Development*, *77*, 262–276.

Evans, D., Kremer, M., & Ngatia, M. (2008). The impact of distributing school uniforms on children's education in Kenya. *World Bank, Mimeo*. https://www.povertyactionlab.org/sites/default/files/publications/The%20Impact%20of%20Distributing%20School%20Uniforms.pdf

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. SAGE Publications.

Finch, H., & Edwards, J. M. (2016). Rasch Model Parameter Estimation in the Presence of a Nonnormal Latent Trait Using a Nonparametric Bayesian Approach. *Educational and Psychological Measurement*, *76*(4), 662–684.

Fisch, S. M. (2004). *Children's learning from educational television: Sesame Street and beyond*. Lawrence Erlbaum Associates.

Fisch, S. M., Kirkorian, H., & Anderson, D. (2005). Transfer of learning in informal education. *Transfer of Learning from a Modern Multidisciplinary Perspective*, 371–393.

Fisch, S. M., Lesh, R., Motoki, E., Crespo, S., & Melfi, V. (2010). Children's learning from multiple media in informal mathematics education. *National Science Foundation-Funded Research Study*. http://kids.pbskids.com:8080/cyberchase/parentsteachers/show/pdf/Cyberchase_ChildrensLearningMultipleMedia.pdf

Fix, B. (2018). *The trouble with human capital theory*.

Friese, S., Soratto, J., & Pires, D. (2018). *Carrying out a computer-aided thematic content analysis with ATLAS.ti* (MMG Working Paper) [Working Paper]. Max Planck Institute for the Study of Religious and Ethnic Diversity.

Gage, N. L. (1989). The paradigm wars and their aftermath a 'historical' sketch of research on teaching since 1989. *Educational Researcher*, *18*(7), 4–10.

Ganimian, A. J., & Murnane, R. J. (2016). Improving Education in Developing Countries Lessons from Rigorous Impact Evaluations. *Review of Educational Research*, *86*(3), 719–755.

Gibbs, G. (2007). *Analyzing Qualitative Data*. SAGE Publications, Ltd.

Glewwe, P., Ilias, N., & Kremer, M. (2010). Teacher incentives. *American Economic Journal: Applied Economics*, *2*(3), 205–227.

Glewwe, P., Kremer, M., & Moulin, S. (2009). Many children left behind? Textbooks and test scores in Kenya. *American Economic Journal: Applied Economics*, *1*(1), 112–135.

Glewwe, P., Kremer, M., Moulin, S., & Zitzewitz, E. (2004). Retrospective vs. prospective analyses of school inputs: The case of flip charts in Kenya. *Journal of Development Economics*, *74*(1), 251–268.

Gomaa, W. H., & Fahmy, A. A. (2013). A survey of text similarity approaches. *International Journal of Computer Applications*, *68*(13), 13–18.

Grbich, C. (2013). *Qualitative Data Analysis: An Introduction* (2nd ed.). SAGE.

Gullason, E. T. (1989). The consumption value of schooling: An empirical estimate of one aspect. *Journal of Human Resources*, 287–298.

Hardman, F., Abd-Kadir, J., Agg, C., Migwi, J., Ndambuku, J., & Smith, F. (2009). Changing pedagogical practice in Kenyan primary schools: The impact of school-based training. *Comparative Education*, *45*(1), 65–86.

Harel, D. (2014). *The effect of model misspecification for polytomous logistic adjacent category item response theory models* [McGill University]. http://digitool.library.mcgill.ca/R/?func=dbin-jump-full&object_id=126933&local_base=GEN01-MCG02

Hart, S., Dixon, A., Drummond, M., & McIntyre, D. (2004). *Learning Without Limits*. Open University Press.

Haskell, R. E. (2000). *Transfer of learning: Cognition and instruction*. Academic Press.

Heary, C., & Hennessy, E. (2006). Focus groups versus individual interviews with children: A comparison of data. *The Irish Journal of Psychology*, *27*(1–2), 58–68.

Heckman, J. J. (2000). Policies to foster human capital. *Research in Economics*, *54*(1), 3–56.

Hendriks, J., Fyfe, S., Styles, I., Skinner, S. R., & Merriman, G. (2012). Scale construction utilising the Rasch unidimensional measurement model: A measurement of adolescent attitudes towards abortion. *The Australasian Medical Journal*, *5*(5), 251.

Hennessy, S. (2014). *Bridging between research and practice: Supporting professional development through collaborative studies of classroom teaching with technology*. Sense Publishers.

Hennessy, S., Rojas-Drummond, S., Higham, R., Márquez, A. M., Maine, F., Ríos, R. M., García-Carrión, R., Torreblanca, O., & Barrera, M. J. (2016). Developing a coding scheme for analysing classroom dialogue across educational contexts. *Learning, Culture and Social Interaction*, *9*, 16–44.

Hogan, D. R., Baltussen, R., Hayashi, C., Lauer, J. A., & Salomon, J. A. (2005). Cost effectiveness analysis of strategies to combat HIV/AIDS in developing countries. *British Medical Journal*, *331*(7530), 1431–1437.

Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods*, *6*(1), 53–60.

Howe, C., & Abedin, M. (2013). Classroom dialogue: A systematic review across four decades of research. *Cambridge Journal of Education*, *43*(3), 325–356.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55.

Hummel-Rossi, B., & Ashdown, J. (2002). The state of cost-benefit and cost-effectiveness analyses in education. *Review of Educational Research*, *72*(1), 1–30.

Jamison, D. T., Klees, S. J., & Wells, S. J. (1978). *The costs of educational media: Guidelines for planning and evaluation* (Vol. 3). SAGE Publications.

Jellison, J. A., & Wolfe, D. E. (1999). Video songs from Sesame Street: A comparison of fifth graders' and adults' opinions regarding messages for preschool children. *Journal of Research in Music Education*, *47*(1), 64–77.

Johnes, G. (1993). Human Capital. In G. Johnes (Ed.), *The Economics of Education* (pp. 5–26). Springer.

Jones, S. (2016). Measuing Learning Outcomes. In I. I. Munene (Ed.), *Achieving Education for All: Dilemmas in System-Wide Reforms and Learning Outcomes in Africa*. Lexington Books.

Jones, S. (2017). *Do schools in low income countries really produce so little learning?* University of Copenhagen. https://economics.handels.gu.se/digitalAssets/1643/1643680_17.-sam-jones-2017-02-20.schoolprod.v2.pdf

Jones, S., & Schipper, Y. (2015). Does Family Background Matter for Learning in East Africa? *Africa Education Review*, *12*(1), 7–27.

Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Scientific Software International.

J-PAL: Abdul Latif Jameel Poverty Action Lab. (2014). *Cost-effectiveness comparisons: Full workbook*. J-PAL. https://www.povertyactionlab.org/research-resources/cost-effectiveness

Kearney, M. S., & Levine, P. B. (2019). Early Childhood Education by Television: Lessons from Sesame Street. *American Economic Journal-Applied Economics*, *11*(1), 318–350.

Kerr, A., & Quinn, S. (2010). Returns to Education in Tanzania: Exploiting a natural experiment. *Centre for the Study of African Economies Conference on Economic Development in Africa*, 21–23.

Kim, S. W. (2015). Effectiveness of a satellite educational television program for Ethiopian secondary education. *Distance Education*, *36*(2), 419–436.

Kirkorian, H. L., Wartella, E. A., & Anderson, D. R. (2008). Media and young children's learning. *The Future of Children*, *18*(1), 39–61.

Klees, S. J. (1995). Economics of Education Tecnology. In M. Carnoy (Ed.), *International Encyclopedia of Economics of Education* (pp. 398–406). Pergamon.

Kremer, M., Brannen, C., & Glennerster, R. (2013). The challenge of education and learning in the developing world. *Science*, *340*(6130), 297–300.

Kremer, M., Duflo, E., & Dupas, P. (2011). Peer effects, teacher incentives, and the impact of tracking. *American Economic Review*, *101*(5), 1739–1774.

Kremer, M., Miguel, E., & Thornton, R. (2009). Incentives to learn. *The Review of Economics and Statistics*, *91*(3), 437–456.

Kuchah, K., & Pinter, A. (2012). 'Was this an interview?'Breaking the power barrier in adult-child interviews in an African context. *Issues in Educational Research*, *22*(3), 283–297.

Lapinid, M., Gustilo, L., Magno, C., Barrot, J., Gabinete, M., & Anito, J. (2017). For the Nth Time, Does Educational Technology Really Make a Difference? - A Large-Scale Investigation of the Effects of Educational TV on Academic Achievement. *Advanced Science Letters*, *23*(2).

Lavigne, H. J., & Anderson, D. R. (2012). Television and Children's Knowledge. In A. Pinkman M., T. Kaefer, & S. Neuman B. (Eds.), *Knowledge Development in Early Childhood: Sources of Learning and Classroom Implications*. Guilford Press.

Lee, J. H. (2009). *Sisimpur's reach and educational impact: Evidence from a national longitudinal survey of a Sesame Street project in Bangladesh*. Sesame Workshop. http://downloads.cdn.sesame.org/sw/SWorg/documents/ACPR+Monograph+Bangladesh+FINAL.pdf

Levin, H. M. (1988). Cost-Effectiveness and Educational Policy. *Educational Evaluation and Policy Analysis*, *10*(1), 51–69.

Levin, H. M., & Belfield, C. (2015). Guiding the development and use of cost-effectiveness analysis in education. *Journal of Research on Educational Effectiveness*, *8*(3), 400–418.

Levin, H. M., & McEwan, P. J. (2001). *Cost-effectiveness analysis: Methods and applications* (Vol. 4). SAGE Publications.

Liamputtong, P. (2010). *Performing qualitative cross-cultural research*. Cambridge University Press.

Linebarger, D. L., & Walker, D. (2005). Infants' and toddlers' television viewing and language outcomes. *American Behavioral Scientist*, *48*(5), 624–645.

Linebarger, Deborah L., McMenamin, K., & Wainwright, D. K. (2008). *Summative evaluation of Super WHY!* Univeristy of Pennsylvania. http://www-tc.pbskids.org/read/files/SuperWHY_Research_View.pdf

Linebarger, Deborah L., & Piotrowski, J. T. (2010). Structure and strategies in children's educational television: The roles of program type and learning strategies in children's learning. *Child Development*, *81*(5), 1582–1597.

Machin, S., & McNally, S. (2008). The literacy hour. *Journal of Public Economics*, *92*(5), 1441–1462.

Mares, M.-L., & Pan, Z. (2013). Effects of Sesame Street: A meta-analysis of children's learning in 15 countries. *Journal of Applied Developmental Psychology*, *34*(3), 140–151.

Marginson, S. (2019). Limitations of human capital theory. *Studies in Higher Education*, *44*(2), 287–301.

Markham, P. (1999). Captioned Videotapes and Second-Language Listening Word Recognition. *Foreign Language Annals*, *32*(3), 321–328.

Mauthner, M. (1997). Methodological aspects of collecting data from children: Lessons from three research projects. *Children & Society*, *11*(1), 16–28.

Maynard, M. (1994). Methods, practice and epistemology: The debate about feminism and research. In M. Maynard & J. Purvis (Eds.), *Researching Women's Lives from a Feminist Perspective* (p. Chapter 1). Taylor and Francis.

McEwan, P. J. (2012). Cost-effectiveness analysis of education and health interventions in developing countries. *Journal of Development Effectiveness*, *4*(2), 189–213.

Medley, A., Kennedy, C., O'Reilly, K., & Sweat, M. (2009). Effectiveness of peer education interventions for HIV prevention in developing countries: A systematic review and meta-analysis. *AIDS Education and Prevention*, *21*(3), 181.

Miguel, E., & Kremer, M. (2004). Worms: Identifying impacts on education and health in the presence of treatment externalities. *Econometrica*, *72*(1), 159–217.

Miltiades, H. B. (2008). Interview as a social event: Cultural influences experienced while interviewing older adults in India. *International Journal of Social Research Methodology*, *11*(4), 277–291.

Mugo, J. K., Ruto, S. J., Nakabugo, M. G., & Mgalla, Z. (2015). A Call to Learning Focus in East Africa: Uwezo's Measurement of Learning in Kenya, Tanznania and Uganda. *Africa Education Review*, *12*(1), 48–66.

Murrar, S., & Brauer, M. (2018). Entertainment-education effectively reduces prejudice. *Group Processes & Intergroup Relations*, *21*(7), 1053–1077.

Nagin, D. S., & Sampson, R. J. (2019). The real gold standard: Measuring counterfactual worlds that matter most to social science and policy. *Annual Review of Criminology*, *2*, 123–145.

National Bureau of Statistics, & Office of Chief Government Statistician. (2013). *2012 Population and Housing Census*. Ministry of Finance. http://www.tzdpg.or.tz/fileadmin/documents/dpg_internal/dpg_working_groups_clusters/cluster_2/water/WSDP/Background_information/2012_Census_General_Report.pdf

Nguyen, T. (2008). Information, role models and perceived returns to education: Experimental evidence from Madagascar. *Unpublished Manuscript*, *6*. http://xxpt.ynjgy.com/resource/data/20091115/U/MIT20091115040/OcwWeb/Economics/14-771Fall-2008/Readings/PaperJM%20TRANG%20NGUYEN%2022jan08.pdf

NVivo. (n.d.). *Coding Query Guidance*. Retrieved 20 October 2019, from http://help-nv11.qsrinternational.com/desktop/procedures/run_a_coding_comparison_query.htm#MiniTOCBookMark10

OECD. (2014). *Education at a Glance 2014: OECD Indicators*. OECD. http://dx.doi.org/10.1787/eag-2014-en

Olaniyan, D. A., & Okemakinde, T. (2008). Human capital theory: Implications for educational development. *Pakistan Journal of Social Sciences*, *5*(5), 479–483.

Oliver, D. E. (2004). Human capital theory and higher education in developing countries. *Journal of Thought*, *39*(1), 119–130.

Olumorin, C. O., Aderoju, M. A., & Onojah, A. O. (2018). Students Awareness and Utilization of Educational Broadcasts to Learn in Ogbomoso, Oyo State Nigeria. *Turkish Online Journal of Distance Education*, *19*(3), 182–192.

Ong, Y. M., Williams, J., & Lamprianou, I. (2013). Exploring differential bundle functioning in mathematics by gender: The effect of hierarchical modelling. *International Journal of Research & Method in Education*, *36*(1), 82–100.

Palmer, E. (1993). *Toward a literate world: Television in literacy education—Lessons from the Arab region*. Westview Press.

Partington, G. (2001). Qualitative Research Interviews: Identifying Problems in Technique. *Issues in Educational Research*, *100*(2).

Paulsen, C. A., & Andrews, J. R. (2014). The Effectiveness of Placing Temporal Constraints on a Transmedia STEM Learning Experience for Young Children. *E-Learning and Digital Media*, *11*(2), 204–213.

Pensiero, N., & Green, A. (2018). The effects of post-compulsory education and training systems on literacy and numeracy skills: A comparative analysis using PISA 2000 and the 2011 survey of adult skills. *European Journal of Education*, *53*(2), 238–253.

Perkins, D., Radelet, S., & Block, S. (2013). *Economics of Development* (7th ed.). Norton & Company.

Perraton, H. (2005). *Open and distance learning in the developing world*. Routledge.

Piotrowski, J. T. (2014). Participatory cues and program familiarity predict young children's learning from educational television. *Media Psychology*, *17*(3), 311–331.

Pradhan, M., Suryadarma, D., Beatty, A., Wong, M., Alishjabana, A., Gaduh, A., & Artha, R. P. (2011). *Improving educational quality through enhancing community participation: Results from a randomized field experiment in Indonesia* (Working Paper No. 5795). The World Bank. https://openknowledge.worldbank.org/bitstream/handle/10986/3559/WPS5795.pdf?sequence=1

Rimal, R. N., Figueroa, M. E., & Storey, J. D. (2013). Character recognition as an alternate measure of television exposure among children: Findings from the Alam Simsim Program in Egypt. *Journal of Health Communication*, *18*(5), 594–609.

Romero, M., Sandefur, J., & Sandholtz, W. (2017). *Can Outsourcing Improve Liberia's Schools? Preliminary Results from Year One of a Three-Year Randomized Evaluation of Partnership Schools for Liberia* (Working Paper No. 462). Centre for Global Development. https://www.cgdev.org/publication/partnership-schools-for-liberia

Ruthven, K., Hennessy, S., & Deaney, R. (2008). Constructions of dynamic geometry: A study of the interpretative flexibility of educational software in classroom practice. *Computers & Education*, *51*(1), 297–317.

Sabates, R., Alcott, B., Rose, P., & Delprato, M. (2018). *Cost-effectiveness with equity: Raising learning for marginalised girls through Camfed's programme in Tanzania* (Policy Brief Policy Paper No. 18/2). REAL Centre, University of Cambridge.

Sabates-Wheeler, R., Sabates, R., & Devereux, S. (2018). Enabling graduation for whom? Identifying and explaining heterogeneity in livelihood trajectories post-cash transfer exposure. *Journal of International Development*, *30*(7), 1071–1095.

Schlesinger, M. A., Flynn, R. M., & Richert, R. A. (2016). US preschoolers' trust of and learning from media characters. *Journal of Children and Media*, 1–20.

Schultz, T. P. (2004). School subsidies for the poor: Evaluating the Mexican Progresa poverty program. *Journal of Development Economics*, *74*(1), 199–250.

Schweisfurth, M. (2013). *Learner-centred Education: Whose pedagogy for whose development?* Routledge.

Schweisfurth, M. (2015). Learner-centred pedagogy: Towards a post-2015 agenda for teaching and learning. *International Journal of Educational Development*, *40*, 259–266.

Scott, D. (2005). Critical Realism and Empirical Research Methods in Education. *Journal of Philosophy of Education*, *39*(4), 633–646.

Shah, S. (2004). The researcher/interviewer in intercultural context: A social intruder! *British Educational Research Journal*, *30*(4), 549–575.

Sheather, S. (2009). *A modern approach to regression with R. New York*. Springer.

Silverman, D. (2011). *Interpreting Qualitative Data*. SAGE Publications.

Singer, J. L., & Singer, D. G. (1994). *Barney and Friends as education and entertainment: Phase 2— Can children learn through preschool exposure to Barney and Friends?* Yale University Family Television Research and Consultation Center.

Sriprakash, A. (2012). *Pedagogies for development: The politics and practice of child-centred education in India*. Springer Science & Business Media.

Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences*, *42*(5), 893–898.

Sternberg, R. J. (2007). Culture, instruction, and assessment. *Comparative Education*, *43*(1), 5–22.

Talabi, J. K. (1989). The Comparative Effects of Televised and Programmed Instruction. *Journal of Educational Television*, *15*(1), 17–24.

Tan, E. (2014). Human Capital Theory: A Holistic Criticism. *Review of Educational Research*, *84*(3), 411–445. JSTOR.

Teddlie, C., & Tashakkori, A. (2012). Common "core" characteristics of mixed methods research: A review of critical issues and call for greater convergence. *American Behavioral Scientist*, *56*(6), 774–788.

Thomas, N., & O'Kane, C. (1998). The ethics of participatory research with children. *Children & Society*, *12*(5), 336–348.

Tikly, L., & Bond, T. (2013). Towards a postcolonial research ethics in comparative and international education. *Compare: A Journal of Comparative and International Education*, *43*(4), 422–442.

Torres, L., Driscoll, M., & Burrow, A. (2010). Racial Microaggressions and Psychological Functioning Among Highly Achieving African-Americans: A Mixed-Methods Approach. *Journal of Social and Clinical Psychology*, *29*, 1074–1099.

Ubongo Kids. (2015a). *Kibena and the Math Rats*. https://www.youtube.com/watch?v=TZ-h1ZtUwgM

Ubongo Kids. (2015b). *The Mosquito Army*. https://www.youtube.com/watch?v=8ZHvCUAr5b0

United Nations. (n.d.). *United Nations (UN) population estimates*. Retrieved 20 October 2019, from https://population.un.org/wpp/Download/Standard/Interpolated/

Usher, R. (1996). A critique of the neglected epistemological assumptions of educational research. In D. Scott & R. Usher (Eds.), *Understanding Educational Research* (p. Chapter 2). Routledge.

Uwezo. (n.d.-a). *About us*. Retrieved 20 October 2019, from http://www.uwezo.net/about-us/

Uwezo. (n.d.-b). *Training Manuals*. Retrieved 20 October 2019, from http://www.uwezo.net/assessment/training/

Uwezo. (2012). *Are our children learning: Report for 2011*. Twaweza. http://www.uwezo.net/wp-content/uploads/2012/08/TZ_2011_AnnualAssessment-Report.pdf

Uwezo. (2013a). *Kenya Training Manual*. http://www.uwezo.net/wp-content/uploads/2012/08/2013-Kenya-Training-Manual.pdf

Uwezo. (2013b). *Kenya Volunteer Workbook*. http://www.uwezo.net/wp-content/uploads/2012/08/2013-Kenya-Volunteer-Workbook.pdf

Uwezo. (2015). *Are our children learning: Report for 2014*. http://www.twaweza.org/uploads/files/UwezoTZ-ALA2014-FINAL-EN.pdf

Uwezo. (2016). *Data Cleaning Protocol*. Twaweza. https://www.twaweza.org/uploads/files/TZ15-Data%20cleaning%20and%20usage%20NOTES%202015.pdf

van der Ark, L. A. (2007). Mokken Scale Analysis in R. *Journal of Statistical Software*, *20*(1), 1–19.

van Leeuwen, L., Renes, R. J., & Leeuwis, C. (2013). Televised Entertainment-Education to Prevent Adolescent Alcohol Use: Perceived Realism, Enjoyment, and Impact. *Health Education & Behavior*, *40*(2), 193–205.

Vandewater, E. A., & Lee, S.-J. (2009). Measuring children's media use in the digital age: Issues and challenges. *American Behavioral Scientist*, *52*(8), 1152–1176.

Villanueva-Mansilla, E., & Olivera, P. (2012). Institutional Barriers to Development Innovation: Assessing the Implementation of XO-1 Computers in Two Peri-Urban Schools in Peru (English). *Information Technologies & International Development*, *8*(4), 177–189.

Warrens, M. J. (2014). New interpretations of Cohen's kappa. *Journal of Mathematics*, *2014*, 1–10.

Wartella, E., Richert, R. A., & Robb, M. B. (2010). Babies, Television and Videos: How Did We Get Here? *Developmental Review*, *30*(2), 116–127.

Webb, N. L. (1982). *Summative Evaluation of Let Me See! An ITV Science Series for Grades 1 and 2* (pp. 1–76). Wisconsin Educational Communications Board. https://archive.org/details/ERIC_ED228990/page/n8

Wolff, L., de Moura Castro, C., Navarro, J. C., & García, N. (2002). Television for secondary education: Experience of Mexico and Brazil. *Technologies for Education: Potentials, Parameters and Prospects*, 144–152.

Wright, J. C., Huston, A. C., Murphy, K. C., St. Peters, M., Piñon, M., Scantlin, R., & Kotler, J. (2001). The relations of early television viewing to school readiness and vocabulary of children from low-income families: The early window project. *Child Development*, *72*(5), 1347–1366.

Yen, W. M. (1984). Effects of Local Item Dependence on the Fit and Equating Performance of the Three-Parameter Logistic Model. *Applied Psychological Measurement*, *8*(2), 125–145.

R packages

Chalmers, P. (2019). mirt: Multidimensional Item Response Theory.

Feinerer, I., Hornik, K. and Artifex Software, Inc. (2018). tm: Text Mining Package.

Gaure, S. (2019). lfe: Linear Group Fixed Effects

Lüdecke, D. (2019). sjPlot: Data Visualization for Statistics in Social Science

McLean, M. W. (2017). RefManageR: Import and Manage BibTeX and BibLaTeX References in R.

Meyer, D. (2019). e1071: Misc Functions of the Department of Statistics, Probability Theory Group.

Rosseel, Y. (2019). lavaan: Latent Variable Analysis.

van der Ark, A., Straat, J. H., Koopman, L. (2018). mokken: Conducts Mokken Scale Analysis.

van der Laan, J. (2018). reclin: Record Linkage Toolkit.

Appendices

Appendix 1: Ethics forms

Appendix 1.1: Interview consent forms

Appendix 1.1 provides the consent forms that were signed by each caregiver and child that participated in interviews. These forms detail the purpose of the study, number of questions and approximate length of the interview. Additionally, the form covers confidentiality and makes it clear that interviewees can choose not to participate in any part of the interview (Section 3.2). Forms were both presented physically and read aloud to caregivers and children.

Caregiver consent form

FOMU YA RUHUSA YA UTAFITI – MAAHOJIANO YA MTOTO

Taarifa ya Utafiti

Joe ni mwanafunzi katika Chuo Kikuu cha Cambridge (Uingereza). Anafanya utafiti kuhusu elimu kwa njia ya televisheni. Mussa anafanya kazi katika Kampuni ya Ubongo Kids na anafanya utafiti kuchunguza elimu kwa njia ya TV, tumesha uliza maswali tayari kwako na mama yako. Sasa wewe na mama wako mmechaguliwa kufanya mahojiano pia.

Kupitia mahojiano haya nina tumaini kupata maelezo zaidi juu ya jinsi unavyoangalia katuni za kielimu na kuelewa vizuri jinsi gani unajifunza kutoka katika katuni hizi. Kama hutokuwa na wasi wasi kuhojiwa wakiwepo wazazi wako, wanakaribishwa kuwepo.

Jina lako halito ripotiwa - Taarifa zote zilizotolewa zitabaki kuwa siri. Huna ruhusa kwako kushiriki katika mahojiano. Zaidi unapaswa wewe kuchagua kukubali sasa, unaweza kujitoa ushiriki wako wakati wowote. Wewe Upo huru kuchagua kutokujibu swali lolote la mahojiano au kuacha mahojiano wakati wowote.Nina furaha kujadili chochote na wewe utakacho kuwanacho baada ya haya mahojiano. Nina patikana katika namba hii 0653654195.

Tafadhali naomba niambie kama una swali lolote

Ruhusa

Nimeelewa dhumuni la utafiti, Nimeruhusiwa kuuliza maswali yoyote na kuelewa kuwa nipo huru kujitoa katika utafiti wakati wowote. Nina toa ruhusa kushiriki katika katika utafiti huu.

Jina: _____   Sahihi: _____   Tarehe: DD/MM/YEAR

Caregiver consent form

FOMU YA RUHUSA YA UTAFITI – MAAHOJIANO YA MZAZI

Taarifa ya Utafiti

Joe ni mwanafunzi katika Chuo Kikuu cha Cambridge (Uingereza). Anafanya utafiti kuhusu elimu kwa njia ya televisheni. Mussa anafanya kazi katika Kampuni ya Ubongo Kids na anafanya utafiti kuchunguza elimu kwa njia ya TV, tumesha uliza maswali tayari kwako na mtoto wako. Sasa wewe na mtoto wako mmechaguliwa kufanya mahojiano pia.

Kupitia mahojiano haya nina tumaini kupata maelezo zaidi juu ya jinsi mtoto wako anavyoangalia katuni za kielimu na kuelewa vizuri jinsi gani anajifunza kutoka katika katuni hizi. Tutazungumza nawe na mtoto wako peke yake, lakini mtoto wako anaweza kuhojiwa nawe ukiwepo kama hii itafanya asiwe na wasiwasi.

Jina lako, na jina la mwanao katika kaya yako hayato ripotiwa - Taarifa zote zilizotolewa zitabaki kuwa siri. Huna ruhusa kwako au mtoto wako kushiriki katika mahojiano. Zaidi unapaswa wewe au mtoto wako kuchagua kukubali sasa, unaweza kujitoa ushiriki wako wakati wowote. Wewe na mtoto wako mpo huru kuchagua kutokujibu swali lolote la mahojiano au kuacha mahojiano wakati wowote. Nina furaha kujadili chochote na wewe au mtoto wako atakacho kuwanacho baada ya haya mahojiano. Nina patikana katika namba hii 0653654195.

Tafadhali naomba niambie kama una swali lolote

Ruhusa

Nimeelewa dhumuni la utafiti, Nimeruhusiwa kuuliza maswali yoyote na kuelewa kuwa mimi na mtoto wangu ambae nina wajibika kwake tupo huru kujitoa katika utafiti wakati wowote. Nina toa ruhusa kushiriki katika katika utafiti huu, na watoto katika kaya yangu kushiriki katika utafiti huu.

Jina: _____   Jina la mtoto: _____

Sahihi: _____   Tarehe: DD/MM/YEAR

Appendix 1.2: Research visa and national research approval

The images below feature a copy of my research visa (Figure A.1) and approval of the study design from Tanzania's Commission for Science and Technology (COSTECH: Figure A.2).

Image redacted

Figure A. 1: Research visa

Image redacted

Figure A. 2: National study design approval

Appendix 1.3: National Bureau of Statistics approval

The figure below features a scanned copy of approval from Tanzania's National Bureau of Statistics
(NBS) regarding the sample frame employed when collecting primary quantitative data.

Figure A. 3: National Bureau of Statistics approval

Appendix 1.4: Contracts between research groups involved in primary quantitative data collection

The below images feature the contracts brokered between the Guluka Kwalala Youth Environment Group (GYEG), the Tanzania Users and Survivors Psychiatry Organisation (TUSPO) and Ubongo, that formed the basis research technician hiring and data collection in Ilala and Temeke.
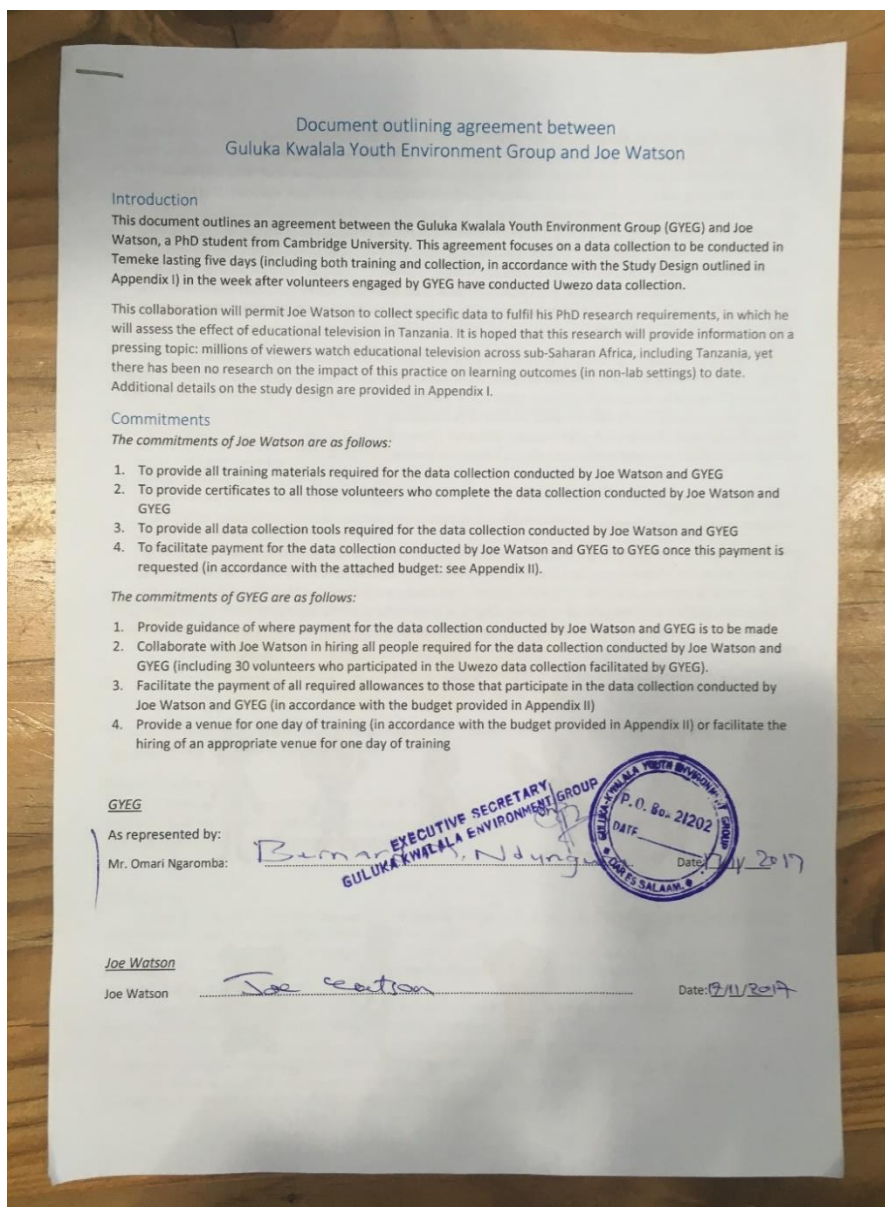
Figure A. 4: Temeke data collection agreement

Figure A. 5: Ilala data collection agreement

Appendix 2: Correspondence between Ubongo Kids episodes and Uwezo assessment

Investigation into the topics covered in *Ubongo Kids* episodes suggested a reasonable degree of consistency with the Uwezo assessment. The mathematics component of the Uwezo test (administered in 2017) assessed number recognition, counting, number patterns, addition and subtraction, all of which were covered in *Ubongo Kids* episodes. A selection of Uwezo tasks is provided in Figure A.6. The link between Uwezo tasks and *Ubongo Kids* cartoons is evidenced by the sampled episode, 'Kibena and the Maths Rats', which covered addition and subtraction. That said, certain *Ubongo Kids* topics have gone beyond the basic concepts covered in the Uwezo assessment (with this point corresponding with the identification in Section 3.3.1.3.1 that Uwezo covers only more fundamental concepts: ACER, 2015). Cartoon topics have included 'converting decimals to fractions' in 'Decimals on the Radio' and 'basic algebra' in 'Math Music'. This suggests that the Uwezo assessment might not capture all features of mathematics capability affected by Ubongo Kids (see Section 7.3.2.1).

Figure A. 6: Uwezo mathematics tasks

Appendix 3: Quantitative data collection tool

1.ii character recognition: Mtoto ataoneshwa picha nakutakiwa kutaja majina ya wahusika au watu/viumbe maarufu, itajulikana kuwa mtoto ametambua majina ya wahusika endapo ameyataja kwa usahihi. (Mtoto hatakiwi kuonyesha muhusika fulani). Usiruhusu mzazi au mtu yoyote kumsaidia mtoto katika kutambua wahusika hawa

Unatakiwa kusema:

"Nina picha nataka kukuonyesha. Unaweza kuniambia majina ya hawa watu/viumbe maarafu?"

2. Character strengths   "Katika sehemu hii nitakusomea maelezo ya wanafunzi tofauti. Tafadhali sikiliza kwa makini maelezo haya na kisha niambie ni mara ngapi unajisikia au unatenda kama mtoto huyo. Ninaposema ni mara ngapi, ningependa uniambie kama hua unatabia kama hizi - sifanyi hivyo, mara chache, mara nyingi, au kila siku.

Nitakupatia maelezo kama mfano. Ngoja nikuambie kuhusu msichana anayeitwa Sarah, Sarah kila siku anakutana na rafiki zake. *Je, kwa kiasi gani unafanya kama Sarah?*
Kwenye kuchagua maneno haya, 'Sifanyi hivyo', 'Mara chache', 'Mara nyingi', 'Kila siku', au 'sijui', chaguo langu litakuwa 'mara nyingi'. Nimechagua neno hili kwa sababu ninakutana na rafiki zangu mara nyingi, lakini sio kila siku.

Unaweza kuchagua jibu lolote unalolitaka. Tafadhali naomba uwe mkweli katika majibu yako. Sasa unajaribu!"

DUMMY QUESTION	Ngoja nikuambie kuhusu kijana anayeitwa Sam. Sam anapendelea kula wali.	Je, kwa kiasi gani wewe unafanya kama Sam?

Je mtoto ameweza kujibu swali la mfano linganisha maelezo yake? Weka alama ya vema katika Hapana/Ndio. Kama Hapana, maliza jaribio

GRIT

G1	Ngoja nikuambie kuhusu kijana anayeitwa Bakari. Bakari anafanya kazi kwa bidii kila mara.	Je, kwa kiasi gani wewe unafanya kama Bakari?
G2	Hawa anamaliza kufanya kazi zote za nyumbani.	Je, kwa kiasi gani wewe unafanya kama Hawa?
G3	Zawadi anapoona kazi ni ngumu huacha kuijaribu.	Je, kwa kiasi gani wewe unafanya kama Zawadi?
G4	Naomi pale anapokuwa ameshindwa kazi kwa mara ya kwanza, huendelea kuijaribu.	Je, kwa kiasi gani wewe unafanya kama Naomi?
G5	Hadija huendelea kujaribu hata kama jambo analofanya ni gumu sana kwake.	Je, kwa kiasi gani wewe unafanya kama Hadija?

MINDSET

M1	Amina anafikiri kwamba wototo wenye akili tu ndio wanafanya vizuri kwenye hisabati.	Je, mara ngapi wewe unafikiri kama hivi?
M2	Daudi anafikiri kwamba watoto wanaweza kufanya vizuri kwenye hisabati kama wakisoma kwa bidi.	Je, mara ngapi wewe unafikiri kama hivi?
M3	Ignas anafikiri kwamba kuna watoto hawatafanya vizuri hisabati hata kama wakijaribu kusoma kwa bidi.	Je, mara ngapi wewe unafikiri kama hivi?
M4	Mashaka anafikiri kwamba unatakiwa kuwa na akili nyingi ili kufanya vizuri kwenye hisabati.	Je, mara ngapi wewe unafikiri kama hivi?

Appendix 4: Quantitative data collection sheet

EA no.: _____ / district name: _____ / ward name: _____ /village name: _____ / HH no.: ____
Mkuu wa kaya (nakili toka katika orodha ya kaya) _____ /Namba ya simu (nakili toka katika orodha ya kaya) _____

Kaya haikufanyiwa tathmini kwa kukusanyiwa taarifa na uwezo		Kaya haina watoto (6-16)	
Kama kaya haikufanyiwa tathmini, usikusanye taarifa		*Kama kaya haina watoto (6-16), usikusanye taarifa*	
Watoto wote (6-16) au mzazi/mlezi hawapo (Jaribu 1)		Watoto wote (6-16) au mzazi/mlezi hawapo (Jaribu 2)	
Kama watoto hawapo, rudi tena baadae		*Kama watoto hawapo tena, usikusanye taarifa*	

Ruhusa ya mzazi/mlezi	Jina la mzazi/mlezi *(mama kama yupo)*	Namba ya simu ya kaya inayotumika	Uhusiano na mtoto	Je mtoto wako ameangalia kipindi cha TV cha Ubongo Kids mara ngapi mwezi uliopita?				
Ndio/Hapana				Haajangalia kabisa	Mara moja	Mara mbili	Mara tatu au zaidi	Sijui

Ruhusa ya mtoto	Ndio / Hapana	Jina la mtoto		Umri wa mtoto		Jinsia ya mtoto	Boy / Girl

1.i – Child TV exposure (all TV)

QUESTION	RESPONSE CATEGORIES	
"Uliwahi kuangalia TV?"	Ndio	
	Hapana	
	Kama mtoto hajawahi kuangalia TV, ruka hadi kipengele 1.ii	
"Huwa unaangalia TV peke yako au na watu wengine?"	Unaangalia peke yako	
	Unaangalia pamoja na rafiki zako	
kama mtoto anaangalia na watu wengine	Unaangalia pamoja na ndugu	
	Unaangalia pamoja na baba au mama	
"Huwa unaangilia TV na nani?"	Watu wengine Tafadhali nitajie (_____)	
	Nyumbani	
	Nyumbani kwa rafiki yako	
"Unaangalia TV wapi?"	Katika banda la video	
	Nyumbani kwa mwalimu wako	
	Mahali pengine Tafadhali nitajie (_____)	

1.ii – Child character recognition

1. Mtoto anaweza kumtambua Mzee Kigo (au "Kigo")	Ndio	Hapana
2. Mtoto anaweza kumtambua Uncle T (au "Anko T")	Ndio	Hapana
3. Mtoto anaweza kumtambua Mama Ndege	Ndio	Hapana
4. Mtoto anaweza kumtambua Dora the Explorer (au "Dora")	Ndio	Hapana
5. Mtoto anaweza kumtambua Kirikou	Ndio	Hapana
6. Mtoto anaweza kumtambua Diamond Platnumz (au "Diamond", au "Platnumz")	Ndio	Hapana
7. Mtoto anaweza kumtambua Akili	Ndio	Hapana
8. Mtoto anaweza kumtambua Fafa (au "Faru Faraja")	Ndio	Hapana
9. Mtoto anaweza kumtambua Kibena	Ndio	Hapana

1.iii – Child Ubongo Kids viewing

| "Unaweza kuniambia ni njia gani unatumia kuangalia au kusikiliza vipindi vya Ubongo Kids?"

Weka alama ya tick kwa majibu ya mtoto

Mtoto asipojibu mpe chaguzi zilizopo	TV	
	YouTube	
	Mkanda wa Video	
	Radio	
	Simu	
	Vingine (Tafadhali nitajie_____)	
	Sijui/Hajajibu	
	Hawakuangalia	

2 – Child grit and mindset

QUESTION	Sifanyi hivyo	Mara chache	Mara nyingi	Kila siku	Sijui/Hajajibu
C1					
G1					
G2					
G3					
G4					
G5					
M1					
M2					
M3					
M4					

C2	Mtoto anakula wali mara ngapi wiki iliyopita?		Taarifa ya mtoto imekamilika na fomu yote imejazwa	Ndio / Hapana

Appendix 5: Trialling of the Ubongo Kids exposure question submitted to Uwezo

Trialling of the exposure question submitted to Uwezo occurred during Uwezo's pilot data collection in Moshi and, subsequently, in the households located near the Ubongo office. During Uwezo piloting, initial trialling began using a single caregiver question (to provide an average estimation of viewing frequency amongst all assessed children). Piloting suggested this question to be unsuccessful, as:

- Caregivers were frequently unsure about what Ubongo Kids was. This led to a high number of "do not know" responses.
- Caregivers sometimes requested information from their children to answer the question. This suggests that two forms of information would be captured: a caregivers' estimation of child viewing activities and a child's estimation of their viewing activities.

As such, Uwezo suggested that any question submitted for inclusion should be child directed. This guidance complied with my experience of Uwezo piloting. Further, this advice corresponded with the literature on measuring educational television exposure. While caregiver questions have been employed previously to give information on child exposure, it has been recognised that one-off questions to caregivers are not entirely reliable proxies for child exposure (Rimal, Figueroa & Storey, 2013). Indeed, caregiver reports of child exposure have only been found to be moderately correlated (.20 to .40) with direct observation of child viewing activities (which were considered to provide the truest measures of exposure by Vandewater & Lee, 2009).

In accordance with piloting in Moshi and Uwezo's subsequent advice that any question included must have two response options (due to the space limitations of the data collection tool), a child-directed binary question was drafted. The child-focused question was piloted amongst children in households near the Ubongo office. This process focused on question comprehension. After establishing a question that was easily understood by children between 6-16 years, a child-focused question was submitted to (and accepted by) Uwezo (see Section 3.3.1.3.2).

Appendix 6: IRT model fit results and interpretation information for exposure to Ubongo Kids

This appendix presents fit statistics for the IRT model concerning exposure to Ubongo Kids. The TLI (0.97), CFI (0.99), RMSEA (0.05) and SRMR (0.03) all fell within acceptable boundaries (adopted in accordance with the literature on latent models, presented in the following subsections of this appendix). These findings provide no evidence that the IRT model functioned improperly. The model was therefore retained in spite of a significant M2 value (15.98, p = 0.01). This M2 finding indicated the contrary. M2 functions in a comparable manner to the chi-squared statistic employed in assessing the fit of structural equation models. These measures' susceptibility to sample size means they can provide a false indication of poor fit amongst larger samples (Jöreskog and Sörbom, 1993).

RMSEA - Root mean square error of approximation

The RMSEA is employed to gauge if "the specified model reasonably approximate the data (as opposed to assessing if it is an exact fit" Beaujean (2014). The measure is typically bound between 0 and 1, with decreasing values suggesting a better fit. Analysis shall employ the upper limit of 0.07 as a cut-off point for models (following Steiger, 2007). This measure is provided by the formula:

$$RMSEA = \sqrt{max([((\chi^2/df) - 1)/(N - 1)] , 0)}$$

SRMR – standardised root mean square residual

SRMR values are also bounded between 0 and 1, with values closer to the lower boundary again suggesting a better fit. Whilst results below 0.05 are indicative of better fitting models, "values as high as 0.08 are deemed acceptable" (Hooper et al., 2008, p. 55). In light of its cumbersome formula, a syntactic description only of the SRMR is provided (Beaujean, 2014, p. 164):

> [SRMR] is calculated using the following steps: (a) subtract the model-implied correlation matrix, C, from the sample correlation matrix, S, to produce the residual correlation matrix; (b) remove the redundant values; (c) square the remaining values; (d) sum all the squared values; (e) divide the sum by the number of non-redundant elements in the matrix; and (f) takes the square root of the resulting number.

TLI – Tucker-Lewis index

Also reported as the non-normed fit index (NNFI), the TLI "assesses the model by comparing the χ^2 value of the model to the χ^2 of the null model" (Hooper et al., 2008, p. 55). Values fall between 0 and 1, with values closer 1 suggesting a better model fit. This research adopts the threshold proposed by Hu and Bentler (1999) of ≥ 0.95 as the lower limit for acceptable model fit. This measure is provided by the formula:

$$[\chi^2/df(Null\ Model) - \chi^2/df(Proposed\ Model)] / [\chi^2/df(Null\ Model) - 1]$$

CFI – comparative fit index

In a similar manner to the TLI (above), this measure "assumes that all latent variables are uncorrelated (null/independence model) and compares the sample covariance matrix with this null model" (Hooper et al., 2008, p. 55). Values again range between 0 and 1, with a better model fit suggested by higher values. The cut-off point of ≥ 0.95 is adopted for good model fit (Hu and Bentler, 1999). CFI is provided by the formula:

$$[d(Null\ Model) - d(Proposed\ Model)]/d(Null\ Model)$$

Appendix 7: Plots of model residuals and random effects for the Panel set model
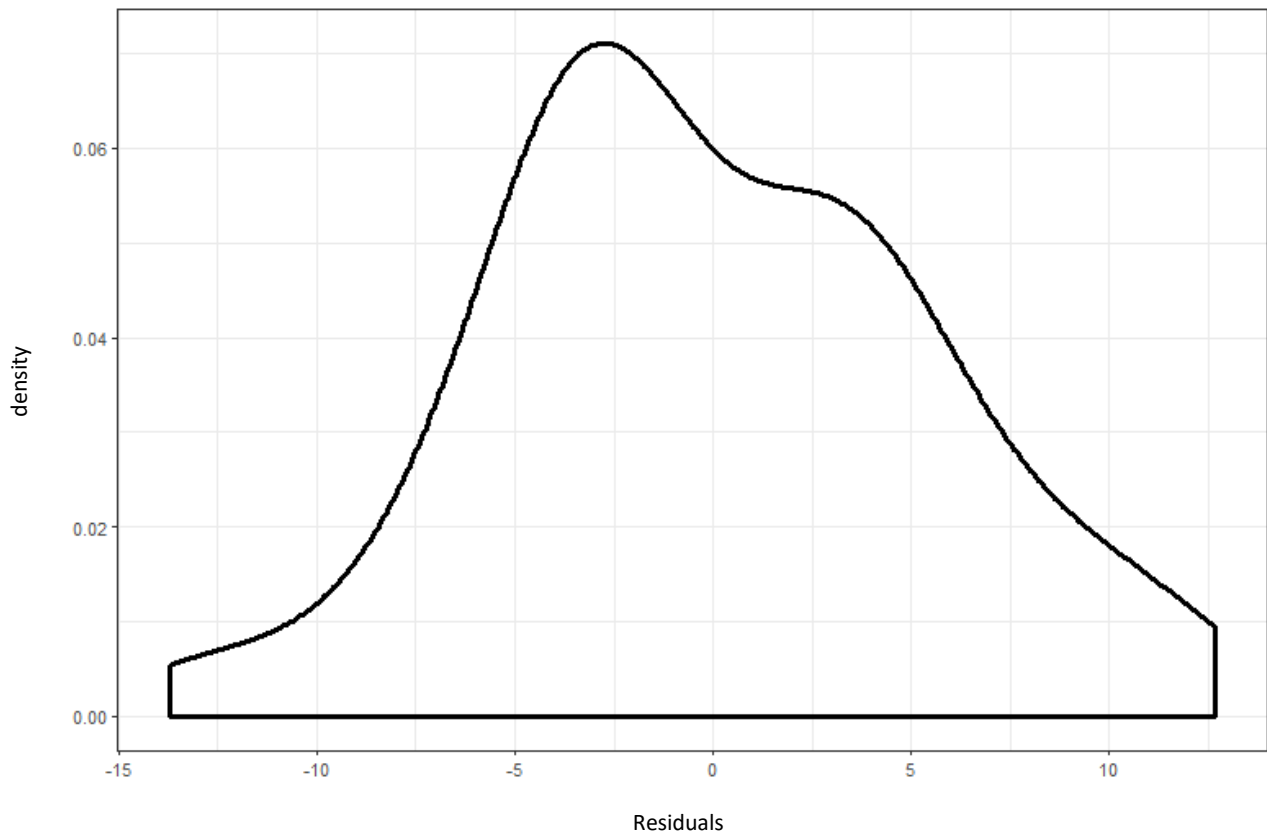
Figure A. 7: Plot of residuals for multilevel model

The above graph presents a density plot for the residuals of all districts in the 'Panel set'. The peak of the density curve is close to zero and there is little suggestion of skewness. As such, this graphic provided tentative support that the model assumption that residuals were normally distributed was satisfied.
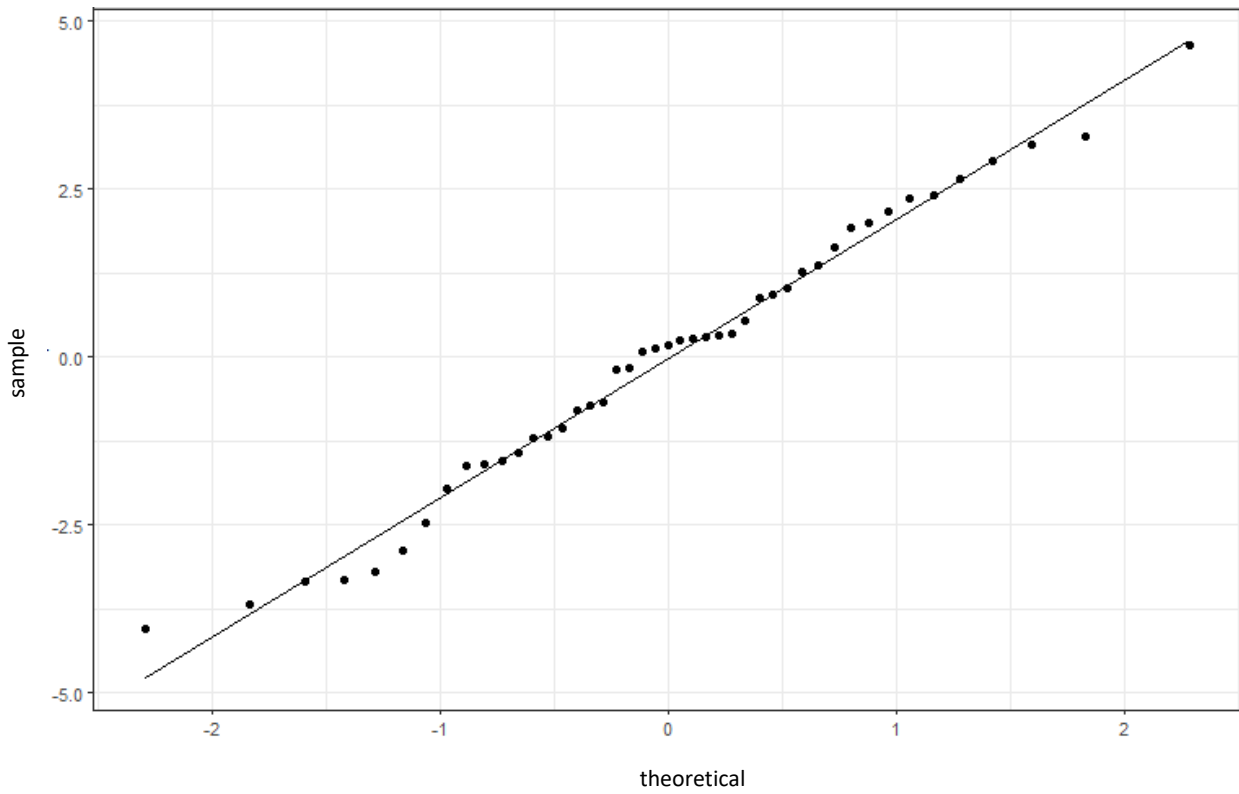
Figure A. 8: Plot of random effects for the intercept of the multilevel model

The above model presents intercept random effects (as points) against a reference line for a normally distributed variable. The closeness of points to the line provides a strong indication that random effects followed a normal distribution. As such, the assumption that random effects are normally distributed was likely satisfied.

Appendix 8: Child interview questions

1. What programmes do you watch on TV? Ni vipindi gani vya television unavyo angalia? //
o Do you watch any different shows on other devices/platforms (e.g., YouTube)? Unaangalia vipindi tofauti kwa kutumia njia gani // (mfanno YouTube)?

2. Out of all the shows you mentioned, which is your favourite? Nje ya vipindi vyote ulivyotaja, kipi unakipenda zaidi?

3. When was the last time you watched the show, Ubongo Kids? Mara ya mwisho uliangalia kipindi cha Ubongo kids lini?
o IF CHILD CAN REMEMBER: Can you tell me what happened in Ubongo Kids that time? Unaweza kuniambia kilitokea nini katika Ubongo kids ulipo angalia?
o PROMPT, IF NEEDED: What did Koba do? Koba alifanya nini?
 OR, What **maths** did Koba use? **Hesabu** gani Koba alitumia?

What platforms do you use to watch Ubongo Kids?
Unaweza kuniambia ni (n)jja gani au vitu gani unatumia kuangalia vitu vya Ubongo Kids?

4. IF MULTIPLE UBONGO KIDS PLATFORMS (suggested by child response to 'back up' survey):
o Do you prefer to watch Ubongo kids on ＿＿＿ or ＿＿＿? Why? Unapenda kuangalia Ubongo Kids katika TV au (ow) YouTube? Kwanini?
o Which platform do you use most? Why? Nini unatumia zaidi kati ya TV au YouTube? Kwanini?
IF ONE UBONGO KIDS PLATFORM (suggested by child response to 'back up' survey):
o Why do you watch Ubongo Kids on [selected platform]? Kwanini unaangalia Ubongo Kids katika [jukwaa lililochaguliwa]?

5.a Do you watch Ubongo Kids with others (and if so, who)?
Unaangalia Ubongo Kids peke yako/watu wengine (na nani)?

5.b IF WATCHES UBONGO KIDS WITH OTHERS: after you watch UK, do you
talk about the show? Baada ya kuangalia Ubongo Kids, je, mnaongelea kuhusu kipindi cha Ubongo Kids?
o What do you discuss? Can you give me an example? Mnajadili nini? Unaweza kunipa mfano?
Can you tell me about the place where you watch Ubongo Kids? Unaweza kuniambia kuhusu sehemu unayoangalia Ubongo Kids

6. If you were not watching Ubongo Kids, what would you be doing? Kama usingekuwa unaangalia Ubongo Kids, ungekuwa unafanya nini?
o PROMPT, IF NEEDED: Would you be watching another television show, or playing? Ungekuwa unaangalia vipindi vingine cha TV au ungekuwa unacheza? Zaidi ya kunakitu kingine ungekiuwa unafanya?

7. Who decides whether you watch Ubongo Kids? Ni nani anae **ku**amulia kuangalia Ubongo kids? //
o PROMPT, IF NEEDED: Do you watch the programme because you want to, or because your parent(s)/caregiver(s) says that you should? Huwa unaangalia kipindi sababu unataka au sababu mzazi/mlezi wako amesema kuwa unapaswa kuangalia?
IF THE CHILD: Why do you choose to watch Ubongo Kids? Kwanini umechagua kuangalia Ubongo Kids?
IF THE PARENT/CAREGIVER: Why do your caregivers want you to watch Ubongo Kids? Kwanini wazazi wako wanataka uangalie Ubongo Kids?

8. What do you enjoy about Ubongo Kids? Nini unafurahia kuhusu Ubongo Kids?

9. Have you learned anything from Ubongo Kids? Umejifunza chochote kutoka Ubongo Kids?
o IF NO: Why not? Kwanini hapana
o IF YES: What did you learn? Umejifunza nini?
PROMPT, IF NEEDED: Can you give me an example? OR, Can you tell me a bit more about that?
Unaweza kunipa mfano? AU: Unaweza kuniambia zaidi kuhusu hilo?
FOLLOW UP QUESTION: Can you tell me about what was happening on Ubongo Kids when you learnt this? Unaweza kuniambia kuhusu kilichotokea kwenye Ubongo Kids wakati unajifunza hilo?

10. Do you think Ubongo Kids is entertainment or education? Unadhani Ubongo Kids ni burudani au (ow) elimu // PROMPT: Kwa nini unafikiri ni elimu/burudhani?
……………………….………….

SHOW CHILD SHORT UBONGO KIDS CLIP Ninapenda kukuonyesha katuni ya UK… Unapenda kuangalia katuni hii?
……………………………………

11. Did you like this cartoon clip? Unapenda katuni hii?
o IF NO: Why not? Kwanini hapana?
o IF YES: What part did you like best? Sehemu gani umeipenda zaidi katika katuni hii? Kwa nini?

12. Did you learn anything from this cartoon clip? Umejifunza chochote toka kwenye katuni hii?
o IF NO: Why not? Kwanini hapana
o IF YES: What did you learn? Umejifunza nini?
PROMPT, IF NEEDED: Can you tell me a bit more about that? OR, Can you give me an example
Unaweza kuniambia zaidi kuhusu hilo? AU: Unaweza kunipa mfano?
o EITHER YES/NO: Was it hard or easy to learn from the clip? Why? Ilikuwa ngumu au rahisi kujifunza katika katuni hii (can point at screen)? kwanini?

END, thanking child: Asante mtoto – tutakuja nyumbani kwako kesho kuona mama yako.
………………………………………….

IF CHILD WATCHES UBONGO KIDS AT HOME, ASK PARENT:
Can you show me where your child watches Ubongo Kids? Naomba kuona ni wapi mtoto wako anaangalia Ubongo Kids?

Appendix 9: Caregiver interview questions

1. Through what platform(s) does your children access media content? Watoto wako wanaangalia vipindi kutumia nija gani?

2. How often do your children watch [media] (for 'media' use caregiver responses from q.1) Mara ngapi watoto wako wanaangalia [TV na YouTube]?
- FOLLOW UP: Do you restrict your children's viewing time: Huwa una wapangia/wachagulia watoto wako muda wa kuangalia [TV na YouTube]
After responses, "Can you help me complete this form ?" (other side of sheet, Figure A.4) – Unaweza kunisaidia kujaza fomu hii?

3. Where do your children watch [media]? Watoto wako wanaangalia [TV] wapi?
- IF AT HOME – We would like to see your TV later… Tunapenda kuona TV yako baadae…
- IF NOT AT HOME – Can you tell me more about this place? Unaweza niambia zaidi kuhusu huko/pale?

4. Do you think watching [media] is good for children's learning? // PROMPT RE SPECIFIC CHILDREN and for them: Do you (as a caregiver) learn anything from watching TV? Unafikiri kuangalia [TV na YouTube] ni vizuri kwa watoto kujifunza? // PROMPT RE SPECIFIC CHILDREN and for them: Kama mzazi unajifunza chochote unapoangalia [TV]

5. Do you watch Ubongo Kids with your children? Unaangalia kipindi cha TV cha Ubongo Kids na watoto wako?
- IF YES – After you watch Ubongo Kids, do you talk about the show // About what?
  Baada ya kuangalia Ubongo Kids, je, mnaongelea kuhusu kipindi cha Ubongo Kids
o What do you talk about? Can you give me an example?
     Mnaongalea kuhusu nini? Uweza kunipa mfano?
- Do you talk during the show? Wakati wa kuangalia Ubongo Kids, je, mnaongelea kuhusu kipindi // kuhusu nini?

6. Do you think Ubongo Kids is education or entertainment? Why… ? Unafikiri Ubongo Kids ni burudani au elimu? Kwa nini… ?

7. Who decides whether your children watches Ubongo Kids? Ni nani anae waamulia watoto wako kuangalia Ubongo Kids?
- PROMPT IF NEEDED: Do your children watch because they want to or because you want them to? Unafikiri watoto wako wanaangalia Ubongo Kids kwa sababu wanataka, au, wewe unataka?
o IF CHILD DECIDES: Why do you think your children choose to watch Ubongo Kids? Unafikiri kwanini watoto wako wamechagua kuangalia Ubongo Kids?
o IF ADULT DECIDES: Why do you choose for your children to watch Ubongo Kids? Kwa nini wewe unataka watoto wako waangalie Ubongo Kids?

8. What would your children be doing if they were not watching Ubongo Kids? Watoto wako wangekuwa wanafanya nini kama wasingekuwa wanaangalia Ubongo Kids?

9. Is your child learning anything from Ubongo Kids? Watoto wako wanajifunza chochote kutoka Ubongo Kids?
- PROMPT: What do they learn // Can you give me an example? Wamejifunza nini? // Unaweza kunipa mfano?

10. CATCH ALL QUESTION
Can you say any other thing? Unaweza kuniambia chochote kingine?
Can you say anything else about learning from TV? Unaweza kuniambia chochote kuhusu kujifunza katika TV/kipindi cha Ubongo kids?
_____

During what hours are your children normally watching [media] on weekdays and weekends?
Watoto wako wanaangalia [TV] muda gani, kwa kawaida katikati ya wiki na siku za mapumziko?

wastani katikati ya wiki		wastani siku za mapumziko	

	MUDA	WANAANGALIA	MUDA	WANAANGALIA
Asabuhi	01:00-02:00	Ndio / Hpana	01:00-02:00	Ndio / Hpana
	02:00-03:00	Ndio / Hpana	02:00-03:00	Ndio / Hpana
	03:00-04:00	Ndio / Hpana	03:00-04:00	Ndio / Hpana
	04:00-05:00	Ndio / Hpana	04:00-05:00	Ndio / Hpana
	05:00-06:00	Ndio / Hpana	05:00-06:00	Ndio / Hpana
Mchana	06:00-07:00	Ndio / Hpana	06:00-07:00	Ndio / Hpana
	07:00-08:00	Ndio / Hpana	07:00-08:00	Ndio / Hpana
	08:00-09:00	Ndio / Hpana	08:00-09:00	Ndio / Hpana
	09:00-10:00	Ndio / Hpana	09:00-10:00	Ndio / Hpana
Jioni	10:00-11:00	Ndio / Hpana	10:00-11:00	Ndio / Hpana
	11:00-12:00	Ndio / Hpana	11:00-12:00	Ndio / Hpana
	12:00-01:00	Ndio / Hpana	12:00-01:00	Ndio / Hpana
Usiku	01:00-02:00	Ndio / Hpana	01:00-02:00	Ndio / Hpana
	02:00-03:00	Ndio / Hpana	02:00-03:00	Ndio / Hpana
	03:00-04:00	Ndio / Hpana	03:00-04:00	Ndio / Hpana
	04:00-05:00	Ndio / Hpana	04:00-05:00	Ndio / Hpana
	05:00-06:00	Ndio / Hpana	05:00-06:00	Ndio / Hpana

Figure A. 9: Child television viewing schedule completed during caregiver interviews

The above image was provided to each caregiver interviewee when considering question 2, above. This was employed to capture information from caregivers on the media exposure of their child's/children's average weekday or weekend-day. (All times are in Kiswahili, meaning that times should be interpreted as 6 hours behind: "01:00 asabhui" translates directly as one o'clock in the morning, and should be interpreted as 07:00 am in English.)

Appendix 10: Cohen's Kappa values for interview coding

Table A. 1: Cohen's Kappa values from inter-rater reliability testing

Code, if features in either coder's coding of either document	File	Kappa
Exposure factors\Media content recall\Akili character recall	103	1
Exposure factors\Media content recall\Akili character recall	112	0.7882
Exposure factors\Media content recall\UK character recall	103	0.8332
Exposure factors\Media content recall\UK character recall	112	0.7497
Exposure factors\Media content recall\UK storyline and song recall	103	0.7911
Exposure factors\Media content recall\UK storyline and song recall	112	0.8495
Exposure factors\Media platforms\Platform - radio negatives	103	1
Exposure factors\Media platforms\Platform - radio negatives	112	0.9787
Exposure factors\Media platforms\Platform - TV positives	103	1
Exposure factors\Media platforms\Platform - TV positives	112	0.6896
Exposure factors\Media platforms\Platform, general	103	0.9475
Exposure factors\Media platforms\Platform, general	112	0.8481
Exposure factors\Programmes and stations viewed\Specific programmes and stations, if viewed and preferences\Specific programme preference	103	1
Exposure factors\Programmes and stations viewed\Specific programmes and stations, if viewed and preferences\Specific programme preference	112	0.6542
Exposure factors\Programmes and stations viewed\Specific programmes and stations, if viewed and preferences\Specific programme viewed	103	0.7835
Exposure factors\Programmes and stations viewed\Specific programmes and stations, if viewed and preferences\Specific programme viewed	112	1
Exposure factors\Programmes and stations viewed\Specific programmes and stations, if viewed and preferences\TV station viewership	103	0.8238
Exposure factors\Programmes and stations viewed\Specific programmes and stations, if viewed and preferences\TV station viewership	112	1
Exposure factors\Programmes and stations viewed\Type of programme viewed or listened to	103	0.0525
Exposure factors\Programmes and stations viewed\Type of programme viewed or listened to	112	0
Exposure factors\Viewing environment	103	0.9822
Exposure factors\Viewing environment	112	0.8416
Exposure factors\Viewing frequency\Viewing frequency, UK, child	103	0.9985
Exposure factors\Viewing frequency\Viewing frequency, UK, child	112	0.7758
Exposure factors\Viewing frequency\Viewing frequency, UK, parent or caregiver	103	1
Exposure factors\Viewing frequency\Viewing frequency, UK, parent or caregiver	112	0.7698
Exposure factors\Viewing frequency\Viewing or listening frequency, general, not UK specific	103	0
Exposure factors\Viewing frequency\Viewing or listening frequency, general, not UK specific	112	0.9525
HK framework\Private benefits of media viewing\Consumption value of viewing UK or other media\Entertainment content,	103	1

TV (not UK specific)		
HK framework\Private benefits of media viewing\Consumption value of viewing UK or other media\Entertainment content, TV (not UK specific)	112	0.8667
HK framework\Private benefits of media viewing\Consumption value of viewing UK or other media\Entertainment content, UK	103	0.6145
HK framework\Private benefits of media viewing\Consumption value of viewing UK or other media\Entertainment content, UK	112	1
HK framework\Private benefits of media viewing\Enhanced skill-set from viewing UK or other media\Educational content, all media\Educational content, radio	103	1
HK framework\Private benefits of media viewing\Enhanced skill-set from viewing UK or other media\Educational content, all media\Educational content, radio	112	0.7936
HK framework\Private benefits of media viewing\Enhanced skill-set from viewing UK or other media\Educational content, all media\Educational content, TV, excluding UK\Educational content, Akili\Educational content, Akili, health or drawing	103	1
HK framework\Private benefits of media viewing\Enhanced skill-set from viewing UK or other media\Educational content, all media\Educational content, TV, excluding UK\Educational content, Akili\Educational content, Akili, health or drawing	112	1
HK framework\Private benefits of media viewing\Enhanced skill-set from viewing UK or other media\Educational content, all media\Educational content, TV, excluding UK\Educational content, Akili\Educational content, Akili, maths	103	1
HK framework\Private benefits of media viewing\Enhanced skill-set from viewing UK or other media\Educational content, all media\Educational content, TV, excluding UK\Educational content, Akili\Educational content, Akili, maths	112	0.8438
HK framework\Private benefits of media viewing\Enhanced skill-set from viewing UK or other media\Educational content, all media\Educational content, TV, excluding UK\Educational content, TV, cooking, cleaning, fashion or dressing, music	103	1
HK framework\Private benefits of media viewing\Enhanced skill-set from viewing UK or other media\Educational content, all media\Educational content, TV, excluding UK\Educational content, TV, cooking, cleaning, fashion or dressing, music	112	0.8464
HK framework\Private benefits of media viewing\Enhanced skill-set from viewing UK or other media\Educational content, all media\Educational content, TV, excluding UK\Educational content, TV, maths, or reading and writing, or stories	103	1
HK framework\Private benefits of media viewing\Enhanced skill-set from viewing UK or other media\Educational content, all media\Educational content, TV, excluding UK\Educational content, TV, maths, or reading and writing, or stories	112	0.9726
HK framework\Private benefits of media viewing\Enhanced skill-set from viewing UK or other media\Educational content, all media\Educational content, TV, excluding UK\Educational content, TV, non-specific	103	1
HK framework\Private benefits of media viewing\Enhanced skill-set from viewing UK or other media\Educational content, all media\Educational content, TV, excluding UK\Educational content, TV, non-specific	112	0.8013
HK framework\Private benefits of media viewing\Enhanced skill-set from viewing UK or other media\Educational content, all media\Educational content, UK\Educational content, UK, growth mindset	103	0.8961
HK framework\Private benefits of media viewing\Enhanced skill-set from viewing UK or other media\Educational content, all media\Educational content, UK\Educational content, UK, growth mindset	112	1
HK framework\Private benefits of media viewing\Enhanced skill-set from viewing UK or other media\Educational content, all media\Educational content, UK\Educational content, UK, growth mindset\Educational content, UK, growth mindset, with specific example	103	1
HK framework\Private benefits of media viewing\Enhanced skill-set from viewing UK or other media\Educational content, all media\Educational content, UK\Educational content, UK, growth mindset\Educational content, UK, growth mindset, with specific example	112	1
HK framework\Private benefits of media viewing\Enhanced skill-set from viewing UK or other media\Educational content, all media\Educational content, UK\Educational content, UK, health	103	1
HK framework\Private benefits of media viewing\Enhanced skill-set from viewing UK or other media\Educational content, all media\Educational content, UK\Educational content, UK, health	112	0.9706
HK framework\Private benefits of media viewing\Enhanced skill-set from viewing UK or other media\Educational content, all media\Educational content, UK\Educational content, UK, maths	103	0.8283

HK framework\Private benefits of media viewing\Enhanced skill-set from viewing UK or other media\Educational content, all media\Educational content, UK\Educational content, UK, maths	112	0.9263
HK framework\Private benefits of media viewing\Enhanced skill-set from viewing UK or other media\Educational content, all media\Educational content, UK\Educational content, UK, maths\Educational content, UK, maths, with specific example	103	0.9475
HK framework\Private benefits of media viewing\Enhanced skill-set from viewing UK or other media\Educational content, all media\Educational content, UK\Educational content, UK, maths\Educational content, UK, maths, with specific example	112	0.756
HK framework\Private benefits of media viewing\Enhanced skill-set from viewing UK or other media\Educational content, all media\Educational content, UK\Educational content, UK, motivation to learn	103	0
HK framework\Private benefits of media viewing\Enhanced skill-set from viewing UK or other media\Educational content, all media\Educational content, UK\Educational content, UK, motivation to learn	112	1
HK framework\Private benefits of media viewing\Enhanced skill-set from viewing UK or other media\Educational content, all media\Educational content, UK\Educational content, UK, non-specific	103	0.2334
HK framework\Private benefits of media viewing\Enhanced skill-set from viewing UK or other media\Educational content, all media\Educational content, UK\Educational content, UK, non-specific	112	0.8108
HK framework\Private benefits of media viewing\Enhanced skill-set from viewing UK or other media\Educational content, all media\Educational content, UK\Educational content, UK, parenting	103	1
HK framework\Private benefits of media viewing\Enhanced skill-set from viewing UK or other media\Educational content, all media\Educational content, UK\Educational content, UK, parenting	112	0.9706
HK framework\Private benefits of media viewing\Enhanced skill-set from viewing UK or other media\Educational content, all media\Educational content, UK\Educational content, UK, reading or alphabet	103	1
HK framework\Private benefits of media viewing\Enhanced skill-set from viewing UK or other media\Educational content, all media\Educational content, UK\Educational content, UK, reading or alphabet	112	1
HK framework\Private benefits of media viewing\Enhanced skill-set from viewing UK or other media\Educational content, all media\Educational content, UK\Educational content, UK, science	103	0.8334
HK framework\Private benefits of media viewing\Enhanced skill-set from viewing UK or other media\Educational content, all media\Educational content, UK\Educational content, UK, science	112	1
HK framework\Private benefits of media viewing\Enhanced skill-set from viewing UK or other media\Educational content, all media\Educational content, UK\Educational content, UK, science\Educational content, UK, science, with specific example	103	1
HK framework\Private benefits of media viewing\Enhanced skill-set from viewing UK or other media\Educational content, all media\Educational content, UK\Educational content, UK, science\Educational content, UK, science, with specific example	112	1
HK framework\Private benefits of media viewing\Enhanced skill-set from viewing UK or other media\Educational content, all media\Educational content, UK\Educational content, UK, socio-emotional	103	1
HK framework\Private benefits of media viewing\Enhanced skill-set from viewing UK or other media\Educational content, all media\Educational content, UK\Educational content, UK, socio-emotional	112	0.9706
HK framework\Private benefits of media viewing\Enhanced skill-set from viewing UK or other media\Educational content, all media\Educational content, UK\Educational content, UK, socio-emotional\Educational content, UK, socio-emotional, with specific example	103	1
HK framework\Private benefits of media viewing\Enhanced skill-set from viewing UK or other media\Educational content, all media\Educational content, UK\Educational content, UK, socio-emotional\Educational content, UK, socio-emotional, with specific example	112	0
HK framework\Private benefits of media viewing\Enhanced skill-set from viewing UK or other media\Factors affecting attainment gains from media\Interaction with others, or lack of it\Co-viewing or lack of it\UK or other media co-viewing with caregiver	103	1
HK framework\Private benefits of media viewing\Enhanced skill-set from viewing UK or other media\Factors affecting attainment gains from media\Interaction with others, or lack of it\Co-viewing or lack of it\UK or other media co-viewing with caregiver	112	0.9357
HK framework\Private benefits of media viewing\Enhanced skill-set from viewing UK or other media\Factors affecting attainment gains from media\Interaction with others, or lack of it\Co-viewing or lack of it\UK or other media co-viewing with peers or siblings	103	0.9895

HK framework\Private benefits of media viewing\Enhanced skill-set from viewing UK or other media\Factors affecting attainment gains from media\Interaction with others, or lack of it\Co-viewing or lack of it\UK or other media co-viewing with peers or siblings	112	1
HK framework\Private benefits of media viewing\Enhanced skill-set from viewing UK or other media\Factors affecting attainment gains from media\Interaction with others, or lack of it\Co-viewing or lack of it\UK or other media viewing alone	103	0
HK framework\Private benefits of media viewing\Enhanced skill-set from viewing UK or other media\Factors affecting attainment gains from media\Interaction with others, or lack of it\Co-viewing or lack of it\UK or other media viewing alone	112	0.706
HK framework\Private benefits of media viewing\Enhanced skill-set from viewing UK or other media\Factors affecting attainment gains from media\Interaction with others, or lack of it\Discussion of UK, or lack of it	103	1
HK framework\Private benefits of media viewing\Enhanced skill-set from viewing UK or other media\Factors affecting attainment gains from media\Interaction with others, or lack of it\Discussion of UK, or lack of it	112	0.825
HK framework\Private benefits of media viewing\TV viewing as a means of avoiding other undesirable activities	103	1
HK framework\Private benefits of media viewing\TV viewing as a means of avoiding other undesirable activities	112	1
HK framework\Private costs of media viewing\Costs of viewing UK cartoons and some, non-UK, inappropriate content for children	103	1
HK framework\Private costs of media viewing\Costs of viewing UK cartoons and some, non-UK, inappropriate content for children	112	0.7405
HK framework\Private costs of media viewing\Opportunity cost of viewing	103	0.9047
HK framework\Private costs of media viewing\Opportunity cost of viewing	112	0.924
HK framework\Private costs of media viewing\Psychic cost of learning, and, ease of learning, from cartoons	103	1
HK framework\Private costs of media viewing\Psychic cost of learning, and, ease of learning, from cartoons	112	1
Average for all sources, with all codes that feature in any coder's coding of either document (unweighted)	ave	0.8545
Average for all sources, with all codes that feature in any coder's coding of either document (weighted by Source Size)	ave	0.8563

Note. In the above table, File 0103 refers to inter-rater reliability testing results concerning a child interview and File 0112 refers to results concerning a parent/caregiver interview. Cohen's Kappa values for individual codes are highlighted green where satisfactory and red where not so. Average Kappa values for all codes and files are presented in grey at the bottom of the above table (and are satisfactory).

Appendix 11: Updated code descriptions following inter-rater reliability testing

Type of programme viewed or listened to

Pre-code check description: This information concerns any piece of information suggesting that a type of TV station is viewed (or even just enjoyed) by the child/parent/caregiver. Information that suggests a specific programme/station is viewed (or enjoyed) is coded elsewhere.

Updated description: This information concerns any piece of information suggesting that a *type* of TV programme is viewed, or even just enjoyed, by a child/parent/caregiver. Types of TV programmes could include 'sports', 'cartoons', 'kids' shows' or 'comedy shows'. Information which suggests that only a *specific programme*/*specific station* is viewed (or enjoyed) is coded under other codes (although information that covers a type of programme *and* a specific programme would be coded under this and another code: e.g., "I like cartoons like *Ubongo Kids*").

Viewing or listening frequency, general, not UK specific

Pre-code check description: This code covers information related to media viewing frequency, or lack of it, including information regarding the times/days that media is viewed (in general, as opposed to related to UK specifically).

Updated description: This code covers information related to media viewing or listening frequency, including information on a lack of viewing/listening (such as, for example, a statement concerning how a child no longer watches DVDs). This code covers information regarding the times/days that media (in general, as opposed to related to UK specifically) is or is not viewed or listened to. (Although specific shows are not normally mentioned when providing information on viewing/listening frequency, non-UK shows can include Dora the Explorer, Kipara and Bruce Lee.)

Educational content, UK, non-specific

Pre-code check description: Any reference to the educational content in UK that does not match a particular topic/subject (or relates to an "abstract" topic such as 'games' or 'riddles'); or, any statement suggesting merely that UK is educational in nature. (When it is possible that the interviewee is confused between UK and another programme, coding is applied on the assumption that they are referring to UK.)

Updated description: Any reference to UK educational content that is generic in nature, or does not match a particular topic/subject covered by other codes (thereby including "abstract" topics such as 'games' or 'riddles'). This code also covers any statement suggesting merely that UK is educational in nature (such as general statements, like "Ubongo Kids teaches me", or, "I think Ubongo Kids is

educational"). (When it is possible that the interviewee is confused between UK and another programme, coding is applied on the assumption that they are referring to UK.)

UK or other media viewing alone

Pre-code check description: Coded here is any reference to viewing UK (or, in rare cases, other media) alone (as a child).

Updated description: Coded here is any reference to child viewing of UK (or, in rare cases, other media) alone. This code could be identified where as a result of a caregiver statement (e.g., "my child watches Ubongo Kids alone") or from a child statement (e.g., "I do not watch Ubongo Kids with anyone else").