Journal of Development Economics
Bridging Educational Gaps Watching TV at Home The Impact of an Educational TV Show on Kenyan Children
--Manuscript Draft--

First Author:	Jasmin Baier
Order of Authors:	Jasmin Baier
	Laura Barasa
	Nicolas Luis Bottan
	Anushka Ghosh
	Johannes Haushofer
	Mark Millrine

Abstract:	Educational television has shown promise for bridging educational gaps in classrooms in developing countries. However, there is less evidence on its effectiveness in a home viewing setting when the show is transmitted free over-the-air, an approach of interest to policymakers because of its scalability and low cost. We are conducting a comprehensive evaluation of the effects of watching a new educational television program at home. The new show, Nuzo & Namia, is produced by Ubongo is broadcasted on a free over-the-air television channel in Kenya. Besides instructional content, key objective of the show is to change children's mindsets about reading, gender attitudes and socio-emotional learning. We recruited 4,300 children in 346 public schools. We employ a randomized SMS-based encouragement design, sending bi-weekly reminders encouraging parents to have their children watch episodes of a new educational show at home. We study mechanisms and provide cost-effectiveness calculations for this scalable intervention.
Response to Reviewers:	

Bridging Educational Gaps Watching TV at Home:

The Impact of an Educational TV Show on Kenyan Children[*]

Jasmin Baier[1], Laura Barasa[2], Nicolas Bottan[3], Anushka Ghosh[4], Johannes Haushofer[5], and Mark Millrine[4]

[1]Oxford University, [2]University of Nairobi, [3]Cornell University, [4]Busara Center for Behavioral Economics, [5]National University of Singapore

This Draft: June 3, 2024.

Abstract

Educational television has shown promise for bridging educational gaps in classrooms in developing countries. However, there is less evidence on its effectiveness in a naturalistic home viewing setting when the show is transmitted free over-the-air, an approach of interest to policymakers because of its scalability and low cost. We are conducting a comprehensive evaluation of the effects of watching a new educational television program at home. The new show, Nuzo & Namia, is produced by Ubongo is broadcasted on a free over-the-air television channel in Kenya. Besides instructional content, a key innovation of the show is its objective to change children's mindsets about reading, gender attitudes and socio-emotional learning. We recruited 4,300 children in 346 public schools. We employ a randomized SMS-based encouragement design, sending bi-weekly reminders encouraging parents to have their children watch episodes of a new educational show at home. Outcomes will be collected in an endline survey in April 2024. We will study mechanisms and provide cost-effectiveness calculations for this scalable intervention.

JEL Classification: I21, I25, J16, C93.

Keywords: Educational TV, home watching, mindset, literacy, gender attitudes, socio-emotional learning.

Timeline

September - November 2022	Preparation, pre-pilot, and pilot of Baseline survey instruments
February - March 2023	Sample selection, randomization, field outreach
April-May 2023	Baseline data collection in 346 schools, with around 4,300 students.
June 2023	Program implementation starts. First episode of Nuzo & Namia aired on June 24th, 2023. The first SMS reminder was sent out a few days earlier.
February 2024	*Submit Proposal to Journal of Development Economics Pre-Results Review*
March 2024	Program implementation ends. Last airing of show for the season on March 27th, 2024. SMS reminders stop a few days prior.
April, 2024	Endline data collection start in all 346 schools with same 4,300 students.
May - June 2024	Endline data collection completed and start data processing
June - September 2024	Data analysis and research output preparation
Fall 2024	Dissemination of results and preparation of publication of research results.

1 Introduction

Although access to primary education has increased drastically in recent decades, recent attention has focused on "the learning crisis" where, in 2019, 57% of children in low- and middle-income countries (86% in Sub-Saharan Africa) did not achieve minimal reading proficiency by age 10 (WorldBank et al., 2022). This has only been exacerbated by the Coronavirus pandemic (Tadesse and Muluye, 2020). Addressing these shortfalls in the short and medium term is difficult in light of the dearth of key resources, particularly quality teachers (UNESCO, 2022). In this context, educational technology (EdTech) could play an important role to bridge the educational gap. Interventions such as computer-aided learning have been shown to be highly effective at improving learning, but at a large cost - particularly in contexts that require investments in additional infrastructure (e.g., Muralidharan et al., 2019; Araya et al., 2019).

Educational television is a promising alternative: It is low cost, can potentially reach millions, does not require guidance, and growing evidence suggests that it can improve children's learning (Mares and Pan, 2013; Cherewick et al., 2021). For example, observational studies have found that TV shows such as Sesame Street improved learning in the United States and other low-income countries (Kearney and Levine, 2019; Mares and Pan, 2013). On the other hand, evidence from small randomized controlled trials of teacher-guided educational TV viewing at school has been found to increase learning in many low-income countries (i.e., Borzekowski, 2018). Although TV access is not yet universal in many low-income countries, access to TV and other electronic media (i.e., smartphones, tablets) has been growing at a steady rate (Center, 2018). Therefore, a natural scale-up with the potential of reaching millions would be to move watching educational shows from classrooms to living rooms. However, there is less causal evidence on the effectiveness of watching educational TV shows in a naturalistic (i.e., not guided) home setting when the show is transmitted over the air.

In this paper, we present evidence from a large randomized controlled trial in Kenya, studying the effects of watching a new children's educational TV show at home. One novel aspect of the show is that, besides being instructional, it is focused on changing children's mindsets about reading, gender attitudes and socio-emotional learning. We recruited 4,300 children from 346 public schools. Students in randomly selected treatment schools receive encouragement to watch the TV show. We send parents biweekly SMS reminders about the time and channel the program airs. Students in control schools do not receive any encouragement or reminders about the show. The TV show, a cartoon called Nuzo & Namia, was created in 2023 by Ubongo.[1] The show targets 6-9 year-olds's and its main educational

---

[1]Ubongo is a non-profit providing a free over-the-air educational television channel with presence in 41 African countries, producing educational content in 12 languages.

objective is improving literacy. Additionally, it aims to change gender norms and encourage different forms of socio-emotional learning (i.e., confidence and curiosity).

The identification of the causal effect of our encouragement treatment relies on random assignment at the school level. Our primary estimates capture intention-to-treat effects. We measure outcomes a year after the baseline data was collected, and after almost 9 months of potential exposure to the TV show.[2] The endline data will be obtained through a household survey. In the survey, besides obtaining measures of our outcomes, we collect self-reported measures of watching the show to estimate the local average treatment effects for treatment compliers.

We study three categories of main outcomes. To measure literacy, we employ the Early Grade Reading Assessment Tool (USAID, 2016) and construct a measure for reading fluency and comprehension. We study three dimensions of gender attitudes: gender stereotype knowledge and flexibility, gender roles, and in- and out-group attitudes and activities. One challenge measuring this outcome is the lack of existing validated measures to capture gender attitudes among young children. One contribution of our study is the development and validation of a novel instrument to capture these preferences. As a measure of socio-emotional learning, we construct a score that combines measures of confidence and curiosity. In section 3.2.2 we describe the characteristics of the show that relates to the different outcomes considered in the study.

We explore mechanisms. Following La Ferrara (2016), we consider three main mechanisms through which educational TV can drive changes in outcomes. First, the show can provide viewers with new *information/skills* that can lead to better literacy either directly (e.g., the show teaches key strategies in early reading comprehension) or indirectly (e.g., learning strategies makes reading easier or more enjoyable, thus increasing practice).[3] Second, the show can mold viewers *preferences* such as positive affect to reading. This could encourage children to read more, thus improving literacy. There are additional behavioral mechanisms associated to changing preferences, such as demanding more books, reading with parents, among others. Third, outcomes can be affected through changes in *time use* where encouraging children to watch the show can crowd-in or out reading time or other behavior (e.g., watching more educational or non-educational TV).

We will also consider indirect mechanisms that could result from our encouragement treatment. For instance, friends can watch the show and be exposed to the aforementioned

---

[2]Nuzo & Namia started airing in June 24, 2023 and is scheduled to finish on March 27, 2024.

[3]While pure information can be important to affect an outcome like gender attitudes (e.g., learning that women can do X), literacy does not have a pure information component. Rather, the show conveys skills that when adopted, can improve reading. This is why we expand on the categorization in La Ferrara (2016) to combine information and skill that is more appropriate for this setting.

mechanisms. In turn, social influence could increase both watching the show and reading behavior. Similarly, teachers could incorporate aspects of the show in their teaching (e.g., examples or better explanations), improving teaching effectiveness. Additionally, because our encouragement treatment is addressed to parents or caregivers, this could change their perception about the importance of reading, which could increase their involvement in their child's education.

This study also explores the importance of different characteristics of the show. These can affect outcomes in two ways. The first is through engagement watching the show (i.e., treatment intensity). The second is related to mediators driving changes in children's beliefs and attitudes. Following social learning theory (Bandura and Walters, 1977), we explore the role played by features like how relatable characters are to children that are important in the adoption of norms and behaviors.

We also test the role that mediators or barriers play when examining mechanisms. For instance, a qualitative study conducted as part of this evaluation found that the presence of a caregiver while watching the show was important to maintain the child's engagement with the show. We elicit measures of caregiver presence during the show to test whether the role this mediator might play. Additionally, if the show does encourage more reading, the presence of reading materials in the household would be a potential barrier to actually changing reading behavior. The baseline survey collected measures of the presence of different types of reading materials at home for heterogeneity analysis.

This paper contributes to a growing literature studying the effect of educational TV shows on children's learning. There are a growing number of experimental studies showing learning gains when children watch educational TV in a controlled setting such as at school under teacher guidance. For example, Borzekowski (2018); Borzekowski et al. (2019) finds positive effects of watching Akili & Me on drawing skills, shape knowledge, number recognition, counting, and English skills in Rwanda and Tanzania. These interventions randomized children to watching shows in a controlled environment (i.e., at school or during special sessions and under supervision). However, existing evidence studying the effect of watching education TV at home is mostly correlational (e.g., Mares and Pan, 2013). In the context of developed countries, (Kearney and Levine, 2019) exploits quasi-experimental variation in broadcast reception showing that Sesame Street improved children's educational attainment and labor market outcomes. Our paper will contribute to this literature by providing one of the first causal estimates of the effects of naturalistic viewing of educational TV through a free over-the-air channel on children's learning outcomes from a large-field experiment.

This paper also will inform a literature documenting how TV shows and representation in the media can change social norms and behaviors. For example, telenovelas in Brazil

have been found to lower fertility and increase divorce rates (Ferrara et al., 2012; Chong and Ferrara, 2009), in the United States the MTV show *16 and Pregnant* lowered teen births (Kearney and Levine, 2015), and *MTV Shuga* changed knowledge and attitudes towards HIV in Nigeria (Banerjee et al., 2019). Furthermore, female role models in a movie led to large learning gains among females students in Uganda (Riley, 2024). This study will contribute to this literature by providing novel experimental evidence on the effectiveness of TV shows to shape children's social norms – particularly, gender attitudes.

Educational technology interventions have received substantial policy and academic attention lately. Recent work has found that interventions such as computer-assisted instruction are highly effective in improving learning (e.g., Muralidharan et al., 2019), however, they may face substantial scaling challenges and large implementation costs. This paper contributes to the policy debate by documenting the effectiveness of an easily scalable and low-cost intervention that relies on existing free-to-air TV programming and SMS messaging to encourage viewing. Additionally, our analysis of mechanisms can more braodly inform the design of educational TV interventions (La Ferrara, 2016).

2 Background

Kenya is one of the largest economies in East Africa, yet it has long struggled to improve childhood education and literacy rates. For example, a Uwezo Learning Assessment Report in 2021 found that "only 2 in 5 grade 4 learners are at least meeting expectations in reading a grade 3 appropriate English text" (Usawa, 2021). The adult literacy rate in Kenya is 82.9%, meaning that 17.1% of the population 15 or above cannot read or write. In particular, the adult literacy rate for women is almost 6 percent lower than the literacy rate for men (UNESCO, 2023).

Gender roles are particularly problematic in the Kenyan education system. Female students are sometimes discouraged from learning due to stereotypes, either internalized or enforced by others. For instance, research interviews of around 166 teacher trainees and 7 tutors from 7 Diploma teacher training colleges found that approximately 15 percent of trainees believed that "female students were better placed to write notes during group discussions, while male students were better placed to handle experiments or make presentations" (Wawire et al., 2023). In this context, Kenya can benefit from effective and low-cost educational programming on television. 81% of the population is reported to have access to TV (Murunga, 2021). This means that educational television has the potential to reach a large share of the population in an equitable manner at relatively low costs.

3   Research Design

3.1   Sample Selection

We are conducting a school-level cluster randomized controlled trial. Our sampling strategy consisted of 3 stages: (1) identify eligible counties and sub-counties, obtain a random draw;[4] (2) identify eligible schools within randomly selected sub-counties and obtain a random draw; (3) obtained final sample of children within eligible schools.

3.1.1   Selecting counties and sub-counties

We use data on public primary school enrollment from the Kenya Basic Education Statistical Booklet 2019 (Kenya Ministry of Education, 2019) and data from the Kenyan Census to obtain sub-county school-level estimates of the number of eligible children. We define an eligible child as one who is enrolled in grades 1-3 and has a TV at home. We then keep sub-counties where we estimated there were at least 40 eligible children enrolled in a public primary school.[5] We further dropped sub-counties that were the only eligible unit within their county to reduce enumerator travel costs. We also excluded Nairobi because as the capital city there are many schools in close geographic proximity and therefore spillovers would be more likely.

We made the final selection of sub-counties by first specifying all possible combinations of 6 counties from the pool of eligible counties identified above, such that each county is in a different region.[6] Each possible combination yields a list of at least 3,000 government primary schools.[7] Finally, we randomly selected one of the combinations of 6 counties yielding a list of over 3,000 government primary schools within their eligible sub-counties.

3.1.2   Selecting schools within sub-counties

We followed a procedure like the one used to select sub-counties to select schools. We first calculated the expected number of eligible boys and girls (i.e., in grades 1-3 and, potentially, with a TV at home). We then defined an eligible school as one that we had estimated at least 12 eligible girls and 12 eligible boys, so that there were at least 24 estimated eligible

---

[4]Kenya is divided into 47 counties and 314 sub-counties

[5]Although we aimed at surveying 15 eligible students in each school, we targeted schools with more than 40 eligible children because: (1) these are based off of rough back-of-the-envelope estimates; (2) participation was voluntary and parents might opt their child out of the study.

[6]There are 8 regions in Kenya.

[7]We targeted combinations of counties yielding at least 3,000 schools because we were hoping to find 500 schools for the study that were at least 4km away from each other. However, schools tend to be quite close together, so we used a conservative factor of 6 when defining the size.

children. We use a threshold of 12 to ensure there was a large enough sample of children with access to TV.

From this group of eligible schools, we randomly sampled schools, such that each additional randomly sampled school was at least 4 kilometers away from *any* other schools already included in our sample.

3.1.3 Sampling children within these schools

Once we had our randomly selected schools, we then obtained samples of children within these schools. Field officers were sent to the locations where our sampled schools were located. With the help of the village chief, the village elders, and the teachers, we were able to obtain household contact information.

We then conducted a phone screening and recruitment exercise to ensure that we had a good number of eligible households that met our study criteria to be included in the study and these were interviewed during baseline. The criteria included children's age and TV access. Due to a fairly low number of households without access to a functioning TV, we included all eligible households in our sample. We also included children in the same household in our sample. Table 1 shows the distribution of the final sample after randomization.

Table 1: Baseline Sample Size

Variable	Treatment	Control	Overall
Number of clusters	173	173	346
Average cluster size	12.71	12.57	12.64
Sample size	2,198	2,175	4,373

3.2 Intervention

To study the causal effect of watching an educational TV show we employ a randomized encouragement design. The show is transmitted free over-the-air TV; therefore, anyone with a TV can potentially watch the show. We take advantage of the fact that this is a brand new show. By design, no advertising for the show was allowed in Kenya and compliance was monitored. Therefore, the only information available about the show was provided through the parent information sessions. This information, along with later SMS reminders, were given to parents in treated schools to encourage their children to watch the show at home.

In this section, we provide details of the encouragement campaign and detail key elements about Nuzo & Namia, the TV show children are encouraged to watch.

3.2.1 Encouragement campaign

We provided the treatment parents with various resources to encourage and remind them about having their children watch the TV show. Before the show aired, treatment parents were invited to a group information session held in the local school or community center. Around 80% of parents attended these group sessions. The remaining parents that were unable to attend received a personal information session. One parent, the mother, typically attended. During the information session parents were presented an overview of the show and of how the show is positioned to support their children's learning. To further increase buy-in amongst the parents, a Ubongo fact sheet, testimonials from other parents who have engaged with past Ubongo shows, and an example of a scene from Ubongo was shown.

The information session also gave parents resources to aide their child watching the show. Parents were instructed to choose a day of the week and a day of the weekend they intended to have their child watch the show. Because there are many competing activities that take place at home when the show airs, parents were introduced to a planning exercise where they identified pre-existing commitments their children could have during the time the show airs. They were then guided on how to reallocate the other activity to a different time or reassign the work to other people in the household. Parents then filled in the planner, highlighting what the children would be doing on the chosen days when the show is airing and plugged in watching the show at the same time. Parents were instructed to hang the planner near the TV or in a visible place in the house to have as a reminder to both the parents and children, and help them build a routine to watch the TV show.

Throughout the show's run, parents were also sent reminder SMS messages twice a week, 30 minutes before the show aired on the days they had chosen to watch. After the show, parents also received an SMS question about the episode that had just aired, designed to elicit responses from the child. Responses to the SMS questions acted as an early warning sign to identify parents that were following through with encouraging their children to watch.

3.2.2 Intervention - Nuzo & Namia

The Nuzo & Namia show was developed by Ubongo and targets 6-9 year-old children with the objective of improving their literacy (reading and comprehension), changing gender attitudes, and socio-emotional learning. The plot revolves around 7-year old twins, Nuzo & Namia, who move into their grandmother's home with their family after the grandmother passes away to

help the twins cope with the loss. In each episode, they take a book from a magic bookshelf and are taken on an exciting adventure to different African countries. In the episodes they learn about different places and cultures, and have to overcome different challenges. The series consists of thirteen 22-minute episodes, was produced in English, Kiswahili and Hausa, and was aired in 5 countries. In Kenya, the show airs 2 times a week on the Akili Kids channel.

The show incorporates several features to reach its objectives:

1. Literacy (Reading and Comprehension):

   (a) Changing children's mindset: An important innovation in this show is the goal of changing children's mindsets about reading, with the objective of encouraging children to read more and thus become better readers. This is conveyed through the creation of socially desirable characters who are relatable role models to children. Through character modeling behaviors such as reading books, they gain agency, confidence, and find motivation to improve their own reading skills. Because viewers perceive characters to be like them, they form parasocial relationships with the characters, leading to stronger engagement, positive emotions and affect towards reading.

   (b) Skill development and practice: The show presents learning strategies in an illustrative and engaging way that captures and holds children's attention. The protagonists face challenges each episode that require reading and comprehension to solve. In the process of solving these challenges, the protagonists model key strategies in early reading comprehension, such as predicting, questioning, clarifying and analysing information.[8] During episodes the protagonists occasionally break the fourth wall, asking the viewer questions and leave time for responses. Thus, the show provides an opportunity for children to engage in informal learning along with the characters.

2. Gender attitudes: The show shapes gender attitudes through its portrayal of characters and storylines. When portraying parental and authority figures, it presents a blend of traditional and progressive gender roles. The mother's character aligns with conventional norms through her emphasis on household organization, a typical trait of female caregivers. In contrast to this, the father character steps out of traditional male roles by engaging in cooking and nurturing activities, such as preparing meals

---

[8]See https://www.youtube.com/watch?v=WiV638jzI84 for a short excerpt.

and encouraging his children in unique ways.[9]

Additionally, the combination of the show's narrative choices, friendship dynamics, character appearances, and problem-solving roles contribute to shaping the viewers' gender perceptions. The storyline and plot of the show actively challenge gender stereotypes in several instances. Namia, in particular, takes on leadership roles when addressing challenges. An example of this is when she leads the interview with characters from Ethiopia while the children investigate what comes first, the chicken or the egg. On the other hand, Nuzo is portrayed as nurturing, particularly after learning about Namia's autism diagnosis; he consistently strives to ensure Namia's well-being. This depiction goes against traditional gender expectations, showcasing diverse and non-stereotypical character attributes.

3. Socio-emotional Learning: The show focuses on two primary dimensions of socio-emotional learning: confidence and curiosity.

   (a) Confidence: The show fosters confidence in viewers through various themes. Characters are shown embracing risks and stepping out of their comfort zones. The characters travel across African countries, learning about different cultures, and face new challenges. Positive self-talk is a key element, with characters using affirming phrases like "believe in yourself" during moments of doubt, reinforcing self-confidence.[10] The series shows both challenges and setbacks, teaching children that these are normal and can be overcome with persistence and problem-solving, thereby boosting confidence. Additionally, the show emphasizes embracing individuality and valuing differences. This is illustrated through Namia's autism and the acceptance and appreciation shown by other characters. This highlights the importance of accepting and celebrating diversity, emphasizing that confidence is crucial, even when people are not the same.

   (b) Curiosity: The show fosters curiosity primarily through the adventures of Nuzo and Namia. These characters are portrayed as innately curious, constantly asking questions and eagerly exploring new environments. Their interactions, such as inquiring about various aspects of the countries they visit or engaging in different cultural dances, highlight their desire for knowledge and discovery. The show also models curiosity through characters like Bubelang (a magical creature that chaperones the children during their adventures), who prompts imaginative thinking

---

[9]Another example is shown in following excerpt, where the chidren discover old photos of their grandmother, showing her graduating and winning a swimming competition: https://www.youtube.com/watch?v=lCsofc7qhUg.

[10]See https://www.youtube.com/watch?v=yzCwToLRJb8 for an example.

with open-ended questions. The show's approach to nurturing curiosity blends factual learning with imaginative exploration and intellectual challenges, with the goal of sparking a sense of wonder in children.

Each episode was originally scheduled to be broadcast twice a week for a period of two weeks (i.e., four broadcasts per episode over a two-week period). However, significant production delays changed the broadcast schedule, repeating episodes a different number of times. Despite these delays, the revised schedule anticipates airing new episodes every two weeks until March 2024 as planned initially.

3.3 Theory of Change

With these show objectives and features in mind, we summarize the theory of change behind the encouragement treatment for learning in Figure 1 below. The Figure presents different potential mechanisms and barriers children could face that could affect a certain mechanisms path. Additionally, it indicates what we will be able to measure directly or indirectly through our endline survey. The first stage of our study would be that receiving encouragement treatment increases the likelihood and intensity of watching the TV show. One barrier to the first stage is not having easy access to a television. As described above, not all children in our sample have access to a television.

A second barrier (and potential mechanism) is the viewing environment. The presence of others while watching, and their human capital, can be important. From a qualitative study conducted as part of the evaluation, we found that caregiver presence and nudges were important to maintain a child's engagement with the show. Additionally, watching the show with other children may help reinforce the content. English fluency in the household could aid in connecting with characters and translating literacy lessons into reading with the child.

A third barrier could be a household's socio-economic status. For example, compliance could be correlated with poverty because of power outages, parents having competing responsibilities, children's nutrition, etc. We will therefore test heterogeneity by socio-economic status. Furthermore, we will test what specific aspects of poverty matter for take up and learning.

We hypothesize that watching the TV show can affect literacy, through channels related to features of the show and also through potential spillover (or unintended) effects. Ultimately, improving literacy is mainly a direct result of reading more (Merga and Roni, 2017). We consider the three main direct mechanisms proposed by La Ferrara (2016) (i.e., preferences, information, and time use), as well as other potential indirect mechanisms.

Preferences: The first channel relates to changing children's preferences for reading,

through increasing their engagement and interest. An important innovation in the show is the goal of changing children's mindset through presenting socially meaningful characters and modeling behaviors. The primary motivation is to change a child's mindset with regard to reading. Therefore, the changes in preferences would be related to the viewer perceiving characters as being socially meaningful, them adopting character's modeling behaviors and watching the show generating positive emotions. As a result, children would change their preferences and be more motivated to read. This could also affect other behaviors related to reading: children can increase their demand for books or reading materials, or increase their demand for having their parents read to them. Both of these factors could lead to a child reading more. It is important to note that there can be one important barrier to having changed preferences over reading result in improved literacy scores: the availability of books and reading materials at home (i.e., if there are zero reading materials at home, the child will not be able to read more regardless of how motivated they are). Although literacy is the primary objective, the show also attempts to influence children's mindsets with regards to gender attitudes and socio-emotional learning. These changes are also driven by the relate-ability of characters and modeling behaviors. It is important to note that changes in gender attitudes and socio-emotional skills could also affect literacy (e.g., Alan et al., 2018, 2019) or vice-versa (e.g., Yu et al., 2023; Deole and Zeydanli, 2021).
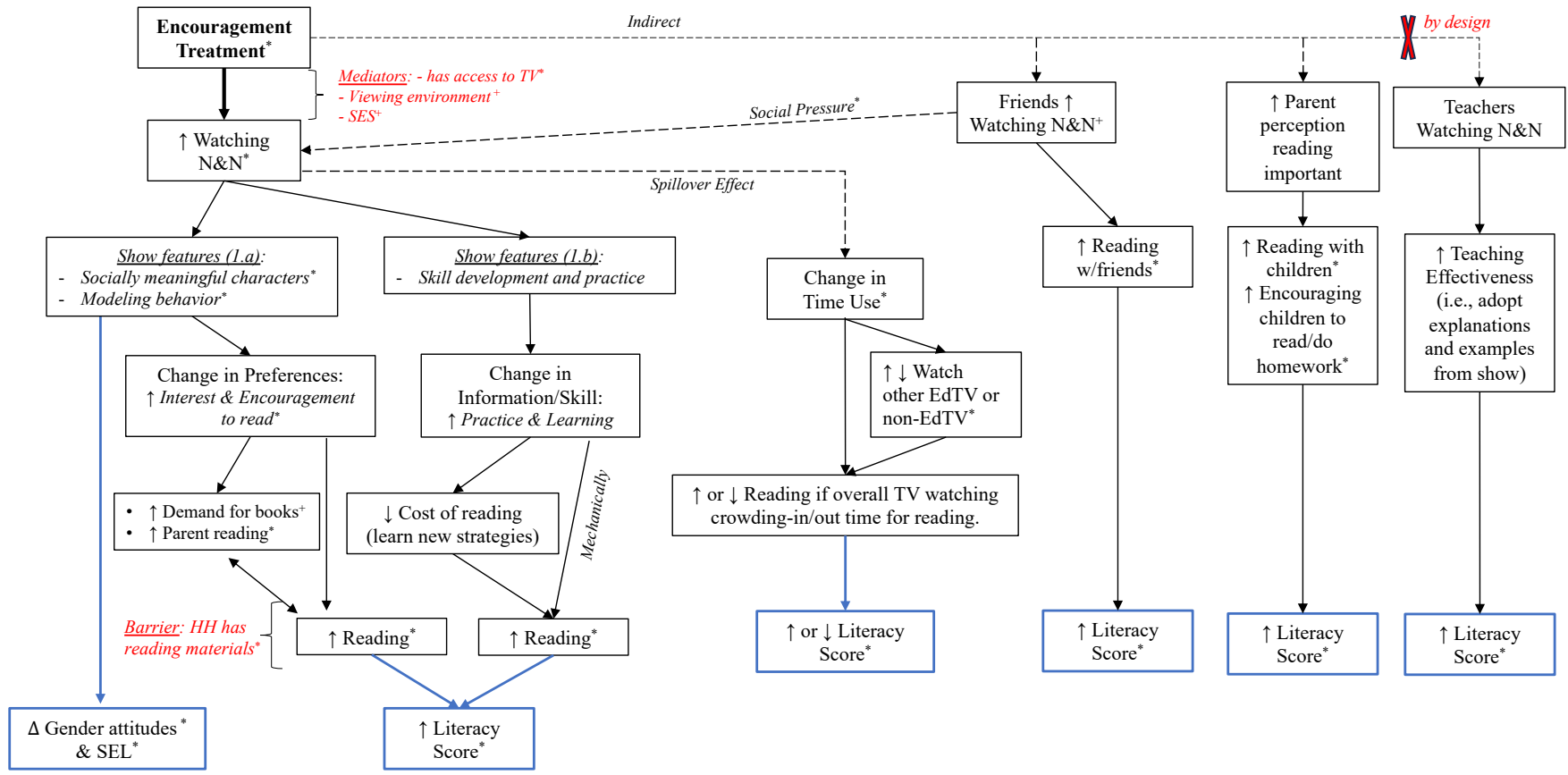
Information/Skills: The second channel is related to the information/skill content provided by the television show to teach key reading and comprehension strategies (show characteristic 1.(a) above). This can affect the literacy score in two ways. The first is by directly applying the skills conveyed on the show (i.e., newly learned reading and comprehension strategies). The second is through exposing the child to more reading practice. Part of this exposure is mechanical through interaction with characters while watching the show (i.e., reading with the characters). Another channel can be through the new skills, which reduce the cost (or difficulty) of reading, thus encouraging more reading off-screen.

Time Use: There is a third channel that is not a direct show characteristic, but could be a potential spillover effect of watching the TV show: changing children's time use through TV watching habits. Watching the show itself can potentially crowd-in or crowd-out other activities as well as time devoted to watching TV. For instance, children might substitute other non-educational (or educational) TV shows with Nuzo & Namia. Alternatively, children could increase time devoted to watching TV shows overall and crowd-out reading time. How changes in time use might affect literacy is uncertain ex-ante. If changes in behavior crowd-out reading, then this could attenuate the benefits of watching Nuzo & Namia. Alternatively, it could boost the effects of watching Nuzo & Namia if it crowds-in reading or if it results in the consumption of additional educational TV shows.

The encouragement treatment can also affect literacy through indirect mechanisms. First, because we employ a school-cluster treatment design, friends within the school could watch Nuzo & Namia. This could generate social pressure to watch the show, further increasing compliance. Additionally, if friends also watch the show and read more, this could lead to opportunities to joint reading, thus practicing more and increasing literacy. Second, the encouragement treatment could affect literacy through a mechanism unrelated to watching Nuzo & Namia. The information session and reminders could signal the importance of reading and comprehension to parents. As a result, parents could make larger investments in educational inputs (e.g., buy books, read with children, encourage reading and doing school homework) which then could lead to higher literacy.

By design, we are minimizing a potential channel that operates in other similar types of educational TV interventions conducted at school. Teachers typically play a central role in these interventions, facilitating viewing and discussions. Teachers could also benefit from watching the show by adopting examples and explanations from the show in their teaching. This could increase their effectiveness in the classroom as research has found that providing teachers with better teaching plans and materials improved learning (International Institute for Educational Planning (IIEP) - UNESCO, 2024). However, our study takes several measures to shut down this channel. First, the recruitment and encouragement design operates exclusively outside of the school. Teachers were not involved at any step of the outreach or informational sessions. Second, also by design, the show is not widely advertised or communicated.

Figure 1: Theory of Change

Encouragement Treatment[*]

Indirect

by design

Mediators: - has access to TV[*]
- Viewing environment[+]
- SES[+]

↑ Watching N&N[*]

Social Pressure[*]

Friends ↑ Watching N&N[+]

↑ Parent perception reading important

Teachers Watching N&N

Spillover Effect

Show features (1.a):
- *Socially meaningful characters*[*]
- *Modeling behavior*[*]

Show features (1.b):
- *Skill development and practice*

Change in Time Use[*]

↑ Reading w/friends[*]

↑ Reading with children[*]
↑ Encouraging children to read/do homework[*]

↑ Teaching Effectiveness (i.e., adopt explanations and examples from show)

Change in Preferences:
↑ *Interest & Encouragement to read*[*]

Change in Information/Skill:
↑ *Practice & Learning*

↑ ↓ Watch other EdTV or non-EdTV[*]

• ↑ Demand for books[+]
• ↑ Parent reading[*]

↓ Cost of reading (learn new strategies)

Mechanically

↑ or ↓ Reading if overall TV watching crowding-in/out time for reading.

Barrier: HH has reading materials[*]

↑ Reading[*]

↑ Reading[*]

↑ or ↓ Literacy Score[*]

↑ Literacy Score[*]

↑ Literacy Score[*]

↑ Literacy Score[*]

Δ Gender attitudes[*] & SEL[*]

↑ Literacy Score[*]

[*] Measured directly
[+] Measured indirectly

3.4 Outcome Variables

We will measure our outcomes through an endline survey following-up with the approximately 4,300 children from our original baseline in April 2024.

3.4.1 Primary and Secondary Outcomes

We have three distinct families of primary outcomes: literacy, gender attitudes, and socio-emotional learning. Whenever applicable, we denote secondary outcomes within these categories below. We study the effects in three steps, to best account for multiple hypothesis testing. First, we construct three indices, one for each family of outcomes (literacy, gender attitudes, SEL). We will follow Anderson (2008) to construct the indices using the `swindex` command in Stata (Schwab et al., 2020). Each index has a mean of zero and standard deviation of one, standardized relative to the control group. Second, we will study effects on the main outcomes (described below), which are components that capture different aspects of interest within families. For instance, reading fluency is a distinct skill from comprehension. When estimating these outcomes we will conservatively calculate q-values across the families of outcomes (8 outcomes) following Benjamini et al. (2006). Finally, when studying secondary outcomes, we will calculate q-values within each of the three families of outcomes to account for multiple hypothesis testing.

1. Literacy: Our literacy module was designed following the Early Grade Reading Assessment Tool (USAID, 2016). The instrument includes sub-sections for letter identification, non-word reading, oral fluency, reading comprehension and listening comprehension. For the each of the first three sub-sections, we calculate the correct words per minute.[11] The two comprehension questions are each scored out of five points. These five components are combined into a single index.

   We construct two primary outcomes to measure literacy that capture distinct skills:

   - Reading index: Aggregates scores into an index following Anderson (2008) for letter identification, non-word reading, and oral fluency.

     – *Secondary Outcomes:* Letter identification, non-word reading, oral fluency.

   - Comprehension score: Aggregates (averages) scores into single measure for reading and listening comprehension.

---

[11] As a robustness check, we will instead use percent correct and percent zero score. Constructing the index using correct words per minute or percent correct likely does not matter as the correlation is over 0.97. These measure capture different aspects of the distribution, thus can be informative of where changes are happening.

– *Secondary Outcomes:* We will examine the treatment effect on each of sub-category (reading and listening comprehension).

2. Gender Attitudes: One contribution of this study is the development and validation of measures to capture gender attitudes among children. We combine the seven measures into a single index for gender attitudes, where a larger score indicates more rigid/traditional attitudes. As main outcomes, we examine three measures that capture different dimensions of gender attitudes.

  ▪ Gender stereotype knowledge and flexibility score. Each child is asked whether they would typically see a boy or a girl performing various actions and exhibiting various characteristics. Following a mapping of which stereotypes are considered "male" or "female" in the literature (see Appendix Table A1), responses are coded with "1" if they follow the stereotype, and "0" otherwise. These items are then aggregated into one score denoting the number of gender stereotypes that the child subscribes to. A higher score means that the child has strong gender stereotype knowledge and low stereotype flexibility.

  ▪ Gender roles: An aggregate score measuring belief in traditional gender roles. The score aggregates individual questions about different behaviors and traits typically associated with either "male" or "female." Each child is asked both how likely they are to perform each behavior and how often they actually perform the behavior. Based on each child's gender and their responses, a final score is constructed denoting the strength of their gender role beliefs. A higher score represents stronger beliefs in traditional gender roles.

    – *Secondary Outcome:* We will explore the score separately for female and male gender roles

  ▪ In- and Out-group attitudes: An aggregate score measuring attitudes toward others with the same and different genders. The score aggregates individual questions about activities with both the in- and out-group. Each child is asked both how willing they are o perform each behavior and how often they actually perform the behavior. Based on each child's gender and their responses, a final score is constructed denoting the strength of in-group and out-group attitudes. A higher score represents strong preference for in-groups while a lower score represents flexibility between in- and out-groups.

    – *Secondary Outcome:* We will explore the score for in-group attitudes and activity for females and males separately.

17

3. Socio-Emotional Learning: We use measures of confidence and curiosity as a proxy for *self-efficacy* using the RTI confidence and curiosity scale, which has been validated in both our East African context and our age group (Jukes et al., 2021). These two scores are combined into a single index for socio-emotional learning (SEL).

   ▪ Confidence and Curiosity: Each score is the simple sum of all items, where relevant items are reverse coded so that a higher final score reflects higher confidence and curiosity.

We thus hypothesize that our randomly assigned encouragement treatment will:

   H1) Increase student literacy scores.

   H2) More progressive (i.e., less rigid/traditional) gender attitudes.

   H3) Improve SEL score.

Appendix tables A1 and A2 present, for each hypothesis, the relevant primary and secondary outcome variable along with details about the source, definitions and variable construction. We discuss additional hypotheses related to mechanisms in the mechanisms section.

3.5 Identification Strategy

Schools were randomly assigned to a pure control or treatment condition where they received encouragement to watch the show (see details below).[12] For this randomization we employ a "multivariate stratified quadruple matching" technique. Traditionally, pair-wise matching was believed to be the most ideal approach to maximize balance and power in field experiments following Bruhn and McKenzie (2009). However, more recent work suggests a more conservative approach of using quadruplets or larger groups (McKenzie, 2022). We follow this approach by employing quadruplets given our clustered design and moderate number of units within each cluster.

To adopt this approach, we find groups of four clusters, based on a list of baseline characteristics, and randomly assign two into the treatment group, and two into the control group. These school-level strata characteristics include:

1. Proportion of female students in the school

---

[12]Having a pure control group ensures that the project's findings are policy-relevant. In reality, children may decide against watching our show for a multitude of reasons. Encouraging children in control to watch a placebo show would artificially change their time-use. It is also unclear what would constitute a suitable placebo show. We therefore opted for a pure control group.

2. Mean asset score

3. Mean grade within school ranging from 1 to 3

4. Caregiver education level

5. Mean reading fluency score at baseline

6. Cluster size or number of sampled students

7. Self-reported measure of how much TV children watch on average

Quadruplets were matched based on clusters with the lowest Mahalanobis distance calculated using these characteristics. Implications on attrition and clustering are discussed below.

3.5.1 Sample and Power Calculations

Table 2 (and Appendix Table A3 for secondary outcomes) shows the minimum detectable effect size at 80% power assuming perfect compliance. The minimum effect size was calculated using the number of clusters, average cluster size, and intra-cluster correlation (ICC). All these measures were generated using the baseline data and assume perfect compliance. Overall, our study is fairly well powered, where the range of minimum detectable effects for our outcomes ranges from 0.09 standard deviations (e.g., for gender role scores) to 0.2 standard deviations (e.g., for in-group attitude and activity score gender). In our main Literacy Index, we are powered to detect effect sizes between 0.168 standard deviations.

Compliance is likely to be imperfect. However, we do not have good estimates from similar designs on what compliance might look like in this setting. The closest estimates we found are from a meta-analysis on the effects of Sesame Street in different developing countries (Mares and Pan, 2013). It is important to note that none of the studies included have similar SMS encouragement designs. Observational studies find viewing rates ranging from 31% in Kosovo to 77% in Bangladesh. Thus, using these bounds would give us a range for the MDE between 0.54 (=0.168/0.31) to 0.22 (=0.168/0.77) standard deviations for the Literacy Index. These effect sizes seem reasonable when contrasted to estimates on cognitive outcomes ranging from 0.165 to 0.403 standard deviations using Mares and Pan (2013) as a benchmark. These estimates may also be conservative. The MDE for most other outcomes are smaller, and our preferred regression specification controls for baseline outcomes further improving precision.

19

Table 2: Minimum detectable effect size at 80% power

Outcome	MDE	Mean	SD	ICC
Literacy Index	0.168	0.009	0.985	0.249
Reading index	0.159	0.014	0.995	0.214
Comprehension (%)	0.170	49.053	29.815	0.259
Gender Attitude Index	0.144	-0.007	0.997	0.161
Gender stereotype (mean)	0.133	0.598	0.178	0.126
Gender role score (mean)	0.147	1.427	4.557	0.173
In-group attitude score (mean)	0.207	2.027	3.577	0.424
In-group acitivity score (mean)	0.204	2.325	4.338	0.409
SEL Index	0.120	0.019	0.989	0.085
Confidence score	0.116	9.559	2.191	0.074
Curiosity score	0.115	9.510	2.174	0.072

Notes: Minimum effect size in standardized units. Intra-cluster correlation calculated at the school level. To adjust for imperfect compliance, divide MDE by net compliance rate.

4 Data

4.1 Pilot Data Collected

An initial pilot was conducted to validate the survey tools. The pilot was designed to be rolled out in two phases:

1. Pre-pilot (September 19th and October 5th) in 2 schools in Nairobi

2. Pilot (October 6th- November 11th) in 4 schools across Kajiado and Muranga

The sample counties were selected so that they were representative of the final study sample in asset ownership, where households would have at least a TV in their homes.[13]

Four schools at least 5 km apart from one another were then manually chosen as part of the pilot sample. We then randomly sampled 10 students per grade in each school for a total pilot sample of 80 students.

---

[13] We picked sub-counties that were close to the sampled sub-counties but missed the selection criteria for the main study by a margin of less than 1%. More specifically, two sub-counties within the study county Kajiado (Mashuuru and Kajiado central) have a calculated percentage of TV ownership of 39.9%, which is just below our threshold of 40%. These sub-counties therefore are very similar to our study sub-counties, but are not in our sample for the final study. To get additional variation, we picked some schools in Murang'a county, which is not a study county but one in which all sub-counties would theoretically be eligible. We chose pilot schools in Gatanga, which directly neighbors sample study sub-counties in Kiambu county and as such are likely to be very similar to future study schools.

4.2 Baseline and Encouragement Data Collected

The initial baseline sample targetted 5,250 children from 350 schools in six counties: Kajiado, Kakamega, Kiambu, Kilifi, Kisumu and Meru. Although we initially planned to conduct surveys in schools, we were unable to obtain approval from the Ministry of Education in time. Thus, we used household surveys instead and redesigned our sampling strategy in a short period of time. Subsequently, a sample of eligible parents was obtained via phone recruitment methods through village chiefs with the support of the County Commissioner's Office.

As a result, our final baseline sample consists of 4,373 children in 346 schools, falling short of our target number of children by 17% and target number of schools by 1%. We do not expect that this shortfall will significantly affect the study design. The shortfall in sample size did not result in a significant loss of power as the intra-cluster correlations (ICC) for our outcomes are relatively low, ranging from 0.07 for some of the social and emotional learning outcomes to 0.40 for some of the gender outcomes. This means that the study is sufficiently powered to detect a minimum effect size of roughly 0.09 to 0.2 standard deviations, as described in the previous section.

Summary statistics from our baseline are presented in Table 3. In our sample, 50% of the children are girls, the sample is almost well balanced between grades 1 through 3. Finally, 91% of the households have a functional TV at home, while the remaining 9% have access to a functional TV outside of their home.

We show that our randomization was successful in Table 4. The treatment and control groups are well balanced and show small differences between the demographic, household characteristics and baseline outcome variables. Not only are point estimates of differences small economically (i.e., the largest difference is for in-group activity score which is about 7% of the control mean) but are also statistically indistinguishable from zero.

We also are able to benchmark our sample's performance on the EGRA tool compared to a nationally representative sample of Grade 1 to 3 students in Kenya for 2021. This is presented in Appendix Table A4. Overall, our sample performs quite similar to the nationally representative sample in 2021. One important measure of foundational reading skills typically used is the share of zero readers (i.e., scoring 0% correct). Zero readers for letter recognition is low in our sample: 2.49% on average. For non-word recognition and oral fluency the shares are higher (18% and 25%), but lower than in the 2021 sample.

21

Table 3: Baseline summary statistics table

Variable	Mean	SD	Min	Max	N	Missing %
Demographics & HH Characteristics						
Child is female	0.50	0.50	0.00	1.00	4373	0.00
Child in Grade 1	0.31	0.46	0.00	1.00	4304	0.02
Child in Grade 2	0.34	0.47	0.00	1.00	4304	0.02
Child in Grade 3	0.35	0.48	0.00	1.00	4304	0.02
Hours watch TV	1.44	0.55	0.25	2.00	4373	0.00
Watch 2+ hours	0.37	0.48	0.00	1.00	4373	0.00
Watch Akili Kids Channel	0.44	0.50	0.00	1.00	4373	0.00
HH size	5.66	2.16	1.00	48.00	4373	0.00
HH has TV	0.91	0.28	0.00	1.00	4373	0.00
HH has smartphone	0.58	0.49	0.00	1.00	4373	0.00
HH Asset score	3.99	1.70	0.00	15.00	4373	0.00
Caregiver has primary or less	0.63	0.48	0.00	1.00	4373	0.00
HH speaks english	0.20	0.40	0.00	1.00	4373	0.00
Nr books	13.98	13.48	5.00	50.00	4373	0.00
Main Outcomes						
Literacy Index	0.01	0.98	-2.21	4.21	4373	0.00
Gender Attitude Index	-0.01	1.00	-3.76	3.92	4373	0.00
SEL Index	0.02	0.99	-2.86	1.32	4353	0.00
Reading index	0.01	1.00	-1.61	5.35	4373	0.00
Comprehension (%)	49.05	29.82	0.00	100.00	4373	0.00
Gender stereotype (mean)	0.60	0.18	0.00	1.00	4373	0.00
Gender role score (mean)	1.43	4.56	-16.00	21.00	4373	0.00
In-group attitude score (mean)	2.03	3.58	-6.00	6.00	4373	0.00
In-group acitivity score (mean)	2.33	4.34	-9.00	9.00	4373	0.00
Confidence score	9.56	2.19	4.00	12.00	4271	0.02
Curiosity score	9.51	2.17	4.00	12.00	4285	0.02

4.3   Planned Endline Data Collection

The development and implementation process for the endline survey is currently underway. We have started the kick-off meetings with village chiefs and other relevant stakeholders. Enumerators will be recruited between February and March 2024, with endline data collection planned for April 2024.

5   Analysis

5.1   Statistical Methods and Models

Our main analysis will estimate the intention-to-treat (ITT) effects of our intervention on the treatment group. Because we employ a randomized school level clustered-treatment

Table 4: Baseline balance table

Variable	Treatment	Control	Difference	p-value
Demographics & HH Characteristics				
Child is female	0.508	0.500	0.008	0.604
Child in Grade 1	0.302	0.314	-0.012	0.503
Child in Grade 2	0.353	0.331	0.022	0.231
Child in Grade 3	0.345	0.354	-0.010	0.614
Hours watch TV	1.430	1.441	-0.011	0.748
Watch 2+ hours	0.357	0.376	-0.019	0.458
Watch Akili Kids Channel	0.442	0.431	0.011	0.652
HH size	5.692	5.636	0.055	0.633
HH has TV	0.924	0.903	0.020	0.252
HH has smartphone	0.593	0.564	0.029	0.226
HH Asset score	4.020	3.960	0.060	0.504
Caregiver has primary or less	0.623	0.638	-0.015	0.523
HH speaks english	0.220	0.184	0.035	0.197
Nr books	14.044	13.914	0.130	0.863
Main Outcomes				
Literacy Index	0.018	0.000	0.018	0.782
Gender Attitude Index	0.014	-0.000	-0.014	0.794
SEL Index	0.037	0.000	0.037	0.361
Reading index	0.028	0.000	0.028	0.653
Comprehension (%)	49.013	49.094	-0.082	0.968
Gender stereotype (mean)	0.601	0.596	0.005	0.605
Gender role score (mean)	1.403	1.452	-0.049	0.847
In-group attitude score (mean)	2.051	2.003	0.048	0.866
In-group acitivity score (mean)	2.404	2.245	0.159	0.642
Confidence score	9.600	9.517	0.083	0.364
Curiosity score	9.530	9.489	0.041	0.637

Notes: Column 3 and 4 shows the estimate and p-value of the coefficient of a regression on the balance variables. Standard errors are clustered at the school level.

approach, we can identify the causal effect by estimating the following model:

$$y_{is} = \alpha_0 + \alpha_1 Treat_s + \beta_1 y_{is,t=0} + \gamma_1 G_s + \gamma_2 \boldsymbol{X_s} + \epsilon_{is} \tag{1}$$

where y_{is} is an outcome for the child i in school s, $Treat_s$ is an indicator for school s being treated. We include the baseline value of the outcome variable $y_{is,t=0}$, when available, to improve statistical precision, set to the mean if missing and an indicator of missingness (M_{is}). G is an indicator for the school quadruplet during randomization, and $\boldsymbol{X}$ is a vector of school-level controls that includes the characteristics used in our stratified randomization.

Our primary parameter of interest is α_1, which captures the direct effect of the treatment on children, as well as any potential spillovers on treated children *within* treatment schools.

Standard errors are clustered at the school level because this is both the level of randomization and stratification. As mentioned in Section 3.4.1, we account for multiple hypothesis testing adopting multiple strategies. First, we construct an index for each of our three families of outcomes following Anderson (2008). Second, we will present q-values following Benjamini et al. (2006) across our 8 main outcomes which together make up the three indices (but not including the main three indices). Third, when studying secondary outcomes (which together make up the 8 main outcomes described in the previous point), we will present q-values adjusting within families of outcomes.

5.1.1 Defining Compliance

Additionally, we will estimate the local average treatment effect for treatment compliers (or TOT). In our context, there is the potential for two-sided non-compliance (i.e. some students assigned to the treatment group fail to take-up the treatment while some students assigned to the control group take-up the treatment), we consider levels of compliance for both the treatment and control groups.

Measuring compliance in the context of watching over-the-air TV at home is difficult. In an ideal world we would like to track the exact number of episodes and engagement with each episode. This would be completely unfeasible in a large-scale field experiment setting. Additionally, tracking viewing with audimeters (say, as Nielsen does with TV ratings) would not be technically feasible as they require cable boxes. This would also be an imperfect measure: we cannot observe whether someone is actually watching.

We adopt a comprehensive strategy employing various measures for compliance. Our primary measures are based on child self-reports and recall. We also obtain alternative measures that rely on parental responses to time use and behavioral measures that do not rely on self-reports. Each has its own benefits and drawbacks. They also capture different dimensions of compliance that we may care about. Using different measures can provide useful bounds of the TOT estimates as they are likely biased in different directions.

1. Child self-report: Prior work on the effect of watching a TV show outside the classroom mostly relied on simple self-reports from parents or children, asking whether they ever watched a show or the frequency of watching the show. These measures are potentially problematic as they can be subject to experimenter demand effects. Our first approach to measuring compliance follows the existing literature. Children are shown an image of the title screen of the show (containing the show's name and image of the two main characters) and are asked *"Have you ever watched this show?"*. If they respond "Yes", then they are asked *"How many episodes do you think you watched?"*. The enumerator

24

never mentions the show by name. Using this measure at face value is potentially problematic as experimenter demand effects could lead children to over-report watching the show. If demand effects are larger among treatment children, estimates of the TOT would likely understate the true effects.

We can construct two compliance variables that likely overstate watching from these questions:

- Ever watched (self-report): an indicator that equals one if replies "Yes" to question *"Have you ever watched this show?"*.

- Consistent watch (self-report): an indicator that equals one if reports watching at least 6 episodes. This is because each skill is covered in two different episodes, for possibly watching each of the skills covered.

2. <u>Child recall</u>: The previous measures have significant shortfalls. Besides potential demand effects, measurement error from imperfect recall is likely a problem. We constructed a novel compliance tool based on show recall that is asked to all children. We have not seen a similar tool in prior studies and we hope it can be useful for future studies. Relying on recall also captures a different dimension of compliance: engagement with the show, rather than simple viewership.

First, there is a character recognition task. Children are shown the image of each of the three main characters and asked to name them. Nuzo and Namia might be easy to infer based on gender and reading the name of the show in the previous question. However, the third main character, who appears in every episode, has a very unique name that is virtually impossible to guess or infer: *Bubelang*. Responses are coded as correct by enumerators if they correctly identify the name or come close to it.

Second, the child is shown brief excerpts from key moments from different episodes. After the brief clip, they are asked if they watched that episode. If they report recognizing it then they are asked about what event happened right before or after (depending on the clip) and what did the characters learn from that key moment. This was repeated for each of the 13 episodes. Enumerators were instructed: *For the following questions, show the child the episode that is going to be played on the screen. Please select "Yes" if the child clearly recognizes the episode and states the content similar to what is noted in the question. Please do not read the content/main takeaway to them.*

One shortfall of this compliance tool is the possibility that higher ability children (defined broadly) likely have better recall. This could potentially downward bias our measure of true compliance. We tried to mitigate this as much as possible by design.

For character recall, enumerators are instructed to mark responses as correct if they come close to the actual name. This task is objectively easier than episode recall. However, it is only an extensive margin measure of compliance. For episode recall, the excerpts were chosen in collaboration with the show's creators to select the most salient and important moment of each episode, thus easier to recall.

We construct the following compliance measures based on recall. These can potentially be downward biased and potentially correlated with child ability, thus overstating the TOT estimates:

- Engaged watching (character recognition): an indicator that equals one if all three characters are correctly identified. Note that to the extent that guessing Nuzo and Namia is trivial (from previously reading the title), guessing the third character would be the sufficient statistic. Because this measure depends on recall, it helps differentiate between casual viewers and those who are genuinely engaged with the show.

- Episode score (episode recognition): takes values of 0-13 based on correctly identifying the event in each episode.

3. <u>Parent/Caregiver self-report</u>: Parents/Caregivers are asked a diary-style question that can be used to construct a very rough proxy for watching the show:

   (a) *"Which days do your children between 6 and 9 years usually watch TV?"*

   We note that these responses are also subject to experimenter demand effects (potentially stronger among adults). However, we attempt to minimize demand effects by asking diary-style questions rather than directly asking about the show. We can construct the measure:

   - Watches TV on days show airs: This indicator equals 1 if parent/caregiver reports child watches TV on Wednesdays or Saturdays.

4. <u>Behavioral</u>: Finally, we obtain a behavioral proxy measure that does not depend on self-reports. This can be learned by evaluating independent items from the literacy test. Some items cover learning objectives covered by episodes of the show, while others do not. While this measure would not directly measure compliance, it would provide suggestive evidence on mechanisms behind learning gains. This is discussed in Section 5.3 on mechanisms.

Using these compliance measures, we plan to employ an instrumental variable approach to obtain TOT estimates.[14] We will present the LATE estimates for the different measures of compliance. As discussed above, different measures have different weaknesses and capture different dimensions of compliance. This will be informative as it will provide a potential range of estimates for the LATE of watching the show. In the next subsection we discuss limitations and interpretation.

5.1.2 Limitations and interpretation

We can only rely on self-reported measures of treatment take-up (i.e., watching the show), which is likely measured with measurement error. To the extent that the nature of measurement error is classical, this would only attenuate our first stage and thus provide a weaker instrument. Similarly, experimenter demand effects can lead to inflated self-reports and thus understate the true TOT. However, some of our measures might be correlated to individual attributes. For example, episode recall could be correlated with child ability, thus overstate the true TOT instead. It is also important to note that intensity of compliance (i.e., number of episodes watched) was not randomized and can be correlated with individual attributes. Heterogeneity analysis on compliance variables can be informative about *who* watches and recalls the show. Ultimately, we plan on presenting alternative measures of take-up, considering their caveats when providing their interpretation.

5.2 Heterogeneous Effects

We conduct several analyses to study heterogeneous impacts in treatment effects. In particular, the following dimensions of (baseline) heterogeneity would be of general interest:

1. Gender of the child

2. Household size

3. Child attendance record

4. Socio-economic status

5. Speaks English at home

6. Grade of the child

7. Caregiver's education level

---

[14]A potential alternative could employ machine learning approaches (such as double/debiased machine learning models proposed by Chernozhukov et al. (2018))

8. Who the child watches television with most frequently:

   ▪ Parent/Caregiver

   ▪ Other children/Siblings

9. Where the child watches television most frequently

10. Baseline level of outcomes

11. Baseline number of books

12. Urban vs rural schools.[15]

We will follow the approach outlined in (Haushofer et al., 2020) for these dimensions of heterogeneity. We may also run analyses for other sources of heterogeneity in a more exploratory manner. For example, we would like to employ quantile treatment effects (QTE) in our primary outcome variables to understand the overall heterogeneity of treatment effect sizes across students (see Ashenfelter, 1978; Firpo, 2007; Powell, 2013).

We also plan to use machine learning-based approaches to estimate heterogeneous impacts. We expect to make use of two different approaches to finding the most important observed characteristics for heterogeneity: Best Linear Predictor (BLP) and Classification Analysis (CLAN) proposed by Chernozhukov et al. (2018), as well as the causal forest algorithm developed by Athey et al. (2019). However, because these methods are a very active area of research, with new tools developed at a rapid pace, we would adopt any new approaches that are considered standard or have desirable properties in our setting.

5.3 Mechanisms

In Section 3.3 we described the Theory of Change, listing the various mechanisms that could mediate the effects of our treatment and the different outcomes. We organize mechanism outcomes into intermediate behaviors and underlying mechanisms.

Intermediate behaviors are behaviors that are often necessary in order to produce a change in the primary outcomes. For example, a positive change in reading behavior is often a necessary condition for improving reading skills. It is difficult for children to improve their reading skills without a significant change in reading behavior. We first measure whether there are differences in reading behavior induced by treatment.

---

[15]At the moment, we were not yet able to gain access to the most up-to-date urban/rural classifications. Recently, Kenya has taken part of the pilot program by the UN for a globally harmonized measure of the Degree of Urbanization. Ideally, we will use that measure if available. If unable, we can use population density or distance to urban centers defined by FAO's Africover dataset as proxies.

- Reading and listening behavior: 8 individual item questions measuring different aspects of the child's reading and listening behavior.

  - How many books are in your house?

  - How many other reading materials (newspapers, comic books, magazines, etc.) are in your house?

  - How often do you borrow reading materials from the library or your friends?

  - How often do you read outside of school, i.e. for enjoyment?

  - How often does your mom, dad, or carer encourage you to read?

  - How often do you look for or ask for books?

  - How often do you listen to stories?

  - How often does your mom, dad, or carer read stories to you?

As discussed in the Theory of Change, there are three main mechanisms that could drive behavioral change and change in outcomes:

1. Preferences: As discussed in the main show characteristics and theory of change, one of the main objectives of the show is to change children's mindsets. This would be accomplished through the show's characteristics constructing socially meaningful characters, providing modeling behavior and generating positive emotions. In turn, these features would change children's mindset regarding reading, gender attitudes and socio-emotional learning outcomes.

   We can directly test the extent to which these show features are important to children:

   - Socially meaningful characters: 3 item question measuring how children identify with characters and their environment:

     - I will show you some characters from the Nuzo & Namia Show. Can you tell me how much you feel like you're like them and their world?
     (a) Nuzo
     (b) Namia
     (c) Bubelang

   Next, we can directly measure how the show changed children's preferences in the domains of our three outcomes (reading, gender attitudes and SEL). Gender and SEL outcomes were described previously as primary outcomes. For changing preferences on reading, we obtain a measure of positive affect towards reading:

▪ Positive affect on reading: 5 individual item questions measuring the child's attitudes towards reading as well as a corresponding reverse-coded question to ensure consistency. For instance, each child is asked if they agree with the statement "I like to learn how to read" as well as the statement "I wish I did not have to learn how to read," with responses ranging from "Strongly Disagree" to "Strongly Agree." A final score is aggregated by taking the sum of these items, where a higher score reflects more positive attitudes about reading. Individual items denoting attitudes toward reading include:

  – Interest in reading
  – Enjoyment from reading
  – Attitudes toward difficult books
  – Self-reported reading ability
  – Self-reported reading speed

2. Information/Skill: Unlike other educational TV interventions where specific pieces of information are conveyed (e.g., on HIV prevention), the show models key strategies for reading and comprehension. As described in the theory of change, part of this effect can be direct through increases in reading. We can use items within the literacy tests to learn to what extent learning these strategies are driving results. In particular, there are questions that are mapped to different curriculum items covered by the show (e.g., recall, scanning, inference and prediction). If children learn the strategies covered in the show, we would expect learning gains to be concentrated in questions that match those topics.

To obtain more direct evidence about how much information and skill are changing, we embed additional questions within our compliance tool described in section 5.1.1. After showing a excerpt from scene depicting a key learning strategy, respondents are asked what happens next in the episode (this next even is both a measure of compliance and illustrates a learning strategy). Next, respondents are asked whether they can describe what the characters are learning to do in that episode - thus directly capturing the child's knowledge of that particular skill. This is repeated for different episodes that cover each of the skills taught during the show.

Finally, we will examine heterogeneity based on ex-ante reading ability to test whether the largest gains would be experienced by children with lowest reading scores (La Ferrara, 2016), suggesting information might play an important role because they are the most likely to have knowledge gaps initially.

3. Time use: we measure changes in time use through various questions. For example, we study changes in reading behavior directly through questions mentioned above when examining changes in reading and listening behavior (e.g., "how often do you read outside of school?"). We try to obtain measures of their opportunity cost of watching the show by asking about what they typically do during non-show days at the time the show airs. Conversely, we ask what they usually did on show days during show times (e.g., Wednesdays at 6.30pm and Saturdays at 10.30am). We also examine changes in TV watching behavior:

   ▪ How much TV do you watch in a day?

   ▪ What shows do you usually watch on TV? Multiple choice answers with broad categories including educational TV and non-educational TV categories such as cartoons, movies, and other TV shows.

We also study the role of the watch environment as a potential intermediary or barrier. In an accompanying qualitative study, we found that caregiver presence and nudges were important mechanisms facilitating engagement with the show. Additionally, engagement and concepts can be reinforced when watching the show with other children or siblings, or if English is usually spoken in the household. Thus, we will explore heterogeneous effects by watching with siblings.

▪ Watch environment: 4 individual item questions characterizing how the child watched the show:

   – Who did you watch the show with most of the time?

   – Where did you watch the show most of the time?

   – Which means did you use to watch the show?

   – How much TV do you watch in a day?

   – What shows do you usually watch on TV? Multiple choice answers with broad categories including educational TV and non-educational TV categories such as cartoons, movies, and other TV shows.

In the Theory of Change we described potential barriers. A potentially important dimension is the household's socio-economic status and problems related to poverty (e.g., power outages, other responsibilities, nutrition). Besides studying heterogeneity by SES status measured at baseline, we can gain further insights to what factors related to poverty might matter. For instance, we will test the extent to which power outages matter using data on

31

actual outages reported by Kenya Power. Assuming there is enough geographic variation in outages during the study period, we will test whether effects vary by incidence of power outages.

A second potential barrier is the availability of reading materials in the household. We will use a baseline measure of the number of reading materials in the household to test the extent to which this could operate as a significant barrier to improving reading skills.

We consider the potential role of indirect mechanisms as described in the theory of change. These could operate through interactions with friends or teachers that relate to the show:

- How often do teachers talk about the Nuzo & Namia show in school?

- How often do you talk about the Nuzo & Namia show with your friends in school?

- How often do you talk about the Nuzo & Namia show with friends in your community?

We also capture dimensions of interactions with others that can potentially be mediators of learning (for literacy and SEL):

- How often do you meet with friends to read?

- How often do you ask friends for help if you have a problem to solve?

- How often do you ask your parents for help if you have a problem to solve?

- How often do you talk to your friends about your feelings?

- How often do you talk to your parents about your feelings?

5.4 Dealing with Attrition and Missing Values

We will identify attrition using the following techniques:

1. We will test for non-random attrition by treatment status by running equation 1 using an indicator for attrition as the outcome.

2. We will regress an attrition indicator on a vector of baseline variables, report the marginal effects, and test if each marginal effect is different to zero. The baseline variables for this analysis will be respondent gender, grade at baseline, school, county, sub-county, and indicators for the field officers who conducted the survey. We will use robust standard errors clustered by county.

3. We will report Lee bounds for all outcomes in order to account for nonrandom attrition.

6 Interpreting results

As stated, this paper contributes to the literature in several ways. We provide the first causal evidence of the effects of home viewing of educational TV on children's literacy, gender attitudes, and socio-emotional learning. We also provide evidence on mechanisms. Learning this can be informative for the design of future educational television shows and encouragement interventions. Of importance to policymakers who could be interested in scaling up such interventions, we can identify barriers to implementing and scaling (e.g., poverty) and will provide a comprehensive cost-effectiveness analysis, as described in detail below.

6.1 Cost-effectiveness Analysis

Relevant for policymakers is the cost of any potential positive impact relative to other interventions. Much of the conversation surrounding education TV and ed-tech in general is the low-cost and easy scale-up. To directly estimate this, we perform a cost-effectiveness analysis (CEA). The CEA will follow the ingredient method discussed in McEwan (2012) and Dhaliwal et al. (2013). Closely collaborating with Ubongo, we plan to track and categorize different costs incurred as a result of producing, airing, and marketing the show. Since the show was produced as part of this experiment, we will extrapolate various costs from Ubongo's other productions as a proxy measure for the cost of the program outside the context of this evaluation. Potential costing tools that may be employed are Brookings' Childhood Cost Calculator (C3) and the World Bank's Standardized Early Childhood Development Costing Tool (SECT) (Gustafsson-Wright et al., 2017; Gustafsson-Wright and Lee, 2021). We will conduct the CEA under the following assumptions:

- Any costs are converted to 2011 US dollars using the exchange rate from the year the costs were incurred.

- Inflation is calculated using GDP deflators.

- A 10% discount rate is assumed as a reasonable rate for discounting in educational programs in developing countries. A sensitivity analysis will be conducted to examine if cost-effectiveness estimates vary with chosen discount rates.

- The time horizon for the project will be the median production timeline of Ubongo shows.

- Any costs associated with the research design (i.e. research staff costs, costs associated with the research design, donor resources, staff time spent communicating with research team, etc.) are excluded.

The following are expected cost categories associated with the program:

▪ Production costs related to the development of the show. This include costs to write, design, film, animate, and edit the show as well as other relevant staff training events.

▪ Distribution costs related to broadcasting the show, including costs to obtain broadcasting license and ministry approval, marketing costs, etc.

▪ Cost of encouragement scheme, including costs associated with in-person encouragement sessions as well as associated SMS systems.

The CEA will focus on the three main outcomes: Literacy Score, Confidence and Curiosity, and Gender Stereotype Knowledge and Flexibility. Most relevant to policy-makers will the cost-effectiveness of the program as it relates to literacy scores. As such, we plan to compare our findings to the cost-effectiveness of other interventions targeted at literacy, with additional comparisons with interventions targeting the remaining two family outcomes as relevant.

References

Alan, S., T. Boneva, and S. Ertac (2019). Ever failed, try again, succeed better: Results from a randomized educational intervention on grit. *The Quarterly Journal of Economics 134*(3), 1121–1162.

Alan, S., S. Ertac, and I. Mumcu (2018). Gender stereotypes in the classroom and effects on achievement. *Review of Economics and Statistics 100*(5), 876–890.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American statistical Association 103*(484), 1481–1495.

Araya, R., E. Arias Ortiz, N. L. Bottan, and J. Cristia (2019). Does gamification in education work? experimental evidence from chile.

Ashenfelter, O. (1978). Estimating the effect of training programs on earnings. *The Review of Economics and Statistics 60*(1), 47–57.

Athey, S., J. Tibshirani, and S. Wager (2019). Generalized random forests. *The Annals of Statistics 47*(2), 1148–1178.

Bandura, A. and R. H. Walters (1977). *Social learning theory*, Volume 1. Englewood cliffs Prentice Hall.

Banerjee, A., E. La Ferrara, and V. H. Orozco-Olvera (2019). The entertaining way to behavioral change: Fighting hiv with mtv. Technical report, National Bureau of Economic Research.

Benjamini, Y., A. M. Krieger, and D. Yekutieli (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika 93*(3), 491–507.

Borzekowski, D., D. Singpurwalla, D. Mehrotra, and D. Howard (2019). The impact of galli galli sim sim on indian preschoolers. *Journal of Applied Developmental Psychology 64*, 101054.

Borzekowski, D. L. (2018). A quasi-experiment examining the impact of educational cartoons on tanzanian children. *Journal of Applied Developmental Psychology 54*, 53–59.

Bruhn, M. and D. McKenzie (2009). In pursuit of balance: Randomization in practice in development field experiments. *American Economic Journal: Applied Economics 1*(4), 200–232.

Center, P. R. (2018). Social media use continues to rise in developing countries but plateaus across developed ones.

Cherewick, M., S. Lebu, C. Su, L. Richards, P. F. Njau, and R. E. Dahl (2021). Promoting gender equity in very young adolescents: targeting a window of opportunity for social emotional learning and identity development. *BMC Public Health 21*(2299).

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018, February). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal 21*(1), C1C68.

Chernozhukov, V., M. Demirer, E. Duflo, and I. Fernández-Val (2018). Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in india.

Chong, A. and E. L. Ferrara (2009). Television and divorce: Evidence from brazilian novelas. *Journal of the European Economic Association 7*(2-3), 458–468.

Deole, S. S. and T. Zeydanli (2021). Does education predict gender role attitudes?: Evidence from european datasets. GLO Discussion Paper 793 [rev.], Global Labor Organization (GLO), Essen.

Dhaliwal, I., E. Duflo, R. Glennerster, and C. Tulloch (2013). Comparative cost-effectiveness analysis to inform policy in developing countries: A general framework with applications for education. *Education Policy in Developing Countries*.

Ferrara, E. L., A. Chong, and S. Duryea (2012). Soap operas and fertility: Evidence from brazil. *American Economic Journal: Applied Economics 4*(4), 1–31.

Firpo, S. (2007). Efficient semiparametric estimation of quantile treatment effects. *Econometrica 75*, 259–276.

Gustafsson-Wright, E., I. Boggild-Jones, and S. Gardiner (2017, September 5). Sect: The standardized early childhood development costing tool.

Gustafsson-Wright, E. and D. Lee (2021, December 23). Getting both costs and effectiveness right to improve decisionmaking in education.

Haushofer, J., R. Mudida, and J. Shapiro (2020). The comparative impact of cash transfers and psychotherapy on psychological and economic well-being. *NBER working paper 7*.

International Institute for Educational Planning (IIEP) - UNESCO (2024). Learning and teaching materials.

Jukes, M. C. H., N. L. Mgonda, J. J. Tibenda, P. Gabrieli, G. Jeremiah, K. L. Betts, J. Williams, and K. L. Bub (2021). Building an assessment of community-defined social-emotional competencies from the ground up in tanzania. *Child Development 92*, 1095–1109.

Kearney, M. S. and P. B. Levine (2015). Media influences on social outcomes: The impact of mtv's 16 and pregnant on teen childbearing. *American Economic Review 105*(12), 3597–3632.

Kearney, M. S. and P. B. Levine (2019). Early childhood education by television: Lessons from sesame street. *American Economic Journal: Applied Economics 11*(1), 318–350.

Kenya Ministry of Education (2019). Basic education statistical booklet. Technical report, Ministry of Education.

La Ferrara, E. (2016). Mass media and social change: Can we use television to fight poverty? *Journal of the European Economic Association 14*(4), 791–827.

Mares, M.-L. and Z. Pan (2013). Effects of sesame street: A meta-analysis of children's learning in 15 countries. *Journal of Applied Developmental Psychology 34*(3), 140–151.

McEwan, P. J. (2012). Cost-effectiveness analysis of education and health interventions in developing countries. *Journal of Development Effectiveness 4*(2), 189–213.

McKenzie, D. (2022, April 19). Why i am now more cautious about using or recommending matched pair randomization and like matched quadruplets instead.

Merga, M. K. and S. M. Roni (2017). The influence of access to ereaders, computers and mobile phones on children's book reading frequency. *Computers & Education 109*, 187–196.

Muralidharan, K., A. Singh, and A. J. Ganimian (2019). Disrupting education? experimental evidence on technology-aided instruction in india. *American Economic Review 109*(4), 1426–1460.

Murunga, J. (2021). Radio, tv and internet audience statistics in kenya q1 2021.

Powell, D. (2013). A new framework for estimation of quantile treatment effects nonseparable disturbance in the presence of covariates.

Riley, E. (2024). Role models in movies: the impact of queen of katwe on students educational attainment. *Review of Economics and Statistics 106*(2), 334–351.

Schwab, B., S. Janzen, N. P. Magnan, and W. M. Thompson (2020). Constructing a summary index using the standardized inverse-covariance weighted average of indicators. *The Stata Journal 20*(4), 952–964.

Tadesse, S. and W. Muluye (2020). The impact of covid-19 pandemic on education system in developing countries: A review. *Open Journal of Social Sciences 8*, 159–170.

UNESCO (2022). Persistent teacher gap in sub-saharan africa jeopardizing education recovery.

UNESCO (2023). Literacy rate, adult male.

USAID (2016). Early grade reading assessment (egra) toolkit. Technical report.

Usawa (2021). Uwezo learning assessment. Technical report, Usawa Agenda.

Wawire, B., F. Kiroro, N. Muhia, C. Asego, R. Perakis, and G. Smarrelli (2023). From evidence to policy-making: What should be in kenyas next education and training gender policy?

WorldBank, UNESCO, UNICEF, USAID, FCDO, and B. . M. G. Foundation (2022). The state of global learning poverty: 2022 update.

Yu, L., J. J. Yu, and X. Tong (2023). Social-emotional skills correlate with reading ability among typically developing readers: A meta-analysis. *Education Sciences 13*(2), 220.

Table A.1: Primary Outcome Description

Hypothesis	Outcome Variable	Survey Question/s	Variable coding	Variable type	Source
H1	Literacy Index	Employs EGRA tool, combining 5 scores for letter recognition (cwpm), non-word reading (cwpm), oral fluency (cwpm), reading comprehension (%) and listening comprehension (%). Definitions for these below.	Employ Anderson (2008) to construct index. Standardized relative to control group mean and sd.	Continuous	Endline survey
H2	Gender Attitude Index	Combines 7 scores for gender stereotype, Gender role (M+F), In-group attitude (M+F), in-group activity (M+F). Definitions below.	Employ Anderson (2008) to construct index. Standardized relative to control group mean and sd.	Continuous	Endline survey
H3	SEL Index	Combines 2 scores (Curiosity Score and Confidence Score) into index. Described below.	Employ Anderson (2008) to construct index. Standardized relative to control group mean and sd.	Continuous	Endline survey
H1	Reading Index	Employs the EGRA Tool. Combines the Literacy section of the tool, includes 3 scores: non-word reading (cwpm), letter recognition (cwpm), and oral fluency (cwpm).	Employ Anderson (2008) to construct index. Standardized relative to control group mean and sd.	Continuous	Endline survey
H1	Comprehension Score	Employing the EGRA Tool. Combines 2 scores: Reading comprehension and listening comprehension	Mean score (% correct)	Continuous	Endline survey
H2	Gender stereotype score	[sterknow_grpinfo] "I will read out some activities for you. Please tell me who you think should perform this activity- a girl, boy, or both girls and boys. If its other, state your response." [ster_doll] "Play with dolls" [ster_clean] "Clean the house" [ster_teach] "Become a teacher" [ster_fight] "Get into a fight" [ster_shop] "Shop for clothes" [ster_doctor] "Become a doctor" [ster_car] "Play with toy cars" [ster_sport] "Play a sport" [ster_doll] "Plays with doll" [ster_clean] "Clean the house" [ster_teach] "Become a teacher [ster_fight] "Get in a fight" [ster_shop] "Go shopping" [ster_doctor] "Become a doctor" [ster_car] "Drive a car" [ster_sport] "Play a sport"	Responses are coded =1 if they follow the stereotype, =0 if not. Then all items are averaged. Higher score means stronger gender stereotype.	Continuous	Endline survey

| H2 | Gender role score | [genrole_info] "I will read out some activities. Tell me how likely you are to perform the activity on a scale of 1-4?"

[role_doll] "Play with dolls"
[role_clean] "Clean the house"
[role_teach] "Become a teacher"
[role_fight] "Get into a fight"
[role_shop] "Go shopping"
[role_doctor] "Become a doctor"
[role_car] "Drive a car"
[role_sport] "Play a sport"

[role_trait] "I will read out some behaviors for you. Can you tell me how often you to show this behavior on a scale of 1-4?"

[role_care] "Caring"
[role_fun] "Funny"
[role_strong] "Strong"
[role_smart] "Smart"
[role_weak] "Weak"
[role_shy] "Shy" | Responses are coded -4 to 4 based on whether it conforms to traditional gender roles of the child respondent. Responses that conform with traditional gender roles are coded as positive, and responses that go against traditional gender roles are coded as negative. Then take the sum of all items. Higher positive score means stronger subscription to gender roles, =0 means flexible gender role knowledge | Continuous | Endline survey |
| H2 | In-group attitude score | [grpattitudes_intro] "I will read out a few activities for you. Can you tell me how much you want to perform that activity on a scale of 1-3? 1 is Unwilling, 2 is somewhat willing, 3 is extremely willing."
[grp_studygirl] "Join a group of girls to study with after school."
[grp_studyboy] "Join a group of boys to study with after school."
[grp_playgirl] "Play a sport with girls."
[grp_playboy] "Play a sport with boys."
[grp_talkgirl] "Talk to a girl in school."
[grp_talkboy] "Talk to a boy in school." | Responses are coded -3 to 3 based on the in- and out-groups of the child respondent. Willingness to perform activities with out-groups are coded as negative, and willingness to engage in activities with in-groups are coded as positive. Then take the sum of all items. Higher positive score means stronger in-group attitudes, =0 means flexible in- and out-group attitudes. | Continuous | Endline survey |

Hypothesis	Outcome Variable	Survey Question/s	Variable coding	Variable type	Source
H2	In-group activity score	[grp4_intro2] "I will read out a few activities for you. Can you tell me how often you perform that activity on a scale of 1-4? 1 is never, 2 is sometimes, 3 is most of the time, 4 is always." [grp_dostudygirl] "Join a group of girls to study with after school." [grp_dostudyboy] Join a group of boys to study with after school." [grp_doplaygirl] "Play a sport with girls." [grp_doplayboy] "Play a sport with boys." [grp_dotalkgirl] "Talk to a girl in school." [grp_dotalkboy] "Talk to a boy in school."	Responses are coded -4 to 4 based on whether the respondent engages with activities with in- and out-group individuals. Activities with out-groups are coded as negative and activities with in-groups are coded as positive. Then take the sum of all items. Higher positive score means stronger likelihood of engaging in activities with in-group, =0 means equal likelihood to engage in activities with in- and out-groups.	Continuous	Endline survey
H3	Confidence Score	[conf_1] "Let me tell you about a child called Bonifasi/Ashura. Bonifasi/Ashura thinks that classwork is much easier for Bonifasi/Ashura than it is for the other children in the class. Are you like Bonifasi/Ashura? [conf_2] "Let me tell you about a child called Frenki/Neema. Frenki/Neema believes that the other children in the class admire Frenki/Neema. Are you like Frenki/Neema? [conf_3] "Let me tell you about a child called Daudi/Furaha. Daudi/Furaha is often first to answer the teacher's questions. Are you like Daudi/Furaha? [conf_4] "Let me tell you about a child called Mariam/Alex. Mariam/Alex enjoys standing in front of the class and doing an exercise. Are you like Mariam/Alex?	Responses are coded =1 if they are like the example character, =0 if not. Then all items are summed. Higher score means higher confidence.	Continuous	Endline survey
H3	Curiosity Score	[curi_1] "Let me tell you about a child called Sharifa/Nickson. Sharifa/Nickson is curious to investigate and understand new things. Are you like Sharifa/Nickson? [curi_2] "Let me tell you about a child called James/Aisha. James/Aisha likes to ask many questions. Are you like James/Aisha? [curi_3] "Let me tell you about a child called Agnes/Kasimu. Agnes/Kasimu enjoys learning new things at schools. Are you like Agnes/Kasimu? [curi_4] "Let me tell you about a child called Tom /Maimuna. Tom /Maimuna expresses himself/herself in class. Are you like Tom /Maimuna?	Responses are coded =1 if they are like the example character, =0 if not. Then all items are summed. Higher score means higher curiosity.	Continuous	Endline survey

Table A.2: Secondary Outcome Description

Hypothesis	Outcome Variable	Survey Question/s	Variable coding	Variable type	Source

H1	Letter Identification Score	[letter_time] Amount of time remaining in seconds [letter_attempted] Total number of letters attempted [letter_incorrect] Number of incorrect letters [letters_correct] Number of correct letters [letters_firstline] Whether the firstline was all incorrect	Score (Correct sounds per minute) = (Total letter sound attempted – Total incorrect) / [(DURATION – Time remaining on device) / DURATION] or % correct (or % zero score)	Continuous	Endline survey
H1	Non-Word Reading Score	[nonwords_time] Amount of time remaining in seconds [nonwords_attempted] Total number of words attempted [nonwords_incorrect] Number of incorrect words [nonwords_correct] Number of correct words [nonwords_firstline] Whether the firstline was all incorrect	Score (correct words per minute) = (Total nonwords attempted – Total incorrect) / [(DURATION – Time remaining on device) / DURATION] or % correct (or % zero score)	Continuous	Endline survey
H1	Oral Fluency	[reading_time] Amount of time remaining in seconds [reading_attempted] Total number of items attempted [reading_incorrect] Number of incorrect items [reading_correct] Number of correct items [reading_firstline] Whether the firstline was all incorrect [reading_sentences] Number of sentences read	Score (correct words per minute) = (Total items attempted – Total incorrect) / [(DURATION – Time remaining on device) / DURATION] or % correct (or =1 if zero score)	Continuous	Endline survey
H1	Reading Comprehension	[comprehension_1] Who wanted to play football? [comprehension_2] What did Baraka not find? [comprehension_3] Where did Baraka look for the ball? [comprehension_4] Why did Baraka smile? [comprehension_5] Why did Baraka run to the field?	% correct (or =1 if zero score)	Continuous	Endline survey
H1	Listening Comprehension Score	[listencomp_1] Who is under a mango tree? [listencomp_2] What does Kenzo see on the tree? [listencomp_3] What does Kenzo want to do with the mango? [listencomp_4] What do you think Kenzo will do next time she wants to pick a mango from the tree? [listencomp_5] What did Kenzo do after the mango fell down?	% correct (or =1 if zero score)	Continuous	Endline survey
H2	Female gender role score	[role_doll] "Play with dolls" [role_clean] "Clean the house" [role_teach] "Become a teacher" [role_shop] "Go shopping" [role_care] "Caring" [role_weak] "Weak" [role_shy] "Shy"	See gender role above	Continuous	Endline survey

H2	Male gender role score	[role_fight] "Get into a fight" [role_doctor] "Become a doctor" [role_car] "Drive a car" [role_sport] "Play a sport" [role_fun] "Funny" [role_strong] "Strong" [role_smart] "Smart"	See gender role above	Continuous	Endline survey
H2	In-group attitude score female	[grp_studygirl] "Join a group of girls to study with after school." [grp_playgirl] "Play a sport with girls." [grp_talkgirl] "Talk to a girl in school."	See in-group attitude above	Continuous	Endline survey
H2	In-group attitude score male	[grp_studyboy] "Join a group of boys to study with after school." [grp_playboy] "Play a sport with boys." [grp_talkboy] "Talk to a boy in school."	See in-group attitude above	Continuous	Endline survey
H2	In-group activity score female	[grp_dostudygirl] "Join a group of girls to study with after school." [grp_doplaygirl] "Play a sport with girls." [grp_dotalkgirl] "Talk to a girl in school."	See in-group activity above	Continuous	Endline survey
H2	In-group activity score male	[grp_dostudyboy] Join a group of boys to study with after school." [grp_doplayboy] "Play a sport with boys." [grp_dotalkboy] "Talk to a boy in school."	See in-group activity above	Continuous	Endline survey

Table A3: Secondary Outcomes: Minimum detectable effect size at 80% power

Outcome	MDE	Mean	SD	ICC
Literacy				
Letter recognition (CWPM)	0.149	24.672	16.892	0.180
Non-word reading (CWPM)	0.145	7.964	8.663	0.166
Oral fluency (CWPM)	0.149	40.226	55.667	0.177
Reading comprehension (%)	0.154	33.300	37.842	0.195
Listening comprehension (%)	0.180	64.807	32.735	0.300
Gender Attitudes				
Female gender role score	0.087	0.161	2.736	0.005
Male gender role score	0.086	0.042	2.830	0.002
In-group attitude score female	0.096	0.339	2.398	0.025
In-group attitude score male	0.100	0.336	2.331	0.033
In-group activity score female	0.098	0.381	2.628	0.028
In-group activity score male	0.100	0.394	2.619	0.034

Notes: Minimum effect size in standardized units. Intra-cluster correlation calculated at the school level. To adjust for imperfect compliance, divide MDE by net compliance rate.

Table A4: Benchmarking EGRA scores with Kenyan nationally representative survey

	Baseline Survey				Kenya 2021 (USAID)		
	Overall	Grade 1	Grade 2	Grade 3	Grade 1	Grade 2	Grade 3
CWPM							
Letter Recognition	24.67	20.30	25.52	27.64	20.70	27.10	30.00
Non-word Recognition	7.96	4.15	7.30	12.02	7.80	14.80	24.10
Oral Fluency	40.23	15.24	34.10	68.14	11.40	27.80	55.70
% Correct							
Letter Recognition	24.64	20.25	25.51	27.60	20.70	27.00	29.90
Non-word Recognition	15.71	7.87	14.54	23.83	15.40	29.30	46.30
Oral Fluency	36.95	18.21	36.74	53.78	15.80	37.90	64.30
Reading Comprehension	33.30	15.63	32.47	50.07	8.60	21.90	42.70
Listening Comprehension	64.81	57.51	64.23	72.13	NA	NA	NA
% Zero score							
Letter Recognition	2.49	3.39	2.58	1.66	21.60	15.40	NA
Non-word Recognition	17.68	28.81	16.62	8.98	52.20	28.70	NA
Oral Fluency	24.79	39.22	25.10	11.70	45.00	26.60	NA
Reading Comprehension	47.79	71.64	47.35	26.93	80.00	53.90	NA
Listening Comprehension	10.95	16.67	10.65	6.05	NA	NA	NA

Notes: Correct words (or sounds) per minute (CWPM) calculated as (Items attempted - Items incorrect)/((duration-time remaining)/duration). Kenya 2021: USAID Kenya Tusome Early Grade Reading Activity Study on Grade 1, 2, and 3 Pupil Learning Outcomes in 2021: Final Report. url: https://pdf.usaid.gov/pdf_docs/PA00ZXKJ.pdf

B. Information session materials

The following pages contain original and translated materials provided to parents at the information session. Note that the first poster was originally distributed in English and all other materials were originally in Kiswahili.
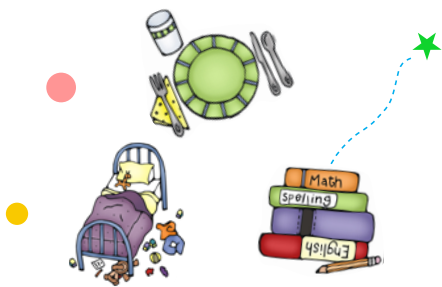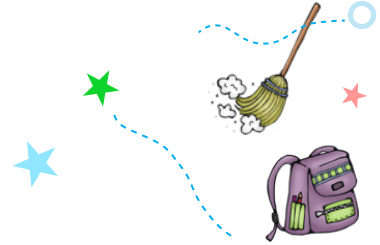
Nuzo and Namia

Time to have fun and learn

**6.30pm – 7.00pm every Wednesday
and 10.30am – 11.00am every Saturday
on the Akili Kids Channel**

**Help Namia find her way to Nuzo
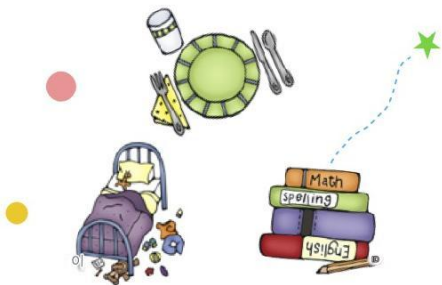in time for the next episode !**
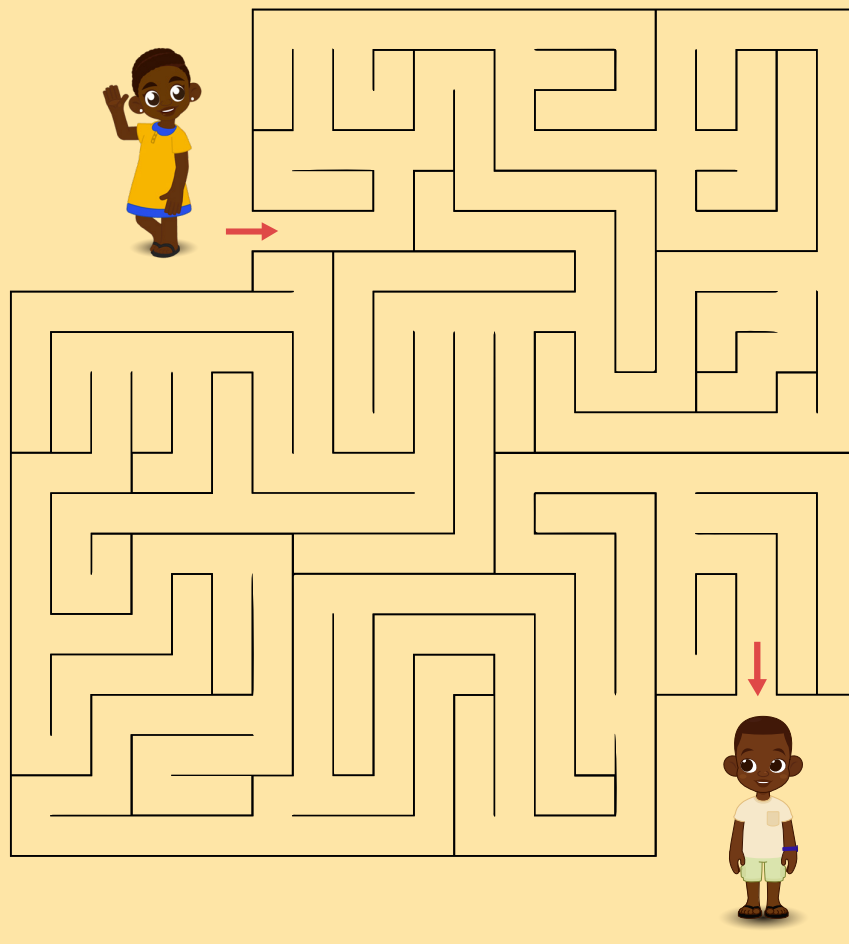
Nuzo and Namia

RATIBA YA KILA JUMAMOSI

Masaa	Kinachofanyika kabla ya kutazama "Nuzo and Namia"	
6:00am → 10:30am		

Masaa	Kufurahia na kusoma na Nuzo na Namia kwenye Akili Kids	Kinachofanyika "Nuzo and Namia" inapotazamwa
10:30am → 11:00am		

Masaa	Kinachofanyika baada ya kutazama "Nuzo and Namia"	
11:00am → 8:00pm		

RATIBA YA KILA JUMATANO

Masaa	Kinachofanyika kabla ya kutazama "Nuzo and Namia"	
4:00pm → 6:30pm		

Masaa	Kufurahia na kusoma na Nuzo na Namia kwenye Akili Kids	Kinachofanyika "Nuzo and Namia" inapotazamwa
6:30pm → 7:00pm		

Masaa	Kinachofanyika baada ya kutazama "Nuzo and Namia"	
7:00pm → 9:00pm		

AKILI KIDS!

NUZO
and
Namia

10-010-010-0

EVERY SATURDAY SCHEDULE	
	What happens before watching "Nuzo and Namia"
Enjoy reading with Nuzo and Namia on Akili Kids	What happens when "Nuzo and Namia" is viewed
What happens after watching "Nuzo and Namia"	

EVERY WEDNESDAY SCHEDULE	
Hours	What happens before watching "Nuzo and Namia"
4:00pm ⬇ 6:30pm	
Masaa	Enjoy reading with Nuzo and Namia on Akili Kids	What happens when "Nuzo and Namia" is viewed
6:30pm ⬇ 7:00pm	
Masaa	What happens after watching "Nuzo and Namia"
2:00pm ⬇ 9:00pm	

mind KIDS!

Fomu ya Mzazi kujaza

93% ya wazazi/wanaotoa huduma kwa wazazi/walezi ambao wanawao huona kipindi cha Nuzo and Namia wamesema kwamba kipindi hiki ni bora!

> *Naweza kusema kipindi cha Nuzo and Namia kimeboresha maisha yangu kwa sababu watoto wangu wanapata elimu bila kulipia masomo ya ziada.*
>
> **Baba ya watoto watatu walio na miaka 3, 5 na 7.**

Mimi _____

na apa kwamba wanangu

_____

_____

_____

watakuwa wakitazama kipindi cha Nuzo and Namia kila Jumamosi saa 10.30 – 11.00 asubuhi katika Akili Kids Channel.

NUZO

Namia

by brain.

★

Parent Form to fill out

93% of parents/caregivers whose children
watch the Nuzo and Namia show
said that this show is the best!

❝

I can say that the Nuzo and Namia
program has improved my life because my
children are getting an education
without paying for extra lessons.

Father of three children
aged 3, 5 and 7.

Me _____

_____ and swear that my sons _____

_____

_____

_____

they will be watching the Nuzo and Namia program
every Saturday at 10.30 - 11.00 am in Akili Kids Channel.

mind
KIDS!

Ratiba ya

Nuzo and Namia

Inafaa utazame Nuzo and Namia angalau **mara moja kila Jumamosi au Jumatano.**

Weka alama kwenye Jumamosi au Jumatano ambazo mtoto/watoto wako wametazama "Nuzo and Namia".

Jumamosi

Juni	Tarehe 24				
Julai	Tarehe 1	Tarehe 8	Tarehe 15	Tarehe 22	Tarehe 29
Agosti	Tarehe 5	Tarehe 12	Tarehe 19	Tarehe 26	
Septemba	Tarehe 2	Tarehe 9	Tarehe 16	Tarehe 23	Tarehe 30
Oktoba	Tarehe 7	Tarehe 14	Tarehe 21	Tarehe 28	
Novemba	Tarehe 4	Tarehe 11	Tarehe 18	Tarehe 25	

Jumatano

Juni	Tarehe 21	Tarehe 28			
Julai	Tarehe 5	Tarehe 12	Tarehe 19	Tarehe 26	
Agosti	Tarehe 2	Tarehe 9	Tarehe 16	Tarehe 23	Tarehe 30
Septemba	Tarehe 6	Tarehe 13	Tarehe 20	Tarehe 27	
Oktoba	Tarehe 4	Tarehe 11	Tarehe 18	Tarehe 25	
Novemba	Tarehe 1	Tarehe 8	Tarehe 15	Tarehe 22	Tarehe 29

Schedule of

NUZO
and
Namia

You should watch Nuzo and Namia at least once every Saturday or Wednesday.

Mark the Saturdays or Wednesdays that your child/children have watched "Nuzo and Namia".

saturday

Month					
June	The 24th				
July	Date 1	The 8th	The 15th	The 22nd	The 29th
August	The 5th	The 12th	The 19th	The 26th	
September	Date 2	The 9th	The 16th	The 23rd	The 30th
October	The 7th	The 14th	The 21st	The 28th	
November	The 4th	The 11th	The 18th	The 25th	

Wednesday

Month					
June	The 21st	The 28th			
July	The 5th	The 12th	The 19th	The 26th	
August	Date 2	The 9th	The 16th	The 23rd	The 30th
September	The 6th	The 13th	The 20th	The 27th	
October	The 4th	The 11th	The 18th	The 25th	
November	Date 1	The 8th	The 15th	The 22nd	The 29th

Kipindi cha

Nuzo and Namia
by ubongo

★

Kinaletwa ndani ya Akili Kids Channel kila Jumatano saa 6.30–7.00 jioni na Jumamosi saa 10.30–11.00 asubuhi.

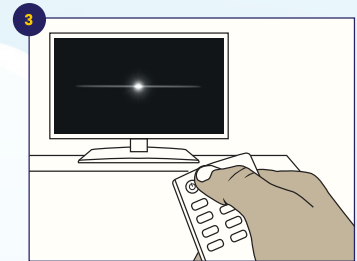## Usipitwe na Kipindi cha Nuzo and Namia.

Fuata hatua zifuatazo ikiwa usajili wako umeisha, na hauna SMART TV ili wanako wasipitwe na Kipindi cha Nuzo and Namia.
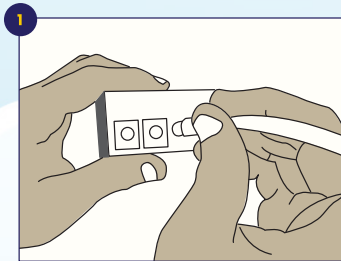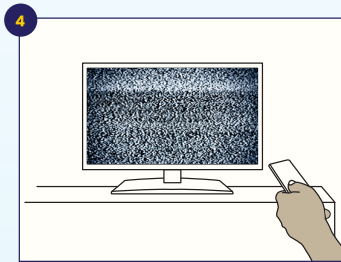
1

Ondoa sanduku la decoda. Sanduku hili linapatikana katikati ya antena na televisheni.
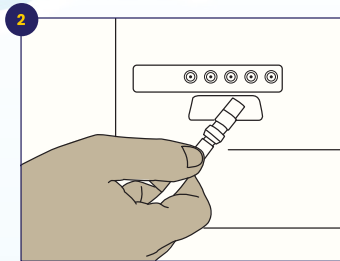
2

Wekelea antenna uliochukua kwenye sanduku kwa antenna ya televisheni

3

Zima televisheni na ungoje.

4

Fungua televisheni na tafuta chaneli za analogue.

5

Chagua Akili Kids.

Tulia na usome, utatue na ufurahie Nuzo and Namia na marafiki!

AKILI KIDS!

# Period of

NUZO
and
Namia

by brain.

It is introduced in Akili Kids Channel
every Wednesday at 6.30-7.00 pm
and Saturday at 10.30-11.00 am.
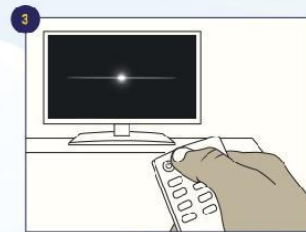

Don't miss the Nuzo and Namia Episode.

Follow the following steps if your subscription has expired, and you do
not have a SMART TV so that your children do not miss the Nuzo and Namia Show.
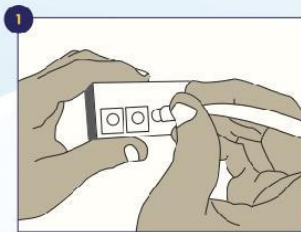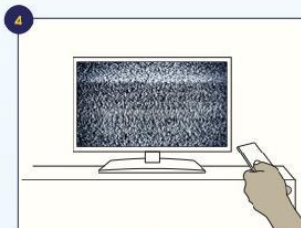

Remove the decoder box.
This box is located between
the antenna and the
television.
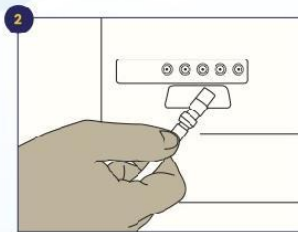
Attach the antenna that you
took in the box to
the television antenna

Turn off the television
and wait.

Turn on the television and search
for analogue channels.

Choose Smart Kids.

Sit back and read,
solve and enjoy
Nuzo and Namia
with friends!

mind
KIDS!

Author Statement: DEVEC-D-24-00428, "Bridging Educational Gaps Watching TV at Home: The Impact of an Educational TV Show on Kenyan Children"

Jasmin Baier: *Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Resources, Validation, Visualization, Writing – original draft, Writing – review & editing.*

Laura Barasa: *Conceptualization, Formal analysis, Investigation, Methodology, Resources, Validation, Visualization, Writing – original draft, Writing – review & editing.*

Nicolas Bottan (Corresponding Author): *Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.*

Anushka Ghosh: *Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.*

Johannes Haushofer: *Conceptualization, Investigation, Methodology, Resources, Supervision.*

Mark Millrine: *Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources.*