

Administration and Scoring of Measurement tools for the Three and Four-year Old's and Women before the RCT Baseline Survey in Tharaka Nithi County, Kenya

Samson Oteyo¹, Esther Gachara¹, Eunice Mvungu¹, Helen Henningham², Simon Onywere², Dorothy Niavasha², Judith Waudu³, James Onditi³, Teresa Mwoma⁴, Hudson Ouko⁴, Margret Mwangi⁴

1. Kenyatta university measurement team
2. Principal investigators
3. KU WEE management
4. Kenyatta university curriculum team

Abstract

The Randomized Controlled Trial (RCT) child-based curriculum intervention was identified as a critical Women's Economic Empowerment (WEE) strategy because women engage in unpaid childcare, which affects their livelihood in rural Kenya. To examine the impact of enrolling 3-year-olds in PPI class under the child-based curriculum and learning environment on both children's developmental milestones and women's economic empowerment, there was a need to adapt existing standardised tools to the local Tharaka Nithi context to assess the participant's at baseline, midline, and endline. The tools which were used in measuring the baseline status of the participants included Daberon-2, Early Childhood Development Index (ECDI-2030), and Preschool Self-Regulation Assessment (PSRA). The household survey was used to assess women's economic status. All the measurement tools used on children were translated into four major dialects spoken in Tharaka Nithi County (TNC), and some pictures of animals, objects, and foods were replaced with local ones. Appropriate and inappropriate responses were identified and revised, informing the administration and scoring of the instruments. To ensure consistency in administration and scoring, fifteen (15) assessors proficient in the dialects spoken within the county were recruited and trained through classroom roleplay, demonstrations, supervised field practice, and interrater supervision. Assessors were provided with administration and scoring guides to boost their proficiency. Using pilot data of 42 children aged between 3 to 4 years, the internal and test-retest reliability of the measurement tools used with the children met the threshold of 0.7. A pool of researchers constructed the household survey for WEE. Fifteen assessors with proficiency in the dialects within the TNC were recruited and trained on administration and scoring using didactical learning, demonstrations by the trainers, and role plays to ensure that the assessment was carried out in a standardised way. Assessors were taken through class training and field practice. The cultural adaptation and training of assessors were expected to boost the objectivity of the assessment because the findings were important in informing whether curriculum-based childcare intervention has an effect on women's economic empowerment and eventually inform policy decisions regarding the curriculum for the three-year-olds.

Key Terms: Adaptation, Administration, Scoring, Fidelity, Reliability.

1.0 Background Information on Childcare as a Hindrance to Women's Economic Empowerment

Childcare responsibilities significantly hinder women's economic empowerment, particularly in rural areas. In many low-income regions, such as Tharaka Nithi in rural Kenya, women predominantly shoulder the burden of unpaid childcare, restricting their ability to engage in formal employment or entrepreneurial activities. Studies have shown that women spend an average of three to six hours more per day on unpaid childcare care work than men, greatly impacting their economic participation and financial independence (UN Women, 2019). This disparity limits women's income generation and career advancement opportunities, perpetuating economic dependency and gender inequality (Ferrant, Pesando, & Nowacka, 2019). A recent study found that addressing the childcare burden through accessible, affordable childcare services could increase women's economic participation by over 20% in low-income regions, reinforcing the need for targeted childcare interventions to support gender equality (Kabeer et al., 2020).

In TNC, the lack of affordable and accessible childcare services further exacerbates this issue, leaving women with few alternatives but to remain out of the workforce. The region faces significant challenges, such as limited infrastructure and sociocultural norms that reinforce traditional gender roles, such as caring for young children, making it difficult for women to seek employment outside their homes. Research suggests that rural settings with strong gender-role expectations often see lower female participation in the labour force, as women are expected to fulfil unpaid caregiving roles, which are intensified by a lack of supportive childcare policies (Blau & Kahn, 2021). Consequently, addressing childcare challenges in Tharaka Nithi is crucial for promoting gender equality and enhancing women's economic empowerment. Community-based childcare solutions, government support, and policy interventions are needed to alleviate the childcare burden and enable women's economic participation, thereby contributing to broader development goals. Evidence suggests that community-driven childcare solutions ease women's workload and opportunities for female employment within these centres, fostering local economic growth and challenging restrictive gender norms (Morrissey, 2021).

Against this backdrop, this study titled “Randomized Evaluation of the Effects of a Preschool Intervention on Child Learning and Female Economic Empowerment in Tharaka Nithi County,

Kenya” was formulated. The study has utilised a randomised control trial design to examine the effects of preschool curriculum-based childcare intervention on children’s readiness for learning/developmental milestones and female economic empowerment. The sample size of 120 schools, and 1500 three-year-old children were randomly selected to participate in the study. There were two groups: the experiment group, whose members were exposed to curriculum-based intervention, and the control group, which was not. Assessments were conducted at baseline (November-December 2023) and later at midline (November-December 2024) and end-line (November-December 2025) for both children and mothers. The hypotheses of the study were twofold: by offering a high-quality, tailored educational intervention to improve child development outcomes and school readiness; by reducing the mothers’ childcare burden, they would have time to engage in productive employment or other economically empowering activities.

1.1 Curriculum-Based Childcare Intervention: Content and Rationale

The child preschool curriculum was an enhancement of the curriculum from the Kenya Institute for Curriculum Development (KICD), which focused on 4-year-olds. Additions were borrowed from preschool curricula in India for 3-5 year-olds (Anganwadi Centres). The proposed enhanced curriculum/intervention includes activities focusing especially on executive function. It addresses cognitive concepts such as size, shape, colour, position, difference/similarity and quantity through various play activities connected to children’s daily experiences. Attention is paid to improving teacher-child interactions by encouraging two-way conversations, using open-ended questions and positive feedback. It promotes socio-emotional development by encouraging sharing, taking turns and helping others. It also teaches empathy through stories. There is also daily singing and story time. It introduces dialogic reading methods while emphasising language skills. This aligns with the Sustainable Development Goals, which entrenched early childhood development in 2015. The emphasis was on ensuring equitable access to quality early childhood development and early learning opportunities (UNICEF,2023).

Implementing curriculum-based childcare interventions offers a promising solution to alleviate the childcare burden on women while promoting early childhood development. Recent evidence from randomised controlled trials (RCTs) in various settings, including rural Kenya, indicates that structured childcare programs can significantly enhance women's ability to engage in economic

activities. For instance, a study by Clark et al. (2020) found that such interventions improve children's cognitive and social outcomes and free up women's time, allowing them to pursue employment or entrepreneurial ventures. These programs provide a dual benefit: support children's developmental needs through structured learning environments and enable women to achieve greater economic independence. By investing in curriculum-based childcare, societies can foster more equitable economic growth and enhance women's and children's well-being.

1.2 Methodology of the Study

When well-designed, conducted properly and using a reasonable sample size, a randomised control trial study can provide evidence of any intervention's effectiveness. The following is a brief description of the study's methodology

1.2.1 Design of the Study

This study is focused on a randomised control trial (RCT) evaluation of child curriculum-based care hinged on children's development milestones/learning and women's economic empowerment. The researchers are cognizant of the fact that different children and their mothers are different in many ways that may influence study outcomes. However, by randomly assigning the participants in the experiment (those exposed to innovative curriculum) and control groups (not exposed), RCT may achieve sufficient control over these influences, such that any difference, especially if it falls on experiment members, can be attributed to the intervention.

1.2.2 Determination of Sample Size and Sampling Procedure

The schools were randomly clustered for treatment and control. Within the catchment area (defined as the area within a 3-kilometre radius of the chosen schools) of each treatment school, 10 children who turned 3 in January 2024 and whose families had expressed a desire to be part of the study with full knowledge of the randomisation were chosen. In the intervention community, 5 of these children were randomly selected through a lottery to be admitted to the intervention school, while the remaining 5 children will stay at home. In addition, in each treatment and comparison school, five 4-year-old children were selected for data collection to measure the impact of the new intervention on the children already eligible to attend PP1. Fifteen children in the catchment area of each treatment school were selected: five 3-year-olds randomly selected to enrol in the new

intervention, five 3-year-olds not selected/enrolled in the new intervention, and five 4-year-olds enrolled into existing PP1: 900 children total ($15 \times 60 = 900$). Ten children in the catchment area of each control school: five 3-year-olds who were selected into the comparison group, five 4-year-olds enrolled in existing PP1: 600 children total ($10 \times 60 = 600$). In each household for each child in the treatment and control group (1 adult male and 1 adult female primary caregiver) of all surveyed children: 1500 households ($900 + 600$). Based on previous findings from childcare interventions in low and middle-income countries, the researchers believe that such sample sizes have high calculations to detect reasonable size impacts.

1.2.3 Instrumentation and Data Collection

Data collection will be done at the baseline to assess the status and school readiness, at mid-line and end-line to measure the effectiveness of the intervention. For this paper, the focus is on the instruments that were used for baseline assessment. Baseline data is crucial in providing a clear starting point for the experiment and a basis for evaluation. It serves as a reference point for measuring the progress and impact of the intervention, which in this case is the modified curriculum. Therefore, it is important to ensure that the instruments and the data collection process are valid and reliable. Capturing accurate information before the experiment starts ensures the experiment's effectiveness and accurate data that can be relied upon in informing policy decisions.

The tools that were to be used to measure the baseline status of the participants included Daberon-2, Early Childhood Development Index (ECDI-2030) and Preschool Self-Regulation Assessment (PSRA). The factors that informed the choice of tools were psychometric properties, ever used in Kenyan or Low-income countries, direct and indirect assessment and being child-friendly. Researchers generated the household survey to collect information on mothers and fathers on the following dimensions: background information, bio data of respondents, work and employment information, household decision-making, children, caregiving, use of time as a caregiver and depression status. It was envisaged that there would be three rounds of data collection: baseline (December 2023), midline (October or November 2024) and end line (October or November 2024).

2.0 Adaptations of the child's direct and indirect assessment to make it suitable for the Tharaka Nithi County context

2.1 Introduction

Different populations perform different behaviours in different social and cultural environmental contexts. That means both individual and contextual environmental factors determine behaviour. In the same vein, research has shown that context-specific psychological instruments have higher predictive validity, accuracy and reliability in assessing psychological constructs than general tools (Pomerance, Converse, 2014).

Against this backdrop, there was a need for the instruments developed in a different context and used in a different population to undergo an adaptation process before being used in the RCT project in TNC. The Daberon-2 is a screening tool developed to sample knowledge and skills (body Parts, colour concepts, number concepts, ordinal numbers, prepositions, following directions, general knowledge and visual perception) of children as being ready to begin school (Danzer, Gerber, Lyons, Voress (1991). It is designed for children aged 4 through 6. However, in this study, it was redesigned and domesticated to screen three and four-year-olds in the Kenyan and Tharaka Nithi context. The original tool had 122 items, but the domesticated version had 105 items as some were dropped as they were not relevant in the local context, and some items did not elicit much information related to the study. The test takes around 40 minutes (Danzer, Gerber, Lyons, Voress, 1991).

The Preschool Self-Regulation Assessment (PSRA) measures young children's self-regulation skills, including emotional, attentional and behavioural domains, using a structured battery of tasks (Smith, Raver, Hayes, & Richardson, (2007). The assessor rates the child using a Likert scale of four (0, 1, 2, and 3) to indicate the child's self-regulation skills. The tool was not translated because it was being used by assessors who are well-versed in English. PSRA was used together with Daberon, and the rating of a child was based on a child's behaviour during the assessment period. However, succinct descriptions were given to assist the assessor in rating the child.

Early Childhood Development Index (ECDI 2030) is used to capture development milestones achieved by children as reported by primary caregivers (UNICEF, 2020). It is usually a 20-item questionnaire asking the primary caregiver, probably the mother, about how their children behave in certain situations and the skills and knowledge they have acquired. It assesses children in four

domains: language/literacy, numeracy, physical, socio-emotional and cognitive development. The tool was translated into Kiswahili, assuming that most mothers can understand and speak Kiswahili, a national language in Kenya. According to Daunert & Seel (2020), accurate measurement of a certain variable requires accurate translation of the instrument into the target group's language.

Researchers generated the household survey to collect information on mothers and fathers in the following dimensions: background information, bio data of respondents, work and employment information, household decision-making, children, caregiving, and use of time as caregivers. This did not require much adaptation as it was generated by researchers.

2.2 Adaptation of Instruments

Adaptation is not just the linguistic translation of instruments; it also includes the cultural and psychometric adaptation of instruments to ensure semantic, conceptual, and content equivalence with the original versions of the tools (Harkness, Villar, Edwards, 2010).

2.2.1 Linguistic Adaptation of the Tools

It was observed that dialects spoken in TNC (Kimuthabi, Kichuka, Kitharaka, Kimwembi and Kimbeere) are similar in most linguistic expressions, but there were specific words or items that differ among dialects. In the translated version, all these variations were captured. Daberon-2, which is a child-direct assessment, was translated into local TNC dialects. To ensure good translation, the following steps were followed

- i) A team of translators from five dialects was formed
- ii) The trainer explained the conceptual and content intent of the original version to the trainee and translators
- iii) Both literal and specific translation was done where some action verbs were replaced with ones that had semantic equivalence with local dialects, as shown in Table 2.1

Table 2.1: Replacement of Action Verbs in Daberon-2

Question	Word in the original version	Reason	Word in the adapted version
Q. 4-7	Put the finger-	Using the translation for 'put' is like saying 'putting an object down'. So instead, we say, 'show'	Show me
Q.32-33	Count for me	Changed to clarify the instructions (i.e., how far to count).	Count to ten for me
Q.34-38	Give me	Changed 'give me' to 'count out' because the children simply gave the number of blocks they could hold in their hands instead of counting the blocks.	Count for me
Q.39-42	How many do I have?	To encourage the children to count out the blocks instead of guessing the number of blocks in front of them, the blocks are put out in a line in front of them.	Count for me
Q.44	Give me all the blocks		Put all blocks here

2.2.2 Cultural Adaptation of the Daberon-2

Cultural adaptation is an extension of linguistic translation techniques because it is a product of pretesting data from already translated tools (Devine et al., 2018). The pretest field practice phase resulted in adaptations/ adjustments that were made to ensure that the children understood the questions. They included adapting pictures, instructions, and questions, adding demonstration items for unfamiliar tasks, and adding context-specific examples of correct and incorrect answers. The sample size for the pretest phase was 77 (school =39; Home visit=38). Some objects, names and words were replaced with locally available/common ones, as shown in Table 2.2. This process

ensured that not only the semantic but also the conceptual equivalence of the adapted version met the intent of the original version.

Table 2.2: Replacement of Objects and Names in Daberon-2

Question	Noun word in the original version	Noun
Q.48	Tricycle	Bicycle
Q.49-53	Changed the picture of the children because the last child's leg is missing.	New picture without defect
Q. 85	A stove	Jiko
Q. 76	Gasoline	Car fuel
Q. 79	Fireman	Policeman
	Dentist	Doctor
Q. 90	Coat	Sweater

2.3 Measurement or Psychometric Adaptations of Tools

To ensure psychometric adaptation, three concepts were considered: ensuring validity, estimation of reliability, and development of administration and scoring guideline protocols.

2.3.1 Validity and Estimation of Reliability

Validity and reliability in research are prerequisites to ensure the integrity and quality of data generated by measurement instruments. (Kimberlin & Winterstein, 2008). As part of the baseline preparation of RCT in TNC, researchers ensured validity and reliability. During the pretesting and Pilot phase, some item domains in Daberon were left out either because all children did not respond to them or what they measured was not an interest of the RCT project. Domain items like plural, categorisation and gross motor were omitted in the assessment. A sample size of 42 children aged between 3 and 4 was used in the Daberon, PSRA and ECDI test-retest from October to November 2023. The interval between the test and retest was two weeks, where the same rater assessed the

same child at two different points in time. The results of internal reliability and test-retest intraclass Correlation Coefficient are as follows.

Table 2. 3: Internal Reliability and Test-Retest of Child Assessments

	Internal Reliability	Test Retest ¹ ICC
Daberon*	0.77	0.85
Preschool Self-Regulation Assessment*	0.93	0.75
ECDI*	0.53	0.63

ICC=Intraclass correlation coefficient; α =Cronbach's alpha

*Test-retest over 2 weeks with 42 children aged 3-4 years.

Internal reliability refers to the extent to which items within the same instrument measure various aspects of the same construct. For example, Daberon contains the following domains (general knowledge, body parts, colour concepts, number concepts, ordinal numbers, following directions, expressive vocabulary, and visual perception) to assess the child's readiness to learn. The internal correlation coefficient indicates interconnectedness in items to yield a global score indicative of a child's school/learning readiness. All the tools met Cronbach alpha coefficient of more than 5 which means they were reliable (Cronbach, 1951; Taber, 2018). Daberon and PSRA had the highest internal reliability of more than 0.7 which was commendable. Each participant was assessed by the same rater at two different point times (Between two weeks). The Test-retest Intraclass correlation coefficient (ICC) of Daberon and PSRA was at a good level threshold, as highlighted by Koo and Li (2016). With scores meeting the recommended threshold, ICC coefficients indicated that there was intra-observer and inter-observer consistency in the administration of the tools hence reliable to be used in the baseline.

2.3.2 Development of Administration and Scoring Guidelines Protocol

The protocol guidelines were developed and included general instructions, probing and repeating, practice questions, scoring guidelines, and working with children with challenging behaviours. The development of an administration and scoring guideline protocol is very important not only

in ensuring the accuracy and consistency of data but also in ensuring the safety of subjects in the assessment process.

2.3.2.1 Children's and Assessor's Instructions

The instructions for participants and assessors were contrasted to ensure that data was collected uniformly across the children by different assessors. For example, on the Daberon assessment sheet, **bold type** indicates what the assessor must say to the child out loud. The *Italic type* indicates instructions for assessors, and they were advised not to read these instructions aloud to the child. These instructions were to completely and exactly be followed as written to ensure consistency in administration.

2.3.2.2 Probing and Repeating

Probing provides access to extra detail and depth through verbal prompts to clarify, elaborate, illustrate or explain a prior answer to an interview question that the participant has already given (Robinson, 2023). Guidelines were given on when and how to probe. Too little or too much probing/repeating might lead to biased results toward a correct or incorrect response. A child could have known the answer but not understood the question or responded to the wrong question. A child who receives many opportunities/extra chances to respond may be helped. Probing should be used to clarify the understanding of a child's response. Young children are still developing their verbal abilities, so sometimes they speak very softly or say unclear things. Probing ensures that the assessor understands the child and their intended response. Assessors were trained that questions could be repeated once at a 5-second rule under the following circumstances: the child does not respond or looks confused; the child demonstrates that he/she has misunderstood the question by talking about something irrelevant, for example, questions that had specific items for comparison, or child requests for the question to be repeated. Situations when not to repeat were also emphasised: after the child gives an incorrect response or when the child responds that he/she does not know.

2.3.2.2 Practice Questions

For some questions, there was a need for a demonstration and practice component prior to administering the question. The purpose of this was first to introduce the concept or the skill being assessed to a child and to ensure that he/she understands how to do the task before conducting the actual assessment. For these practice sessions, the assessor is required to indicate the child's understanding on the assessment sheet by circling 'yes' or 'no'.

2.3.2.3 Scoring Guidelines

It was emphasised that the assessor was to score as she administered each question. Also, at the end of the assessment, while still with the child, the assessor was to check the entire scoresheet and ensure that no responses were omitted. All questions were to be scored either a 0 or 1, where 1 represents a "correct response" and signifies that the child demonstrated the skill being assessed. 0 stands for an "incorrect response". That is, 0 means the child does not have or did not demonstrate the skill being assessed. Children have different ways of expressing that they do not know an answer verbally or through nonverbal cues. The assessors were exposed to some of them like: some children use words, while others shrug their shoulders, and yet others stay quiet and do not respond to the question. All these instances are scored as 0. If a child corrects him/herself while still focused on the item or just after the assessor moves on to the next item, the assessor is required to record the response as correct.

2.3.2.4 Working with the Child Throughout the Assessment

There were guidelines on how to work with children throughout the assessment that include: the establishment of a relaxed and playful rapport, arrangement of assessment material for ease transition between activities, consistent, neutral encouragement to the child whether correct and incorrect responses, avoiding leading or hinting cues when the child is responding, practising 5 seconds rule before repeating or proceeding to next question, and being keen on children's non-verbal cues that are indicative of discomfort for example when the child wants a break or wants to use bathroom. Various guidelines were given when interacting with children with delayed speech or challenging behaviours.

2.3.2.5 Training Guide and Fidelity Checklist

A training guide manual was developed to guide how to train assessors to administer instruments consistently and accurately. The training was scheduled to take three weeks, and it included classroom training, field practice, and inter-observer reliability, as elaborated later. As part of quality control, a fidelity checklist was prepared to assist trainers and inter-observers in rating the assessors' competence and proficiency throughout the assessment period. During training, corrective feedback was given to assist the assessor in improving administration and scoring competence and proficiency.

3.0 Training of the Child Assessors

All research data collection and data handling tasks involve personal skills and knowledge that deeply determine data quality, accuracy and study credibility. Training assessors for the current study was, therefore, an important aspect of ensuring that they have the necessary skills and abilities to conduct assessments and interact with the children. According to Ponizovsky-Bergelson, Deyan, and Strier (2019), Interviews with young children may share similarities with adult interviews; however, they pose unique challenges that need to be addressed in training and planning for research. One such challenge is child-adult interaction, which is affected by power-based relations, including children's inhibitions and adults' lack of trust that young children are able to report important information. Being friendly and flexible in various assessment contexts and respectfully persuading individuals who are reluctant to participate in the study are key attributes of an assessor (Harwood, Eileen & Hutchinson, Elizabeth (2009).

Fifteen assessors with proficiency in the dialects within the TNC were recruited. Training was necessary to ensure that the assessment was carried out in a standardised way as possible. Harwood, Eileen, Hutchinson, and Elizabeth (2009) emphasise that some level of training should be considered when there is more than one person involved in carrying out the research project. The approach needed to be consistent among different assessors in the administration of instruments and scoring procedures. The training, therefore, was intended to ensure that the assessment and test procedures were harmonised among different assessors.

3.1 Goals of the Training

The overall purpose of the training was to impart skills to the assessors on how to administer the following assessments: Daberon-2, Preschool Self-Regulation Assessment (PSRA) and the Early Childhood Development Index (ECDI). The specific objectives included:

- a) To introduce the assessor to various assessment tools
- b) The assessor to learn in the class how to administer each assessment with the support of the trainer through demonstration and role play
- c) The assessors practice in the field by administering each assessment tool to children and caregivers with the support of the trainer
- d) Trainers conduct inter-rater reliabilities with each assessor to ensure that they are conducting the assessments with fidelity.

3.2 Training Methods

Mohajan (2017) suggests that various training approaches may be used, as long as it can be demonstrated that reliable and accurate assessor judgments are obtained. The training utilised didactical learning, demonstrations by the trainers, and role plays for the trainees to practice the assessment. There were three groups to allow the trainers and child specialists to focus on individual trainee's needs effectively. Each group was led by a trainer and supported by a child specialist.

3.2.1 Explanation and Demonstrations by the Trainer

The first demonstrations were done among the trainers, where one trainer took the role of a child while another trainer acted as an assessor. The other group members were expected to learn vicariously by observing the assessment process and expected interaction behaviours among children. This was followed by discussions between the trainers and the group members on what they observed, and clarifications/explanations were made. The assessors were divided into three groups of five members: one trainer and a child psychologist. During the demonstration, the following were clearly explained and demonstrated in the small groups: having all the materials needed for assessment before starting, choosing a conducive environment for assessment, sitting arrangement and creating rapport with children using various strategies. Also, in demonstrations, the following administration procedures were emphasised: following the protocol for each task,

transitioning quickly and efficiently between tasks, using an appropriate pace throughout the assessment to keep the child's interest, praising the child with the same tone of voice throughout the assessment, having the child's attention before asking a question, laying out the materials before asking a question and how to hold the clipboard so that the child cannot see the scores.

3.2.2 Role-Playing Method

Cabral, 1987 as cited by Carmel. (2017) describes role play as a valuable technique that has been adopted in academic research. He further states that role-play requires participants to step into an 'as if fiction in which they are not only acquiring knowledge but also making it', practising it and embodying it. It was important to ensure that the assessors acquired accurate information, skills and attitudes before role play. Before role-playing, the assessors were taken through assessment sheet of each instrument and the following were highlighted: general information on top of assessment sheet, the instructions on how to use the sheet (bold type vs italic type), and the instructions on how to score responses, where to score responses (on the line to the left of each question), and how to score the demonstration items (as a 'yes' or 'no'), where to find the list of materials for each activity, where to find the different translations for each question, the special instructions for different parts of the assessment among others. Then, the assessors were taken through the administration and scoring guide to have clear guidelines on administering and scoring the assessment with each child and ensure that everyone will conduct assessments the same way, regardless of the child being assessed. The assessors were also taken through appropriate and inappropriate responses as given in the scoring guide.

Based on vicarious learning from demonstrations and explanations of assessment sheets and guides, the assessor role played as a child and an assessor among themselves with the support of the trainers. Each person practised the administration of the various parts of assessment tools while others and trainers used fidelity checklists to check each person's skills. The checklist had the protocols to be followed when administering the assessment. Intermittently, the trainer will pause role play to clarify certain concepts, make corrective feedback and assist play assessors practice again in areas they erred. Then assessors worked in pairs to role practice (child and assessor interchangeably) as trainers moved in between pairs to check on the progress in practice. Practising in small groups is motivating since the inhibitions are lowered, and it encourages assessors, even

the socially inhibited, to speak, ensuring all assessors have equal chances to practice (Abdessallam, Abdessallam & Azize, 2023).

The assessors also practised how to respond to common challenging child behaviours that occur during assessments. Young children are still in the process of physical and cognitive development. Therefore, they are likely to get tired easily, become fussy, and lose interest in the assessment process. This is likely to pose challenges to the assessors during the assessment. It is important for assessors to use techniques that would facilitate a child's participation. Coyne, Mallon, & Chubb (2021) state that when children are given space, opportunity and means, and with skilled facilitation, they can clearly share their perspectives and meanings. To learn how to handle various challenges that occur during the assessment of children, demonstrations were made by the trainers to the whole group on how to handle different challenging scenarios. Afterwards, the assessors' roles played the same scenarios, with each assessor playing both roles of an assessor and a child. The other assessors observed and gave feedback while the trainer and child specialist gave support on adherence to the protocols. The effectiveness of role plays as a training method was demonstrated by Mamaghan, Hosseinian, and Maghsoodi (2024), who observed that role-play led to enhanced acquisition of skills when compared with traditional methods and an increase in self-esteem and practical skills in assessment among nursing students.

3.2.3 Field Practice Technique

In the field practice, each assessor practised by conducting at least one (1) complete assessment of a child and the caregiver's view of the child on various developmental milestones per day with the support of a trainer. Each assessor conducted at least five (5) assessments by the end of the week. Also, assessors practiced on scoring of PSRA. Throughout the practice, trainers supported assessors to administer assessments as per the protocols. In between the assessments, the trainer could pause the assessment to discuss any problem to prevent the assessor from practising wrongly.

3.2.4 Inter-Rater Reliabilities Technique

In this stage, the trainer scored along with the assessor as they administered the assessment. The aim was to conduct five weekly inter-rater reliabilities assessments per assessor. Each trainer was to work with five assessors. This was done to ensure scoring fidelity as the trainers compared their

scores with those of the assessors. This is a common practice in tool development as it helps the assessors to align their judgement and scoring criteria (Hodwitz, Kuper, and Brydges 2019).

4.0 Conclusions

Like any RCT intervention, the findings of this study will play a major role in decision-making about the appropriate age when a child needs to begin school, whether it is 3 or 4 years old. This is based on its twofold outcomes: Improved children's readiness to learn and women's economic empowerment resulting from the implementation of curriculum-based intervention. For this to happen, well-designed RCT needs to minimise sources of error, including a data quality plan, which is the focus of this paper. Adaptation of standardised instruments through pretest, pilot testing and training was achieved, and the following were the conclusions

- i) All standardised instruments were linguistically and culturally adapted to ensure local equivalence of the original version in terms of content and construct intent.
- ii) All modified versions of standardised instruments meet the reliability threshold hence they could provide consistent scores
- iii) All assessors went through in-depth classroom training and field practice to ensure their proficiency and competence in the administration and scoring of the instruments
- iv) All assessors were fluent in local dialects spoken in Tharaka Nithi, and instruments were translated and back-translated to avoid bias
- v) All trainers had administration and scoring guides to ensure consistent administration and scoring procedure
- vi) There were Inter-observation raters for three groups as part of quality checks, and corrective feedback was given to assessors and data collection firms to boost data quality
- vii) There were random spot checks from the Ku measurement team to ensure fidelity to administration and scoring protocol

5.0 Recommendations

The researchers are cognizant of the fact that linguistic, culturally adapted and psychometrically sound instruments will boost the objective assessment of research participants against the backdrop

of well-trained assessors on administration and scoring protocols to inform whether curriculum-based child care intervention affects women's economic empowerment. There was a need, as shown in the paper, to

- i) Document the processes of pre-pilot and pilot for future researchers to replicate the process
- ii) To elaborate that adaptation of instruments is a continuous field process as experience has shown the instruments require continuous revision of items, administration and scoring as researchers interact with diverse participants.

References

- Axford, S. N. (1992). Review of the DABERON-2: Screening for School Readiness. In J. J. Kramer & J. C. Conoley (Eds.). *The eleventh mental measurement yearbook* (pp. 256–259). Lincoln, NE: The Buros Institute of Mental Measurements.
- Carmel. O. (2017). Role-play and research. 10.4324/9781315456539-30.
- Clark, S., Kabiru, C. W., Laszlo, S., & Muthuri, S. (2020). The Impact of Childcare on Poor Rural Women's Economic Empowerment in Kenya. *World Development*, 136, 105146.
- Coyne, I., Mallon, D.& Chubb, E.(2021). Research with young children: Exploring the methodological advantages and challenges of using hand puppets and draw and tell.*The International Journal of Childhood and Children's Care*.<https://doi.org/10.1111/chso.12452>
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334. doi:[10.1007/bf02310555](https://doi.org/10.1007/bf02310555)
- Danzer, V.A., Gerber, M. F., Lyons, T. M., Voress J. K. (1991). *Daberon-2: Screening for School readiness*. Pro-ed, Texas
- Daunert &Seel (2020). *Translating and Adapting Survey Instruments: Important Steps That Will Define Quality and Optimize Validity of Adapted Research Instrument*.

Devine J., Klasen F., Moon J., Herdman M., Hurtado M.P., Castillo G., Haller A.C., Correia H., Forrest C.B., Ravens-Sieberer U. (2018). Translation and cross-cultural adaptation of eight pediatric PROMIS® item banks into Spanish and German. *Qual. Life Res.*; 27:2415–2430. doi: 10.1007/s11136-018-1874-8.

Ferrant, G., Pesando, L. M., & Nowacka, K. (2019). Unpaid Care Work: The missing link in the analysis of gender gaps in labour outcomes. OECD Development Centre. Retrieved from OECD iLibrary.

Hodwitz, K., Kuper, A., & Brydges, R. (2019). Realising One's Own Subjectivity: Assessors' Perceptions of the Influence of Training on Their Conduct of Workplace-Based Assessments. *Academic Medicine: Journal of the Association of American Medical Colleges*, 94(12), 1970–1979.

<https://doi.org/10.1002/9780470609927.ch7>. Harkness J.A., Villar A., Edwards B. (2010). Translation, adaptation, and design. *Surv. Methods Multinat. Multiregional Multicult. Context*. 2010:115– 140. doi: 10.1002/9780470609927.ch7.

Khamouja, Abdessallam & Benmhamed, Abdessallam & el Ghouati, Azize. (2023). The Importance of Role-Playing Activities in Developing Students' Speaking Competence. Vol. 66. 225-230.

Kimberlin, C. L., & Winterstein, A. G. (2008). Validity and Reliability of Measurement Instruments Used in Research. *American Journal of Health-System Pharmacists*, 65(1), 2276- 2284.

Koo T.K, Li M.Y (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*.15(2)155 doi:10.1016/j.jcm.2016.02.012. PMC 4913118. PMID 27330520

Mamaghan, E.A., Hosseinian, E. & Maghsoodi, E. (2024) Role Playing is an Effective Method for Training Physical Examinations: A Mixed-methods Study. <https://doi.org/10.1016/j.teln.2024.06.005>

- Mohajan, H. (2017). *Two Criteria for Good Measurements in Research: Validity and Reliability*. Published in: *Annals of Spiru Haret University*, Vol. 17, No. 4 (24 December 2017): pp. 56–82.
- Ponizovsky-Bergelson, Deyan & Strier (2019). A qualitative Interview with young children: What Encourages or inhibits young children’s participation? *International Journal of Qualitative Methods*. <https://doi.org/10.1177/1609406919840516>
- Pomerance M.H., Converse P.D. (2014). Investigating context specificity, self-schema characteristics, and personality test validity. *Personal. Individ. Differ.* 2014; 58:54–59. doi: 10.1016/j.paid.2013.10.005.
- Robinson.O.C.(2023).Probing in qualitative research interviews: Theory and practice.*Qualitative Research in Psychology*,20(3), 382-397.<https://doi.org/10.1080/1478088.2023.223862>
- Smith-Donald, R., Raver, C. C., Hayes, T., & Richardson, B. (2007). *Preschool Self-Regulation Assessment (PSRA)* [Database record]. APA PsycTest.<https://doi.org/10.1037/t71117-00>
- Taber, K. S. (2018). The use of Cronbach’s alpha when developing and reporting research instruments in science education. *Research in Science Education*, 48(6), 1273–1296. <https://doi.org/10.1007/s11165-016-9602-2>
- United Nations International Children's Emergency Fund (2023). Early Childhood Development Index 2030
- United Nations International Children's Emergency Fund (2020). Early childhood development index
- UN Women. (2019). Progress of the World’s Women 2019-2020: Families in a Changing World. Retrieved from UN Women website.