Collaboration on endogenous development and institutionalization of mathematics curriculum in Afirica

Project Period (FY) : 2021-10-07 – 2026-03-31

Final Report on Research Results

March 2026

Prof. Takuya Baba, Hiroshima University

Dr. Bentry Nkhata, University of Zambia

Table of Contents

List of Tables

List of Figures

Preface

In 2003 I began my involvement with the Zambia Special Education Program (hereafter referred to as "ZAMPRO") by supervising two ZAMPRO students. More than twenty years have passed since then. The first student I encountered through this program is now over fifty years old. At that time, I was confronted with a level of low academic achievement among Zambian students that I had never experienced before.
In Japan, it is not uncommon to see errors such as calculating 1/2＋1/3 by separately adding numerators and denominators to obtain 2/5. However, what I observed in Zambia went further: the calculation was reduced to 1+1+2+3, leaving no trace of fractions at all.

Sometime later, a JICA technical cooperation project in the field of mathematics and science education was launched in Zambia. ZAMPRO students began to collaborate with the project, and faculty members from Hiroshima University also gradually became involved. Through these activities, Zambian researchers in mathematics and science education, education officers in the ministry, teacher educators, and school teachers were accepted as trainees and international students in Hiroshima University. The JICA technical cooperation projects at that time placed particular emphasis on the professional development of teachers, and issues related to teacher capacity were a central concern.

Subsequently, there were several turning points in my engagement with mathematics and science education in Zambia. One was related to curriculum reform, corresponding to what is called the Course of Study in Japan. Another was closely connected to the issue of low academic achievement, namely, children's learning processes. Regarding the former, I was invited to a workshop in 2013 when the Republic of Zambia undertook a curriculum revision. Later, from 2012 to 2015, a JICA country-focused training program was conducted at Hiroshima University to strengthen the capacities of staff from the Curriculum Development Centre (CDC), which is responsible for curriculum development in Zambia. Through these engagements, I visited the Zambian Ministry of Education in 2024 as part of the JICA Project Research "Comparative Analysis of Mathematics and Science Curricula in Primary and Secondary Education in the African Region," during which the quality of the curriculum developed in 2014 was appreciated highly by the officials in the Ministry.

Regarding the issue of low academic achievement, although it was identified at an early stage and led to the doctoral dissertation by Toyomi Uchida, *A Study on Diagnostic Assessment of Computational Ability in Basic Education in Zambia: Focusing on Discriminability and Instructional Implications* (2009), systematic research efforts were not pursued immediately thereafter. One reason is that investigating low academic achievement requires surveys at the lower primary level, which in turn necessitates a deep understanding of local languages and cultures—an endeavor with clear limitations for Japanese researchers alone. Another reason is that low academic achievement manifests under diverse conditions and collecting only a

small number of individual cases makes it difficult to determine whether findings reflect incidental occurrences or broader patterns among Zambian students. International educational assessments (such as SACMEQ and PISA-D) and national assessments have indicated extremely low achievement levels. Therefore, understanding the actual situation more precisely requires systematic collection of data from a sufficient number of cases in order to identify broader patterns.

Under these circumstances, I undertook the JICA Zambia Project Research "Analysis of Primary Mathematics Tasks" from November 2017 to 2021 (extended by one year due to the COVID-19 pandemic). For this project, we formed a team which consists of long-standing collaborators from the University of Zambia, their master's students, Hiroshima University graduates of ZAMPRO, and alumni who had obtained doctoral degrees in mathematics education and possessed extensive knowledge of Africa and especially Zambia. As a result, the team had not only a strong understanding of local languages and cultures but also solid expertise in mathematics education. Through JICA, we requested cooperation from the Zambian Ministry of Education, and ten primary schools in Lusaka District, Lusaka Province, meeting conditions of comparability, participated in the survey.

Based on these outcomes, the JICA Project Research Report *JICA Zambia Project Research: Analysis of Primary Mathematics Tasks* (2021) was produced. While the results were confirmed, remaining issues were identified, including the need to expand the range of numbers addressed and to clarify the construction of meaning in multiplication and division.

Building on these findings, the present study was conducted with support from the Fund for the Promotion of Joint International Research (Strengthening International Collaborative Research (B)), entitled "International Collaborative Research on Endogenous Development and Institutionalization of Mathematics Curricula in Africa" (2021–2026). This research is namely a continuation of the previous study. As before, it addresses the issue of low academic achievement; however, the mathematical content examined was expanded to include larger numbers (up to 1,000) and decimals (up to the first decimal place), with particular emphasis placed on multiplication and division. Apart from the mathematical content, the research framework remained the same: the Zambian and Japanese teams consisted of the same members, and the research was conducted in the same ten primary schools in Lusaka District, Lusaka Province.

As outlined above, this report represents the culmination of my engagement with educational development in Zambia since I first became involved at Hiroshima University in 2003. It is my hope that this work will also serve as a foundation for future initiatives.

January 2026
Takuya Baba

Chapter 1 Project Overview

1.1 Extreme Learning Poverty and Approaches to Addressing It

At the World Conference on Education for All held in Jomtien, Thailand, in 1990, the international community shared a common concern that a large number of children around the world were deprived of educational opportunities and unable to attend school. As an outcome of the conference, the *World Declaration on Education for All* was adopted, expressing a collective commitment to addressing this challenge. Subsequently, in 2000, the *World Education Forum* on Education for All was held in Dakar, Senegal, where the international community reaffirmed the importance of universal primary education and emphasized the setting of deadlines and the importance of monitoring progress (Ogawa et al., 2005). In the same year, the Millennium Development Goals (MDGs), which covered development issues more broadly, were also adopted, and the expansion of primary education was positioned as one of their key goals. As a result, substantial quantitative improvements in education, particularly in enrollment, were achieved.

Amid these developments, the World Bank introduced a new perspective on educational challenges. By combining *schooling deprivation*—children who do not attend school—and *learning deprivation*—children who attend school but do not acquire the minimum level of proficiency set by national standards—the World Bank proposed the concept of *learning poverty* and introduced the Learning Poverty Index. This framework highlighted that learning deprivation is the more serious issue and that ensuring school attendance alone, although essential, represents only a first step toward resolving the problem.

Table 1. Learning Poverty in Selected Countries in Asia and Africa

Region	Country	Learning Poverty	Learning Deprivation	Schooling Deprivation
Asia	Bangladesh	58.2	56.0	5.0
	Cambodia	90.0	89.0	9.3
	Indonesia	52.8	49.4	6.8
	Thailand	23.4	21.9	1.9
Africa	Ethiopia	90.4	88.7	14.8
	Senegal	68.6	58.9	23.5
	Uganda	81.9	81.1	4.4
	Zambia	98.5	98.2	14.9

Source: World Bank (2022)

Note: $LP = SD + [(1 - SD) \times LD]$

As shown in Table 1, learning poverty is more severe in Sub-Saharan Africa than in Asia. In countries where learning poverty is extremely serious, various sociocultural factors exert strong influence, including the fragility of domestic institutions and expertise, the impact of language of instruction on learning, insufficient teacher capacity, inadequate or insufficient teaching materials, and constraints on prioritizing educational issues. It is difficult to use approaches developed in industrialized countries to adequately address these challenges without fully accounting for such sociocultural contexts.

Accordingly, this study adopts an evidence-based endogenous approach (Baba & Nkhata, in press), which seeks solutions that are appropriate for Zambian children, with particular attention to expertise, language of instruction, teacher capacity, and teaching materials. In other words, the approach aims for developing countries themselves such as Zambia to identify the essence of extremely low achievement in mathematics education and to examine and improve practices based on empirical data (Baba, 2025).

1.2 Review of Phase I

This section is based on JICA & Hiroshima University (2021) and Baba et al. (2019). Since 2000, Zambia has achieved substantial improvements in the quantity of education, particularly in enrollment rates, through the efforts of the Ministry of Education. As shown in Table 1, the schooling deprivation rate is 14.9, which implies a net enrollment rate of 85.1%. However, the learning deprivation rate is as high as 98.2%, indicating that 98.2% of Zambian children who attend school do not reach the minimum proficiency level set by the government.

In addition, national, regional, and international educational assessments—including the National Assessment Survey (NAS) (MoGE, 2017), the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) (Awich,2021; Hungi et al.), and PISA (OECD, 2019)—have pointed out extremely serious problems in the quality of mathematics education. Previous studies conducted in Zambia (Nakawa, 2016; Uchida, 2009) have also revealed that children predominantly rely on counting strategies in computation.

In light of this situation, Phase I of the project (2017–2021) conducted field studies as outlined below, with the aim of developing diagnostic assessment tools for basic arithmetic abilities of Zambian children, as well as intervention tools to improve those abilities.

[Overview of the Study]

- **Purpose**: To develop and refine assessment and intervention tools for basic arithmetic abilities of Zambian children
- **Period**: 2017–2021

    - Preparatory phase: 2018–2020
    - Final phase: 2021
- **Methods**: Interviews and written tests
- **Target area**: Lusaka City, Lusaka Province, Zambia

Target schools: For research convenience, Lusaka City was selected, and within the city, two schools with similar socioeconomic characteristics were recommended by the Ministry of Education from each of five zones. The schools were as follows:

- Chipata Primary School
- New Kabanana Primary School
- Muchinga Primary School
- George Central Primary School
- Chamba Valley Primary School
- Kaunda Square Primary School
- Twashuka Primary School
- John Laing Primary School
- New Kamulanga Primary School
- Chimwemwe Primary School

Target grades: Grades 1 to 4

Definition of mathematical competence:
Understanding of the following:

- Natural numbers up to 100
- The decimal number system
- Four basic operations (addition, subtraction, multiplication, division)

In addition:

- Ability to represent and explain their thoughts mathematically

[Research Process]

To appropriately assess the abilities of Zambian children, test items must be valid in terms of the following points related to content validity:

children's actual learning conditions
the perspective of mathematical development and curriculum
Zambian and Japanese experts' interpretation and judgement

It was necessary to develop test items based on collected data and experts' views to ensure such points. Therefore, a design experiment methodology (Gravemeijer & Cobb, 2006) was adopted. Diagnostic assessment items were initially developed, then repeatedly administered and revised to enhance validity. In practice, preliminary surveys were conducted four times—in March and September 2018, and March and September 2019.

Based on literature and curriculum analysis, initial items were developed and piloted during the preparatory phase. Children's responses were examined to refine the overall number of items, wording, validity of items, and assessment rubrics. Interview-based assessments were conducted one-on-one, allowing detailed capture of children's thinking processes.

After interview judgments stabilized, group-administered written tests were developed to increase sample size. Although a full-scale survey planned for March 2020 could not be conducted due to the COVID-19 pandemic, an alternative hybrid method was developed, with the Japanese team participating online. A pre-test was conducted in March 2021 and a post-test in June 2021.

During the interval between the pre- and post-tests, intervention activities using intervention tools were conducted in five experimental schools. The effectiveness of the intervention was examined through comparison with control schools. The intervention tools were designed to correspond item-by-item with the assessment tools and included student workbooks and instructional guidelines to address the difficulties identified through assessment.

Table 2. Developed Tools

Category	**Contents**
Assessment tools	Interview tasks, assessment rubrics
	Written tests
Intervention tools	Pupils' workbooks
	Teachers' Guide

[Findings and Remaining Issues of Phase I study]

Interview-based assessments revealed two main findings. First, children in the experimental group shifted from counting objects one by one to recognizing quantities in groups. Second, all children in the experimental group showed improvement, along with reduced response time, with differences much larger than those observed in the control group.

Written test results also revealed two main findings. First, statistically significant differences were found between the experimental and control groups in tasks related to recognizing

numbers as groups. Second, no statistically significant differences were found in computation tasks. This was likely because all tasks involved numbers up to two digits, allowing children to still rely on counting strategies.

Two major issues remained. The first issue concerned multiplication and division. Although overall effects were observed in the experimental group, little effect was found for multiplication and division. The "ten-frame" (Figure 1) was introduced as a key intervention tool for recognizing groups, but it proved insufficient for addressing multiplication and division.

Figure 1. Ten-frame

The second issue concerned understanding place-value relationships after recognizing grouped quantities. While the ten-frame enabled children to recognize groups of ten, this represents only the first step toward "structural understanding of number." If the ten-frame becomes a fixed way of viewing numbers, it risks becoming another form of counting. It is necessary to understand number structure across place values—units, tens, hundreds, and thousands (1, 10, 100, 1000).

1.3 Purpose and Methods of the Phase II Study

In Zambia, the National Assessment Survey is the only nationwide test that assesses arithmetic ability, while SACMEQ serves a similar role at the regional level. As described above, these assessments reveal extremely low achievement among Zambian children. However, to identify and make use of latent abilities hidden within severe learning poverty, diagnostic assessment is required (Uchida, 2009). Therefore, as in Phase I, an interview-based diagnostic approach was adopted.

Based on the issues identified in Phase I, Phase II focused on multiplication and division—particularly multiplicative reasoning (Simon, 1995; Vergnaud, 1983, 2009)—to clarify children's understanding of the meaning of operations. In addition, with regard to place-value system, emphasis was on structural understanding of the places of 0.1s, 1s, 10s, and 100s, and their interrelationships, to clarify children's understanding of number and the meaning of representation and calculation. Since multiplication, division, and decimals were the main mathematical concepts, the target grades were set as Grades 3 to 6.

The research methodology followed the same design experiment approach (Gravemeijer & Cobb, 2006) as in Phase I. Three pilot studies were conducted during the development stage to refine assessment and intervention tools. The same ten schools as in Phase I were chosen. In the final stage, pre- and post-tests using assessment tools (interviews and written tests) were conducted in all ten schools. Intervention activities using workbooks and instructional guidelines were conducted in only five experimental schools between the pre- and post-tests to examine differences from control schools and to evaluate effectiveness of intervention.

- **Purpose**: To develop and refine assessment and intervention tools for basic arithmetic competence of Zambian children
- **Period**: 2022–2025
- **Methods**: Same as Phase I
- **Target area**: Same as Phase I
- **Target schools**: Same as Phase I
- **Target grades**: Grades 3 to 6

Definition of basic arithmetic competence:

- Place values of 0.1s, 1s, 10s, and 100s, for natural numbers and decimals
- Relationships among these place values
- Four basic operations (with particular emphasis on multiplication and division)
- Ability to represent ideas mathematically

1.4 Research Team Structure and Management

To adopt an evidence-based endogenous approach (Baba, 2022) in a country facing extreme learning poverty, it is essential to understand local sociocultural contexts and the realities and challenges of mathematics education. The Japanese team consisted of five core members (Table 3), all of whom had participated in the Zambia Special Education Program[1] or international cooperation practices, possessed deep knowledge of Zambian education and international cooperation, and held doctoral degrees in mathematics education. Two graduate students also participated, making significant contributions while conducting field research, thereby bringing valuable insights to the project.

The Zambian team, excluding the leader Nkhata, comprised six members (Table 4) selected from current and former master's students at the University of Zambia during Phase I. They

[1] This is an educational program for developing international cooperation personnel, jointly implemented by Hiroshima University and the Japan International Cooperation Agency (JICA) Overseas Cooperation Volunteers Secretariat since 2002. Master's degree students are dispatched to Zambia for two years as Overseas Cooperation Volunteers, utilizing their experiences there for their master's thesis.

demonstrated strong competence and motivation in mathematics education, a deep understanding of the assessment approach, and a thorough grasp of the learning realities of Zambian children. Their continued participation after completing their degrees, with the understanding of the Ministry of Education and their respective institutions, was indispensable to the success of this study.

One coordinator was appointed on both the Japanese and Zambian sides to manage communication within the team, coordinate with the Ministry of Education and participating schools, and arrange logistics such as transportation and accommodation. Although two temporary assistants were hired during the final pre- and post-tests in Phase II, the core team remained unchanged from Phase I, enabling smooth and effective project implementation.

Table 3. Japanese Research Team

Name	Affiliation	Role
Takuya **Baba**	Hiroshima University	Principal Investigator / Team Leader
Nagisa **Nakawa**	Kanto Gakuin University	Research Member
Satoshi **Kusaka**	Naruto University of Education	Research Member
Koji **Watanabe**	Miyazaki International University	Research Member
Masato **Kosaka**	University of Fukui	Research Member
Kanae **Minakoshi**	Hiroshima University	Project Coordination
Natsuko **Furukawa**	Hiroshima University	Report Preparation
Yoshitaka **Abe**	Graduate Student, Hiroshima University	Research Assistant
Shiori **Yuda**	Fukushima Prefecture (then Graduate Student)	Research Assistant

Table 4. Zambian Research Team

Name	Affiliation	Role
Bentry **Nkhata**	University of Zambia	Team Leader
Patricia Nalube	University of Zambia	
Shelly Sikwale	Ministry of Education	
Arthur **Mungalu**	Charles Lwanga College of Education	Research Member
Barbara **Mudenda**	Yolo Secondary School	Research Member
Emmanuel **Kaabo**	David Livingstone College of Education	Research Member
Bareford **Mambwe**	Mukando Secondary School	Research Member
Doye **Chikola**	St. Mary's College of Education	Research Member
Spiwe **Tafeni**	Mukushi Day Secondary School	Research Member
Jane Njovu **Takuzwa**	—	Project Coordination

1.5 Target Schools and Participants

With permission and cooperation from the Zambian Ministry of Education, continuous research was conducted at the same ten public primary schools in Lusaka District, Lusaka Province, that had participated in Phase I. In Phase II, a liaison officer (Ms. Shelly Sikwale) was appointed by the Ministry to coordinate with the research team.

From each school, two students per grade (Grades 3 to 6) were selected for interviews: one high-achieving and one mid-achieving student, selected by the classroom teacher (Table 5). Written tests were administered to 20 students per grade per school, including interview participants. Grades 3 and 4 (40 students total) and Grades 5 and 6 (40 students total) were tested jointly (Table 6).

Table 5. Interview Survey Plan (Target Schools and Number of Participants)

		Development Stage			Final Stage	
		2022	2023		2024	
		Sept.	March	Sept.	March	June
		Round 1	Round 2	Round 3	Pre-test	Post-test
Number of schools		5	5	5	10	10
Number of students	Grade 3	10	10	10	20	20
	Grade 4	10	10	10	20	20
	Grade 5	10	10	10	20	20
	Grade 6	10	10	10	20	20

Table 6. Written Test Implementation Plan (Target Schools and Number of Students)

		Final Stage	
		2024	
		March	June
		Pre-test	Post-test
Number of schools		10	10
Number of students	Grade 3	100	100
	Grade 4	100	100
	Grade 5	100	100
	Grade 6	100	100

Chapter 2 Research Procedures

2.1 Development stage

2.1.1 Overview

In Phase II, three rounds of surveys were conducted at the development stage, following procedures similar to those in Phase I. In each round, we employed interviews to observe pupils' problem-solving processes and mathematical thinking from multiple perspectives—by both Japanese and Zambian team members—and engaged in repeated discussions on interpretation and evaluation of the processes. Because subjective elements inevitably enter into the interpretation of observational data, we verified the validity of interpretations through team discussions. Incorporating multiple perspectives (across persons and cultures) enabled a multifaceted view and led to more objective and valid evaluations. We also placed emphasis on ensuring consistency across tasks.

All interviews were video-recorded. The recordings were used in conjunction with observers' on-site notes and impressions for discussion, and were revisited when analyzing specific pupils' responses in detail or when opinions diverged within the team, thereby enabling reliable evaluations that did not depend on a one-time or single-person interpretation.

Objectivity: Agreement in observation and interpretation across multiple observers, cultures, and occasions.
Validity: Alignment between the intended constructs of tasks developed with reference to the curriculum and prior research, and the data obtained in interviews.
Reliability: Stability of measurement results across different implementers and timings.
Consistency: Agreement of interpretations across consecutive tasks.

Tasks and Reaction Levels.
In problem-solving interviews, the tasks used are crucial. We developed tasks that decomposed pupils' basic arithmetic competence into finer elements. Interviews were conducted individually, and observers evaluated how pupils responded to each task. Pupils' responses were rated on five levels (details in **2.2** "Refinement of the Developed Tools" and **2.1.3** "Structure of the Interview Guide"). In the first preliminary survey, the judgment criteria did not always match pupils' actual states, and observers sometimes diverged; criteria also varied across tasks. Through discussion, distinctions between levels were clarified, inter-observer differences diminished, and common criteria applicable across tasks were established.

The surveys followed the procedures below:

[Step 1] We analyzed prior research and earlier surveys to enhance content validity.

[Step 2] We developed or revised the evaluation tools—tasks and reaction levels—based on data from three preliminary surveys, debating levels and criteria while judging levels against the criteria; we also considered response time and measurement error.

[Step 3] We reduced the number of tasks to fit an appropriate response time (approx. 30–40 minutes per pupil) and to suppress measurement error, and revised criteria wherever ambiguity caused unstable judgments. Interviews were conducted by an interviewer, an observer, a camera operator, and Japanese members (who sometimes joined from Japan via Zoom) as shown in Figure 2.

Figure 2. Scene from the field survey

With the minimum number of local staff ensured, the camera operator also served as "Observer 2." Observation and discussion by multiple members, including Japanese staff, enhanced objectivity and consistency.

2.1.2 Refinement of the Developed Tools

The evaluation tools developed for this study decomposed pupils' basic arithmetic competence into finer elements and implemented tasks corresponding to each **sub-competence**. This allowed a detailed grasp and staged evaluation of pupils' abilities. Rather than merely judging correctness, we aimed to analyze understanding and sources of error more precisely by examining pupils' thinking processes and the backgrounds of incorrect responses.

To improve the content validity and reliability of the interview guide and test, we iteratively piloted and refined them in schools. As a result, the tools were refined along three principal dimensions—**structure**, **interview methods**, and **response levels**. The definition of **basic arithmetic competence** (reiterated) comprised: (i) places of 0.1s, 1s, 10s, 100s and natural/decimal numbers; (ii) relations among these places; and (iii) the four operations (addition, subtraction, multiplication, division).

A) Refining task content and structure.

To appropriately measure each sub-competence, we selected and refined tasks along the following **three focal areas**:

1. **Place Value and Magnitude Comparison**
    - To evaluate how well pupils understand the structure of numbers, tasks targeted at **relations among places**—**100s, 10s, 1s, and 0.1s**—and required comparing magnitudes at these places.
    - These items examine whether pupils can relate units across places (e.g., understanding that ten 0.1 make 1; ten 10 make 100) and make magnitude judgments accordingly.
2. **Representation such as Tiles and Grids**
    - We incorporated problems that use multiple diagrammatic representations to assess pupils' capacity to visually grasp quantity and relations**.**
    - Tasks varied the unit quantity assigned to a single grid (e.g., "if one whole grid represents 100" or "represents 1"), prompted pupils to **represent various size of numbers with tiles**, and examined the **depth of conceptual and relational understanding** behind these representations.
    - Because numbers can extend beyond 100 or involve decimals such as 0.1, we explicitly used **hypothetical framings** (e.g., "**if** one large box represents 100") to surface pupils' conceptual flexibility.
3. **Use of the Number Line**
    - Recognizing that diverse representations ultimately converge on the continuity of numbers as expressed on a number line**,** we included tasks that require pupils to locate positions and compare values on number lines. This is because the continuity of numbers is important characteristics to acquire by the end of primary and secondary education.
    - We varied tick-marking schemes to help pupils grasp that **numbers exist even between ticks**, where there are no ticks, thereby assessing whether they can reason about interval structure**.**

Through these three focal areas, we enabled detailed analysis of how pupils conceptualize and operate on numbers across discrete (tiles, grids) and continuous (number line) representations, and across both natural and decimal number systems so that natural numbers and decimals numbers are to be connected in future

B) Refining interview methods in the local language.

To unify procedures and strengthen cross-task consistency, we refined the interview methods along the following three measures:

1. **Standardization of Question Content and Wording**
    - We compiled a detailed interview guide that standardizes **key terms** and **question phrasing** so that wording differences across interviewers (person) and items (item) would not influence responses. The wording was carefully calibrated to avoid hinting at answers or leading pupils in a particular direction, thereby preserving the diagnostic value of each task. This standardization ensured that task intent and content were consistent among persons and items and that data remained comparable across sessions and schools.
2. **Strengthening Training and Procedural Alignment among Interviewers**
    - To minimize interviewer-dependent variability, we conducted **role-play rehearsals** in advance, confirming the step-by-step procedure (instruction, timing, acceptable prompts). We also discussed **appropriate follow-up questions** responsive to typical pupil reactions (e.g., requests for clarification without hinting at solution methods). These shared protocols improved inter-rater reliability and stabilized the administration across sites and days.
3. **Refinement of Expressions in the Local Language**
    - We reviewed and adjusted the **local-language wording** so that pupils could readily grasp the mathematical intent, especially for concepts such as number lines and grids. Terminology and sentence patterns were tuned for **clarity and non-ambiguity**, ensuring that pupils' understanding of tasks reflected their mathematical competence rather than their linguistic proficiency. This linguistic refinement helped align the intended construct with the observed response.

Result. Through these three measures, we enhanced procedural consistency**,** reduced measurement error attributable to language and administration, and supported objective, comparable interpretation of pupils' responses across tasks and grades.

C) Sharpening criteria for reaction levels.

To evaluate pupils' thinking more precisely—beyond mere correctness—we refined the **response-level criteria** along the following three enhancements:

1. **Recognition of Errors and Self-Correction**
    - We incorporated into the criteria whether, while explaining their solution or responding to follow-up prompts, pupils **notice their own errors** and make **self-corrections**.

- When a pupil initially gives an incorrect response but, through explanation or minimal, non-leading prompts, revises the method and answer, that process is reflected in the level judgment (e.g., distinguished from a fluent, explanation-free correctness). This acknowledges diagnostic evidence about how understanding is formed, not only what answer is produced.

2. **Explicit Criteria for Justification (Levels 4 and 5)**
   - We clarified that Level 4 denotes smoothly correct representations/answers without sufficient justification, whereas Level 5 requires smooth correctness with explanation—for example, using concrete or schematic representations (*tiles, grid, number line*) to justify *why* the procedure works or the value is correct.
   - By making the **presence and quality of explanation** explicit, the criteria separate pupils who can execute procedures from those who can articulate underlying relations/structures (e.g., place-value composition, unit relations, interval reasoning on a number line).
3. **Finer-Grained, Objective Indicators across Stages**
   - We introduced **stage-specific indicators** that move beyond dichotomous right/wrong.
     - *Example (number line):* not only "point to a location," but also relate the target value to neighboring ticks/intervals, explain **sub-interval partitioning** (e.g., between 4 and 5), and justify comparisons.
   - These indicators were aligned across tasks so that the same conceptual moves (e.g., coordinating units, justifying comparisons) map to the same reaction level, reinforcing consistency and inter-rater reliability.

2.1.3 Structure of the Interview Guide

After three field trials, measurement errors were largely eliminated. Even when observers made different on-site judgments, agreement was reached by returning to the criteria during the judgement. In this way, the interview guide was finalized.

Table 7. Sub-competences and Tasks.

Sub-competence	Task	
	Natural number	Decimal number
I. Number Representation	1.1 Counting	1.1 Understanding the structure of a number line (a hierarchy of 100's and 10's unit) and read the numbers struck on the number line for a given number

	1.2 Understanding the structure of a number line (a hierarchy of 100's and 10's unit) and type the given number on the number line	1.2 Counting
	1.3 Understanding the structure of a number line (a hierarchical relation between 100's and 10's places) and read the numbers struck on the number line for a given number	1.3 Understanding the structure of a number line (a hierarchical relation between 100's, 10s, 1s and 0.1s places) and read the given number on the number line
	1.4(1) Assuming that a unit of the grid (with line) represents 100, answer the enclosed numbers	1.4(1) Assuming that a unit of the grid (with line) represents 1, answer the enclosed numbers
	1.4(2) Grid to number lines	1.4(2) Grid to number lines
	1.5(1) Assuming that a unit of the grid (without line) represents 100, enclose number	1.5(1) Assuming that a unit of the grid (without line) represents 1, enclose number
	1.5(2) Assuming that a unit of the grid (with line) represents 100, enclose a specific number	1.5(2) Assuming that a unit of the grid (with line) represents 1, enclose a specific number
Ⅱ. Comparison numbers	2.1 Comparison using number	2.1 Comparison using number
	2.2 Representing numbers and comparing numbers represented by tiles	2.2 Representing numbers and comparing numbers represented by tiles
	2.3 Representing numbers and comparing numbers represented by a number line	2.3 Representing numbers and comparing numbers represented by a number line
Ⅲ. Calculation	3a Addition	3a Addition
	3s Subtraction	3s Subtraction
	3m (1) Multiplication	3m (1) Multiplication
	3m (2)(3) Representing multiplication using tiles	3m (2)(3) Representing multiplication using tiles
	3d (1) Division	3d (1) Division
	3d (2)(3) Representing division using tiles	3d (2)(3) Representing division using tiles

In the final interview guide, considering task fundamentality and extensibility, three sub-competences were structured (see Table 7). Across tasks, five **response levels** were standardized so that judgments could be both **diagnostic**—(i) identifying the pupil's level of understanding and (ii) suggesting intervention points to deepen understanding—and consistent. For example, a pupil is judged as **Level 4** "answers correctly but cannot explain why using concrete representations (tiles, number line)"; thus, an intervention of prompting **Level 5** (explain with concrete materials) is indicated. Consistency refers to the relationships among criteria across tasks.

Table 8. Five Response Levels with Examples.

Response Level	Level depiction		Notes
Level 1	Cannot say the answer.	Neither of them is correct	
Level 2	The represented numbers are incorrect	Only former (1) or latter (2) is correct	
Level 3	The numbers represented are correct. (A pupil/ interviewer asks some questions.)	Both former and latter are correct. (A pupil/ interviewer asks some questions.)	Represented numbers are correct, with questions or prompts; include cases where pupils notice an error during explanation or upon interviewer's prompts.
Level 4	The represented numbers are correct smoothly	Both former and latter are correct smoothly	
Level 5	The numbers represented are correct smoothly with explanation	Both former and latter are correct smoothly with explanation	

2.2 Final Stage: Tools and Implementation

2.2.1 Correspondence between Evaluation and Intervention Tools

As noted earlier, to **diagnostically** assess abilities hidden within severe learning poverty among Zambian pupils (Uchida, 2009), we adopted problem-solving interviews approach with tasks; the core of this study is the development of diagnostic assessment using interviews. Based on this, we developed **evaluation tools** (interview guide, tests) and **intervention tools** (pupil workbook, teacher's guide) to confirm final outcomes through intervention activities and pre-/post-assessments. The test used in pre- and post-surveys and the intervention tools were developed in correspondence with the interview guide (Table 9; Figure 3). The pupil workbook, essentially a practice book based on the interview guide, includes **Extra Practice** items as preparatory problems (e.g., representing with 1s/10s/100s tiles, using number lines and grids, expressing calculations with tiles). The teacher's guide explains sub-competences, task intentions, and instructional notes to support teacher learning and adaptation to a new way of teaching. During the intervention, several teachers were observed using the workbooks highly effectively.

Table 9. Correspondence among Interview Guide, Test, and Workbook

Natural Number

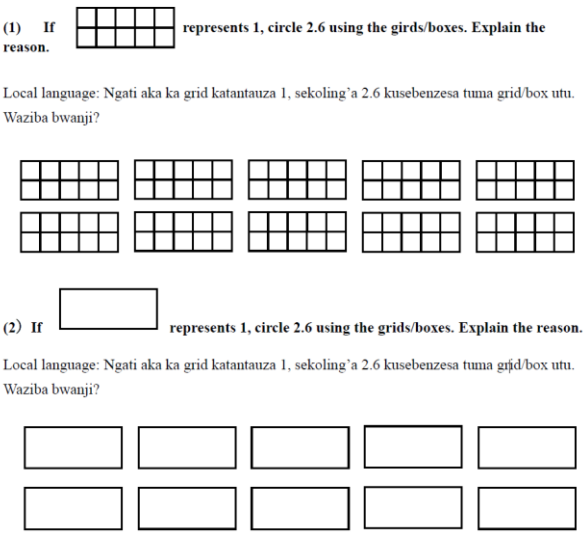
	Competence/sub-competence	Task	Test	Workbook
Ⅰ. Number Representation	Counting	1.1		
	Understanding the structure of a number line and read the numbers struck on the number line for a given number	1.3	Q1	(G3,4) Extra B Practice 3 (G4,5) Extra NB Practice N2
	Understanding the structure of a number and type the given number on the number line	1.2	Q2(G3,4) Q1(G5,6)	Practice 4
	Assuming that the grid (with line) represents 100, and reading a number	1.4(1)	Q3(G3,4) Q2(G5,6)	(G3,4) Extra A Practice 1 (G4,5) Extra NA Practice N1
	Assuming that the grid (with line/ without line) represents 100, and enclosing a number	1.5	Q4	Practice 2
	Representing a number by tiles		Q5(G3,4) Q3(G5,6)	Extra C.D
	Understanding the structure of grids to number lines	1.4(2)	Q6	Practice 5
Ⅱ. Comparison numbers	Representing numbers and compare numbers represented by tiles	2.2	G3,4 Q7	Extra E Practice 6
	Judging by number lines/ Representing numbers and comparing numbers represented by a number line	2.3	Q8	Extra G Practice 8
	Comparison using number	2.1	Q9	Extra F Practice 7
Ⅲ. Calculation	Addition	3a(1)	Q12(a)	Practice 9
	Representing addition using tiles	3a(2)(3)		Extra H
	Subtraction	3s(1)	Q12(b)	Practice 10
	Representing subtraction using tiles	3s(2)(3)		Extra I
	Multiplication	3m(1)	Q12(c)	Practice 13
	Representing multiplication using tiles	3m(2)(3)	Q10(2)	Extra J
	Representing addition formula and multiplication formula		Q10(1)	Practice 11
	Division	3d (1)	Q12(d)	Practice 14
	Representing division using tiles	3d (2)(3)	Q11(2)	Extra K
	Representing the division sentence for the diagram		Q11(1)	Practice 12
	Word problem		Q13	Extra L

Decimal number

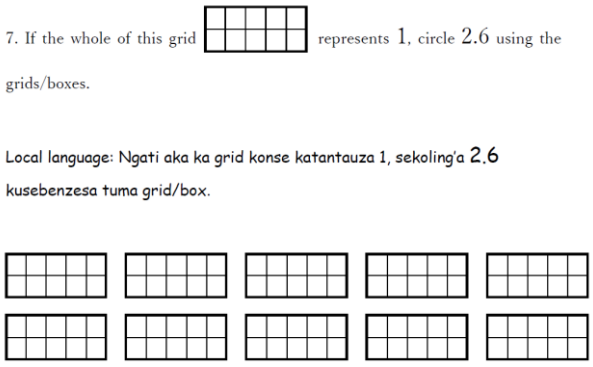
		Task	Test	Workbook
I. Number Representation	Counting	1.2		
	Understanding the structure of a number line and read the numbers struck on the number line for a given number	1.1	Q5	Extra B Practice 3
	Understanding the structure of a number and type the given number on the number line	1.3	Q4	Practice 4
	Assuming that the grid (with line) represents 1, and reading a number	1.4(1)	Q6	Extra A Practice 1
	Assuming that the grid (with line/ without line) represents 1, and enclosing a number	1.5	Q7	Practice 2
	Representing a number by tiles		Q8	Extra C.D
	Understanding the structure of grids to number lines	1.4(2)	Q9	Practice 5
Ⅱ. Comparison numbers	Representing numbers and compare numbers represented by tiles	2.2	Q10	Extra E Practice 6
	Judging by number lines/ Representing numbers and comparing numbers represented by a number line	2.3	Q11	Extra G Practice 8
	Comparison using number	2.1	Q12	Extra F Practice 7
Ⅲ. Calculation	Addition	3a	Q15(a)	Practice 9
	Representing addition using tiles	3a(2)(3)		Extra H’
	Subtraction	3s	Q15(b)	Practice 10
	Representing subtraction using tiles	3s(2)(3)		Extra I’
	Multiplication	3m(1)	Q15(c)	Practice 13
	Representing multiplication using tiles	3m(2)(3)	Q13(2)	Extra J’
	Representing addition formula and multiplication formula		Q13(1)	Practice 11
	Division	3d (1)	Q15(d)	Practice 14
	Representing division using tiles	3d (2)(3)	Q14(2)	Extra K’
	Representing the division sentence for the diagram		Q14(1)	Practice 12
	Word problem		Q16	Extra K’

Figure 3. Correspondence among pre-survey and post-survey tools and intervention tools.

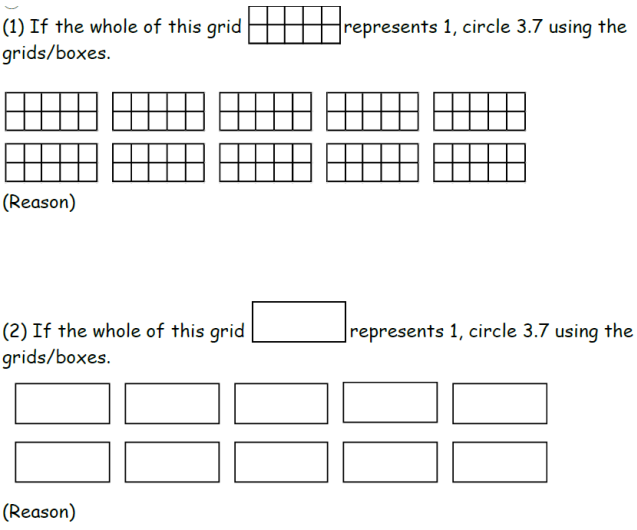
Excerpts from Interview Guides for Pre- and Post-Surveys



Test excerpts from pre- and post-surveys



Excerpt from a Children's Workbook for Educational Intervention



2.2.2 Structure of the Final Tool (Test)

Table 10 presents an overview of the test items. For **natural numbers**, Items 1–6, and for **decimal numbers**, Items 1–7, aimed to verify understanding of number representations and decimal structure by clarifying correspondences across representations—specifically among **number lines**, **grids and numbers**. For natural numbers, Items 7–9, and for decimal numbers, Items 10–12, addressed **number comparison** using numerals, tiles, and number lines. For natural numbers, Items 10–13, and for decimals, Items 13–16, addressed **calculation**, including items probing the meanings of the four operations through formulas and tiles, written calculations, and word problems.

Table 10. Overview of Test Items.

Natural Numbers

Item No	Outline
1	Answer the number indicated by the arrow on the number line.
2	Draw an arrow on the number line that represents **420**.
3	When the entire grid divided into ten parts represents **100**, answer the number represented by the enclosed area.
4 (1)	When the entire grid divided into ten parts represents **100**, **enclose 260**.
4 (2)	When the entire grid **without** divisions represents **100**, **enclose 490**.
5	Represent **234** using tiles of **1**, **10**, and **100**.
6	When the entire grid divided into ten parts represents **100**, indicate **on a number line** the number represented by the enclosed area.
7	Of A and B shown with tiles of **1**, **10**, and **100**, answer which is larger.
8	Of the two arrows shown on a number line, answer which represents the larger number.
9 (1)	Between **240** and **280**, answer which is larger.
9 (2)	Between **350** and **620**, answer which is larger.
10 (1)	Convert $120+120+120+120=480$ into a **multiplication** expression.
10 (2)	Represent $3¥times 210$ using tiles of **1**, **10**, and **100**.
11 (1)	Express the given diagram as a **division** sentence.
11 (2)	Represent $390¥div 3$ using tiles of **1**, **10**, and **100**.
12a	178 + 213$
12b	286 - 147$
12c	230X3$
12d	480 ÷ 4
13	Word problem on 360 ÷ 3 and its solution method.

Decimal numbers

Item No	Outline
1	Draw an arrow on the number line that represents **4.2**.
2	When the entire grid divided into ten parts represents **1**, **enclose 2.6**.
3	Represent **23.4** using tiles of **0.1**, **1**, and **10**.
4	Draw an arrow on the number line that represents **4.2**.
5	Answer the number indicated by the arrow on the number line.
6	When the entire grid divided into ten parts represents **1**, answer the number represented by the enclosed area.
7	When the entire grid divided into ten parts represents **1**, **enclose 2.6**.
8	Represent **23.4** using tiles of **0.1**, **1**, and **10**.
9	When the entire grid divided into ten parts represents **1**, indicate **on a number line** the number represented by the enclosed area.
10	Of A and B shown with tiles of **0.1**, **1**, and **10**, answer which is larger.
11	Of the two arrows shown on a number line, answer which represents the larger number.
12 (1)	Between **2.4** and **2.7**, answer which is larger.
12 (2)	Between **1.8** and **3.4**, answer which is larger.
13 (1)	Convert $1.2+1.2+1.2+1.2=4.8$ into a **multiplication** expression.
13 (2)	Represent $3¥times 2.1$ using tiles of **0.1**, **1**, and **10**.
14 (1)	Express the given diagram as a **division** sentence.
14 (2)	Represent $390¥div 3$ using tiles of **0.1**, **1**, and **10**.
15a	17.8 + 21.3$
15b	28.6 - 14.7$
15c	2.3 X 3
15d	4.8÷4
16	Word problem on 18.6÷3 and its solution method.

2.2.3 Test Administration

Because interview participants were limited in number, test administration to a larger group of pupils was also considered in the final phase to examine intervention effects. Although the format was a written test, our aim was not only to judge only correctness of answers but also to reveal calculation processes and ways of thinking—for instance, whether pupils could treat semi-concrete materials (tiles on paper) as grouped units and count them, and whether they could describe their calculation processes.

Given (i) many lower-grade pupils in Zambia face difficulty learning mathematics in English; (ii) Grades 1–4 are taught in a local language (Nyanja in Lusaka); (iii) interpreting written word problems in the local language can be difficult; and (iv) the desire to exclude reading-comprehension differences from scores, we adopted the following procedures: the examiner reads each item aloud, proceeding in order; after two readings, pupils answered; after the allotted time, the examiner instructs pupils to move to the next item and repeats the same procedures. Pupils are seated with sufficient spacing, and multiple researchers monitor to prevent pupils from cheating. This method exemplifies the "evidence-based endogenous approach in countries experiencing severe learning poverty influenced by sociocultural factors" (Baba, 2022).

2.2.4 Implementation of Workshops

Two workshops were conducted prior to the survey at the final stage.
The first workshop targeted all 10 schools in both the experimental and control groups, and covered an overview of the project as well as a review of the Phase I intervention activities.
The second workshop targeted only the five experimental schools and provided explanations of the Phase II intervention activities. Details are provided below.

＜Workshop Ⅰ＞

- **Date:** 23 February 2024
- **Venue:** University of Zambia
- **Purpose:** To review the intervention activities of Phase I and to provide an overall overview of the project
- **Main Participants:**
  Dr. Bentry Nkhata (Zambia Team Leader, UNZA),
  Dr. Liberty Mweemba (Dean, School of Education, UNZA),
  Ms. Shelly Sikwale (Representative, Ministry of Education, Zambia),
  Prof. Takuya Baba (Project Leader, Hiroshima University),
  Ms. Barbara Mudenda & Mr. Bareford Mambwe (Research Assistants),
  Dr. Patricia Nalube (Zambia Team Sub-leader, UNZA)
- **Content:** A review of the Phase I activities was conducted, and points for improvement and issues to be addressed going forward were shared. Participants confirmed the outcomes of the activities implemented and engaged in discussions toward the next phase.

＜Workshop Ⅱ＞

- **Date:** 25 March 2024
- **Venue:** University of Zambia
- **Purpose:** To explain the content of the Phase II intervention activities to teachers in the experimental group
- **Main Participants:**
  Dr. Bentry Nkhata (Zambia Team Leader, UNZA),
  Ms. Shelly Sikwale (Representative, Ministry of Education, Zambia),
  Ms. Barbara Mudenda, Mr. Bareford Mambwe, Mr. Emmanuel Kaao,
  Ms. Doye Chikola, Ms. Spiwe Tafeni (Research Assistants),
  Dr. Patricia Nalube (Zambia Team Sub-leader, UNZA),
  Dr. Satoshi Kusaka (Team Member),
  Principals and responsible teachers from the five experimental schools
- **Content:** Explanations were provided to the target teachers in the experimental group regarding the workbook to be used for the educational intervention. First, Kusaka presented an overview of the survey and the overall structure of the workbook. Participants were then divided into groups, and the research assistants explained the details of each task. By allocating sufficient time to explain each task and using the actual teaching materials (tiles, grids, number lines), teachers' understanding was deepened.

Figure 4. Scene from the Workshop Ⅱ

2.2.5 Pre- and Post-Surveys

The pre- and post-surveys, initially scheduled for January 2024, were postponed due to school closures resulting from a cholera outbreak. After schools reopened in February, the pre-survey began in late March (Tables 11 and 12). Interviews were conducted face-to-face by members of the Zambian and Japanese teams; the Japanese team also joined online, and each day concluded with joint discussions of evaluation results. Classroom teachers remained in the room to alleviate pupils' anxiety but did not directly participate.

Table 11. Pre-survey schedule (Mar–Apr 2024)

Sun	Mon	Tue	Wed	Thu	Fri	Sat
3/24	25	26 School A (I) School B (T)	27 School B (I) School A (T)	28 School C (I) School D (T)	29	30
31	4/1	2 School D (I) School C (T)	3 School E (I) School F (T)	4 School F (I) School E (T)	5 School G (I) School H (T)	6
7	8 School H (I) School G (T)	9 School I (I) School J (T)	10 School J (I) School I (T)			

Table 12. Post-survey schedule (Jun–Jul 2024)

Sun	Mon	Tue	Wed	Thu	Fri	Sat
6/23	24 School C (I) School B (T)	25 School B (I) School C (T)	26 School I (I) School J (T)	27 School J (I) School I (T)	28 School E (I) School D(T)	29 School D (I)
30	7/1	2 School G (I)	3 School F (I) School E (T)	4 School A (I) School F (T)	5 School H (I) School A (T)	6
7	8 School H (T)	9 School Gb (T)				

Figure 5. Scene from the Pre-survey

Figure 6. Scene from the Post-survey

2.2.6 Relationship between Participants and Interview Tasks

Due to time constraints, not all tasks were administered to every grade. To enable cross-grade comparisons, identical tasks from Tasks 1.1 to 1.5, were used regardless of grade, while the breadth of tasks administered varied by grade (Table 13). The identical are for natural numbers.

Table 13. Alignment of survey tasks with grades

	Natural Number												Decimal Number											
	1					2			3				1					2			3			
	1	2	3	4	5	1	2	3	a	s	m	d	1	2	3	4	5	1	2	3	a	s	m	d
G3																								
G4																								
G5																								
G6																								

2.2.7 Educational Intervention

An educational intervention was scheduled between the pre- and post-surveys. Whereas the three field visits and the pre/post interviews documented pupils' states without hints, the intervention was implemented by classroom teachers using the intervention tools, which was important in view of the trust relationship between teachers and pupils. The tools consisted of a **pupil workbook** and a **teacher's guide**. The period fell partly during a school vacation; in consultation with the Ministry of Education, we set the schedule (Table 14). Teachers followed the plan and were able to work intensively during the vacation. The plan was to complete the workbook two or three times over 38 days. The Zambian team regularly visited schools to monitor implementation. Although teachers had grasped the general usage in the workshops, some key points required further clarification; through meetings, interviews, and use, teachers learned the methods over time.[1] The same tools (test, interview guide) were used for pre- and post-surveys; the intervention tools were also closely aligned with them.

[1] We observed via Zoom from Japan, and while some teacher interventions were excellent, we believe they can be utilized in future educational activities.

Table 14. Implementation dates of the educational intervention

G3,4

Competence	Task	Date (1st rotation)		Date (2nd rotation)		Date (3rd rotation)
1. Representation	Extra A Practice 1 Practice 2	1	Apr 15	14	May 17	
	Extra B Practice 3	2	Apr 16	15	May 20	
	Practice 4 Practice 5	3	Apr 17	16	May 21	
	Extra C Extra D	4	Apr 18	17	May 22	
3. Calculation	Extra H	5	Apr 19	18	May 23	DAY 32-38 (Jun13-Jun21) Review Teacher selects units that pupils lack understanding of and reviews them.
	Extra I	6	Apr 29	19	May 24	
	Practice 9	6	Apr 29	20	May 27	
	Practice 10	6	Apr 29	21	May 28	
	Extra J	7	Apr 30	22	May 29	
	Practice 11	7	Apr 30	23	May 30	
	Extra K	8	May 2	24	May 31	
	Extra L	8	May 2	25	Jun 3	
	Practice 12	8	May 2	26	Jun 4	
	Practice 13	9	May 3	27	Jun 5	
	Practice 14	9	May 3	28	Jun 6	
2. Comparison	Extra E Practice 6	10	May 13	29	Jun 7	
	Extra F Practice 7	11	May 14	30	Jun 10	
	Extra G	12	May 15	31	Jun 11	
	Practice 8	13	May 16			

G5,6

Competence	Task	Date (1st rotation)		Date (2nd rotation)	
1. Natural number representation	Extra NA Practice N1	1	Apr 15	15	May 21
	Extra NB Practice N2	2	Apr 16	16	May 22
2. Decimal number representation	Extra A Practice 1 Practice 2	3	Apr 17	17	May 23
	Extra B Practice 3	4	Apr 18	18	May 24
	Practice 4 Practice 5	5	Apr 19	19	May 27
	Extra C Extra D	6	Apr 30	20	May 28
4. Calculation	Extra H (Natural)			21	May 29
	Extra H' (Decimal)	7	May 2	22	May 30
	Extra I (Natural)			23	May 31
	Extra I' (Decimal)	7	May 2	24	Jun 3
	Practice 9	7	May 2	25	Jun 4
	Practice 10	7	May 2	26	Jun 5
	Extra J (Natural)			27	
	Extra J' (Decimal)	8	May 3	28	Jun 7
	Practice 11	8	May 3	29	Jun 10
	Extra K (Natural)			30	
	Extra K' (Decimal)	9	May 13	31	Jun 12
	Extra L	9	May 13	32	Jun 13
	Practice 12	10	May 14	33	Jun 14
	Practice 13	10	May 14	34	Jun 17
	Practice 14	11	May 15	35	Jun 18
3. Comparison	Extra E	12	May 16	36	Jun 19
	Practice 6	12	May 16	36	Jun 19
	Extra F	13	May 17	37	Jun 20
	Practice 7	13	May 17	37	Jun 20
	Extra G	14	May 20	38	Jun 21
	Practice 8	14	May 20	38	Jun 21

Chapter 3 Analysis Results

3.1. Methods of Data Processing

3.1.1. Interview Data

Following the interview guide, interviews were conducted. In addition to the interviewer, multiple raters were assigned to evaluate children's responses to each problem (task). When raters' scores differed, the interview content was reviewed and discussed, and a final score was determined. The finalised scores were entered into Excel to create an analytical dataset for the interviews.

3.1.2. Test Data

To extract as much information as possible from the children's answers, we did not simply record correctness; instead, we entered their actual answers and the codes used to classify them. Table 15 shows a subset of the coding scheme.

Table 15. Subset of Input Codes

item	Correct answer	Codes		
q1	270	Enter the answer as given	88: no response	99: illegible
q2	420	Enter the number indicated by the arrow	88: no response	99: illegible
q3	180	Enter the answer as given	88: no response	99: illegible
q4(1)	260	Enter the number enclosed by the frame	88: no response	99: illegible

3.2. Results of the Interview Analysis

As shown in Table 16, the experimental group comprised 40 children for both the baseline-pre-survey and post-survey interviews.

Table 16. Information on Interview Participants

	Total	Experimental group	Control group
Grade 3	20	10	10
Grade 4	20	10	10
Grade 5	20	10	10
Grade 6	20	10	10
Total	80	40	40

3.2.1. Score Changes by School; Changes by Higher- and Middle-Performing Groups

The mean scores from the pre-survey and post-survey interviews for each school are shown in Table 17. In the experimental group, there was little change in the averages for School E2 and School E1. At School E2, teacher absenteeism was frequent, and the Grade 3 homeroom teacher was absent for an extended period for some reason, leading to significant delays in activities. This situation, together with delays in distributing the workbook and insufficient collaboration among teachers, indicated challenges in information sharing within the school. Such instability may have affected the results. At School E1, the Pre-survey mean was already high, and thus little change was observed. One possible reason is that School E1 has been in the experimental group since the early stage of Phase I, and the school may have already achieved an overall enhancement of basic academic skills. This suggests that not only Phase II but also the cumulative efforts since Phase I positively influenced student learning outcomes.

Table 17. Interview Scores (Pre-survey vs. Post-survey) by School

Group	School	Pre-survey	Post-survey	Difference
control	School C1	2.09	2.68	0.59
control	School C2	2.67	3.32	0.65
control	School C3	3.24	3.48	0.24
control	School C4	2.23	2.67	0.44
control	School C5	2.75	3.40	0.65
experimental	School E1	3.58	3.99	0.41
experimental	School E2	2.36	2.84	0.48
experimental	School E3	3.06	4.08	1.02
experimental	School E4	2.51	3.89	1.38
experimental	School E5	2.53	3.48	0.95

Changes in interview scores for the higher- and middle-performing groups are shown in Table 18. Children with lower baseline pre-survey scores were categorized as the middle-performing group, and those with higher baseline pre-survey scores as the higher-performing group. In the control group, no substantial differences in change were observed between performance groups. In contrast, the experimental group showed greater gains than the control group, with the middle-performing group exhibiting a larger difference between the baseline pre-survey and endline post-survey means.

Table 18. Interview Scores (Pre-survey vs. Post-survey) by Performance Group

Group	Performance	Pre-survey	Post-survey	Difference
control	Middle	2.11	2.73	0.62
	Higher	2.97	3.54	0.57
experimental	Middle	2.24	3.31	1.08
	Higher	3.18	3.89	0.71

3.2.2. Changes in Interview Scores by Grade

To compare changes in pre-survey and endline post-survey interview scores between the experimental and control groups for each grade, a difference-in-differences (DiD) analysis (multiple regression) was conducted. Because five items on natural numbers (1.1 to 1.5) were common across Grades 3 to 6, results for these common tasks are also presented.

Table 19. Analysis of the Mean and Difference of Differences in Interview Scores Across Grades (Results of Multiple Regression Analysis)

	Cronbach's α coefficient			Control group		Experimental group		Results of multiple regression analysis			
Grade	Total (40)	Pre (20)	Post (20)	Pre (20)	Post (20)	Pre (20)	Post (20)	Partial regression coefficient	Standard error	T-value	P-value
3	0.85	0.76	0.85	2.41	3.34	2.00	3.28	0.33	0.39	0.85	0.40
4	0.80	0.80	0.71	2.50	3.08	2.63	3.37	0.44	0.58	0.76	0.45
5	0.84	0.80	0.80	2.73	3.38	3.02	4.05	0.38	0.50	0.75	0.45
6	0.92	0.91	0.92	2.49	2.75	3.19	3.45	0.00	0.66	0.00	1.00
3-6 (common 5 questions)	0.79	0.80	0.73	2.70	3.20	3.01	3.84	0.33	0.39	0.85	0.40

Cronbach's alpha coefficients were around 0.8, adequate. The interview scores and analytical results for each grade are presented in Table 19. The multiple regression results indicated that the intervention effects in the experimental group were not statistically significant.

3.2.3. Changes in Interview Scores by Task

Next, we focus on score changes and means for each task. In Table 20, estimated intervention effects of 1.0 or greater are shown in red, and those of 0.1 or less in blue. Tasks with estimated intervention effects of 1.0 or greater are considered to show large changes, whereas those with 0.1 or less are considered to show small changes

Table 20. Scores for Each Interview Task Item (Natural Numbers) and Difference-in-Differences Analysis

	Qestion	Control Group		Experimental Group		Difference		Estimated Effect Size of Intervention
		Pre-survey	Post-survey	Pre-survey	Post-survey	Control Group	Experimental Group	
整数	1.1	2.23	2.45	2.85	3.41	0.23	0.56	0.34
	1.2	2.63	2.64	3.05	3.75	0.02	0.70	0.68
	1.3	3.28	4.38	3.15	4.10	1.09	0.95	-0.14
	1.4	2.23	3.30	2.30	3.50	1.08	1.20	0.13
	1.5	3.33	3.43	3.68	4.45	0.10	0.78	0.68
	2.1	3.00	4.00	2.70	3.50	1.00	0.80	-0.20
	2.2	3.40	4.30	2.80	4.00	0.90	1.20	0.30
	2.3	2.80	3.30	2.10	2.70	0.50	0.60	0.10
	3a	3.40	4.00	3.60	4.30	0.60	0.70	0.10
	3s	2.30	3.50	2.80	3.80	1.20	1.00	-0.20
	3m	1.20	1.80	1.20	2.00	0.60	0.80	0.20
	3d	1.20	1.40	1.00	1.89	0.20	0.89	0.69
Decimals	1.1	2.95	4.05	3.30	4.35	1.10	1.05	-0.05
	1.2	1.15	1.55	1.50	3.30	0.40	1.80	1.40
	1.3	3.00	3.15	3.65	4.05	0.15	0.40	0.25
	1.4	2.15	3.25	2.45	3.40	1.10	0.95	-0.15
	1.5	3.50	4.10	4.10	4.70	0.60	0.60	0.00
	2.1	2.27	3.20	2.70	3.30	0.93	0.60	-0.33
	2.2	3.45	3.30	2.67	3.60	-0.15	0.93	1.09
	2.3	2.55	3.00	2.40	3.40	0.45	1.00	0.55
	3a	1.90	2.40	2.30	3.10	0.50	0.80	0.30
	3s	2.40	2.50	3.30	3.80	0.10	0.50	0.40
	3m	1.40	1.80	2.00	2.10	0.40	0.10	-0.30
	3d	1.40	1.30	1.90	1.78	-0.10	-0.12	-0.02

Based on the estimated intervention effects by task, decimal tasks 1.2 and 2.2 are considered likely to show intervention effects. Conversely, no intervention effects were observed for natural number tasks 1.3, 2.1, 2.3, 3a, 3s and for decimal tasks 1.1, 1.4, 1.5, 2.1, 3m, 3d.

Examining the tasks where intervention effects appeared: decimal task 1.2(1) asked students to identify the number represented by each tick mark on the number line shown in Figure 7. Decimal task 2.2 required representing 3.5 and 1.6 using tiles for tens, ones, and tenths, as shown in Figure 8, and answering which number is larger. Number lines for decimals are included in the Grade 6 textbook, whereas tiles are not. Therefore, these may have been unfamiliar to Zambian children, potentially leading to larger intervention effects.

Figure 7. Number line for decimal task 1.2(1)

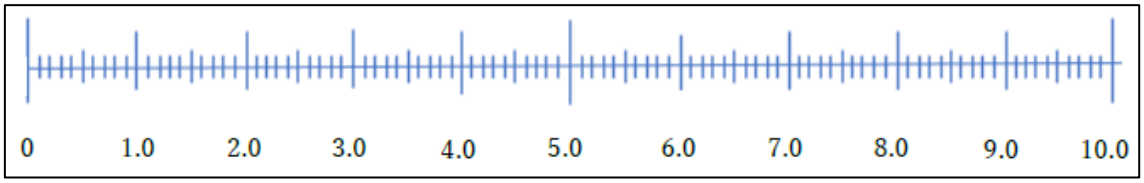
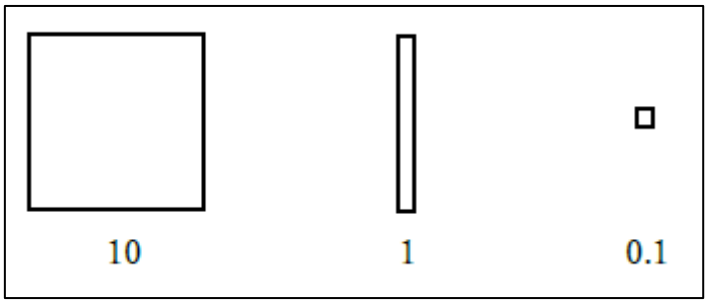


Figure 8. Tiles for decimal task 2.2



Considering tasks where no intervention effects were observed: natural number task 1.3 involved identifying the number indicated by an arrow on the number line in Figure 9. Natural number task 2.1 asked students to order [400, 80, 210] from smallest to largest. Natural number task 2.3 asked students to indicate [50, 370, 160] on a number line. Natural number task 3a asked students to solve 330 + 135 using column arithmetic, and natural number task 3s asked students to solve 385 − 152 using column arithmetic. Decimal task 1.1 required identifying the number indicated by an arrow on the number line in Figure 10. Decimal task 1.4 asked, when a ten-frame represents 1, to state the number enclosed by a red line in Figure 11. Decimal task 1.5 asked, when a ten-frame represents 1, to enclose 2.6. Decimal task 2.1 asked to order [1.2, 0.8, 2.1] from smallest to largest. Decimal task 3m asked to solve 2.4 × 3 using column arithmetic, and decimal task 3d asked to solve 3.6 ÷ 3 using column arithmetic.

Because both the control and experimental groups showed score improvements on natural number tasks 1.3, 2.1, 2.3, 3a, 3s and decimal tasks 1.1, 1.4, 1.5, 2.1, it is possible that students' understanding deepened through regular lessons over the three months. On the other hand, for decimal tasks 3m and 3d, neither group showed improvement, suggesting that the intervention did not necessarily function effectively.

Figure 9. Number line for natural number task 1.3

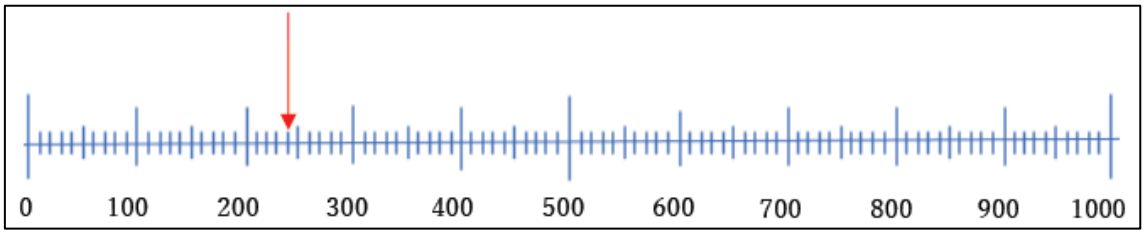


Figure 10. Number line for decimal task 1.1

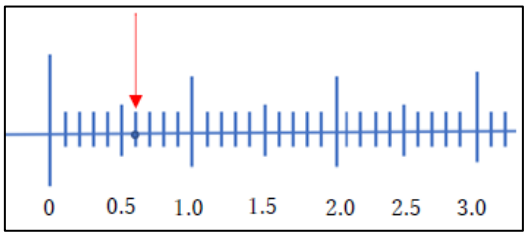
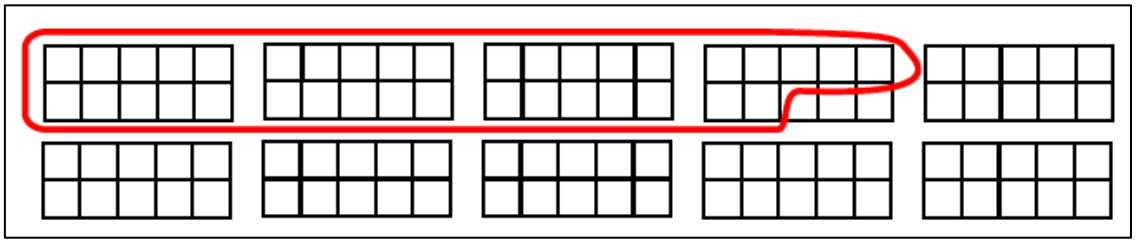


Figure 11. Decimal task 1.4



3.2.4. Changes in Individual Children Who Received Both Pre-survey and Post-survey Interviews

We examine changes in individual children who received both the pre-survey and post-survey interviews. Here, for each grade, we focus on the child whose score increased the most.

(1) Grade 3

Comparing scores between the control and experimental groups, two children in the experimental group recorded lower scores at post-survey. Their scores changed from 1.38 to

1.25 and from 4.0 to 3.0, respectively, indicating they were unable to solve the problems. Among the other children, four in the experimental group increased by 2.0 points or more, while none did so in the control group. Children who increased by 1 point or more numbered five in the control group and six in the experimental group. Thus, children in the experimental group exhibited greater gains than those in the control group and increases of 2.0 points indicate an intervention effect. In the experimental group, two children (Tables 21 and 22) showed the largest score increases, from 1.63 to 4.13 and from 1.8 to 4.3, respectively. Both children changed from incorrect to correct on all questions.

Table 21. Child A in Grade 3 with the largest improvement (all natural number items)

Grade 3	1.1	1.2	1.3	1.4	1.5	2.1	2.2	2.3
Pre	1	4	2	1	1	1	1	2
Post	5	5	5	2	4	5	4	4

(Column: Task number)

Table 22. Child B in Grade 3 with the largest improvement (all natural number items)

Grade 3	1.1	1.2	1.3	1.4	1.5	2.1	2.2	2.3
Pre	2	2	1	1	1	4	1	2
Post	5	2	5	5	5	4	5	3

(Column: Task number)

For Child A (overall increase of 2.5 points), although numerical scores suggest improvement, we highlight areas with limited change. On task 1.3, the score remained at 2.0 points; this item asked the child to read 240 on a number line, and the response was incorrect. Tasks related to the number line are 1.1 and 1.2, where the child scored 5.0 points, indicating a basic grasp of number lines. Other items with scores of 4.0 points included representing quantities using grids (task 1.5) and comparing magnitudes of quantities.

For Child B (overall increase of 2.5 points), we also focused on tasks that showed no improvement. Task 1.2 remained at 2.0 points; this item asked the child to indicate the position of 260 on a number line, and no improvement was observed, even though other number-line items were correct, leaving the cause unclear. In addition, the child scored 4.0 points on task 2.1 and 3.0 points on task 2.3—both tasks involved comparing the magnitudes of three numbers. Although verbal explanation proved difficult, the child correctly answered task 2.3 at post-survey. Thus, even when number lines are understood, some items may still be answered incorrectly. This resembles findings by Shibuya (2009) regarding Zambian children's calculation and number-pattern comprehension, where correct answers could regress or understanding remained unstable. Children who showed large improvements were able to provide explanations, indicating learning outcomes; however, instability and regression were also observed, reaffirming characteristics and difficulties in Zambian

children's learning in this study. This shows the importance of consolidating newly learned topics.

(2) Grade 4

Comparing scores between the control and experimental groups, one child in the experimental group showed a decrease at post-survey (from 2.33 to 2.22). Among the others, one child in the experimental group increased by 2.0 points or more, while none did so in the control group. Children who increased by 1.0 point or more numbered three in the control group and four in the experimental group. Therefore, although gains in the experimental group exceeded those in the control group, the magnitude of gains in Grade 4 appears smaller than in Grade 3. The child in the experimental group with the largest increase (Table 23) changed from 2.3 to 4.6.

Table 23. Child B in Grade 4 with the largest improvement (all natural number items)

Grade 4	1.1	1.2	1.3	1.4	1.5	3a	3s	3m	3d
Pre	2	5	1	1	3	5	2	1	1
Post	5	5	5	5	5	5	5	1	5

(Column: Task number)

Based on this child's results, verbal explanations were possible for all items except task 3m. Task 3m involved column multiplication of 240 × 3, and manipulation with concrete materials and explanation for 240 × 3. All were incorrect, with no change between pre-survey and post-survey. While more children in the experimental group than in the control group scored highly (≥3.0 points at post-survey: one in control, three in experimental), many scores clustered around 1.0 or 2.0 overall. Therefore, including this child, the interview did not reveal intervention effects for (3-digit) × (1-digit) multiplication. Possible reasons include increased difficulty in manipulation as numbers grow larger, and the need to move beyond repeated addition towards multiplicative reasoning (doubling). Nakawa (2012) reported on mastery of multiplication tables at the primary stage; even for (1-digit) × (1-digit), Zambian children in primary to secondary education stages have not fully consolidated the facts. Thus, it is plausible that a certain proportion of Grade 3–6 children experience difficulty with (3-digit) × (1-digit) multiplication. The tools used in the study help visualize number magnitudes and are effective for nurturing number sense. However, to grasp and consolidate calculation skills, additional measures are needed, and the intervention's effects should be verified empirically.

(3) Grade 5

Comparing scores between the control and experimental groups, one child in the control group showed a decrease at post-survey (from 2.46 to 2.15). Among the others, one child in the experimental group increased by 2.0 points or more, while none did so in the control

group. Children who increased by 1.0 point or more numbered two in the control group and four in the experimental group. Thus, although gains in the experimental group exceeded those in the control group, the magnitude of gains appears smaller than in Grade 3 and similar to Grade 4. The child in the experimental group with the largest increase (Table 24) changed from 2.3 to 4.6. Although the experimental group shows better results than the control group in terms of pre-survey and post-survey differences, the differences are smaller than those in Grades 3 and 4, suggesting that the intervention effects were not large.

Table 24. Child with the largest improvement in Grade 5

Grade	Natural numbers					Decimals							
4	1.1	1.2	1.3	1.4	1.5	1.1	1.2	1.3	1.4	1.5	2.1	2.2	2.3
Pre	1	1	1	1	1	5	1	2	1	1	4	1	4
Post	5	5	5	4	5	5	5	2	1	4	4	5	5

(Column: Task number)

This child's pattern resembles that of the Grade 4 child with the largest improvement. Except for tasks 1.3 and 1.4, the child answered all items correctly, scoring 4.0 or 5.0 points, indicating the ability to give explanations. Comparing natural numbers and decimals shows little difference, indicating that performance was not affected by number type—an encouraging outcome. However, task 1.4 for decimal numbers remained at 1.0 point. This item asked for a ten frame of 2 × 5 cells where the whole frame represents 1, to find the value of one cell. This requires proportional reasoning and understanding of units, where no improvement was observed. Meanwhile, the child answered all other decimal items correctly, suggesting that although understanding of decimals partly improved through the intervention, there remains some weakness and room for improvement in proportional perspectives and unit awareness.

(4) Grade 6

Comparing scores between the control and experimental groups, two children in the experimental group and four in the control group showed decreases at post-survey. Their changes were 4.57 to 4.14 and 4.64 to 4.57 in the experimental group, and 3.57 to 3.50 and 2.43 to 1.92 in the control group. Thus, this grade had the largest number of children with decreased scores. However, the two experimental-group children and one control-group child exhibited only small changes. The two experimental-group children had averages above 4.0 points, indicating many correct answers and the ability to explain even at pre-survey.

Among the others, no child in either group increased by 2 points or more. Two children in the control group increased by 1 point or more, while none did so in the experimental group, with differences under 1 point. Thus, compared with other grades, the intervention effects appear small. At the pre-survey, they have already obtained generally high scores. The child in the experimental group with the largest increase (Table 25) changed from 3.93 to 4.86—

an exceptional case whose pre-survey average was already high, with an increase of about 1.0 point.

Table 25. Child B in Grade 6 with the largest improvement

Grade 6	Natural numbers					Decimals								
	1.1	1.2	1.3	1.4	1.5	1.1	1.2	1.3	1.4	1.5	3a	3s	3m	3d
Pre	5	5	5	5	5	5	2	4	4	5	2	5	2	1
Post	5	5	5	5	5	5	5	5	5	5	4	2	2	4

(Column: Task number)

This child was able to provide explanations for all items except tasks 3a and 3d, indicating learning gains. For tasks 3(a) and 3(d), which involved decimal addition and division and manipulation with concrete materials, the score was 4.0 points. Despite some remaining areas for improvement, the child improved relative to pre-survey. While children - especially in Grades 5 and 6 - showed sizable gains, rather than a comprehensive understanding, individual weaknesses remained, with instability noted as above. This suggests the need for consolidation and long-term memory in the Japanese sense (Yuzawa, 2019).

3.3. Results of the Test Analysis

Tests were administered to Grades 3–6. The test for Grades 3–4 dealt with natural numbers and the four operations, while that for Grades 5–6 dealt with natural and decimal numbers and the four operations.

There were five schools each in the control and experimental groups; Table 26 shows the sample sizes for each school. About 40 students from each school participated. The total numbers of participants in the pre- and post- surveys were 798 and 803, respectively, for Grades 3–4 and Grades 5–6. The sample sizes for students who participated in both the pre- and post- surveys were 324 and 317, respectively.

Table 26. Sample Sizes

		G3&4			G5&6		
Group	School	pre	post	common	pre	post	common
control	School C1	40	40	32	40	40	33
	School C2	40	40	34	41	40	33
	School C3	40	40	35	40	40	35
	School C4	40	40	33	40	40	26
	School C5	38	40	33	40	40	33
experimental	School E1	40	40	36	40	40	32
	School E2	40	40	30	40	40	32
	School E3	40	40	26	41	41	32
	School E4	40	40	34	40	40	33
	School E5	40	40	31	40	40	28
Total		398	400	324	402	401	317
		798			803		

Analysis for Grades 3–4

An overview of the test items is shown in Table 27. Here, excluding items 10 and 11 whose correct-answer rates were 5% or less and considerably lower than other items, we focus on the remaining 18 items and analyze correctness.

Table 27. Test Items for Grades 3–4

Item	Content
1	Answer the number shown on the number line.
2	Indicate the position of 420 on the number line.
3	Based on a 100-grid with partitions, answer the number of the enclosed part.
4(1)	Enclose cells on a 100-grid with partitions to represent 260.
4(2)	Enclose cells on a 100-grid without partitions to represent 490.
5(1)	Draw two hundreds-squares.
5(2)	Draw three tens-rectangles.
5(3)	Draw four ones-squares.
6	Indicate on a number line the 360 represented on a partitioned 100-grid.
7	Compare magnitudes using hundreds-squares, tens-rectangles, and ones-squares.
8	Compare the magnitudes of two numbers shown on a number line.
9(1)	Compare the magnitudes of 240 and 280.

9(2)	Compare the magnitudes of 350 and 620.
10(1)	Express 120+120+120+120=480 using multiplication.
10(2)	Draw a diagram for 3×210.
11(1)	Express dividing 480 into four parts using division.
11(2)	Draw a diagram for 390÷3.
12(a)	178+213
12(b)	286−147
12(c)	230×3
12(d)	480÷4
13	Formulate and compute 360÷3.

First, each item was converted to a binary variable (1 for correct, 0 for incorrect) to create a binary dataset. Cronbach's alpha for these items was 0.865, suggesting that the reliability of the 18 items was reasonably ensured.

Figure 12. Distribution of test scores for control vs. experimental (Grades 3–4)

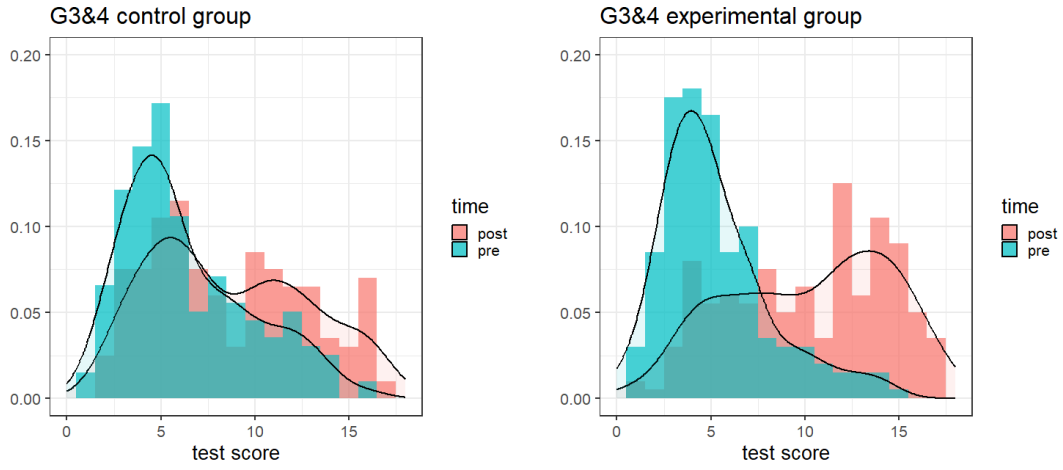


Table28. Descriptive Statistics of Score Distributions

	control		experimental	
time	average	sd	average	sd
pre	6.369	3.402	5.210	2.943
post	8.685	4.160	10.210	4.286

Analysis for Grades 5–6

An overview of the test items is shown in Table 29. Here, excluding items 13–16 whose correct-answer rates were 5% or less and considerably lower than other items, we focus on the remaining 17 items and analyze correctness, as with Grades 3–4.

Table29. Test Items for Grades 5–6

Item	Content
1	Indicate the position of 420 on the number line.
2	Enclose cells on a partitioned 100-grid to represent 260.
3(1)	Draw two hundreds-squares.
3(2)	Draw three tens-rectangles.
3(3)	Draw four ones-squares.
4	Indicate the position of 4.2 on the number line.
5	Answer the number shown on the number line.
6	Based on a partitioned 1-grid, answer the number of the enclosed part.
7	Enclose cells on a partitioned 1-grid to represent 2.6.
8(1)	Draw two tens-squares.
8(2)	Draw three ones-squares.
8(3)	Draw four tenths-rectangles.
9	On a partitioned 1-grid representing 3.6, indicate it on a number line.
10	Compare magnitudes using tens-squares, ones-rectangles, and tenths-squares.
11	Compare the magnitudes of two numbers shown on a number line.
12(1)	Compare magnitudes of 2.4 and 2.7.
12(2)	Compare magnitudes of 1.8 and 3.4.
13(1)	Express 1.2+1.2+1.2+1.2=4.8 using multiplication.
13(2)	Draw a diagram for 3×2.1.
14(1)	Express dividing 4.8 into four parts using division.
14(2)	Draw a diagram for 3.9÷3.
15(a)	17.8+21.3
15(b)	28.6−14.7
15(c)	2.3×3
15(d)	4.8÷4
16	Formulate and compute 18.6÷3.

First, each item was converted to a binary variable (1 for correct, 0 for incorrect) to create a binary dataset. Cronbach's alpha for these items was 0.889, suggesting that the reliability of the 22 items was reasonably ensured.

Figure 13. Distribution of test scores for control vs. experimental (Grades 5–6)

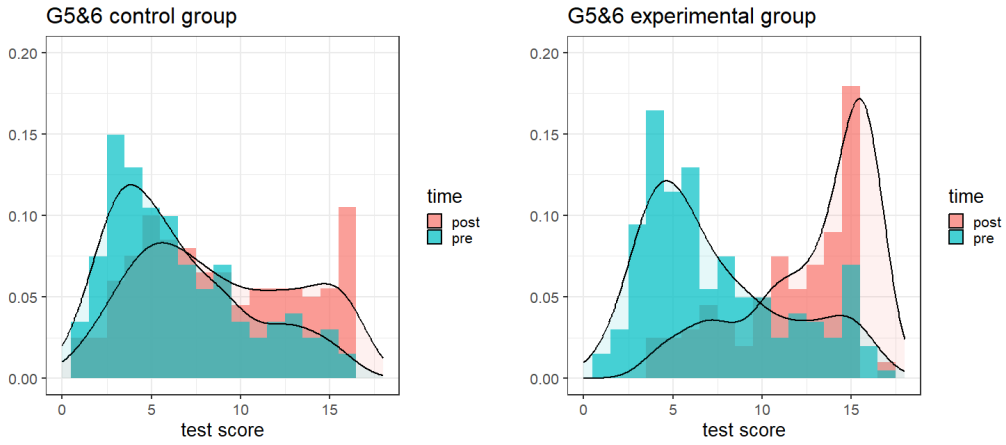


Table30. Descriptive Statistics of Score Distributions

	control		experimental	
time	average	sd	average	sd
pre	6.413	3.885	7.289	4.032
post	8.965	4.439	12.861	3.510

Changes in Mean Test Scores in Control vs. Experimental Groups

Figure 13 shows the changes in average test scores between pre-survey and post-survey for the control and experimental groups, as in Table 30. In both Grades 3–4 and Grades 5–6, the experimental group appears to have larger gains than the control group.

We then conducted a difference-in-differences analysis to examine the magnitude of the intervention effects. As shown in Table 31, the intervention effects for Grades 3–4 and Grades 5–6 were 2.684 and 3.120, respectively. This can be interpreted as an effect size equivalent to approximately three additional items answered correctly in the experimental group relative to the control group from pre-survey to post-survey.

Figure 14. Changes in average test scores (pre-survey vs. post-survey)

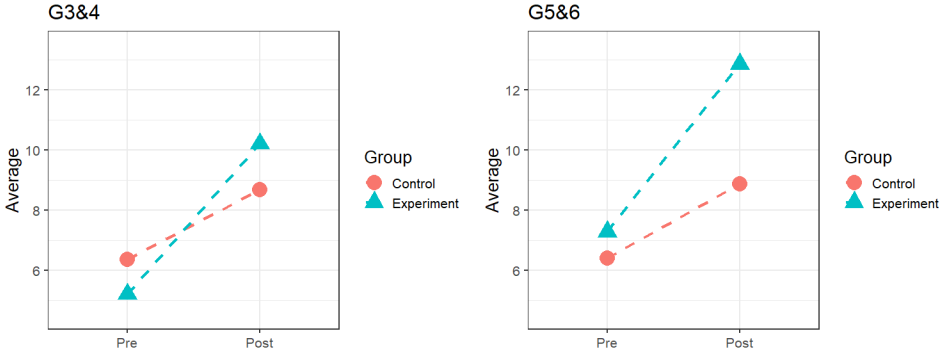


Table 31. Results of Difference-in-Differences Analysis

	3–4 graders			5–6 graders		
	Estimate	SE	p-value	Estimate	SE	p-value
Constant	6.369	0.266	<0.00	6.413	0.281	<0.00
Post	2.316	0.375	<0.00	2.452	0.398	<0.00
Experimental group	-1.159	0.375	<0.05	0.876	0.397	<0.05
Intervention effect	2.684	0.530	<0.00	3.120	0.562	<0.00

We also examined which items showed intervention effects. As seen in Tables 32 and 33, many items exhibited intervention effects. For example, for Item 1 in Grades 3–4, the proportion of correct answers increased four-fold at post-survey in the experimental group—indicating an intervention effect.

Table 32. Logistic regression results for Grades 3–4

Question	Accuracy rate				Regression coefficient				
	Control group		Experimental group						
	Pre	Post	Pre	Post	Constant term	Post	Experimental group	Intervention effect	
1	17%	26%	11%	46%	0.200	1.757	0.587	4.051	*
2	27%	36%	15%	46%	0.366	1.506	0.464	3.336	*
3	23%	41%	19%	57%	0.303	2.249	0.775	2.513	*
4(1)	44%	52%	34%	64%	0.784	1.355	0.657	2.547	*
4(2)	18%	34%	18%	48%	0.215	2.399	1.022	1.753	
5(1)	19%	39%	9%	72%	0.238	2.636	0.391	10.502	*
5(2)	22%	42%	12%	68%	0.286	2.534	0.477	6.009	*
5(3)	24%	40%	13%	69%	0.320	2.083	0.446	7.307	*
6	20%	36%	14%	53%	0.253	2.174	0.616	3.257	*
7	78%	85%	72%	92%	3.605	1.572	0.696	2.729	*
8	91%	94%	88%	88%	10.000	1.567	0.733	0.609	
9(1)	93%	96%	87%	96%	14.231	1.686	0.450	2.221	
9(2)	84%	87%	74%	92%	5.387	1.189	0.515	3.263	*
12(a)	30%	47%	22%	33%	0.435	2.040	0.649	0.856	
12(b)	10%	30%	10%	19%	0.106	3.942	1.047	0.536	
12(c)	15%	40%	12%	28%	0.172	3.805	0.757	0.787	
12(d)	7%	20%	6%	23%	0.070	3.558	0.828	1.443	
13	15%	28%	10%	31%	0.172	2.210	0.648	1.787	

Table 33. Logistic regression results for Grades 5–6

Question	Accuracy rate				Regression coefficient				
	Control group		Experimental group						
	Pre	Post	Pre	Post	Constant term	Post	Experimental group	Intervention effect	
1	34%	53%	40%	71%	0.511	2.162	1.320	1.649	
2	37%	47%	50%	79%	0.583	1.492	1.699	2.563	*
3(1)	27%	48%	30%	82%	0.377	2.450	1.130	4.396	*
3(2)	25%	49%	28%	80%	0.331	2.902	1.195	3.504	*
3(3)	23%	47%	28%	82%	0.297	2.929	1.334	3.823	*
4	23%	37%	27%	64%	0.297	1.937	1.238	2.518	*
5	3%	4%	1%	4%	0.036	1.005	0.420	3.078	
6	22%	36%	26%	68%	0.280	1.964	1.278	3.044	*
7	49%	61%	52%	82%	0.951	1.610	1.150	2.517	*
8(1)	34%	55%	28%	87%	0.511	2.343	0.774	7.258	*
8(2)	34%	50%	34%	86%	0.523	1.913	0.978	6.064	*
8(3)	27%	48%	30%	86%	0.367	2.513	1.158	5.547	*
9	20%	37%	30%	69%	0.256	2.292	1.700	2.194	*
10	73%	88%	85%	96%	2.722	2.571	2.014	1.513	
11	77%	82%	80%	83%	3.370	1.307	1.158	0.963	
12(1)	79%	91%	91%	97%	3.786	2.516	2.530	1.348	
12(2)	54%	59%	68%	72%	1.161	1.239	1.843	0.952	

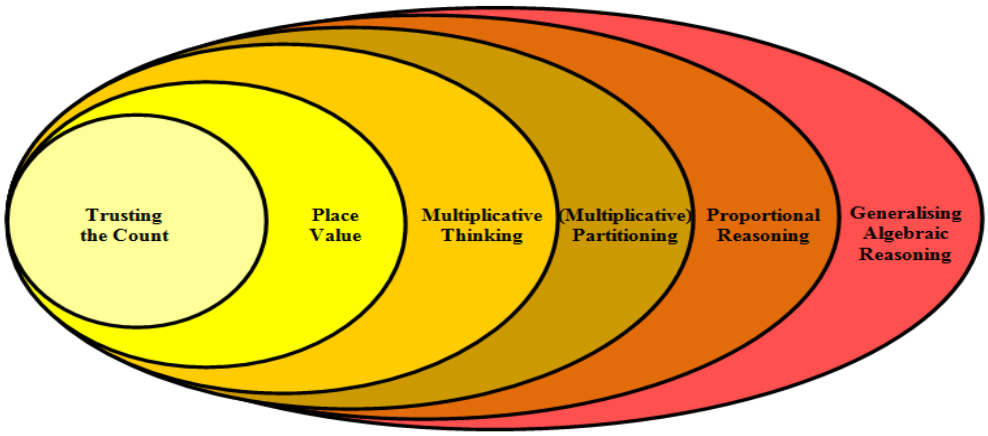
Chapter 4 Discussion

4.1 Considerations on Interview Analysis

From the interview results, when organizing the growth brought by intervention, it can be said that the intervention had an effect because numerical differences in growth between the experimental and control groups were observed in Grades 3 and 4. On the other hand, although some pupils improved by 2.0 points, it became clear that growth was smaller in Grades 5 and 6. In particular, Grade 6 showed little improvement. This suggests a tendency for growth to diminish when mathematical concepts become more complex for children—here referring to decimals in number range and multiplication/division in operations. In any case, this relates to the achievements and issues from Phase I. Previous studies have pointed out that Zambian children have difficulty linguistically explaining the reasons behind mathematical operations, mathematical concepts themselves, or their own mathematical thinking (Shibuya, 2008; 2009). The fact that Grade 3–4 children became able to linguistically explain natural number operations and Grade 5–6 children became able to explain operations involving natural numbers and decimals can be considered a major achievement of the intervention. On the other hand, not all pupils improved uniformly in decimal operations, and individual differences were observed, suggesting that prescriptions for improving calculation skills may differ by pupil. Regarding intervention tools, particularly for multiplication and division, it will be necessary to consider and propose further options (Mungalu, 2025).

From a research perspective, for example, in Japan many studies and teaching materials on multiplication and division have been developed and implemented. It is necessary to examine diverse materials from various perspectives, which will clarify diverse Zambian children's actual learning and contribute to the accumulation of research. In particular, multiplicative thinking, which underlies multiplication and division, forms the foundation of mathematical concepts in secondary education and beyond. From this perspective, further enrichment is required as shown in Figure 15.

Figure 15. Multiplicative Thinking and Its Development (Hurst &Hurrel, 2016)



In recent African mathematics education research, advanced research such as STEM/STEAM and modeling has been conducted (Ogunniyi & Iwuanyanwu, 2024). However, children's computational ability and the logical thinking that accompanies calculation are deeply related to learning in other mathematical areas, so further research accumulation is required. Going forward, it would be beneficial to implement and examine mathematical activities utilizing ICT and individually optimized learning, and investigate whether difficulties stem from challenges in acquiring skills, sustaining skills, or language factors—and construct strategies accordingly.

As another perspective, Matsuo & Sato (2024) investigated the effects of an integrated program combining arithmetic and the arts among children in Zambia and Japan, clarifying characteristics of Zambian children's learning. Considering continuity of learning, it would be beneficial to implement initiatives that clarify what children can and cannot do at the early-childhood stage and link these findings to primary education.

4.2 Considerations on Test Analysis

Test scores were calculated and the intervention effect was examined using a difference-in-differences analysis. As a result, the overall intervention effect was confirmed. Logistic regression analysis was then used to examine the intervention effect for each test item.

In Grades 3 and 4, for example, Item 9(1) asks pupils to compare 240 and 280, and Item 9(2) asks them to compare 350 and 620. While an intervention effect was observed in 9(2), this was not necessarily the case for 9(1). Thus, pupils face difficulties when the highest place value is the same and the next place value must be compared. This suggests remaining challenges in understanding the structure of the decimal place value system. Also, Items

12(a)–(d) in computation appear to be difficult for many pupils, indicating a need for further learning of the decimal place value system. In Grades 5 and 6, pupils also showed difficulties with Items 12(1) and 12(2), which involve comparing decimal magnitudes. Based on analysis of individual items, understanding of the structure of the decimal system appears to be a challenge. While students' achievement is too low in some items, the items are excluded from analysis. These items should be studied further to develop different approaches to identify students' difficulties.

Chapter 5 Recommendations

5.1 Policy Recommendations

◆ Basic Perspective for Curriculum Development

Through this study, the concept of an evidence-based endogenous approach (Baba, 2022) was concretized. This refers to re-examining a country's own educational issues through data—particularly important in tackling severe low academic achievement. Although research not grounded in prior studies risks becoming arbitrary, Zambia's extremely low achievement levels differ significantly from those of other countries, making it necessary to collect and interpret data based on Zambia's sociocultural context.

In comparative education, the concept of policy borrowing (Steiner-Kamisi, 2016) exists in curriculum development. Policy borrowing refers to the process by which educational ideas, policies, or practices developed in one context are adopted and adapted in another. Many developing countries have introduced curriculum reforms inspired by international trends and educational models from other countries. However, the effectiveness of such reforms depends not only on the quality of the borrowed policy itself but also on the extent to which it aligns with local educational needs, cultural values, and implementation conditions. However, when underlying conditions differ significantly, borrowed policies may fail to take root. Moreover, "localization," the final stage of the borrowing cycle, concerns the compatibility between the borrowed policy and the local context. The endogenous approach proposed in this report is expected to develop further.

◆ Organizing Learning in Curriculum Development

The Zambian and Japanese teams learned many practical and experiential points during Phases I and II, based on the evidence-based endogenous approach. One key method was design experiments (Gravemeijer & Cobb, 2006), in which evaluation tools were revised repeatedly while collecting data, deepening insights into children's potential. Perspective from Zambian researchers was indispensable, and it is important to foster a group of Zambian researchers who mutually elevate one another (Baba & Nakai, 2011).

Another recommendation is that the Curriculum Development Centre (CDC) should conduct research that clarifies the reality of extremely low achievement and examines ways to improve it. Incorporating research components into curriculum development and adding the term "research" to the institutional name—such as Curriculum Development and Research Centre (CDRC)—would reflect a culture of research-based learning.

◆ Improving Teacher Education

Diagnostic assessment tools were used to create intervention tools. Workshops were conducted so teachers at the experimental schools could effectively use these tools. During the intervention, team members visited experimental schools, answered teachers' questions, and provided advice based on the characteristics of the tools. As a result, some teachers were able to incorporate child-centered interventions even within typical Zambian teaching styles.

Such initiatives have a huge potential to change teachers' instructional practices and enhance their ability to notice children's thinking (Yuda, 2025). Mathematics education textbooks for teacher training are currently being created by the team for the University of Zambia, using interview data of Zambian children responding to tasks—highly effective for preservice and in-service training. The data provides (1) pupils have different levels of achievement, (2) teachers can employ different approaches according to the levels of students, and (3) all these are targeting that pupils are able to explain their mathematical ideas. This is related to the next point.

◆ Improving Children's Learning: Importance and Potential of Explanation Activities

Many Zambian teachers tend to interpret children's understanding dichotomously as "being able to answer correctly or not." However, Phases I and II showed that children's mathematical understanding has multiple stages. Encouraging pupils to read, represent, and explain numbers and operations using models promotes mathematical understanding. This indicates potential for improving learning among Zambian children, who have tended to rely heavily on counting.

The structural understanding of numbers addressed in Phase II is an entry point and pillar for the decimal system (Arthur, 2025). Given the severe learning deficits identified through SACMEQ and similar assessments, these initiatives are significant for future development.

◆ Importance of Language

This study developed test items with attention to diagrammatic, operational, linguistic, and symbolic representations (Nakahara, 1995; Lesh, Post, and Behr, 1987). Explanation activities are clearly linked to language. Senior Lecturer Bentry Nkhata and the Zambian team continually refined linguistic expressions during interviews to find more child-friendly wording (Nkhata & Baba, 2024). As a result, children better understood interview questions and became able to explain their reasoning.

This has value beyond mere correctness of answers. In curriculum development, further knowledge is needed not only about types of languages used but also about noteworthy

linguistic features, expressions difficult to translate into English, and vocabulary that hinders children's mathematical understanding.

5.2 Research Recommendations

Based on the three-tier curriculum of TIMSS and referencing Baba (2025), the following considerations are offered.

Evidence-Based Endogenous Curriculum Development

This study identified understanding of number structure and multiplication/division as issues. Some of the prominent topics are the hierarchical nature of the decimal system and multiplicative reasoning which extends to higher mathematical concepts—the pillars of school mathematics today. One team member, Arthur (2025), took the second topic and implemented his doctoral study on array structures. This approach should be deepened and expanded across the curriculum.

Teacher Education and Use of Evidence

In the final stage of this study, intervention activities were implemented, and some participating teachers provided high-quality instruction. Interview data and intervention data yielded valuable insights. Together with project members and mathematics education researchers at the University of Zambia, textbooks and videos are being developed for teacher education and school use. Future studies are waited for evaluation of their effectiveness.

Use and Analysis of Data on Children

This study analyzed and recorded children's thinking. Although severe learning deficits were known, pupils demonstrated various abilities during the study. This can serve as a starting point for addressing low academic achievement. With cooperation from many stakeholders, valuable data were collected, which should be treated as common properties. Trends in open science and open data highlight both research reliability (e.g., reproducibility) and usefulness for local researchers and citizens. Ethical relationships between researchers and participants remain essential. Researchers must ensure participants understand the meaning of data collection and must return insights gained from the data.

◆ Formation of Expert Communities

Finally, promoting this research requires forming expert communities (Baba & Nakai, 2011). These communities should cultivate shared knowledge regarding research ethics and the sociocultural contexts of developing countries.

References

Awich, M. (2021). *SACMEQ IV international report.* Southern and Eastern Africa Consortium for Monitoring Educational Quality

馬場卓也（2025）「アフリカにおける数学教育研究の過去、現在、未来」『アフリカ教育研究』第 16 号、3–20 頁。

Baba, T. (2022, February 12). Evidence-based and Endogenous Curriculum Development [Invited lecture]. International Cooperation Development in Mathematics Education, University of Tsukuba, Japan.

Baba, T., Nakawa, N., Nkhata, B., Arthur. M. Barbara, M., Chikola, D., Emmanuel, K., Nambwe, B., Kosaka, M., Kusaka, S., Minakoshi,K. Spiwe, T., Takuzwa J.N., & Watanabe, K. (2021). Project Research for "Analysis of Challenges and Solutions for Improving Numeracy Competence in Zambia" Project Completion Report November 2021 Japan International Cooperation Agency Hiroshima University.

Baba, T., Nakawa, N., Nkhata, B, Arther, M., Barbara.M., Emmanuel, K., Kosaka, M., Kusaka, S., Mambwe, B., Nkhalamo, C.J., Watanabe, K. (2019). The Development of an Assessment Instrument for Numeracy Competence and its Application to Selected Primary Schools in Zambia. Zambia Journal of Teacher Professional Growth (ZJTPG), 5(2), 72-92. The University of Zambia Press.

Baba, T., & Nakai, K. (2011). Teachers' institution and participation in a lesson study project in Zambia: Implication and possibilities. In Proceedings of seminars held at the Center for the Study of International Cooperation in Education (CICE), Africa-Asia university dialogue for educational development: Report of the International Experience Sharing Seminar (2); Actual status and issues of teacher professional development (pp. 85-102). Hiroshima, Japan: Hiroshima University.

Gravemeijer, K. & Cobb, P. (2006) Design research from a learning design perspective. In Van den Akker, J., Gravemeijer, K., McKenney, S., & Nieveen, N. (Eds.). Educational Design Research (1st ed.). Routledge. (pp. 17-51)
https://doi.org/10.4324/9780203088364

Hungi, N., Makuwa, D., Ross, K., Saito, M., Dolata, S., Cappelle, F. V. (2010). SACMEQIII Project Result: Pupil Achievement Levels in Reading and Mathematics. Working Document Number 1. Paris: SACMEQ

Hurst, C. & Hurrell, D. (2014). Developing the Big Ideas of Number. International Journal of Educational Studies in Mathematics. 1. 1-18. 10.17278/ijesim.2014.02.001.

Hurst, C. & Hurrell, D. (2016). Investigating Children's Multiplicative Thinking: Implications for Teaching. European Journal of STEM Education. 1. 1-11. 10.20897/lectito.201656.

JICA & Hiroshima University (2021). Project Research for "Analysis of Challenges and Solutions for Improving Numeracy Competence in Zambia" Project Completion Report. https://openjicareport.jica.go.jp/pdf/12362984.pdf

Lesh R., Post, T. & Behr, M. (1987). Representations and translations among representations in mathematics learning and problem solving. In C. Janvier (Ed.), Problems of Representation in the Teaching and Learning of Mathematics (pp. 33-40) Hillsdale, NJ: Lawrence Erlbaum Associates.

松尾七重・佐藤 真帆(2024). 算数と造形の総合的なプログラムの効果　日本とザンビアの場合」『日本科学教育学会年会論文集』248 (1) https://doi.org/10.14935/jssep.48.0_449

MoGE (Ministry of General Education, Zambia) (2017). Zambia's National Assessment Survey Report 2016: Learning Achievement at the Primary School Level. MoGE.

Mungalu, A. (2025). The Use of Array Patterns to Enhance Primary School Pupils' Multiplicative Thinking in Zambia. Unpublished Doctoral Dissertation. Hiroshima University.

中原忠男(1995)『算数数学教育における構成的アプローチの研究』聖文社.

Nakawa, N. (2016) Children's Recognition of Numbers in the Republic of Zambia: Focusing on Subitising and Counting toward Concrete and Semi-Concrete Objects, 7, 155-172. Africa Education Research Journal, 7, 155-172.

中和渚(2012) 「ザンビア共和国の数学教育開発のための基礎的調査（２）－中央州カブウェのある基礎学校における子どもたちの１桁の乗法の習熟度と課題－」. 第 45 回日本数学教育学会研究発表会発表資料. 2012 年 11 月 12 日.

Nkhata, B., Baba, T. (2024). The Influence of Non-Mathematical Impediments on Development of Diagnostic Tools in Numeracy Competence Project in Zambia. ICME 15th Congress on Mathematical Education. July7-14, 2024, ICC, Sydney, Australia.

OECD (2019), PISA 2018 Results (Volume I): What Students Know and Can Do, PISA, OECD Publishing, Paris, https://doi.org/10.1787/5f07c754-en.

小川啓一、江連誠、武寛子(2005)『万人のための教育(EFA)への挑戦 ： 日本の ODA に対する提言』.国際協力機構国際協力総合研修所.

Ogunniyi, M., & Iwuanyanwu, P. N. (2024). Analysis of teachers' perspectives towards the use of IKS to improve STEM education for sustainable development. *African Journal of Research in Mathematics, Science and Technology Education, 28*(3), 319–329. https://doi.org/10.1080/18117295.2024.2352980

澁谷渚（2009）「本質的学習環境（SLE）に基づく数学科授業開発研究（２）–ザンビアのある基礎学校における生徒の数のパターンの認識に関する記述の分析」『数学教育学研究』, 15(1), 136-146.

澁谷渚（2008）「本質的学習環境（SLE）に基づく数学科授業開発研究（１）–ザンビア基礎学校における生徒の活動の分析–」『数学教育学研究』, 14, 187-197.

Simon, M. A. (1995). Investigating the Development of Multiplicative Reasoning [Review of The Development of Multiplicative Reasoning in the Learning of Mathematics, by G. Harel & J. Confrey]. Journal for Research in Mathematics Education, 26(3), 282–288. https://doi.org/10.2307/749132

Steiner-Khamsi, G. (2016). New directions in policy borrowing research. Asia Pacific Education Review, 17. 381-390. https://doi.org/10.1007/s12564-016-9442-9

Uchida, T. (2009). The diagnostic evaluation of pupils' performance in basic school in Zambia: Focusing on discrimination of performance and educational suggestiveness. Journal of International Cooperation in Education, 12(2), 1–12.

Vergnaud, G. (2009). The Theory of Conceptual Fields. Human Development 52(2). 83-94. https://doi.org/10.1159/000202727

Vergnaud, G. (1983). Multiplicative structures, in Lesh, R. and Landau, M. (Eds), Acquisition of Mathematics Concepts and Processes, Academic Press. 127-174.

World Bank (2022) The State of Global Learning Poverty: 2022 Update. https://thedocs.worldbank.org/en/doc/e52f55322528903b27f1b7e61238e416-0200022022/original/Learning-poverty-report-2022-06-21-final-V7-0-conferenceEdition.pdf

湯澤 正通 (2019). ワーキングメモリの発達と児童生徒の学習：読み書き・算数障害への支援, 『発達心理学研究』, 30 巻, 4 号, p. 188-201. https://doi.org/10.11201/jjdp.30.188, https://www.jstage.jst.go.jp/article/jjdp/30/4/30_188/_article/-char/ja

湯田しおり（2025）『ザンビアにおける数学教師の気づきに関する研究』、広島大学大学院修士論文(未公刊).

This work was supported by JSPS KAKENHI Grant Number JP21KK0038.