Bull. Grad. Sch. Humanit. Soc. Sci. Hiroshima Univ : Stud. Educ., Vol.4, 2023, 257－266

Using Item Response Theory to Analyze Student Ability to Acquire Lower Primary Curriculum Content: A Case of Rural Malawi

Kyoko Taniguchi
(Received. October 6, 2023)

Abstract: Improving student achievement is a significant issue in Malawi. This study uses item response theory (IRT) to analyze students' ability to acquire curriculum content in rural Malawi' s primary education system. Using data from a sample of 1,476 grade 5 students across 30 public primary schools, the analysis computes two parameters of IRT—item difficulty and item discrimination—on selected test questions in the areas of English and mathematics. In English, students demonstrated skills in vocabulary and in asking and answering questions, while they had difficulty with tenses and stories. In mathematics, students' strengths were concentrated in counting, writing numbers, and addition, while their weaknesses were concentrated in solving common fractions and in problems involving time and addition of money. Although IRT is useful for analyzing student ability, the test items being studied must meet certain assumptions before IRT can be used with them—posing a challenge in low-income countries. The findings of this study are intended to contribute to improving student skills in Malawi.

Key words: item response theory, student ability, academic achievement, curriculum, sub-Saharan Africa,

1. Introduction

Ensuring a better quality of education is a fundamental matter in every society. Improving basic academic achievement is a challenging issue, especially in low-income countries. Approximately 87 percent of students living in such countries cannot reach the minimum proficiency level in primary and lower secondary education (UNESCO IES, 2018). In sub-Saharan Africa, ample studies have reported low student achievement (e.g., Hungi et al., 2010; Uwezo, 2014). However, research on the factors that affect achievement has been thin. In other words, almost all existing studies have neglected" to analyze student achievement in a concrete way.

There are currently two main theories in test development and analysis: classical test theory (CTT) and item response theory (IRT). IRT has been widely used in test development in the United States since the 1970s (Hambleton and Swaminathan, 1985). Statistically, IRT is a more robust theoretical approach than CTT. IRT has been employed in test development and analysis in large-scale studies,

International Education Development Program, Division of Educational Sciences, Graduate School of Humanities and Social Sciences

including some in low-income countries. However, especially in low-income countries, IRT is rarely used in a detailed and comprehensive analysis of student test results.

In Malawi, a low-income country located in sub-Saharan Africa, student achievement has been and still is a significant issue in primary education. The Southern and Eastern African Consortium for Monitoring Educational Quality (SACMEQ) III reported that in Malawi, only 26.7 percent and 8.3 percent of grade 6 students, respectively, had surpassed the basic level of reading and numeracy skills (Hungi et al., 2010). Equally concerning, the country's primary school completion rate is approximately 50 percent (MoEST, 2017): only half of students enrolled in primary school complete their education. Completion rates for grades 5 and 8 were 60 percent and 40.5 percent, respectively (MoEST, 2017). According to a report by the Malawi government, grade repetition and dropout rates during primary education were 23.9 percent and 4.1 percent, respectively (MoEST, 2017). SACMEQ III found that in Malawi, the repetition rate for grade 6 was 60.3 percent, the highest among the 15 sub-Saharan African countries that were compared (Hungi et al., 2010).

Drawing on data from a rural area of Malawi, this study uses IRT to explore students' ability to acquire the intended curriculum content in lower primary education. More specifically, the study aims to answer two complementary basic questions: What were the specific areas of difficulty in targeted questions? What were the specific areas of discrimination in targeted questions? This paper's chief contribution lies in its attempt to answer these questions by applying them to the way IRT measures student ability.

2. IRT and its utilization

2.1. Shortcomings of CTT

Before the widespread adoption of IRT in test development and analysis, CTT was the conventional approach. However, CTT has been associated with several limitations, including three in particular, highlighted by Hambleton and Swaminathan (1985): group-dependent indices, dependence between observed and actual scores, and an assumption that measurement errors are equal. First, CTT relies on group-dependent item difficulty and item discrimination indices. These values are contingent on the specific group of examinees under consideration. Second, CTT assumes a dependence between observed test scores and actual scores. Yet in reality, as test difficulty changes, observed and actual scores fluctuate accordingly. Last, CTT's assumption of equal measurement errors for all examinees lacks a solid foundation. In reality, ability estimates are less precise for low- and high-ability students than for average-ability students. In response to these limitations, IRT emerged as a new measurement theory. Its development addressed these concerns and provided a more suitable test development and analysis framework.

2.2. Basic concepts of IRT

IRT is a statistical framework that takes into account how examinees' item and test performance relate to their abilities. Unlike CTT, IRT provides item characteristics that are not dependent on specific groups, ability scores that are independent of the test itself, and a similar measure of precision for each level of ability (Hambleton and Swaminathan, 1985). By the early 1990s, IRT had emerged as the most promising approach for measuring achievement through testing (Keeves, 1994). IRT operates on two fundamental assumptions: trait-based predictions and item characteristic functions. IRT assumes that examinee test performance can be predicted by factors referred to as *traits*, *latent traits*, or *abilities*. The relationship between examinee item performance and the traits that influence item performance is described by a monotonically increasing function known as an *item characteristic function* (also called an *item characteristic curve* in the one-trait, or one-dimensional, model). It quantifies the probability of an examinee's answering an item correctly at various points on an ability scale. While IRT encompasses several models, this study employs a two-parameter model that assesses both item difficulty and item discrimination.

2.3. Utilization of IRT in low-income countries

Although IRT has been widely utilized in high-income countries, only a few studies have employed this method in the analysis of test results in low-income countries. Watanabe (2011) analyzed the Second Regional Comparative and Explanatory Study (SERCE) in Ecuador using IRT. Also employing IRT, Watanabe (2014) also drawing on Programme for International Student Assessment (PISA) data for seven countries (Brazil, Indonesia, Mexico, Thailand, Tunisia, Turkey, and Uruguay) (2014).

In sub-Saharan Africa, a great number of studies have been conducted to evaluate student achievement, carried out by organizations such as the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) and the Programme d'Analyse des Systèmes Éducatifs des Pays de la Confemen (PASEC). Although SACMEQ and PASEC use a one-parameter model of IRT, known as the Rasch model, for constructing test question items, they have never used IRT to analyze test results. Moreover, they have not yet explored utilizing the two-parameter model of IRT. They examine test results in the framework of CTT, which is based on a raw score, or a number-right score, analyzed using mean, standard deviation, variance, and Cronbach's alpha. Thus, in spite of the advantages of IRT over CTT, and the additional knowledge to be gained through use of two-parameter IRT to analyze test results, the present study seems to be the first in sub-Saharan Africa to apply two-parameter IRT to test results.

2.4. Analyses of student achievement in Malawi

Three studies have been conducted in Malawi to identify levels of student achievement. In the first of these, Kunje, Selemani-Meke, and Ogawa (2009) surveyed student achievement in mathematics, Chichewa (a local language), and English in grades 5 and 7 in the South Western Education Division of the county. The stated intent was to explore the relationship of school-, classroom-, and student-level factors with student achievement in mathematics. Although the authors obtained data on three subject areas, they did not analyze student achievement in a granular way but instead simply calculated the mean score and standard deviation for each subject in each grade.

In the second study, Tomita and Muta (2010; 2012) analyzed data from the Monitoring Achievement in Lower Primary (MALP) survey, conducted by the United Nations Children's Fund (UNICEF). The study investigated which had a greater effect on student achievement in Chichewa, mathematics, English, and life skills in grade 4: school factors or students' family backgrounds.

Finally, the SACMEQ assessment has been conducted periodically in the whole country since 1995. SACMEQ III used a one-parameter setup of IRT to construct test question items (Hungi et al., 2010).

Thus, as shown above for the sub-Saharan region, no studies to date have used two-parameter IRT to analyze academic achievement in Malawi. The present study sets out to do so.

3. Methodology

3.1. Framework

This study uses as its framework Keeves's model (1994) of the context and components of the curriculum. As shown in Figure 1, the curriculum itself (right-hand column) can be viewed as, in reality, three distinct curricula: the intended curriculum, the implemented curriculum, and the achieved curriculum. This study focuses on measuring students' ability to acquire the intended curriculum. In Malawi, the intended curriculum is reflected in syllabuses and textbooks developed at the national level (MIE, 2005a, 2005b, 2005c, 2005d).

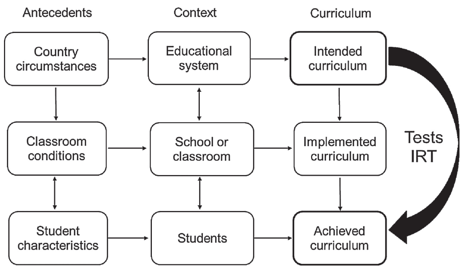

Figure 1. The context and components of the curriculum
Source: Keeves (1994).

3.2. Instruments

The study developed its own research instruments, English and mathematics tests based on the national curriculum. The first step was to analyze the Malawi Primary School Syllabuses (MIE, 2005a, 2005b, 2005c, 2005d) from grades 1 to 4, noting core elements and the associated curriculum topics. The second was to classify each element by its level and cognitive domain, according to the SACMEQ III framework (Hungi et al., 2010). The third was to select question items from textbooks and past papers of the national examination for pretests that were conducted in one school. Referring to the results, teachers in the school, along with primary education advisers in the district education office, modified the tests. Finally, the tests were administered to grade 5 students. The contents of these English and mathematics tests are shown in Appendices 1 and 2, respectively. There were 35 question items for each subject.

3.4. Analysis

The IRT analysis required binary response data. Therefore, correct answers were coded as 1, and double marks and incorrect answers as 0.

In addition, before conducting the IRT analysis, the IRT assumptions needed to be tested. There are two main assumptions in IRT: local independence and unidimensionality. Local independence means that when the abilities influencing test performance are constant, the examinee's response to any pair of items is separately independent. This assumption was clear in the tests. Since all question items were independent, they could not answer the other question items.

Unidimensionality means that only one ability is measured by each item in a test. To test this assumption, we calculated the mean score on each question item, a correlation between the rate of the correct answers on each question item and the total score, and factors loading each question item. Question items that met the following conditions were selected for inclusion in the analysis: (1) a mean score of between 0.010 and 0.900; (2) a correlation of greater than 0.240 between the rate of the correct answer on each question item and the total score; and (3) factor loading of greater than 0.150. As a result, the final English and mathematics tests, as analyzed, contained 25 and 28 question items, respectively.

Finally, this study analyzed the selected question items using a two-parameter logistic model of IRT. The formula of the two-parameter logistic model is as follows:

$$P_i(\theta) = \frac{e^{Da_i\,(\theta - b_i\,)}}{1 + e^{Da_i(\theta - b_i)}}\ (i = 1, 2, \dots, \mathrm{n})$$

$P_i(\theta)$ *is the probability that an examinee with ability level* θ *answers item i correctly.*

D *is a scaling factor, which is* 1.702.

a *is the item discrimination parameter.*

b *is the item difficulty parameter.*

Item difficulty demonstrates to what degree students have difficulty with a question item. Item discrimination indicates a question item's ability to distinguish between high-ability and low-ability students. These parameters are discussed more fully below.

4. Findings

Figures 2 and 3 show the results on student ability in English and mathematics, respectively. Item difficulty is on the horizontal axis and item discrimination on the vertical axis. Item difficulty, as mentioned, rates how much difficulty students have with a question item. Low item difficulty (left-hand side of each figure) indicates an easy item, while high item difficulty (right-hand side) indicates a difficult item. Item discrimination indicates how well a question item distinguishes between high-ability and low-ability students. A high item discrimination (i.e., one that lies toward the top of the figure) means the probability of a student's giving the correct answer to that question varies with small differences in student ability, so that it can analyze academic achievement in a granular way. Low item discrimination (a low position in the figure) indicates the opposite.

As shown in Figures 2 and 3, respectively, the difficulty of all question items in both English and mathematics tests was between -3.000 and 3.000. The discrimination of all question items in both the English and mathematics tests was above 0.300. Thus, all question items were at an acceptable level of difficulty for test takers and could be expected to distinguish adequately between students of high and low achievement.

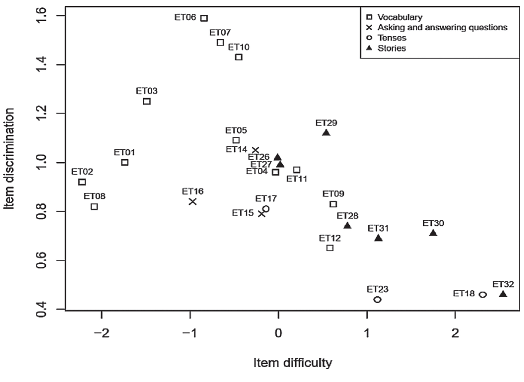

Figure 2. Item difficulty and item discrimination of each question item on the grade 5 English test

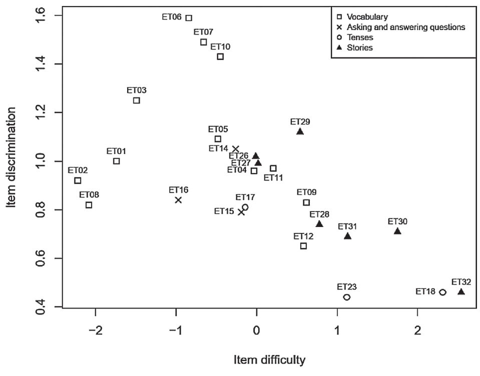

Figure 3. Item difficulty and item discrimination of each question item on the grade 5 mathematics test

4.1. Results of the English test

As shown in Figure 2, when the item difficulty was higher, the item discrimination was lower, and vice versa. Item difficulty was higher for questions on stories and tenses, while it was lower for those involving vocabulary and asking and answering questions. Item discrimination was lower for items

ET32 (i.e., English test item 32), ET18, and ET23, all on stories and tenses, while it was higher on the vocabulary items (ET06, ET07, and ET10).

The item difficulty of ET32, "interpret sentences to match words and phrases," in the stories category, and in "present form" (ET18), under tenses, was high, over 2.000; and item discrimination was the lowest for these, below 0.500. The next-highest item difficulty was on "use context and simple sentence structure to match sentences" (ET30 and ET31) in stories, falling between 1.000 and 2.000. The discrimination of these question items was relatively low, around 0.700. The "interpretation"measure for these items was higher than it was for "use context and simple sentence structure to match sentences," showing that students had difficulty interpreting the contents.

In contrast, the difficulty of vocabulary items "match words and pictures involving concrete concepts and everyday objects" (ET01 and ET02) and "match words and pictures in sentences involving concrete concepts and everyday objects" (ET08) was on the low end, below -1.700. The item discrimination of these question items was moderate, between 0.800 and 1.000. Interestingly, the discrimination of vocabulary items "match words and pictures involving concrete concepts and everyday objects" (ET06 and ET07) and "match words and pictures in a sentence involving concrete concepts and everyday objects" (ET10) was higher, above 1.400, while the item difficulty of these item questions was between -0.400 and -0.900. These question items distinguished well between high-achieving and low-achieving students, compared with question items ET01, ET02, and ET08.

A few potential reasons exist for different levels of item discrimination. First, giving the correct answer on a vocabulary item depends on the student's familiarity with the word in daily life. The item discrimination of question item ET06 was highest among all of the vocabulary question items. This question item asked about a farmer. Since the research was conducted in a rural area, students there were familiar with the vocabulary word in question. On the other hand, the item discrimination of question ET08, about a boy who was fishing, was lower. Some schools are located far from the lake, so some students had little knowledge of fishing. Therefore, word familiarity is likely to coincide with a student's environment and that student's ability to recognize the given picture or word.

Second, student performance was related to the format of the question items. In items ET07 and ET10, students selected pictures to match given words, but the other question items were the opposite: students selected words to match pictures. As shown in Figure 1, students' ability to choose the correct answer was reflected in the question format.

Last, student performance is likely affected by the probability of careless mistakes in the test itself. Question item ET02 asked about a hand, but a head was depicted in the selection items. Students tended to select "head" in this instance instead of "hand."

4.2. Results of the mathematics test

The tendency in the mathematics test results was slightly different from that of the English test results. When the item difficulty was relatively high, the item discrimination was relatively low in some question items, but this trend was not applicable to all question items. For example, the item difficulty of "addition of money," "time," and "common fractions" questions was high but the item discrimination was low. In contrast, the item difficulty of "counting and writing numbers" and "addition of numbers" was low but the item discrimination was high.

The item difficulty of "addition of money" (MT33), "time" (MT29), and "common fractions" (MT25) was high, between 1.700 and 2.100, and item discrimination was low, below 0.510. All students had difficulty with the "time" question item (MT29) because students in rural areas of Malawi rarely see an analog clock. For the question item "addition of money" (MT33), students could not select the answer by reading and understanding sentences. Except for these three questions, the item difficulty was below 0.700. This means that many question items were only moderately difficult, or even easy.

As for item discrimination, however, "addition of money" (MT32), "common fractions" (MT27), and "shapes" (MT03) were relatively low, around 0.500. In question item MT32, "addition of money," students needed to select a formula after reading sentences. Even high-achieving students had difficulty deriving a formula through understanding the sentences' meaning, and then finding the right answer. In question item M27, "common fractions," students needed to calculate a decimal fraction, but even high-achieving students could not do so. Question item MT03 asked students to select the name, "rectangle," of an illustrated figure. Most students knew the figure itself but did not know what it was called in English.

In contrast, the item discrimination of "multiplication" (MT16 and MT17) was higher, although the item difficulty of these questions was between -0.900 and -0.700. These question items were good at classifying students into high-achieving and low-achieving groups.

The item difficulty of both "common fractions" (MT24) and "counting and writing numbers" (MT23) was low, below -1.800. The item discrimination differed: around 0.700 for MT24 and around 1.110 for MT23. In "addition of numbers," the item difficulty of question items MT04 and MT05 was low, and item discrimination was high. However, the item difficulty of questions MT08 and MT09 was moderate, around 0, and item discrimination was also moderate, around 1. Items MT08 and MT09 involved carrying two- or three-digit numbers, while items MT04, MT05, MT06, and MT07 involved adding one-digit numbers without carrying.

5. Discussion

Incorporating item difficulty and item discrimination makes it possible to analyze students' abilities in a more granular way than with other methods. In English, for example, the analysis shows that students had strengths in the vocabulary of everyday life and were able to answer questions involving self-introduction and daily life. On the other hand, they had difficulty with tenses and stories.

Digging deeper reveals that with tenses, students had difficulty with the present tense (e.g., "Florence likes chicken.") (MT18). Only about 27 percent of students could choose the correct answer, "likes." Also, they faced difficulty with the past progressive form (e.g., "Erika was crying for her book.") (MT23). Approximately 38 percent of students could choose the correct answer, "crying." On the other hand, they were good with the present progressive form ("I am eating meat."), with approximately half choosing the correct answer, "eating" (ET17). In stories, students showed the ability to use context and simple sentence structure to match words. Still, as the cognitive requirements increased—applying these skills to match sentences, and interpreting sentences to match words and phrases—students had more difficulty. This result proves the necessity of improving students' skills in understanding and interpreting English sentences.

In mathematics, students had good skills in counting, writing numbers, addition, subtraction, and multiplication, whereas they could not perform tasks related to common fractions, time, and addition of money. In the question item "add fractions of the same denominators (1/6 + 4/6)" (MT25), approximately 38 percent of students could choose the correct answer of "5/6." However, approximately 33 percent of students chose the answer of "5/12," which adds the numbers in both numerator and dominator, showing that they did not understand the concept of fractions. This result is consistent with the finding of low item discrimination in question item MT24. Students tended to remember how to calculate the answers to given question items rather than understanding the meaning of the question items. In the question item "add decimal fractions without carrying (0.2 + 0.3)" (MT27), approximately 45 percent of the students could choose the correct answer of "0.5." Yet 36.4 percent chose the answer of "5," which ignores the decimal.

In the category of "time," question MT29 depicted an analog clock face showing 3:00 and asked students to choose the correct time. Only 27.0 percent of students chose the correct answer of "3:00," while approximately 43.0 percent answered "3:12," showing a lack of understanding of how to read an analog clock. Digital watches have become popular in Malawi, and almost no schools have clocks in the classrooms. Students rarely see an analog clock in their daily lives. This explains why the percentage of students choosing the correct answer was low, despite low item difficulty and low item discrimination. In "addition of numbers" (MT32 and MT33), students needed to read and understand a sentence before solving the mathematical problem it implied. That is, students required reading skills before using mathematical skills. Thus, item difficulty was high and item discrimination was low for question items MT32 and MT33. This issue was evident not only in word problems but also in other question items in mathematics. When students do not know the name of a geometric figure in English, for example, they cannot choose its name from a list. Hence, improving reading and understanding of English is a chief prerequisite for primary Malawi students to correctly answer mathematical question items.

6. Conclusion

This study employed IRT to analyze students' ability to acquire the Malawi primary curriculum. IRT helps reveal student strengths and weaknesses by classifying both the level of question items and student ability. Methodologically, the assumptions of IRT are stricter than those of the alternative analysis method, CTT. In particular, only questions that meet the criterion of unidimensionality (i.e., measuring only one ability) can be evaluated using IRT. Therefore, certain question items cannot be analyzed by IRT, even though they may have been designed to measure specific skills in the curriculum contents. In low-income countries, question items on skills tests often fail to be unidimensional and therefore cannot be analyzed using IRT. Given these obstacles, it is challenging to utilize IRT in the context of low-income countries.

Yet the findings of this study contribute to improving student skills. In English, students were found to have skills in the vocabulary of everyday life and to be able to answer questions involving self-introduction and daily life. On the other hand, they had difficulty with tenses and stories. It is therefore imperative to improve students' skills in syntax, reading sentences, and interpreting sentences into other words or phrases. Reading and understanding sentences also affected students' skills in solving word problems in mathematics. In Malawi, from grade 5 on, the instructional language is English. Therefore, improving English skills before students matriculate to higher grades, or simultaneously as they go through primary school, is necessary in order for them to acquire the knowledge and skills they need to learn in upper primary education.

In mathematics, students had skills in counting and writing numbers, addition, subtraction, and multiplication, but they were unfamiliar with common fractions, time, and addition of money. It is also necessary, then, to improve students' skills in these areas. In addition, skill improvement needs to focus not only on how to calculate but also on number sense and the meaning of fractions.

Acknowledgments

For financial support of this research, I am grateful for the Sasakawa Scientific Research Grant from the Japan Science Society extended for this research, the Hiroshima University Alumni Association, and the Japan Society for the Promotion of Science. I would like to thank officers in the Nkhata Bay District and teachers and students in Malawi who cooperated with this research.

References

Hambleton, R. K., Swaminathan, H. 1985. *Item Response Theory: Principles and Applications*. Kluwer-Nijhoff Publishing.

Hungi, N., Makuwa, D., Ross, K., Saito, M., Dolata, S., van Cappelle, F., Paviot, L., and Vellien, J. 2010. *SACMEQ III Project Result: Pupil Achievement Levels in Reading and Mathematics*. Working Document number 1. Paris: Southern and Eastern Africa Consortium for Monitoring Educational Quality.

Keeves, J. P. 1994. "Testing and the Curriculum," in *The International Encyclopedia of Education* (2nd ed.), 6321-6328. Oxford, UK: Elsevier Science Ltd.

Kunje, D., Selemani-Meke, E., and Ogawa, K. 2009. "An Investigation of the Relationship between School and Pupil Characteristics and Achievement at the Basic Education Level in Malawi." *Journal of International Cooperation in Education* **12**(1), 33-49.

Malawi Institute of Education (MIE). 2005a. *Malawi Primary School Syllabuses Standard 1*. Domasi, Malawi: MIE.

Malawi Institute of Education (MIE). 2005b. *Malawi Primary School Syllabuses Standard 2*. Domasi, Malawi: MIE.

Malawi Institute of Education (MIE). 2005c. *Malawi Primary School Syllabuses Standard 3*. Domasi, Malawi: MIE.

Malawi Institute of Education (MIE). 2005d. *Malawi Primary School Syllabuses Standard 4.* Domasi, Malawi: MIE.

Malawi Ministry of Education, Science, and Technology (MoEST). 2011. Educational Statistics, Educational Management Information System (EMIS) 2010. Lilongwe: MoEST.

Malawi Ministry of Education, Science, and Technology (MoEST). 2012. Educational Statistics, Educational Management Information System (EMIS) 2011. Lilongwe: MoEST.

Malawi Ministry of Education, Science, and Technology (MoEST). 2013. Educational Statistics, Educational Management Information System (EMIS) 2012. Lilongwe: MoEST.

Malawi Ministry of Education, Science, and Technology (MoEST). 2017. Educational Statistics, Educational Management Information System (EMIS) 2016. Lilongwe: MoEST.

Tomita, M., Muta, H. 2010. "Seitonogakuryokunieikyowoataeruinsinikansurukenkyu: Malwaikyowakoku・MALPwozireitoshite [Factors Influencing Pupils' Learning Achievements: From an MALP Survey in Malawi]." *Journal of International Development Studies* **19**(1), 67-80.

Tomita, M., Muta, H. 2012. "Gakkoyointokateikankyogagakuryokuniataerueikyoryokunohikaku: TeisyotokukokuniokeruHeyneman-Loxleysetunokensyotosonokaisyaku-Malawiwozireitoshite [The Comparison of the Influence of School Factors and Family-Related Factors on Learning Achievements: Confirmation and Policy Implication of HL Effect in Low-Income Countries from MALP Example in Malawi]." *Journal of International Cooperation in Education* **15**(1), 23-38.

UNESCO Institute for Statistics (UIS). 2018. *Data to Nurture Learning*. SDG 4 Data Digest 2018. Quebec: UNESCO-UIS.

Uwezo East Africa. 2014. *Are Our Children Learning? Literacy and Numeracy across East Africa 2013*. Nairobi, Kenya: Uwezo East Africa.

Watanabe, K. 2011. "Ekuadoruniokerusugakuwonichizyodekatuyosurunoryokunohennkanikansurukenkyu: PISA 2003 'Sugakutekiriterashi' nokomokuhannorironnniyorugakunensanityumokushitabunsekikara [A Study for the Variation of Achieving Mathematical Application Ability in Ecuador: An Analysis in PISA 2003 'Mathematical Literacy' Focusing on the Difference between Two Grades by Using Item Response Theory]. *Sugakukyoikukenkyu* **17**(2), 45-54.

Watanabe, K. 2014. "Sugakukyoikuniokerudaikibokyoikucyosanonizibunsekinikansurukenkyu: PISAcyoosawozireitoshite." Unpublished doctoral dissertation, Hiroshima University, Japan.

Appendix 1. Contents of the English test

Core elements	Curriculum topic	Level	Cognitive domain	No. of items
Reading	Vocabulary	Level 1	Match word and picture (Noun)	6
		Level 2	Match word and picture (Verb)	2
		Level 3	Match short sentence and picture (Three words)	3
		Level 4	Match short sentence and picture (More than three words)	2
Critical thinking and reasoning	Asking and answering questions	Level 1	Introduction	2
		Level 2	Daily life	1
Structure and use of language	Tenses	Level 1	Present form	4
		Level 2	Present progressive form	2
		Level 3	Past form, past progressive form, future form	3
		Level 4	Adverbs, prepositions, conjunctions	0
Reading	Stories	Level 1	Use context and simple sentence structure to match words	5
		Level 2	Use context and simple sentence structure to match sentences	3
		Level 3	Interpret sentence to match words and phrases	2
Total				35

Source: Developed by the author based on the Malawi Primary School Syllabuses (MIE, 2005a, 2005b, 2005c, 2005d) and SACMEQ framework (Hungi et al., 2010).

Appendix 2. Contents of the mathematics test

Core elements	Curriculum topic	Level	Cognitive domain	No. of items
Numbers, operations and relationships	Counting and writing numbers	Level 1	Count illustrated object	1
		Level 2	Understand number and digit	1
	Addition of numbers	Level 1	Add two or three-digit numbers and two-digit numbers without carrying	3
		Level 2	Add one-digit number and one-digit number with carrying	2
		Level 3	Add two or three-digit numbers and three-digit numbers with carrying	1
	Subtraction of numbers	Level 1	Subtract from two or three-digit numbers to two-digit numbers without borrowing	3
		Level 2	Subtract from two-digit numbers to one-digit number with borrowing	1
		Level 3	Subtract from three-digit numbers to three-digit numbers with borrowing	1
	Multiplication	Level 1	Multiplication of one-digit number by one-digit number	2
		Level 2	Multiplication of one-digit number by two-digit numbers	1
	Division	Level 1	Division of one-digit number by one-digit number	1
		Level 2	Division of two-digit numbers by one-digit number	2
	Common fractions	Level 1	Understand meaning of fraction	1
		Level 2	Add or subtract fractions of the same denominators	4
		Level 3	Add or subtract fractions of the different denominators	0
Space and Shape	Shapes	Level 1	Recognize shape	2
Patterns, functions and algebra	Patterns	Level 1	Sequences (from small numbers to large numbers)	1
		Level 2	Sequences (from large numbers to small numbers)	1
Measurements	Time, volume, areas	Level 1	Understand time, Measure volume	1
		Level 2	Find areas	2
Accounting and business studies	Addition of money, multiplication of money	Level 1	Addition of money	2
		Level 2	Multiplication of money	2
Data handing	Picture graphs	Level 1	Read picture graphs (One information)	0
		Level 2	Read picture graphs (Multiple information)	0
Total				35

Source: Developed by the author based on the Malawi Primary School Syllabuses (MIE, 2005a, 2005b, 2005c, 2005d) and SACMEQ framework (Hungi et al., 2010).