SMERU Technical Report

Comprehensive Reading and Mathematics Assessment Tool (CERMAT)

Niken Rarasati

Goldy Dharmawan

Arya Swarnata

Anisah H. Zulfa

Delbert Lim

smeru RESEARCH INSTITUTE

SMERU TECHNICAL REPORT

Comprehensive Reading and Mathematics Assessment Tool (CERMAT)

Niken Rarasati

Goldy Dharmawan

Arya Swarnata

Anisah H. Zulfa

Delbert Lim

Reviewed by
Thomas Coen and Daniel Suryadarma

Editor
Alia An Nadhiva

The SMERU Research Institute

March 2020

Comprehensive Reading and Mathematics Assessment Tool (CERMAT)

Authors: Niken Rarasati, Goldy Dharmawan, Arya Swarnata, Anisah H. Zulfa, Delbert Lim

Reviewed by Thomas Coen and Daniel Suryadarma
Editor: Alia An Nadhiva

Cover photo: Goldy Dharmawan

The SMERU Research Institute Cataloging-in-Publication Data

Niken Rarasati

Comprehensive Reading and Mathematics Assessment Tool (CERMAT)/ Niken Rarasati, et al: Editor Alia An Nadhiva.

--Jakarta: Smeru Research Institute, 2020
--67 p.; 27 cm.
ISBN 978-623-7492-28-3
ISBN 978-623-7492-27-6 [PDF]

1. Literacy 2. Numeracy 3. Learning assessment
I. Title

370.7 –ddc 23

Published by:
The SMERU Research Institute
Jl. Cikini Raya No.10A
Jakarta 10330
Indonesia

First published in March 2020

ACKNOWLEDGEMENTS

The development of the Comprehensive Reading and Mathematics Assessment Tool (CERMAT) is part of RISE Programme in Indonesia's research activities, which receive generous funding from the UK Department for International Development, Australia's Department of Foreign Affairs and Trade, and Bill and Melinda Gates Foundation. We would like to thank our partner, KIAT Guru, especially Sharon Lumbanraja and Indah Ayu Prameswari, for their generosity in providing us the first-stage pilot data and items so that we did not have to start this project from scratch. We would also like to express our appreciation to INOVASI for providing us their resources and expertise regarding numeracy and literacy. Our special thanks go to Ariyadi Wijaya from Yogyakarta State University; Budi Poniam, Dhitta Puti Sarasvati, Desyarti Safarini Deshinta Argaswari, and Namirah Fatmanissa from Faculty of Education, Sampoerna University, for their valuable inputs in our numeracy framework and items; and Dwi Firli Azhari from Singapore School, Jakarta, as our literacy specialist.

ABSTRACT

Comprehensive Reading and Mathematics Assessment Tool (CERMAT)

Niken Rarasati, Goldy Dharmawan, Arya Swarnata, Anisah H. Zulfa, and Delbert Lim

RISE Programme in Indonesia seeks to learn whether the education policy reforms carried out in Indonesia over the past twenty years have been successful in improving the country's student learning outcomes. To evaluate how well particular education policies help schools conduct better learning, we develop Comprehensive Reading and Mathematics Assessment Tool (CERMAT), a student learning assessment (SLA) tool that can assess specifically the reading and mathematics skills of students in Grades 1–9. CERMAT uses a framework based on the revised Bloom's Taxonomy for cognitive domains; stages of numeracy development; Fountas and Pinnell's Text Level Gradient that has been adapted into Indonesian literacy context; and Indonesia's 2006 and 2013 national curricula. We use the Rasch model to evaluate the quality of the psychometric properties of CERMAT. After the instrument was piloted and underwent revisions for three cycles, it finally reached a sufficient reliability score and contained items with a wide range of difficulty levels. Thus, CERMAT is sensitive enough to detect an increase in student abilities. There are two methods used for administering CERMAT: (i) individual and oral tests for students in Grades 1–3 and (ii) classical and written tests for students in Grades 4–9.

Keywords: literacy, numeracy, learning assessment

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

CERMAT		Comprehensive Reading and Mathematics Assessment Tool
COMPASS		Computer Adaptive Placement Assessment and Support System
DAPODIK	*Data Pokok Pendidikan*	The Ministry of Education and Culture's Education Data Centre
EGMA		Early Grade Mathematics Assessment
EGRA		Early Grade Reading Assessment
GRM		Graded Response Model
IB		International Baccalaureate
IEA		International Association for the Evaluation of Educational Achievement
INAP		Indonesian National Assessment Programme (the Indonesian version is AKSI or *Asesmen Kompetensi Siswa Indonesia*
INOVASI		Innovation for Indonesia's School Children
KIAT Guru	*Kinerja dan Akuntabilitas Guru*	Teacher Performance and Accountability
KTSP	*Kurikulum Tingkat Satuan Pendidikan*	2006 Curriculum
MoEC		Ministry of Education and Culture
NAEP		National Assessment of Educational Progress
PCA		Principal Component Analysis
PCM		Partial Credit Model
PIRLS		Progress in International Reading Literacy Study
PISA		Programme for International Student Assessment
PRIORITAS		Prioritizing Reform, Innovation and Opportunities for Reaching Indonesia's Teachers, Administrators, and Students
RISE		Research on Improving Systems of Education
SLA		student learning assessment
TIMSS		Trends in International Mathematics and Science Study
TNP2K		Team for the Acceleration of Poverty Reduction

UN	*Ujian Nasional*	National Examination
USAID		The United States Agency for International Development
USBN	*Ujian Sekolah Berstandar Nasional*	National-Based School Examination

I. INTRODUCTION

1.1 Background

Research on Improving Systems of Education (RISE) Programme in Indonesia focuses on (i) evaluating the impacts of education policy reforms at the national level, especially those relating to the improvement of teacher and teaching quality; and (ii) understanding how education policy reform initiated in innovative districts[1] can improve learning in the respective districts and spread to other districts. With these two focus areas, we aim to learn whether policies in question can successfully improve the country's student learning outcomes.

In Indonesia, the student learning outcomes are assessed using both high-stakes and low-stakes tests. The high-stakes test, known as the National Examination (UN), is carried out nationally at every end of junior and senior high school years (Grades 9 and 12, respectively). The UN contains questions related to topics covered in the curriculum. The questions tend to test the students' memory on the knowledge taught in school rather than their ability to apply the concept in everyday life or to think of it critically. Such examination encourages the teacher to "teaching to test" rather than focus on the learning process. The UN is designed to assess whether or not students have passed the minimum standard of the curriculum. Consequently, the test is dominated by items with low and medium difficulty levels. With only a few difficult items, this test is ineffective to capture a variety of abilities to detect score increase sensitively.

The low-stakes test is the Indonesian National Assessment Program (INAP); it is used to map, diagnose, and evaluate education progress in each province in Indonesia by looking at the performance of sampled Grade 4 students. The INAP, administered by the Indonesian Ministry of Education and Culture (MoEC) and was launched in 2016, is the only nationally standardized tool that assesses middle and higher order thinking skills. However, the tool only contains items for Grade 4 students. A test designed for Grade 8 and 10 students is still under development.

In addition to the assessment tools administered by the MoEC, there are other student learning assessment (SLA) tools adapted or developed by several development projects. In early 2016, KIAT Guru[2] and INOVASI[3] developed a numeracy and literacy assessment tool that covers materials for students in Grades 1 to 5. The tool follows the cognitive domain of TIMMS (Trends in International Mathematics and Science Study) and PIRLS (Progress in International Reading Literacy Study). The development of the two projects' tool involved three stages of instrument try-outs in 256 schools in the provinces of West Java, Special Region of Yogyakarta, Banten, and West Nusa Tenggara. Each of the try-out was followed by experts' review on psychometric properties and content of the items. Since RISE Programme in Indonesia focuses its research on basic education (Grades 1 to 9), we developed Comprehensive Reading and Mathematics Assessment Tool (CERMAT) by adopting the KIAT Guru and INOVASI's SLA for students in Grades 1 to 5 and added items for students in Grades 6 to 9.

[1]RISE Programme in Indonesia defines innovative districts as cities or regencies with innovative policies specifically aimed at improving their student learning outcomes. The districts are selected based on their positive trends of the National Examination (UN) score.

[2]KIAT Guru (an Indonesian abbreviation of Teacher Performance and Accountability) is a World Bank and the National Team for the Acceleration of Poverty Reduction's (TNP2K's) education project conducted in Indonesia's remote areas.

[3]INOVASI (Innovation for Indonesia's School Children) is a partnership between the governments of Indonesia and Australia to improve student learning outcomes in literacy and numeracy in diverse elementary schools and districts across Indonesia.

CERMAT aims to evaluate how well the particular education reforms or policies help schools in providing their students with the skills that could prepare them to live in this fast-changing world. Accordingly, the development of the CERMAT considers the following aspects:

a) It assesses not only procedural knowledge, but also the abilities to apply, analyze, and evaluate.
b) It has items that cover a wide range of students' abilities.
c) It is sufficiently sensitive to capture students' literacy and numeracy improvements.
d) The test has vertical continuity across grades to inform whether a student's ability is within the ability range of other students in his/her grade.

This technical report outlines the process of developing CERMAT, including the manual on how to implement and score the tests.

1.2 Theoretical Background

As stated in the previous section, there are four aspects to consider in the development process of CERMAT. An overview of the theory that we use as the basis of our instrument adaptation to meet the four aspects is presented in this section.

1.2.1 Cognitive Domain

In assessing the cognitive dimension of students' numeracy and literacy abilities, CERMAT refers to the revised Bloom's Taxonomy (Krathwohl, 2002). Krathwohl divided the taxonomy into two dimensions: knowledge and cognitive processes. This new classification makes the taxonomy easier to operationalize into a learning outcome assessment tool. The taxonomy provides six levels of cognitive skills, which starts from retrieving the relevant knowledge that a student has remembered to synthesizing the elements together to create a new original idea or product. Cognitive skills from the taxonomy that are relevant for elementary and junior high school students have been adapted by the International Association for the Evaluation of Educational Achievement (IEA) to the TIMSS' and PIRLS' frameworks (Mullis and Martin, 2013). The cognitive domain framework of CERMAT also adopted the TIMSS' and PIRLS' frameworks. For comparability and practical reasons, we use the TIMSS' and PIRLS' cognitive domain frameworks in CERMAT, respectively. The cognitive process dimension of Bloom's Taxonomy and its equivalent to the TIMSS' and PIRLS' frameworks (Mullis and Martin, 2013) is presented in the following table.

Table 1. Bloom's Cognitive Process Dimension and Its Equivalent to TIMSS and PIRLS

Bloom's Cognitive Process Taxonomy	Definition	TIMSS Framework for Numeracy	PIRLS Framework for Literacy
Remember • Recognizing • Recalling	Retrieving relevant knowledge from long-term memory	Knowing	Focus on and retrieve explicitly stated information
Understand • Interpreting • Exemplifying • Classifying • Summarizing • Inferring • Comparing • Explaining	Determining the meaning of instructional messages, including oral, written, and graphic communication	Knowing	Make straightforward inferences
Apply • Executing • Implementing	Carrying out or using a procedure in a given situation	Applying	Make straightforward inferences
Analyze • Differentiating • Organizing • Attributing	Breaking material into its constituent parts and detecting how the parts relate to one another and to an overall structure or purpose	Reasoning	Interpret and integrated ideas and information
Evaluate • Checking • Critiquing	Making judgments based on criteria and standards	Reasoning	Evaluate and criticise content and textual elements
Create • Generating • Planning • Producing	Putting elements together to form a novel, coherent whole, or make an original product	Not assessed in TIMSS and SLA	Not assessed in PIRLS and SLA

1.2.2 Content Domain

a) Numeracy

The numeracy assessment tool contains three primary mathematics content domains for elementary and junior high schools: numbers, geometry and measurement, as well as data and statistics. The tool's framework covers both the scope of mathematical content and also the complexity of mathematical skills. Referring to the Indonesian national curricula[4] and theories on numeracy development (Education Scotland, 2017; Booker et al., 2014; Van de Walle, Karp and Bay-Williams, 2013), we created a numeracy staircase that implies skills that a student needs to achieve before acquiring other skills with broader scope or higher complexity. In general, a student's ability in dealing with numbers is categorized into two broad levels.

[4]We specifically refer to two Minister of Education and Culture Regulations: No. 67/2013 on Basic Framework and Curriculum Structure of Elementary School/*Madrasah Ibtidaiah* and No. 68/2013 on Basic Framework and Curriculum Structure of Junior High School/*Madrasah Sanawiah*.

(1) **Level 1: The number concepts.** At this stage, children are required to have the sense of quantities and sizes and how they are represented by numbers; identify the place value of each digit in a number; and identify sequences or patterns.

(2) **Level 2: The number operations.** This level requires skills of using numbers to solve mathematical problems, such as addition, subtraction, multiplication, and division.

Apart from the above categories, the type of number also determines the complexity of the number concepts. We introduce numbers starting from whole numbers, followed by fractions, decimals, and negative integers. In each type of number, we always start with items asking about the concept (level 1), then followed by mathematical problems that include those types of numbers (level 2).

With respect to skills, the leveling framework includes comparing, ordering, counting, adding, and subtracting numbers; multiplying and dividing numbers; and solving word problems. The levels of number domain are presented in the following staircase.

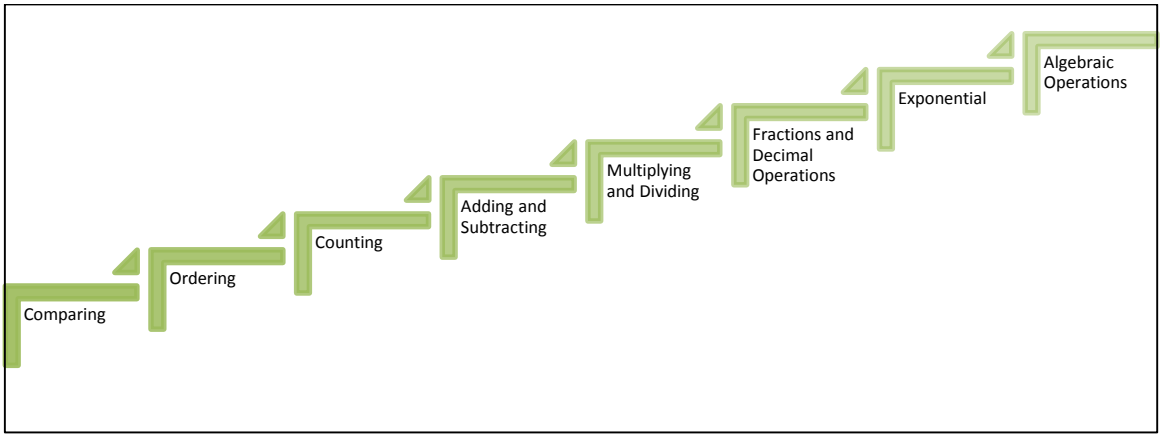


Figure 1. Number domain staircase

The geometric materials in CERMAT are put into levels based on van Hiele's five levels of geometric thinking (Van de Walle, Karp and Bay-Williams, 2013). The levels are labeled visualization, analysis, informal deduction, formal deduction, and rigor.

(1) **Level 0: recognition or visualization.** In this stage, children recognize shapes in terms of what they resemble (e.g., triangle resembles mountain).

(2) **Level 1: analysis.** In this level, children start analyzing and naming the geometric figures.

(3) **Level 2: abstraction (informal deduction).** Children start to see the relationships between properties and figures, categorize the shapes based on their characteristics, and formulate meaningful definitions (e.g., all squares are rectangles, but not all rectangles are squares).

The third and fourth levels, formal deduction and rigor, are too advanced for elementary and junior high school students and thus are not included in the test. Meanwhile, the measurement domain is categorized into three levels as illustrated in the following staircase.

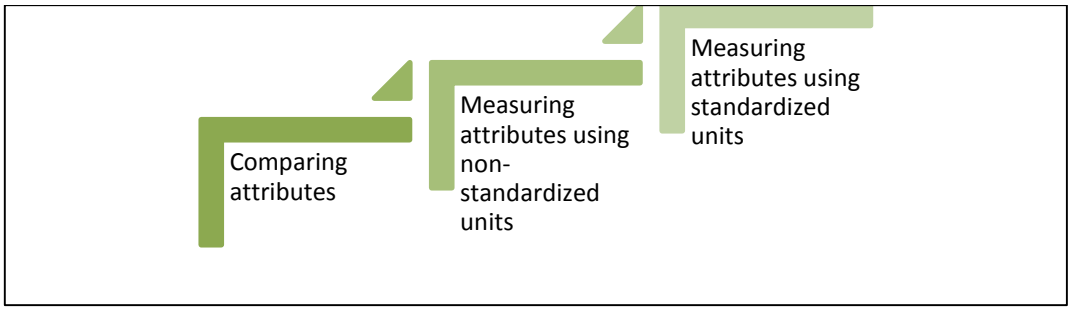


Figure 2. Levels in measurement

Data and statistics domain covers items on data representation in the form of pictograms and simple tables for elementary school level and central tendency as well as the basic concept of probability for junior high school level. The third and fourth levels for each content domain guide us in deciding materials to be assessed in each booklet of the test. The detailed leveling for each domain can be seen in Appendix 3.

b) Literacy

This literacy assessment tool is designed to capture students' ability to construct meaning from various forms of texts. During our review of the Indonesian national curricula[5], we found that they focus on two aspects: reading skills and text types. Since the function of the SLA is to assess students' ability to construct meaning from the text, we only adopt the skill's aspect of the curricula into the framework. The reading skills aspect in the national curricula for the first semester of Grade 1 focuses on early literacy skills, such as syllable recognition and word recognition. From the second semester of Grade 1 up to Grade 3, the students learn to comprehend the content of short narrations and be able to get insights from them. From Grades 4 to 6, students learn how to create linkages between ideas, organize information, evaluate texts, and give opinions. Domains related to reading skills for junior high school level (Grades 7 to 9) are similar to those for higher grades of elementary level. It is the complexity that increases in terms of the length of the passages and the complexity of the text genre (e.g., experiment report, popular science article, news report, and classic literature).

The domain in the national curricula that is related to text types only focuses on the genre, but not specifically on the complexity level of the text. Without the explanation of the complexity of the text, the text genre on the curricula draws questions on what type of text is appropriate for a particular level. Take the Grade 1 curricula as an example. It is written in the curricula that one of the texts to be given to students is a poem. Comprehending typical poems requires a skill to understand the implicit message of the text. As referred to Bloom's Taxonomy, interpreting the implicit meaning of a text is the second stage of the pyramid. Reflecting on this case, we refer to the international benchmark of text levelling, which has been adapted into the Indonesian context.

The Fountas and Pinnell Text Level Gradient is a widely used framework to determine the complexity of the text appropriate to a particular grade (Fountas & Pinnell Literacy™, 2016). The levels are constructed based on several factors:

(1) **Genre.** Each genre has characteristics and features. Understanding the features can help us to determine in which level the text belongs to.

[5]The National Curricula in this document refers to domains included in *both* the 2006 and 2013 Curricula.

(2) **Text structure.** The more complex the information is structured in the text, the higher the level of the text.

(3) **Content.** The content factor refers to how complex the topic is presented in the text.

(4) **Themes and ideas.** The number of ideas and themes presented in the text is one of the factors determining the difficulty level of the text.

(5) **Language and literary features.** This aspect refers to the style of the written language used in the text. The children's familiarity with the style appropriate to their age is considered to be the factor that determines the difficulty level of the text.

(6) **Sentence complexity.** Texts with simpler and more natural sentences are considered to be easier to process.

(7) **Vocabulary.** The difficulty level of the vocabulary is determined by the children's familiarity with the words. The more commonly used the words in the text are, the easier the text will be.

(8) **Words**. This factor refers to the number of difficult words used or commonly used words are repeated in the text.

(9) **Illustrations.** For young readers, illustrations that provide information related to the text can make the text easier to understand.

(10) **Book and print features**. This factor refers to the physical aspects of the text, such as layout, font size, and length.

PRIORITAS (Prioritizing Reform, Innovation and Opportunities for Reaching Indonesia's Teachers, Administrators, and Students), an education project in Indonesia funded by The United States Agency for International Development (USAID), has adapted the Fountas and Pinnell Text Levelling by taking into account the context relating to the above ten factors of text difficulty level (USAID PRIORITAS, 2015). The level ranges from A for texts with only one to two words per page to Z for complex texts, such as classic literature or popular science articles (see Appendix 2 for a more detailed leveling). Based on the review by literacy experts and findings of our SLA pilot study, we decided to use the range of reading levels that would be suitable for assessing students in Indonesia. The following table shows a range of text levels in each of the SLA booklets.

Table 2. Text Level Presented in the SLA Test Booklets

Grade Level	Level	Appear in the Test Booklet for						
		Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6	>=Grade 7
Pre-school	A–D							
Grade 1	E–J							
Grade 2	K–M							
Grade 3	N–P							
Grade 4	Q–S							
Grade 5	T–V							
Grade 6	W–Y							
Grades 7–9	Z							

1.2.3 The Use of the Rasch Model in Designing the Test

We used the Rasch model to evaluate the quality of psychometric properties of each item included in the test as well as the overall set of instruments (Boone, Staver and Yale, 2014). The model predicts the probability of a person with a particular ability to give a correct answer to an item with a particular difficulty level. The model can be written as follows:

$$p\left(x_j = 1 \middle| \theta, \beta_j\right) \frac{e^{(\theta - \beta_j)}}{1 + e^{(\theta - \beta_j)}}$$

where $p\left(x_j = 1 \middle| \theta, \beta_j\right)$ is the probability of the response of 1, θ is the person's ability level, and β is the item's difficulty level. Based on the above model, we understand that Rasch examines two attributes that can describe the quality of a test: item measure and person measure. The person measure is considered as a latent ability of the person, which is the closest to the true score of the person. This latent ability cannot be predicted if we estimate the score using a raw score calculation. This model assumes that the difficulty level of items is the most relevant parameter to predict the person's latent ability as well as to evaluate the quality of an instrument. Hence, difficulty level should be the focus of the Rasch analysis.

A good test is a test that has a distribution of difficulty level that is relevant to the purpose of the test. If a test is designed for selection purposes, it should contain difficult items. On the other hand, if a test is designed for measuring people's level of intelligence, for example, it should contain items with a wide range of difficulty levels to be able to assess low-intelligent and high-intelligent people.

Increasing the precision of a measurement is the focus of a test development using the Rasch approach. Using the Rasch model, we wanted to create a test that is adaptive to our targeted students' ability. To see how to fit the distribution of our items' difficulty level with the students' ability, we created an item-to-person map (known as the Wright Map) using the Rasch model. In the following figure, we can see the distribution of item difficulties in a *Bahasa Indonesia*[6] test for Grade 1 (on the right side) mapped against the distribution of the Grade 1 students' ability.

[6]Hereinafter referred to as Indonesian.

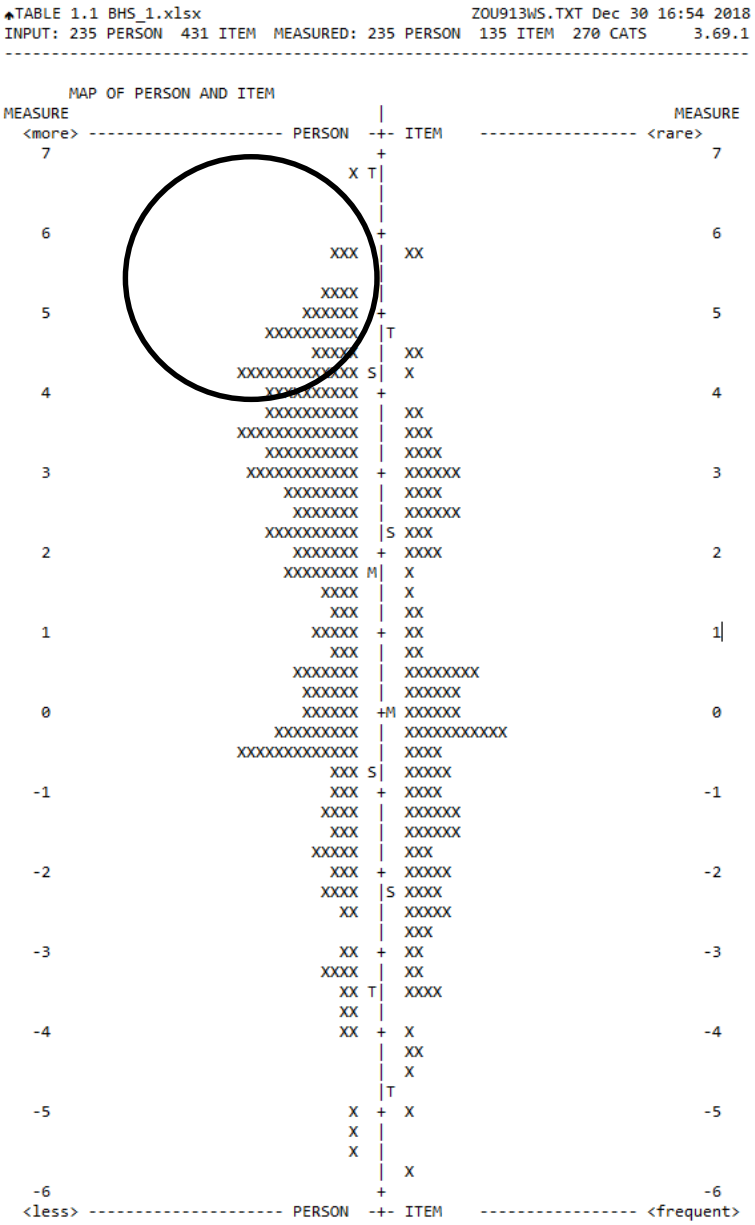


Figure 3. Item-person map for Indonesian - Grade 1

Figure 3 shows that the test items have a difficulty level that can sensitively assess students with various abilities, except a small group of students who have high ability (see the circle on the above figure). This figure suggests adding more difficult items to make the instrument more sensitive in assessing the high-ability students.

1.2.4 The Link of Booklet, Grade, and Other Assessment Instruments

To be able to compare the score of this SLA with the score obtained from another instrument, the two instruments need to have items in common, known as anchor items. The function of these anchor items is to equate the psychometric properties of a test with another test. The proportion of anchor items in each test booklet is approximately 30% of the total number of items in the booklet. The anchor items allowed us to equate the results among different grades using one metric. When choosing the anchor items, we made sure that there were no subdomains omitted from the booklet.

Apart from the anchor items between the modules, initially, we designed our instrument to be comparable with other general assessments, such as Programme for International Student Assessment (PISA), TIMSS, and INAP. Therefore, we adopted some items from other tests as an anchor to compare our instruments' result with the result of other assessments. However, there was insufficient information on the other assessments for the junior high school level. Consequently, our junior high school level instruments did not adopt the items from other assessment instruments.

a) Horizontal Equating Test Strategy

In CERMAT, we have several identical items to those of INAP. We can equate our test with INAP by putting the anchor items' psychometric properties obtained from the Rasch analysis of INAP to CERMAT as a fixed parameter. In the Rasch analysis of CERMAT, the psychometric properties of the rest of the items will be adjusted to these properties.

b) Vertical Equating Test Strategy

Similar to the horizontal equating test strategy, this strategy uses the same procedure to compare a test with a higher or lower test. In CERMAT, we have several items that anchor the Grade 1 test to the preliteracy or numeracy test, Grade 2 test to the Grade 1 test, Grade 3 test to Grade 2 test, and so on. The aim of employing the vertical equating test strategy is to be able to track students' progress across grades to see whether a particular student's ability is below, similar to, or above his/her age group.

II. TEST DESIGN

2.1 CERMAT's Cognitive and Content Domain Framework

2.1.1 Numeracy

To design a valid and efficient instrument, we studied instruments used by other institutions to know the composition of knowing, applying, and reasoning items. Some instruments, such as the UN and the National-Based School Examination (USBN), emphasize more on procedural knowledge than on middle or higher order thinking skills. On the other hand, international assessments such as PISA and TIMSS try to balance the composition between knowing, applying, and reasoning.

We have conducted multiple pilots to find the best-fit combination of knowing, applying, and reasoning in each grade based on Indonesian students' ability. The instrument for the lower-level elementary school is designed to include the domains of knowing and applying. Moreover, the instrument also emphasizes more on the knowing domain than the applying domain. Although the instrument for Grades 1, 2, and 3 items domain is similar, we increased the proportion of items to assess the applying skill in the Grade 3 instrument. This increase is based on a child's cognitive development theory.

For the higher-level elementary school, we designed half of the items to cover the knowing domain, while the rest are divided into the applying and reasoning domains. The small proportion of higher order thinking items in the test was decided based on findings from the pilot study where there are so many higher order thinking items that the students could not answer. At the same time, the findings also suggested that we add more variation of easy items. In the Grade 4 instrument, we introduced the reasoning item. The difficulty of the item increases gradually following the topic learned at a specific curriculum level. Thus, although the domain compositions for the higher-level elementary school are rather similar, the difficulty levels have been adjusted to be at the right level.

The junior high level school numeracy instrument emphasizes more on the applying domain following the Indonesian curricula properties. This is aligned with what is stated in the curricula, that the junior high school students are required to be able to apply mathematical concepts that have been learned during elementary schools in more complex problems. The cognitive domain framework of our numeracy instrument can be seen in the following table.

Table 3. Cognitive Domain Framework for Numeracy Test

Level	Grade	Knowing (%)	Applying (%)	Reasoning (%)
Elementary School	1	76	24	
	2	76	24	
	3	67	33	
	4	52	30	18
	5	42	39	18
	6	52	36	18
Junior High School	7	35	40	25
	8	35	40	25
	9	35	40	25

In constructing the framework for content domain for our numeracy test, we mapped two Indonesian curricula that are currently being used in schools: the 2006 Curriculum (KTSP) and the 2013 Curriculum. We only picked content domains that appear in both curricula and arranged those domains into grade level according to the numeracy staircase and literacy level (see section 1.2.2). By only including contents that are supposed to be taught in the classroom, the test will have good content validity. The following table explains the content domain framework of our numeracy test.

Table 4. Content Domain Framework for Numeracy Test

Level	Grade	Numbers (%)	Geometry and Measurement (%)	Data and Statistics (%)
Elementary School	1	85	15	
	2	85	15	
	3	60	30	10
	4	75	15	10
	5	70	20	10
	6	70	25	5
Junior High School	7	70	15	15
	8	60	20	20
	9	60	25	15

2.1.2 Literacy

In designing the literacy assessment instrument, we tried to adopt the framework used in PIRLS. We consider the compatibility and comparability of the framework in adopting PIRLS.

To develop valid and efficient instrument, we conducted multiple pilots. From our pilots, we learned that there were many students from Grades 1 to 3 in our sample group who could not recognize letters and words. Therefore, we decided to insert the early literacy components not only in the module for Grade 1 but also in modules for Grades 2 and 3. To assess the reading comprehension, we included three cognitive-process domains:

a) Focus on and retrieve explicitly stated information
b) Making straight-forward inferences
c) Interpret and integrate ideas and information

In addition to the three cognitive-process domains, we added another domain for the higher-level elementary school instrument, which designated students in Grades 4, 5, and 6. The cognitive-process domain is an ability to evaluate and criticize content and textual elements. We also added more proportion of items (Table 5) to assess the ability to interpret and integrate ideas and information. Following the Indonesian curricula, the length of the passages for the higher grades is similar, but the vocabulary richness as well as the topics' difficulties are gradually increasing.

The last group of the instrument covers the modules for Grades 7, 8, and 9. The passages belonging to this group are relatively longer and have a more advanced vocabulary. Most of the items in these booklets assess students' ability in making straightforward inferences from the given passages (40%–50%) and only contain a small proportion of items that ask students to retrieve explicitly stated information (10%–20%). We picked text genres according to how the Indonesian curricula

introduce various genres for each grade. We put more advanced vocabulary and more difficult items for the later grades.

Table 5. Cognitive Framework for Literacy Test

Level	Grade	Early Literacy Components (%)	Focus on and Retrieve Explicitly Stated Information (%)	Making Straightforward Inferences (%)	Interpret and Integrate Ideas and Information (%)	Evaluate and Criticize Content and Textual Elements (%)
Elementary School	1	10	40–50	25–30	15–20	
	2	10	40–50	25–30	15–20	
	3	10	40–50	25–30	15–20	
	4		10–20	15–20	30–60	10–30
	5		10–20	15–20	30–60	10–30
	6		10–20	15–20	30–60	10–30
Junior High School	7		10–20	40–50	30–40	10–20
	8		10–20	40–50	30–40	10–20
	9		10–20	40–50	30–40	10–20

2.1.3 Item Presentation

a) Numeracy

In designing our test items, we chose two types of item presentation for our numeracy section: multiple-choice and closed constructed-response items. The proportion of the two types of item presentation is presented in Table 6.

(1) Multiple-Choice Items

The multiple-choice problems require the students to choose the most correct and relevant answers from the choices provided. The multiple-choice problems in both elementary and junior high school modules have four choices. Most of the items in the elementary school modules use multiple-choice problem type, with a few closed constructed-response items. Learning from the experience in constructing items for Grades 1 to 6, we would need considerable resources to score the answers of the closed constructed-response item. Hence, when creating the items for Grades 7 to 9, we tried to minimize the number of constructed-response items. We then designed the mathematics tests to only consist of multiple-choice items.

(2) Closed Constructed-Response Items

In the numeracy instrument, the closed constructed-response problems require the students to write their answers in the provided fields. This type of item presentation may reduce guessing bias by the students. In our pilots, we found that many students tend to repeatedly answer A or D on a different item based on their guesses. Therefore, to minimize correct answers made by guessing, we introduced the close constructed-response item type.

Table 6. Allocation of Item Presentation in Numeracy Assessment by Grade

Level	Grade	Multiple Choice (%)	Closed Constructed-Response (%)
Elementary School	1	50	50
	2	50	50
	3	50	50
	4	50	50
	5	70	30
	6	70	30
Junior High School	7	100	
	8	100	
	9	100	

b) Literacy

In the literacy assessment instrument, we presented three types of item presentation: multiple-choice, closed constructed-response, and open constructed-response items. The proportion of the three types of item presentation is presented in Table 7.

(1) Multiple-Choice Items

The multiple-choice problems require the students to choose the most correct and relevant answers from the choices provided. The multiple-choice problems in both elementary school and junior high school modules have four choices.

(2) Closed Constructed-Response Items

The closed constructed-response problems require the students to give short and straightforward answers. For Grades 1 to 3, the students are required to answer verbally or in a written form. For Grades 7 to 9, the students are required to provide short and concise answers in a written form. Questions with this type of presentation usually have specific correct and explicit answers. In scoring this type of answers, we compared the correctness and relevance of the answers to the rubric. Different students may choose different words or terminology to answer the questions; however, if the meaning of the answer is relevant and correct, it can get full score.

(3) Open Constructed-Response Items

The open constructed-response problems require the students to provide concise and elaborative answers in a written form. To answer the questions that fall into this type of presentation, the students would have to explain or describe a process, situation, or object. Moreover, the questions may also require the students to mention items questioned in the problem set.

Table 7. Allocation of Item Presentation in Literacy Assessment by Grade

Level	Grade	Multiple-Choice (%)	Closed Constructed-Response (%)	Open Constructed-Response (%)
Elementary School	1	10–20	75–85	10–20
	2	10–20	75–85	10–20
	3	10–20	75–85	10–20
	4	80–85		15–20
	5	80–85		15–20
	6	80–85		15–20
Junior High School	7	80–90	10–15	10–15
	8	40–60	10–15	30–45
	9	40–60	10–15	30–45

2.2 Test Format and Scope

In general, our instrument comprises two types of assessment: literacy and numeracy. Both assessments cover modules of all grades in elementary and junior high school levels. We conducted several pilots to ensure that our instruments are valid and efficient. In practice, our instruments are designed to be concise and able to explain the learning outcome of the students at a specific grade using as few items as possible. The table below shows the number of items for each instrument type and specific grade.

Table 8. Number of Items by Grade and Instrument Type

Instrument Type	Level	Grade	Number of Items
Numeracy	Elementary School	1	20
		2	25
		3	25
		4	23
		5	23
		6	25
	Junior High School	7	12
		8	12
		9	12
Literacy	Elementary School	1	17
		2	23
		3	23
		4	19–21
		5	19–21
		6	19–21
	Junior High School	7	12
		8	12
		9	12

Table 8 shows the number of items in each booklet. All booklets were designed to be completed in less than one hour. For example, a Grade 7 student will have two hours to finish both numeracy and literacy modules. Since more time is needed to complete the junior high school problem set than the elementary school problem set, we assigned fewer items per module for the junior high school module.

2.3 Test Scoring

2.3.1 Item Scoring

For multiple-choice and close constructed-response items, the score is as simple as 1 for a correct answer and 0 for an incorrect one. In the Rasch model, we name this type of scoring Graded Response Model (GRM) where the variables are dichotomous (either 0 or 1). In the open constructed-response items, some questions require lengthy or multiple responses to have a full answer. In this case, sometimes the students can provide a partially correct response. To accommodate this condition, we use Partial Credit Model (PCM) where the variables can be polytomous (for our SLA, it is either 0, 1, or 2).

2.3.2 Estimating Students' Score

As discussed in the previous section, the Rasch model enabled us to estimate the true score or latent ability of a student. In estimating the score, the model predicted the maximum likelihood of a person to get a full score in one particular item by looking at the pattern of the responses to items in the whole test. Hence, this model could still predict the likelihood of someone getting a full score for a missing item. To create a test that is comprehensive yet still has an efficient number of items in each module, we divided items for each grade in several test booklets with anchor items that link one booklet to another.

III. DEVELOPMENT OF THE TEST

3.1 School Index for Sample Selection

As mentioned before, we aimed at creating a test that is sufficiently sensitive to capture students' literacy and numeracy improvement in RISE's targeted population. Hence, we designed each booklet to have items with a wide range of difficulty levels that can capture the performance of students with low and high ability.

To obtain a sample of students with a high variation of abilities that could represent the range of abilities of RISE studies' targeted population, we generated a school index using Principal Component Analysis (PCA) based on the available schools' characteristics data that potentially have linkage to the learning outcome. Given the limited national data available, we only had teacher-student ratio, classroom quality, Internet access, the proportion of civil servant and contract-based teachers, library quality, access to clean water, accreditation score, and number of special education teachers as variables relating to learning outcome. The data was taken from *Data Pokok Pendidikan* (DAPODIK, The Ministry of Education and Culture's Education Data Centre). We excluded schools in two regions—Maluku and Papua—that would not be selected as RISE Indonesia study area. From the PCA, we took the first component as the school quality index. We grouped the schools into three categories: high-, medium-, and low-quality schools. The high-quality schools are those with the 20% highest index. It was followed by the medium-quality schools, whose indexes are in the fourth and third quintiles, and low-quality schools that are in the first and second quintiles. The mean of the quality index in each category is presented in the table below.

Table 9. Mean of the Index in Each Quality Group

School Category	Mean of Quality Index	Min	Max
High quality	0.937	0.496	2.835
Medium quality	0.003	-0.526	0.496
Low quality	-1.655	-27.230	-0.527

For each of our pilots, we picked schools that had the closest index to the mean of each of the category. Our data shows that the means of SLA scores resemble the quality index of the schools (Figure 4). It can be implied that the chosen variables from DAPODIK can describe the quality of the schools.

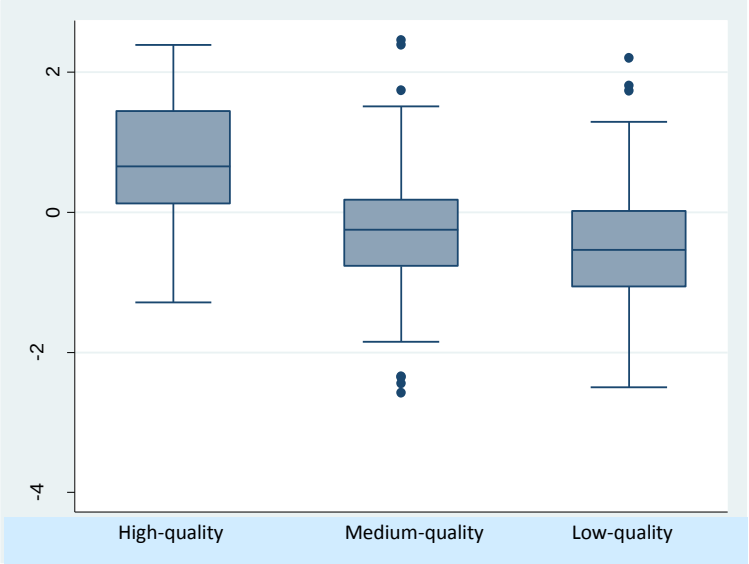


Figure 4. SLA scores by school quality category

3.2 Item Development

Revisiting and Revising the SLA Framework

The development of CERMAT for Grades 1 to 9 was an adaptation process from the SLA tools developed by INOVASI and KIAT Guru. We started by mapping the 2006 and 2013 National Curricula as the currently used curricula in Indonesian public schools. To enrich our reference on content domain, we also reviewed content domain from other standardized tests, e.g., National Assessment of Educational Progress (NAEP), the Computer Adaptive Placement Assessment and Support System (COMPASS), PIRLS, and INAP. With the help of literacy and numeracy experts, we decided the proportion of content and cognitive domain for each grade level. Then, we mapped the good-quality items from INOVASI's and KIAT Guru's SLA tools to our framework and created new items as needed. INOVASI's and KIAT Guru's item banks consist of items designed for Grades 1 to 5 and draft items designed for Grade 6; all were piloted in six schools in West Lombok in April 2018. The schools represent the three categories of quality.

Findings from the first pilot show that the Grades 1 to 3 tests have good quality. The difficulty level of the items is widely distributed and fits with the ability of the targeted sample. Hence, we only did minor revisions to the tests for Grades 1 to 3 and the items were not included in the subsequent pilot. We used the Wright Map (Figure 5) to help us determine which items needed revising or what kind of items we needed to add. The left side of the graph shows the distribution of students. The higher the position of a student, the higher his or her ability is. The right side of the graph shows the distribution of items based on their difficulty. The items placed at the bottom of the graph are easier than those placed at the top. The following graph exemplifies the distribution of Indonesian items for Grade 3 students that captures all levels of students' ability, even though there are still items that are too easy and need to be removed from the test.

Figure 5. Wright Map of Indonesian items and students' abilities - Grade 3

In contrast, we found that the distribution of item difficulty in tests for Grade 4 to 6 had yet to capture the range of students' ability. Take a look at the following graph for example.

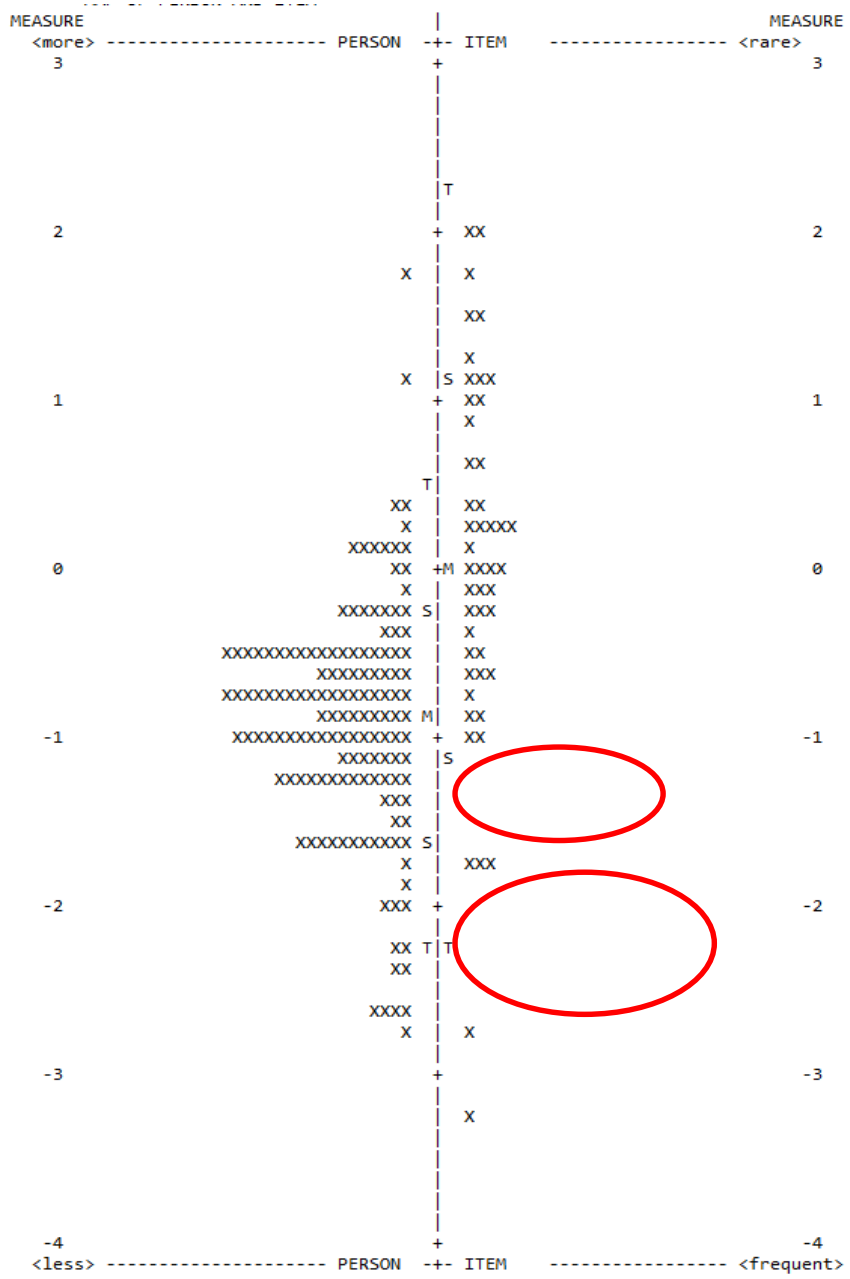


Figure 6. Wright Map of mathematics items and students' abilities - Grade 4

Figure 6 shows that the overall test was too difficult for the students. There are also blank areas (marked in red circles) where we can see a group of students whose abilities could not be captured by items with the appropriate level of difficulty. After looking at these findings, the numeracy and literacy expert revised the items to make their difficulty level appropriate to the ability of the targeted students.

In July 2018, we conducted the second cycle of the pilot study. This cycle also included booklets for students in Grades 7 to 9 with mathematics items constructed by a mathematics expert from Sampoerna University. From the second cycle, we found that booklets for Grades 4 and 5 needed minor revisions; while booklets for Grades 6 to 9 required more revisions. After the booklets were revised, we tested them again in September 2018. The last cycle of the pilot study was followed by the finalization of the whole SLA instrument.

IV. PSYCHOMETRIC PROPERTIES OF THE TEST

This chapter describes the quality of the test in terms of its psychometric properties. The following subsections present data that shows how valid and reliable the test is.

4.1 Test Validity

4.1.1 Content Validity

Validity tells us how accurate our measurement is to assess a particular aspect. The validity of a test relies more on a judgment of whether the content of the test is already in line with the purpose of the test rather than the outputs of the statistical test. To make sure that our SLA tool can validly assess what students are supposed to learn at school, we referred the content domain of the test to the national curricula.

All of the items that we created went through a content review process by literacy and numeracy experts. The experts are teachers and lecturers of Indonesian and mathematics education who have experience in teaching Grades 1 to 3, 4 to 6, or 7 to 9. They also have at least a formal training certificate or experience in creating higher order thinking items for internationally standardized assessments or curricula, e.g., Early Grade Reading Assessment (EGRA), Early Grade Mathematics Assessment (EGMA), and International Baccalaureate (IB) Assessment. During the development process, they were involved in formulizing the content and cognitive domain framework, creating items, and reviewing items based on findings from the psychometric analysis.

4.1.2 Outfit and Infit Test

To support the above judgment, we also checked the validity of the SLA tool by estimating the outfit and infit mean squares of the items. The outfit test shows how much noise the items measure. The expected value is 1.0. The value that is too far below 1.0 indicates that the observations are too predictable, while if it is too far above 1.0, it indicates unpredictability. Meanwhile, the purpose of estimating the infit index is to see whether there is an unpredictable pattern captured in the test (e.g., there is a group of people with low ability who have higher probability to answer difficult items). The good infit index ranges from 0.5 to 1.5. The following table shows that the results of the outfit and infit tests of our SLA tool indicate good validity.

Table 10. Outfit and Infit Indices

Test	Outfit		Infit	
	MNSQ[a]	SD[b]	MNSQ	SD
Numeracy				
Grade 1	1.15	0.64	1.01	0.12
Grade 2	1.15	0.85	0.99	0.14
Grade 3	1	0.46	0.98	0.15
Grade 4	1.09	0.48	1	0.2
Grade 5	1.03	0.32	0.98	0.17
Grade 6	1.11	0.75	1.02	0.24
Grade 7	1	0.27	1.01	0.14
Grade 8	1.06	0.45	1	0.16
Grade 9	1.08	0.39	1.04	0.2
Literacy				
Grade 1	1.12	0.91	0.99	0.26
Grade 2	1.06	0.96	1	0.24
Grade 3	0.97	0.76	0.99	0.26
Grade 4	1.22	1.09	1.01	0.26
Grade 5	1.01	0.53	1	0.45
Grade 6	0.94	0.24	1	0.1
Grade 7	1.11	0.85	1	0.19
Grade 8	1.02	0.42	0.98	0.14
Grade 9	1.04	0.5	0.99	0.16

[a]MNSQ: mean square
[b]SD: standard deviation

4.2 Test Reliability

We used the Rasch model to estimate the reliability of the test. In the classical approach, reliability is only seen as how consistent the sample could answer the items correctly by performing a correlation between an item with the rest of the items. In the CERMAT design, where we intentionally put items with extremely low and high difficulty levels, the classical reliability estimation will result in low item-total correlation for the extreme items. In Rasch model, the person ability and item difficulty were taken into account in estimating the consistency of the pattern. The item reliability score indicates the consistency of the difficulty level of the items after they were tested in different levels of students' ability. The person reliability score indicates that the consistency of students with a specific ability will always get a particular latent ability score (e.g., high-ability students will consistently have a high score). The expected reliability score is above 0.75.

In addition to reliability score, we also have the separation index. Person separation index shows how well the test can differentiate students into groups of ability. For example, the separation index of 2.9 could divide students into three groups of ability: high, medium, and low. Likewise, the item

separation index shows how well the items can be separated into a particular number of levels (e.g., index of 5.47 shows five levels of difficulty).

The following table shows that the items in our tests are reliable (0.8–0.98). However, a low person reliability score can still be found in numeracy tests for Grades 1, 3, 7, 8, and 9; and literacy test for Grade 8. By looking at the person separation index, we can see that this may be due to the small variation of the sampled students' ability (separation indexes are around 1). The implication of low person reliability score and separation index is that the test cannot be used to classify students who attend the same grade into many levels of competence. However, the wide range of item difficulties with a high item separation index make this test sufficiently sensitive for evaluative purposes.

Table 11. Item-Person Reliability Scores and Separation Indices

Test	Person		Item	
	Reliability	Separation Index	Reliability	Separation Index
Numeracy				
Grade 1	0.68	1.45	0.98	6.55
Grade 2	0.75	1.72	0.98	6.51
Grade 3	0.66	1.4	0.97	5.47
Grade 4	0.71	1.56	0.95	4.47
Grade 5	0.73	1.66	0.95	4.15
Grade 6	0.75	1.71	0.96	4.83
Grade 7	0.68	1.44	0.92	3.32
Grade 8	0.72	1.59	0.92	3.3
Grade 9	0.65	1.35	0.8	2.02
Literacy				
Grade 1	0.98	6.39	0.98	8.02
Grade 2	0.97	5.5	0.98	6.91
Grade 3	0.94	3.92	0.96	5.13
Grade 4	0.8	1.98	0.95	4.59
Grade 5	0.79	1.93	0.82	2.12
Grade 6	0.57	1.16	0.94	3.8
Grade 7	0.78	1.89	0.93	3.6
Grade 8	0.7	1.52	0.92	3.29
Grade 9	0.71	1.56	0.89	2.92

LIST OF REFERENCES

Booker, George, Denise Bond, Len Sparrow, and Paul Swan (2014) *Teaching Primary Mathematics*. 5th ed. Frenchs Forest: Pearson Australia.

Boone, William J., John R. Staver, and Melissa S. Yale (2014). *Rasch Analysis in the Human Sciences*. 1st ed. Dordrecht: Springer Netherlands.

Education Scotland (2017). *Benchmarks: Numeracy and Mathematics*. Livingston: Education Scotland.

Fountas & Pinnell Literacy™ (2016). *Fountas & Pinnel*. [online] <https://www.fountasandpinnell.com/Authenticated/ResourceDocuments/FP_FPL_Chart_Factors-Related-to-Text-Difficulty.pdf> [4 December 2018].

Krathwohl, David R. (2002) 'A Revision of Bloom's Taxonomy: An Overview.' *Theory Into Practice*, 41 (4): 212–218.

Mullis, Ina V.S. and Michael O. Martin (eds.) (2013) *TIMSS 2015 Assessment Frameworks*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College and International Association for the Evaluation of Educational Achievement.

USAID PRIORITAS (2015) *Buku Sumber untuk Dosen LPTK: Pembelajaran Literasi di Sekolah Dasar/ Madrasah Ibtidaiyah* [online] <https://prioritaspendidikan.org/file/Pembelajaran_Literasi_SD_di_LPTK1.pdf> [4 December 2018]

Van de Walle, John A., Karen S. Karp, and Jennifer M. Bay-Williams (2013) *Elementary and Middle School Mathematics: Teaching Developmentally*. 8th ed. Boston: Pearson.

Government Laws and Regulations

Regulation of the Minister for Education and Culture No. 67/2013 on Basic Framework and Curriculum Structure of Elementary School/Madrasah Ibtidaiah.

Regulation of the Minister for Education and Culture No. 68/2013 on Basic Framework and Curriculum Structure of Junior High School/Madrasah Sanawiah.

APPENDICES

APPENDIX 1

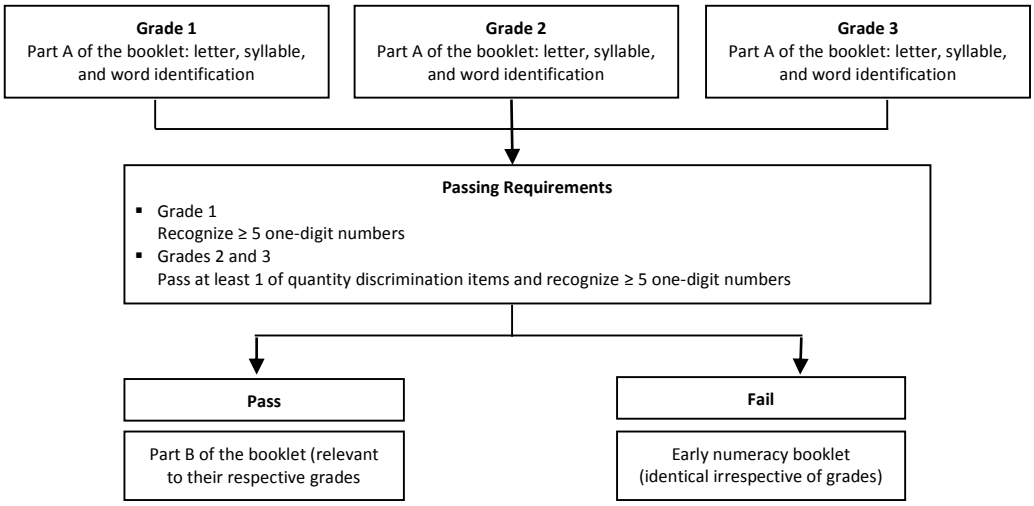
Test Administration Manual

1. Test Flow

We used three booklets when we administered the tests for each of Grades 1 to 3; each test had a different purpose. Booklet A serves as a screening tool to ensure the students have the required skills to attempt Booklet B, which will then measure their skills respective to their grade. Failure to pass Booklet A indicates that the students do not possess the required literacy skills to be able to understand the questions in Booklet B. To be able to capture the skills possessed by the students, we gave them a distinct booklet applicable to their capability; the early numeracy and early literacy booklets serve the purpose.

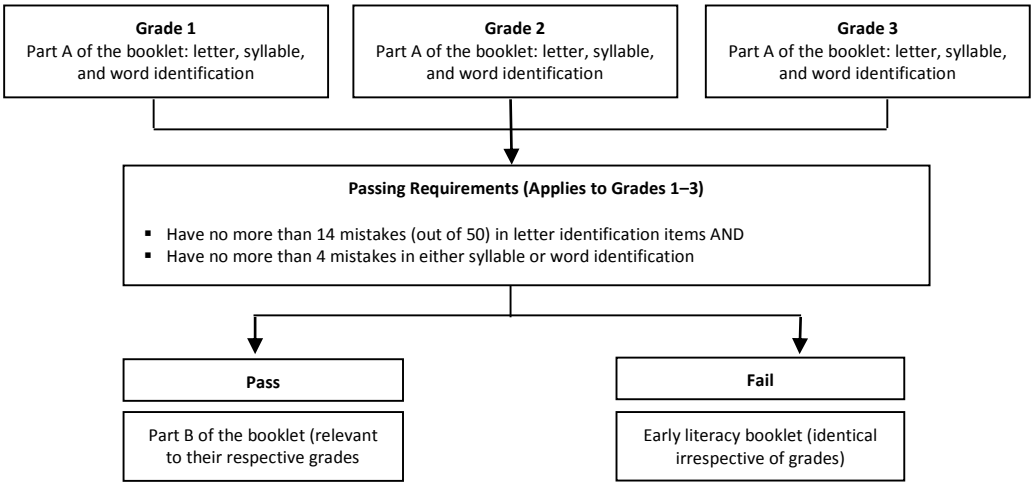
a) Numeracy for Grades 1 to 3

The following chart outlines the flow of the mathematics tests for Grades 1 to 3.



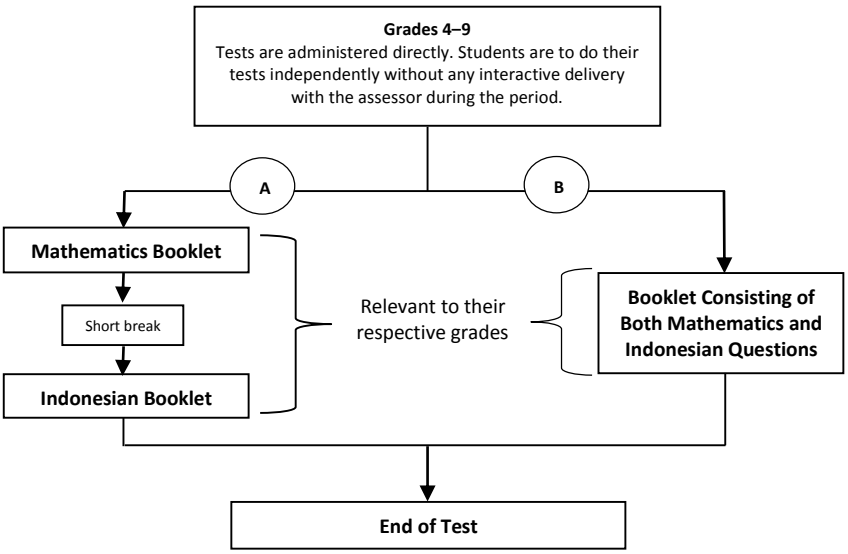
b) Literacy for Grades 1 to 3

The following chart outlines the flow of the Indonesian tests for Grades 1 to 3.



c) Grades 4 to 9

The following chart outlines the flow of the mathematics and Indonesian tests for Grades 4 to 9.



Flow A is used when the total number of questions may be overwhelming for the students, impacting their capacity to answer the questions. Mathematics generally precedes Indonesian due to the likelihood of Indonesian being more taxing for the students because of the higher word count. Hence, the order prescribed is aimed at minimizing any adverse effects between the two tests.

Flow B is used when the total number of questions is appropriate to be bundled in one single test. Bundling together the two disciplines allow flexibility for the students to decide which subject to do first according to their individual capabilities.

2. Test Procedure

a) Test Preparation

(1) General

In conducting the exam, a conducive environment should be arranged to the best of the circumstances. An ideal condition includes

(a) an uncrowded room or a semiclosed area (if a room is unavailable),

(b) suitable lighting,

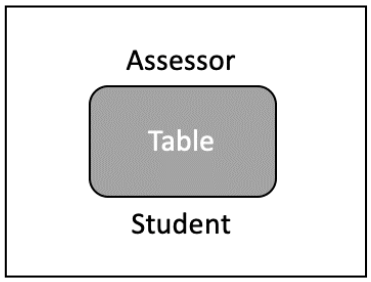
(c) good ventilation, and

(d) a venue free of noise and interference.

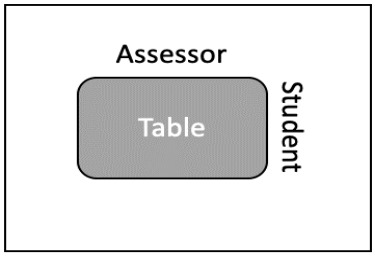
The assessor should also make sure that there are no materials, such as posters, relevant to the exam visible in class; coverage or removal may be necessary.

(2) Grades 1 to 3

As early grade tests are conducted one-on-one, the assessors should arrange their seating appropriate for such type of assessment, as outlined by the following layout options. Also, they should ensure that the distance between any two students doing the assessment is not intrusive of each other, as shown in the following layout.



Layout 1



Layout 2

The following equipment should be prepared and ensured to be correct:

(a) Tablet with test materials

(b) Tests' show card

(c) Required materials for each subject
   i) Mathematics: Blank paper and pencil
   ii) Indonesian: Dictation sheet

Considering the interactive nature of the tests, the assessors should build rapport with the students prior to the test; this can be achieved by asking several simple questions and making warm eye contact. Finally, they should ensure that the students are fully prepared to start the test.

(3) Grades 4 to 9

The assessors should coordinate directly with the relevant field observers for scheduling purposes (assessment to be done prior to or after the classroom observation). Permit should also be obtained prior to the assessment from the teachers whose teaching period is to be replaced by the assessment.

The following equipment should be prepared and ensured to be correct:

(a) Question sheets. Several versions may be employed to prevent cheating.

(b) Answer sheets

(c) Consent form to be filled by the homeroom teacher

(d) Stationary

b) During the Test

(1) Grades 1 to 3

Upon starting, the assessors should consult with the students of their preferences on which exam they would like to take first. Then, they should ensure that the show card is faced directly to the student.

During the Mathematics Assessment

Part A: Reading Numbers
Instruction for students: Clearly read aloud the following numbers from left to right.
Instructions for assessors:

(a) Always begin with the assumption that students are unable to read the text.

The students are to be shown figures or numbers on the show card, while the questions are being read out.

(b) The sample page should be shown to the students prior to starting the actual test and the following procedures should be followed.

i) Ask the students to read the sample questions.

ii) Emphasize correct answers with phrases such as "Yes, that is correct! This is the number…"

iii) Correct their mistakes when wrong with phrases such as "This number is called…"; emphasize the pronunciation of "tens" and "hundreds" ("*puluh"* and "*ratus"*).

(c) Give the following instruction: "I will keep counting the time you take to read out the numbers. So please attempt to read as fast and clearly as possible. You are allowed to point at the numbers as you are reading them".

Early Numeracy Test: there are two parts involved in the assessment.

(a) Reading numbers with illustrations.

i) The students are requested to read the numbers pointed by the assessors.

ii) The assessors are allowed to express the number with their fingers.

(b) Pointing numbers read out by the assessors

Assessors are allowed to provide an example of the instruction by saying phrases such as "Can you point at a pencil? Good. Please do the same for the numbers that I am about to read out".

Part B: The students should attempt to answer the questions independently, the assessors may only assist in reading out the questions.

During the Indonesian Assessment

Part A: Reading Letters, Words, and Syllables

Instruction for the students: Clearly read aloud the following letters from left to right.

Instructions for the assessors:

(a) The sample page should be shown to the students prior to starting the actual test and the following procedures should be followed:

i) Ask the students to read the sample questions.

ii) Emphasize correct answers with phrases such as "Yes that is correct! This is the letter…".

iii) Correct their mistakes when wrong with phrases such as "This letter is…".

(b) Give the following instruction: "I will keep count the time you take to read out the letters. So please attempt to read as fast and clearly as possible. You are allowed to point at the letters as you are reading them".

Note: The students are allowed to pronounce each alphabet differently from the Indonesian pronunciation rules, EXCEPT for F, P, and V.

Early Literacy Test: there are three parts involved in the assessment.

(a) Reading letters with illustrations.

i) The students are requested to read the numbers pointed by the assessors.

ii) The assessors are allowed to express the number with their fingers.

(b) Pointing letters read out by the assessors.

The assessors are allowed to provide an example of the instruction by saying phrases such as "Can you point at a pencil? Good. Please do the same for the numbers that I am about to read out".

(c) Listening

i) The assessors should read out the passage with a clear articulation. Do not read too fast.

ii) The assessors should ensure that the students are ready to pay attention before reading.

iii) After the passage has been read, the assessors should immediately instruct: "Now please answer the following question…"

The assessors should avoid giving out too much instruction between the passage reading activity and the question asked.

Part B

Dictation instructions to be read to the students and to be followed by the assessors:

(a) "I will read out a sentence THREE times."

(b) "You should listen carefully on the first reading. After that, I will provide you with a pencil and paper."

(c) "Then, I will read out the sentence for the second time. Please write what you heard on the paper. I will provide you with a writing time of 15 seconds."

(d) "Finally, I will read out the same sentence for the third and final time for you to check on your answer."

(2) Grades 4 to 9

Instruction for the assessors:

(a) Introduce yourself.

(b) Explain clearly that the assessment is meant to be a form of practice for mathematics and Indonesian, and WILL NOT influence their school grades in any way.

(c) Distribute the answer sheets.

(d) Guide students in filling their identity and other relevant pieces of information in the answer sheet. Make sure that all of the students have filled all the required fields before distributing the question booklets.

(e) Explain that they are about to be given a booklet consisting of both mathematics and Indonesian questions, of which they are free to choose the particular questions to do first according to their preference and capabilities. They are allowed to immediately move on to the next set of questions after they are finished on a subject.

(f) Clearly explain how to fill the answer sheets and emphasize that the question booklet MUST NOT be used for any writing or rough work.

(g) Distribute question booklets. If different versions are used, make sure that no two students sitting side by side receive the same version of questions.

c) Rules

(1) The questions are to be attempted individually without any form of cooperation between students.

(2) The students are allowed to ask clarificatory questions by raising their hand towards the assessor.

(3) The students are allowed to go to the bathroom.

(4) The students cannot leave the room before the whole duration of the exam has elapsed. They should quietly remain in their seats even if they finish early.

d) Test Duration

The SLA is designed as a power test rather than a speed test. In a power test, we try to eliminate factors that can reduce the ability of the test to capture the students' latent score. In this case, the ability to read or count faster is not a part of the skills that the instrument aim at assessing. Hence, the test has no time limit. However, for practical matter in a classical test (Grades 4 to 9), we give the students time limit that is ten minutes more than the time that they would actually need to complete the whole test booklet. If a booklet consists of fifteen reading questions, the average time of students to complete the test is 30 to 35 minutes. Hence, the given time limit is 45 minutes. The same amount of time is also applied to a booklet consisting of 20 mathematics questions. It takes around 30 minutes for students to finish the 20 item mathematics test.

APPENDIX 2

The Fountas and Pinnel Text Levelling

Text Level	Indicators
A	• Recognize letters and their sounds • Point to words while reading • Use picture to support understanding • Know the difference between words and pictures • One sentence per page with simple words • Read easy, high frequency words
B	• Follow a sentence over 2 lines of text • Continue to point to words while reading • Recognize pattern throughout story • Reread to fix reading mistakes • Read, easy high frequency words
C	• Read simple stories with 2–6 lines of text on page • Notice repeated lines and phrases • Begin to follow text with eyes, rather than pointing • Use strategies to help understanding • Begin to correct reading mistakes • Read easy, high frequency words
D	• Read fiction and simple nonfiction • Continue to follow text with eyes, rather than pointing • Read text with fewer lines of repeated words • Read compound words (e.g., newspaper, sandbox) • Continue to correct reading mistakes • Read easy, high frequency words
E	• Read books with 3–8 lines of text • Follow text with eyes, rather than pointing • Read texts that require more attention for understanding • Follow punctuation correctly • Take apart long words • Rely on meaning from the text, rather than pictures • Read fluently • Read easy, high frequency words
F	• Begin to understand genres (fiction, nonfiction, etc.) • Read and understand dialogue in text • Read words with multiple syllables • Automatically read high frequency words
G	• Continue to understand different genres • Read 3–8 lines of text per page; text is smaller • Read difficult words by using letter/sound information, thinking of familiar words, taking apart words • Read text with a few challenging vocabulary words • Automatically read high frequency words

Text Level	Indicators
H	• Read longer text with more challenging vocabulary • Read difficult words by using letter/sound information, thinking of familiar words, taking apart words • Automatically read high frequency words
I	• Read short texts (8–16 pages) and easy chapter books (40–60 pages) • Understand longer sentences of more than 10 words • Read many texts silently, without pointing to words • Automatically read a large number of high frequency words (from all previous levels and more)
J	• Read many types of texts (informational texts, short chapter books, simple biographies) • Understand a large number of longer sentences • Use strategies to figure out hard words (go back and reread, use picture clues, find smaller words inside the bigger word, etc.) • Automatically read a large number of high frequency words (from all previous levels and more)
K	• Read many types of texts (biographies, informational texts, realistic fiction stories, fantasy stories, traditional literature, simple texts) • Read many illustrated chapter books • Must remember many details • Understand dialogue and the use of quotation marks (“ “) • Books have many characters that change a little in the story • Read stories with diverse cultures • Use strategies to figure out hard words (go back and reread, use picture clues, find smaller words inside the bigger word, use word parts like prefixes/suffixes, etc.) • Automatically read a large number of high frequency words (from all previous levels)
L	• Read easy chapter books with less pictures • Read short informational and fiction books • Read slower or faster depending on the book • Learn new concepts through reading • Use what they already know to help their reading • Use pictures and text to help understand • Connect known facts to new information • Understand difficult ideas • Understand a large number of words (plurals, contractions, possessives, multi-syllable words, content-specific words, technical words) • Understand difficult sentences
M	• Know the characteristics of different genres (realistic fiction stories, fantasy stories, informational text, traditional literature, biographies, etc.) • Read fiction chapter books, such as series books (e.g., Junie B. Jones) or mysteries • Read fiction texts that have many characters that change in the story • Read shorter nonfiction texts on one topic • Understand difficult sentences
N	• Process short fiction stories, chapter books, short informational texts, series books (e.g., most Amber Brown books) or mysteries • Read fiction texts that have many characters that change in the story • Read nonfiction texts on many related topics • Automatically use strategies (find smaller words inside the bigger word, use word parts like prefixes/suffixes, etc.) • Read and understand descriptive words

Text Level	Indicators
	• Understand difficult sentences
O	• Know the characteristics of most genres • Read chapter books, shorter informational texts, mysteries, series books, books with sequels, or short stories • Read fiction stories with many characters that change throughout the story • Read nonfiction texts that give information on many related topics • Understand difficult sentences and words • Figure out new vocabulary words by using clues
P	• Know the characteristics of most genres • Read chapters books, shorter informational texts, mysteries, series books, books with sequels, short stories • Read fiction stories with many characters that change throughout the story • Read nonfiction texts that give information on new topics • Understand mature themes (race, language, culture, etc.) • Make sense of new vocabulary words
Q	• Automatically read and understand characteristics of most genres, including biographies on new topics, chapter books, shorter informational texts, mysteries, series, books with sequels, short stories. • Read fiction stories with many characters that change throughout the story • Make sense of new vocabulary words • Look for information in pictures, photographs, maps, charts, etc. • Can break words into syllables • Understand texts with different layouts • Look for information in pictures, photographs, maps, charts, etc.
R	• Automatically read and understand characteristics of most genres, including biographies on new topics, fantasies, chapter books, shorter informational texts, mysteries, series, books with sequels, short stories, diaries, and logs. • Read fiction stories with many characters that change throughout the story • Make sense of new vocabulary words • Look for information in pictures, photographs, maps, charts, etc. • Can break words into syllables • Use strategies to figure out difficult words • Understand texts with different layouts
S	• Automatically read and understand characteristics of most genres, including biographies on new topics, fantasies, chapter books, shorter informational texts, mysteries, series, books with sequels, short stories, diaries, and logs. • Read fiction stories with many characters that change in the story • Understand hard sentences and words • Can break words into syllables • Understand texts with different layouts • Look for information in pictures, photographs, maps, charts, etc.
T	• Automatically read and understand characteristics of most genres, including biographies on new topics, fantasies, chapter books, shorter informational texts, mysteries, series, books with sequels, short stories, diaries, logs, fantasies, myths, and legends. • Read longer texts with many lines of print that require the reader to remember lots of information • Can break words into syllables • Use strategies to figure out difficult words • Look for information in pictures, photographs, maps, charts, etc. • Use what they already know to understand a text

Text Level	Indicators
U	• Automatically read and understand characteristics of most genres, including biographies on new topics, fantasies, chapter books, shorter informational texts, mysteries, series, books with sequels, short stories, diaries, logs, fantasies, myths, and legends. • Read longer texts with many lines of print that require the reader to remember lots of information • Can break words into syllables • Use strategies to figure out difficult words • Search for and use information in a text • Look for information in pictures, photographs, maps, charts, etc.
V	• Read and understand characteristics of most genres, including biographies on new topics, fantasies, chapter books, shorter informational texts, mysteries, series, books with sequels, short stories, diaries, logs, fantasies, myths, and legends • Read texts that are longer and involve remembering information • Can break words into syllables • Search for and use information in a text • Look for information in pictures, photographs, maps, charts, etc.
X, Y, and Z	• Read and understand characteristics of all genres, including biographies on new topics, fantasies, chapter books, shorter informational texts, mysteries, series, books with sequels, short stories, diaries, logs, fantasies, myths, and legends. • Use critical thinking skills • Read long texts with long sentences and paragraphs • Understand mature themes (abuse, poverty, war, etc.) • Read texts with many characters that change in the story • Use what they know to understand a text • Search for and use information in a text • Look for information in pictures, photographs, maps, charts, etc. • Read texts that require knowing about history and science

Source: Fountasandpinnell.com, 2014; Sachem.edu, 2019.

APPENDIX 3

Numeracy Staircase

Domain	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	Level 7	Level 8	Level 9
Number Concept	Determining quantity: presented in iconical or pictorial representation	Recognizing simple fractions represented in concrete form	Representing fraction as a part of whole and a part of a collection by using pictorial form (e.g., shaded area)	Making equivalent fractions	Representing power by two as repeated multiplication and understanding square root as the inverse of power by two	Recognising negative integers			
	Recognizing numbers up to two digits	Recognizing numbers up to three digits		Recognizing prime numbers					
	Comparing the quantity of objects	Comparing and ordering whole numbers	Comparing and ordering like fractions	Comparing and ordering unlike fractions, decimals, and percentages		Comparing and ordering positive and negative integers			
	Identifying place values up to two digits	Identifying place values up to three digits							
	Determining a missing number in an ordered pattern on a number line	Determining a missing number in a skipped pattern on a number line	Determining two missing numbers in a skipped pattern on a number line	Determining two sequential missing numbers in a skipped pattern on a number line					
Number Operation	Adding and subtracting whole numbers up to 99	Adding and subtracting whole numbers up to 999	Performing mixed math operations of whole numbers: addition and subtraction	Performing mixed math operations of whole numbers: multiplication and division		Performing number operations of positive and negative integers			

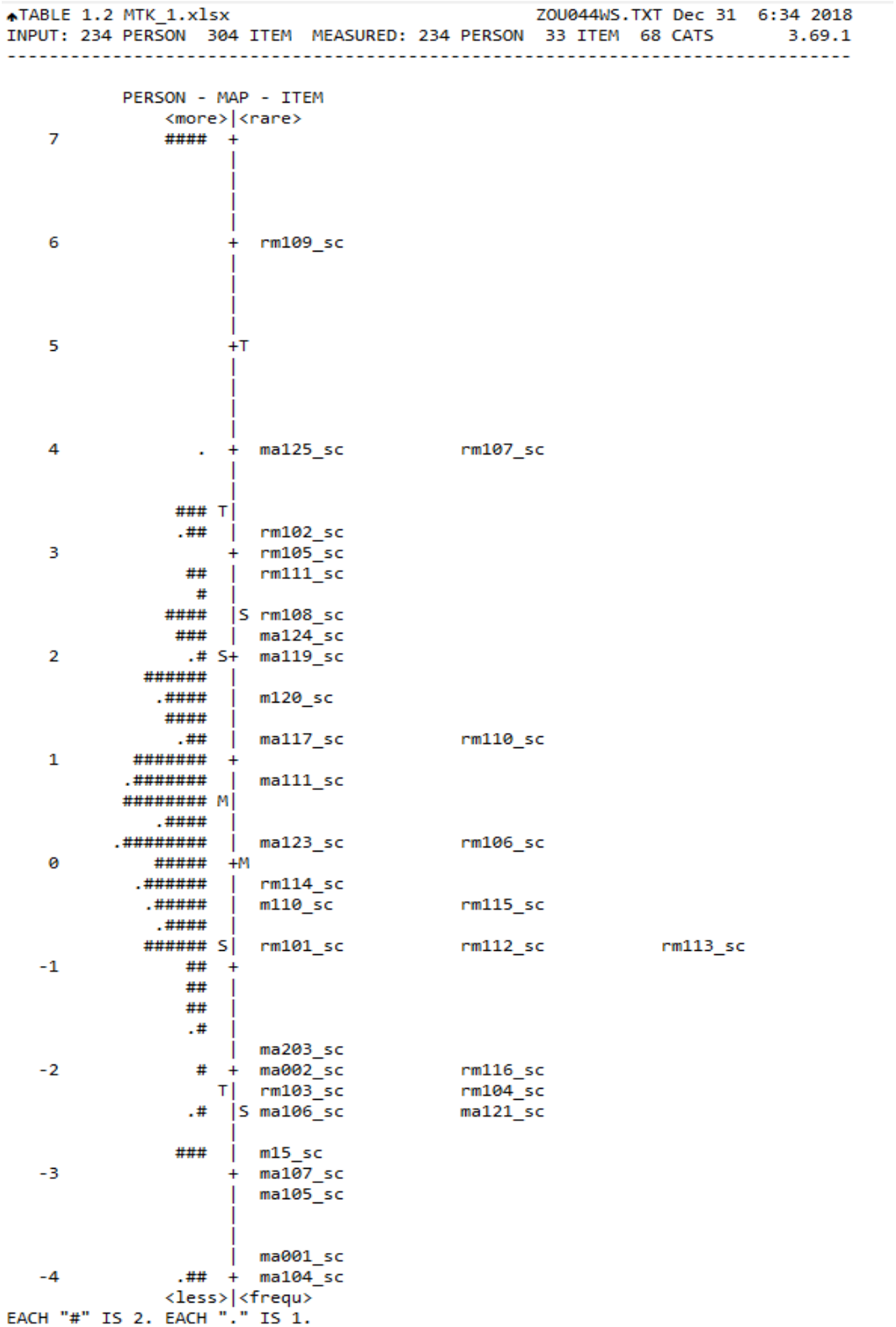
Domain	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	Level 7	Level 8	Level 9
		Multiplying and dividing whole numbers up to 100		Estimating and rounding					
				Least common multiples and greatest common divisors	Multiplying and dividing fractions				
			Adding and subtracting like fractions	Converting fractions from and into decimals and percentages	Adding and subtracting unlike fractions	Performing mixed math operations of whole numbers and fractions			
Word Problem	Simple word problem supported by illustration	Simple word problem **without** illustration: two available numbers and one missing number	Simple word problem without illustration: more than two available numbers and one missing number; requiring single-step solution process	Simple word problem without illustration: more than two available numbers and one missing number; requiring simple multi-step solution process					
Geometry and Measurement	Recognizing simple plane figures (triangle, rectangle, square) and simple solid figures (cube, cuboid)	Categorizing simple plane figures (triangle, rectangle, square) and simple solid figures (cube, cuboid)	Describing the properties and characteristics of simple plane figures (triangle, rectangle, square)	Recognizing and categorizing polygon		Describing the properties and characteristics of a circle			
			Categorizing simple solid shapes (cube, cuboid)	Describing the characteristics of simple solid shapes (cube, cuboid)	Making the nets of cube and cuboid	Recognizing and categorizing prism, cylinder, pyramid, cone, and sphere			

Domain	Level 1	Level 2	Level 3	Level 4	Level 5	Level 6	Level 7	Level 8	Level 9
	Comparing length and weight	Measuring length and weight using nonstandardized unit	Measuring length, weight, and time using standardized unit	Unit conversion for length, weight, and time	Recognising and comparing distance, time, and speed				
		Comparing time and size	Recognizing and comparing angles	Measuring angles					
			Measuring area using nonstandardized unit such as unit square	Determining the area of square, rectangle, triangle		Measuring the area and perimeter of circle			
				Measuring volume using nonstandardized unit: cube, cuboid	Determining the volume of cube and cuboid	Determining the volume of a combination of cuboids			
			Determining the perimeter of square, rectangle, triangle		Coordinate point, map, and scale				
			Point and linear symmetry		Point of compass				
Data and Statistics			Identifying information from pictogram and 2 x n table	Identifying information from table, bar, and pie chart	Using information from charts and tables to solve mathematical problems	mean, median, mode			

APPENDIX 4

Wright Map

Math - Grade 1

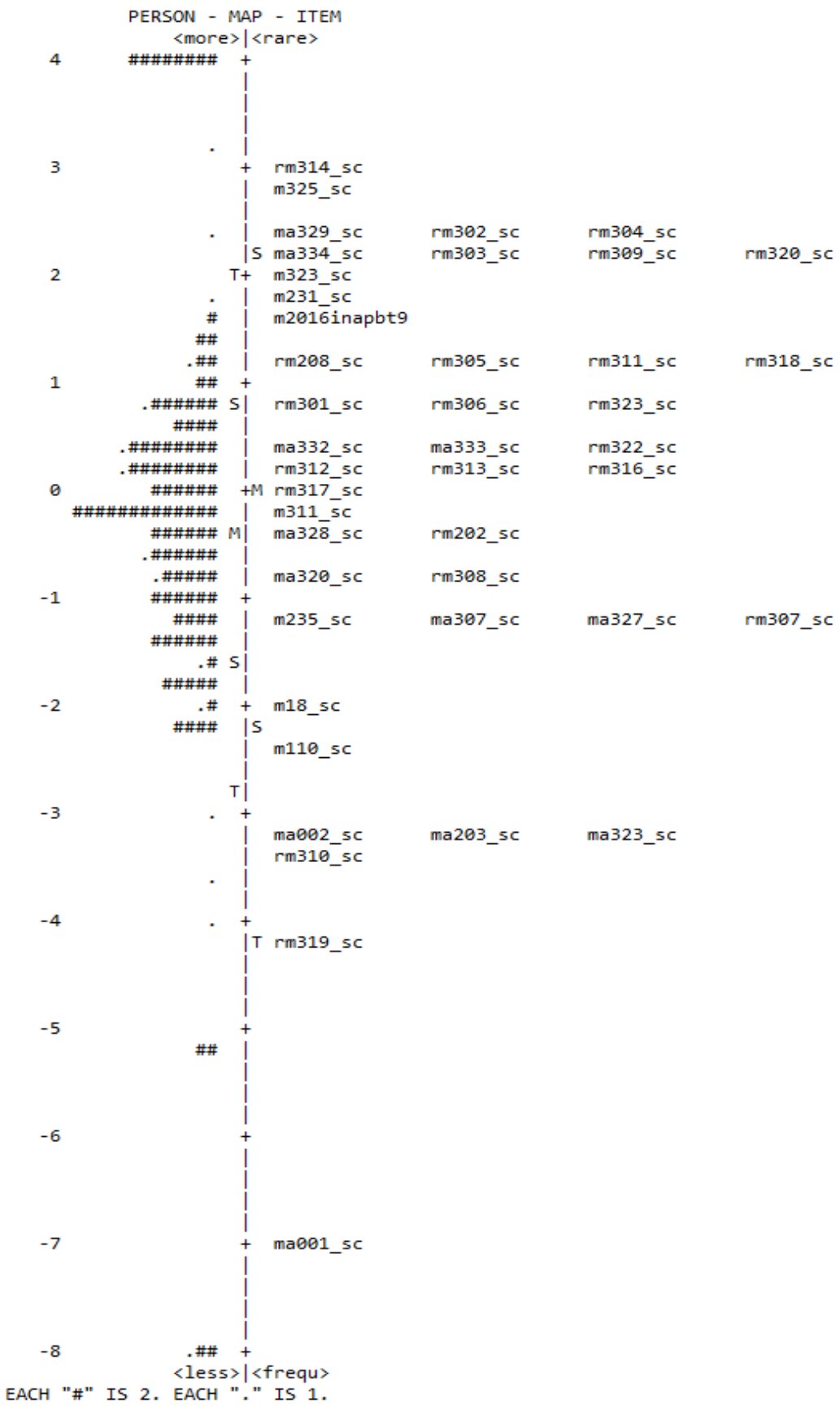


Math - Grade 2

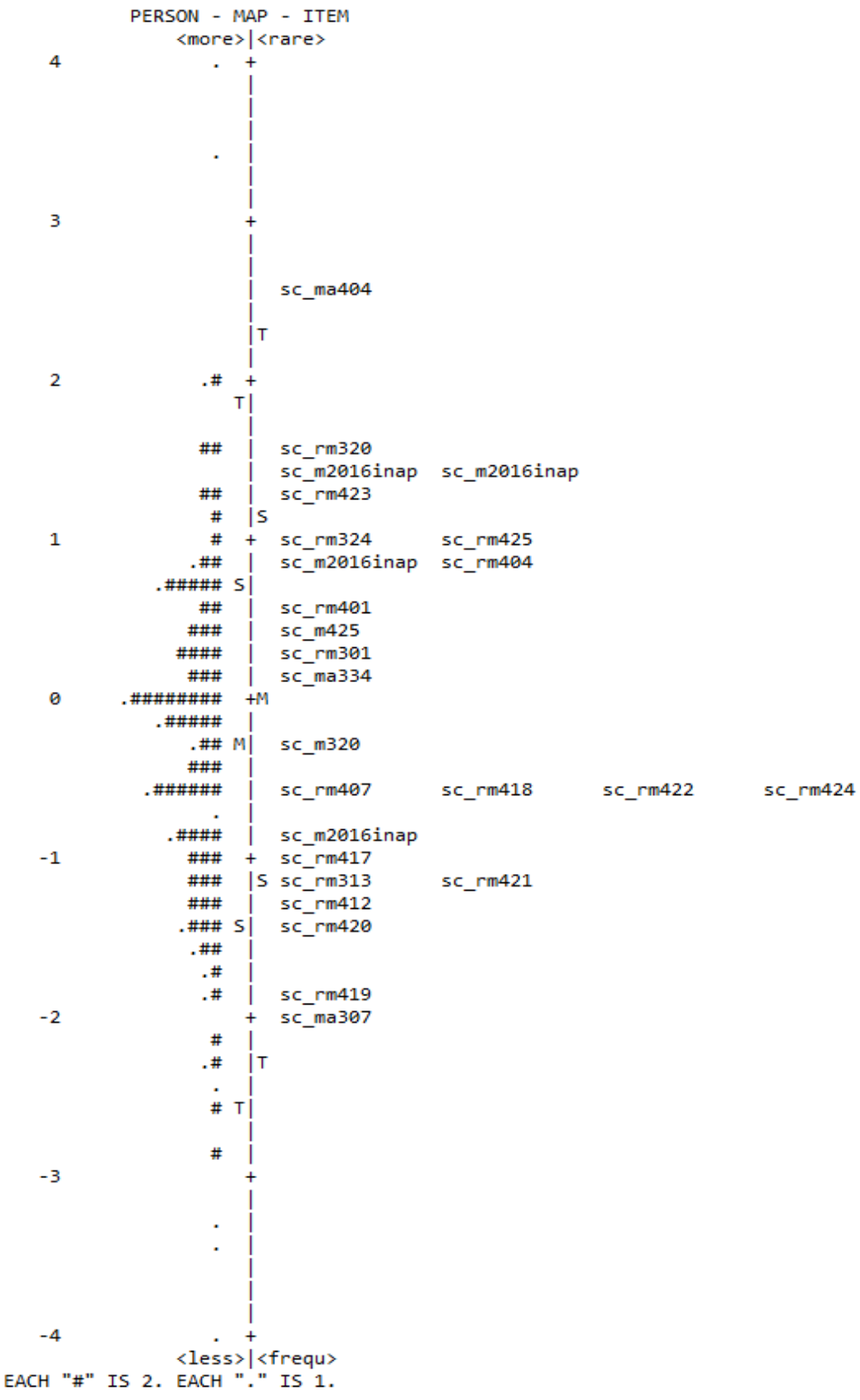
```
↑TABLE 1.2 MTK_2.xlsx                          ZOU424WS.TXT Dec 31  7:11 2018
INPUT: 233 PERSON  304 ITEM  MEASURED: 233 PERSON  42 ITEM  84 CATS       3.69.1
-------------------------------------------------------------------------------

           PERSON - MAP - ITEM
               <more>|<rare>
    5        .#####  +
                     |
                     |
                     |
                  .  |
                  .  |T ma125_sc   rm205_sc
                     |
    4                +
                     |
                     |  rm215_sc
                     |
                     |
                     |  rm203_sc
    3             # T+
                  #  |  ma326_sc
                  .  |
                  #  |
                 .#  |  rm105_sc   rm206_sc
                ###  |S rm208_sc
    2            ##  +
                 #   |
                ##   |  rm202_sc   rm209_sc
              .#### S|  rm204_sc   rm207_sc   rm214_sc
                ###  |
              .####  |
    1            ##  +
              .#####  |  rm217_sc
            .#######  |  m126_sc    ma117_sc   rm218_sc
              .#####  |  ma320_sc
            .#######  |  m120_sc    ma322_sc   rm106_sc
               .#### M|  ma112_sc   ma321_sc
    0         .#####  +M rm212_sc
             ######   |  rm201_sc
                .##   |  ma323_sc   rm211_sc
             #######  |
                .##   |
           #########  |
    -1               +
               .####  |  m110_sc
                 .## S|
                  .#  |  ma123_sc   rm210_sc   rm216_sc
                  #   |  m18_sc
                  .   |
    -2            #  +
                     |S
                     |  ma203_sc
                     |
                  # T|  rm213_sc
                  .  |  ma002_sc   ma106_sc   ma107_sc   ma121_sc   rm104_sc
    -3            .  +  m15_sc
                     |
                #####|
                     |
                     |
                     |  ma001_sc   ma105_sc
    -4            ##  +
                <less>|<frequ>
EACH "#" IS 2. EACH "." IS 1.
```

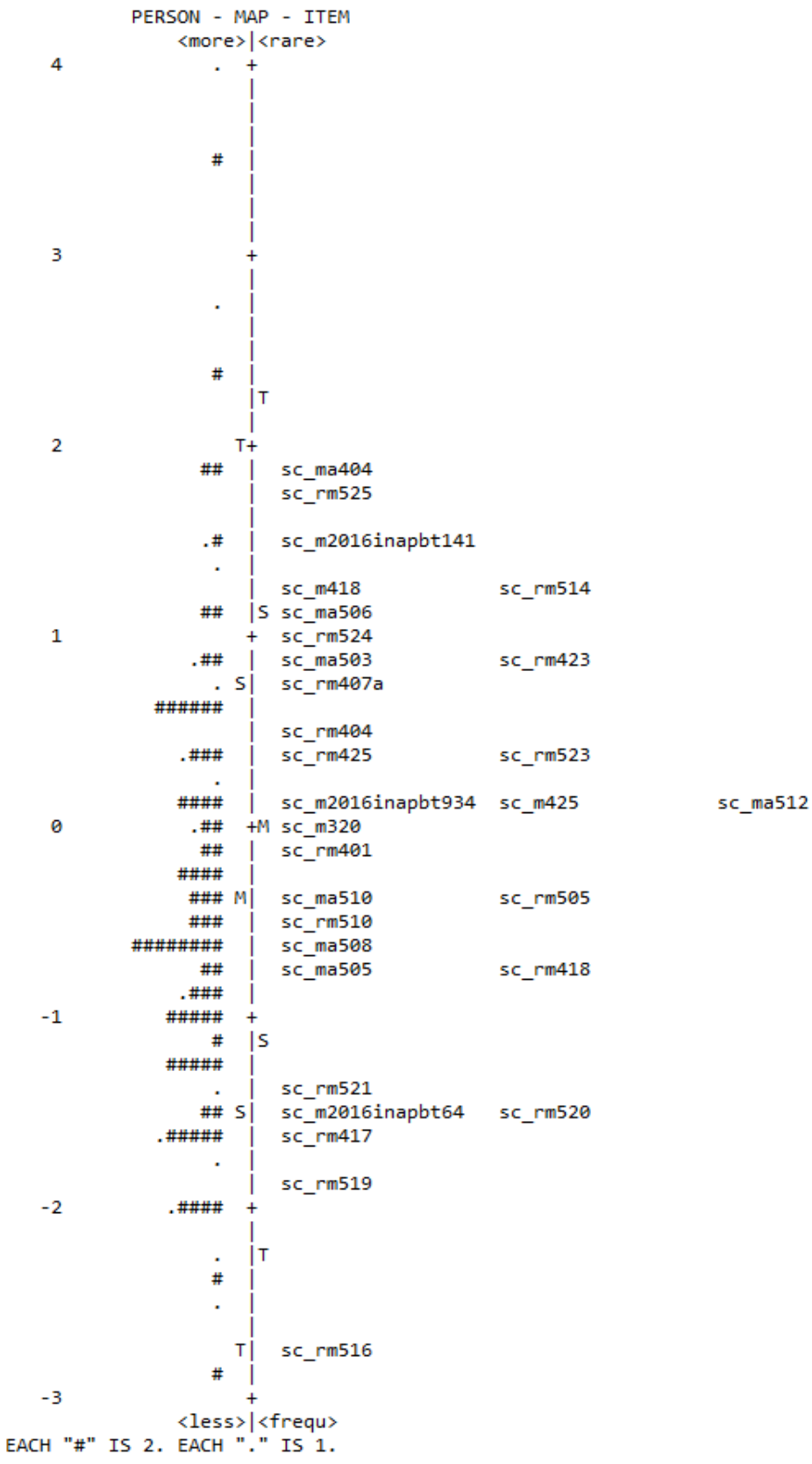
Math - Grade 3


PERSON - MAP - ITEM
<more>|<rare>
4 ######## +
3 +
rm314_sc
m325_sc
ma329_sc rm302_sc rm304_sc
S ma334_sc rm303_sc rm309_sc rm320_sc
2 T+ m323_sc
m231_sc
m2016inapbt9
rm208_sc rm305_sc rm311_sc rm318_sc
1 +
.###### S rm301_sc rm306_sc rm323_sc
ma332_sc ma333_sc rm322_sc
rm312_sc rm313_sc rm316_sc
0 ###### +M rm317_sc
m311_sc
M ma328_sc rm202_sc
ma320_sc rm308_sc
-1 +
m235_sc ma307_sc ma327_sc rm307_sc
S
-2 +
m18_sc
S
m110_sc
T
-3 +
ma002_sc ma203_sc ma323_sc
rm310_sc
-4 +
T rm319_sc
-5 +
-6 +
-7 +
ma001_sc
-8 +
<less>|<frequ>
EACH "#" IS 2. EACH "." IS 1.

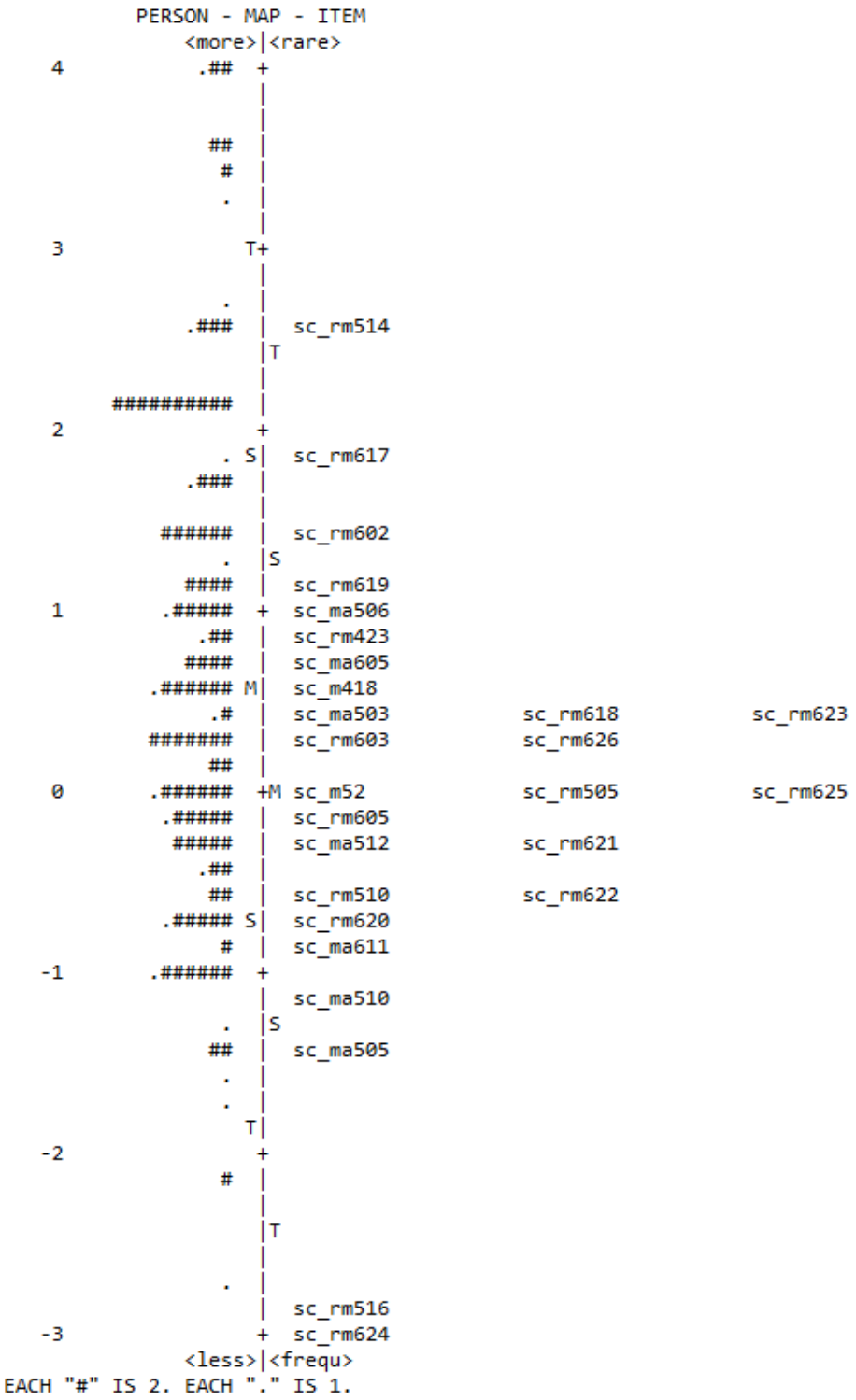
Math - Grade 4


PERSON - MAP - ITEM
<more>|<rare>
4
3
sc_ma404
T
2
sc_rm320
sc_m2016inap sc_m2016inap
sc_rm423
S
1
sc_rm324 sc_rm425
sc_m2016inap sc_rm404
sc_rm401
sc_m425
sc_rm301
sc_ma334
0
M
sc_m320
sc_rm407 sc_rm418 sc_rm422 sc_rm424
sc_m2016inap
-1
sc_rm417
S sc_rm313 sc_rm421
sc_rm412
S sc_rm420
sc_rm419
-2
sc_ma307
T
-3
-4
<less>|<frequ>
EACH "#" IS 2. EACH "." IS 1.

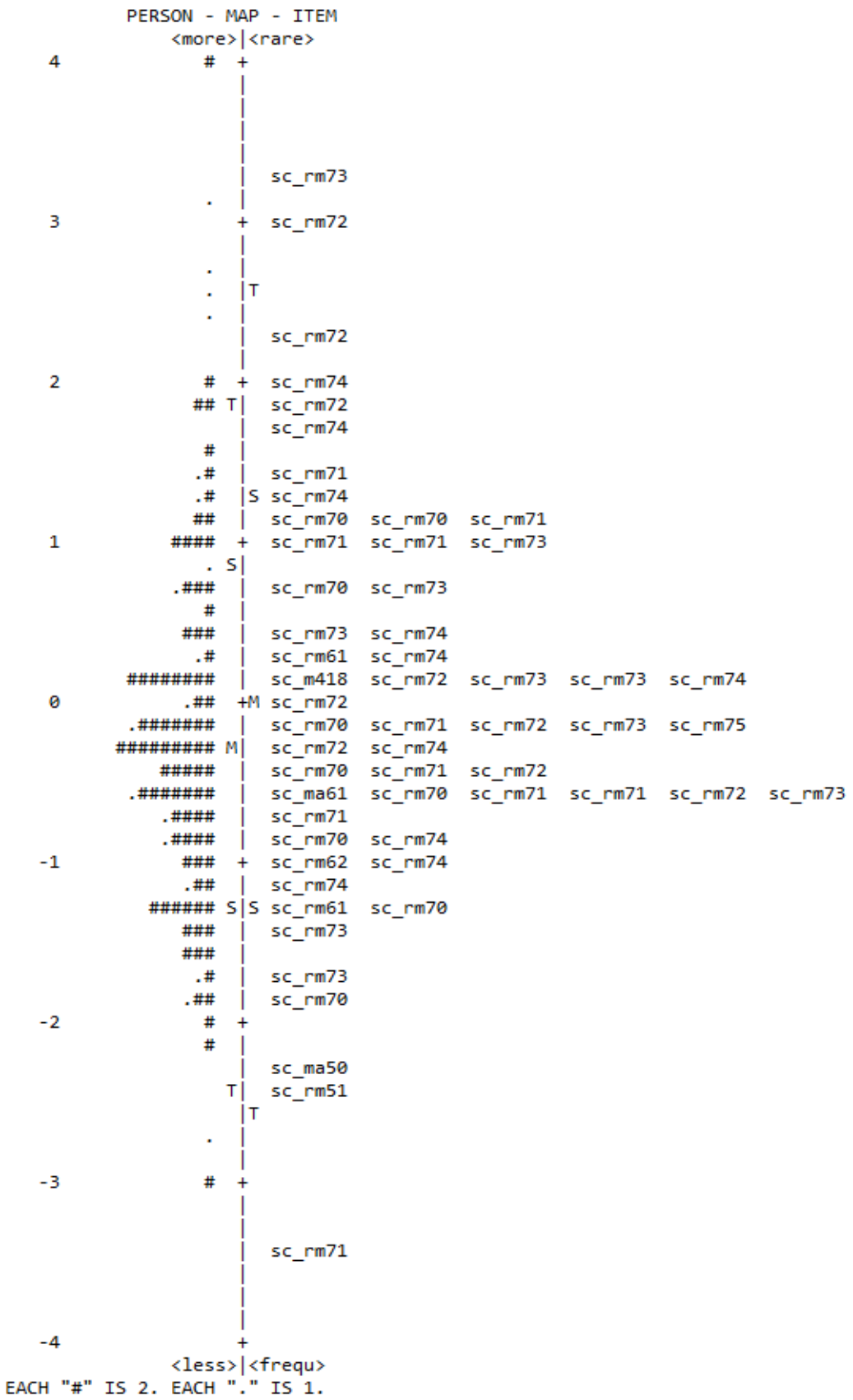
Math - Grade 5


PERSON - MAP - ITEM
<more>|<rare>
4
3
2
1
0
-1
-2
-3
sc_ma404
sc_rm525
sc_m2016inapbt141
sc_m418
sc_rm514
sc_ma506
sc_rm524
sc_ma503
sc_rm423
sc_rm407a
sc_rm404
sc_rm425
sc_rm523
sc_m2016inapbt934
sc_m425
sc_ma512
sc_m320
sc_rm401
sc_ma510
sc_rm505
sc_rm510
sc_ma508
sc_ma505
sc_rm418
sc_rm521
sc_m2016inapbt64
sc_rm520
sc_rm417
sc_rm519
sc_rm516
<less>|<frequ>
EACH "#" IS 2. EACH "." IS 1.

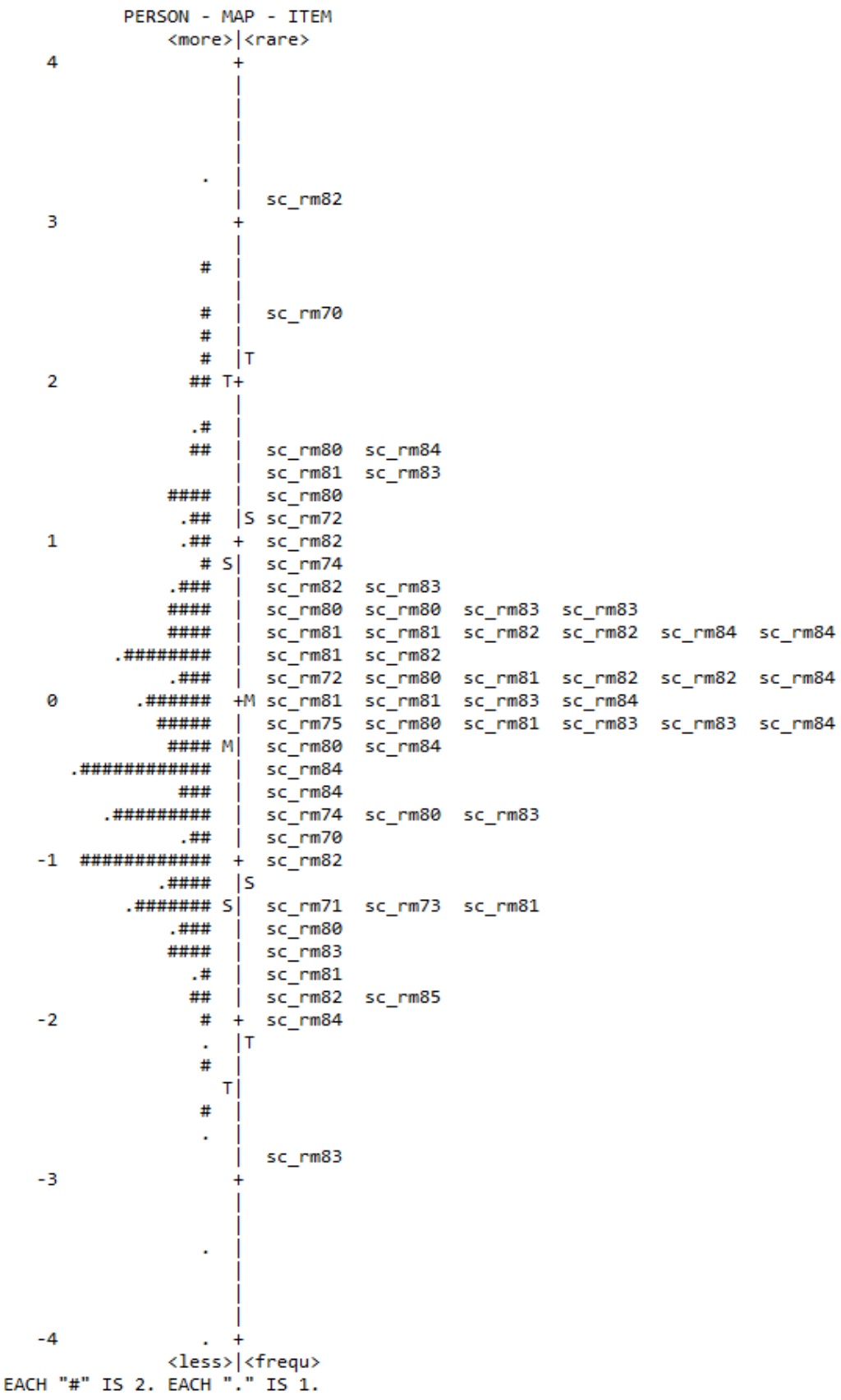
Math - Grade 6


PERSON - MAP - ITEM
<more>|<rare>
4
3
2
1
0
-1
-2
-3
sc_rm514
sc_rm617
sc_rm602
sc_rm619
sc_ma506
sc_rm423
sc_ma605
sc_m418
sc_ma503
sc_rm603
sc_m52
sc_rm605
sc_ma512
sc_rm510
sc_rm620
sc_ma611
sc_ma510
sc_ma505
sc_rm516
sc_rm624
sc_rm618
sc_rm626
sc_rm505
sc_rm621
sc_rm622
sc_rm623
sc_rm625
<less>|<frequ>
EACH "#" IS 2. EACH "." IS 1.

Math - Grade 7


PERSON - MAP - ITEM
<more>|<rare>
4
3
2
1
0
-1
-2
-3
-4
sc_rm73
sc_rm72
sc_rm74
sc_rm71
sc_rm70 sc_rm70 sc_rm71
sc_rm71 sc_rm71 sc_rm73
sc_rm70 sc_rm73
sc_rm73 sc_rm74
sc_rm61 sc_rm74
sc_m418 sc_rm72 sc_rm73 sc_rm73 sc_rm74
sc_rm72
sc_rm70 sc_rm71 sc_rm72 sc_rm73 sc_rm75
sc_rm72 sc_rm74
sc_rm70 sc_rm71 sc_rm72
sc_ma61 sc_rm70 sc_rm71 sc_rm71 sc_rm72 sc_rm73
sc_rm62 sc_rm74
sc_rm61 sc_rm70
sc_ma50
sc_rm51
<less>|<frequ>
EACH "#" IS 2. EACH "." IS 1.

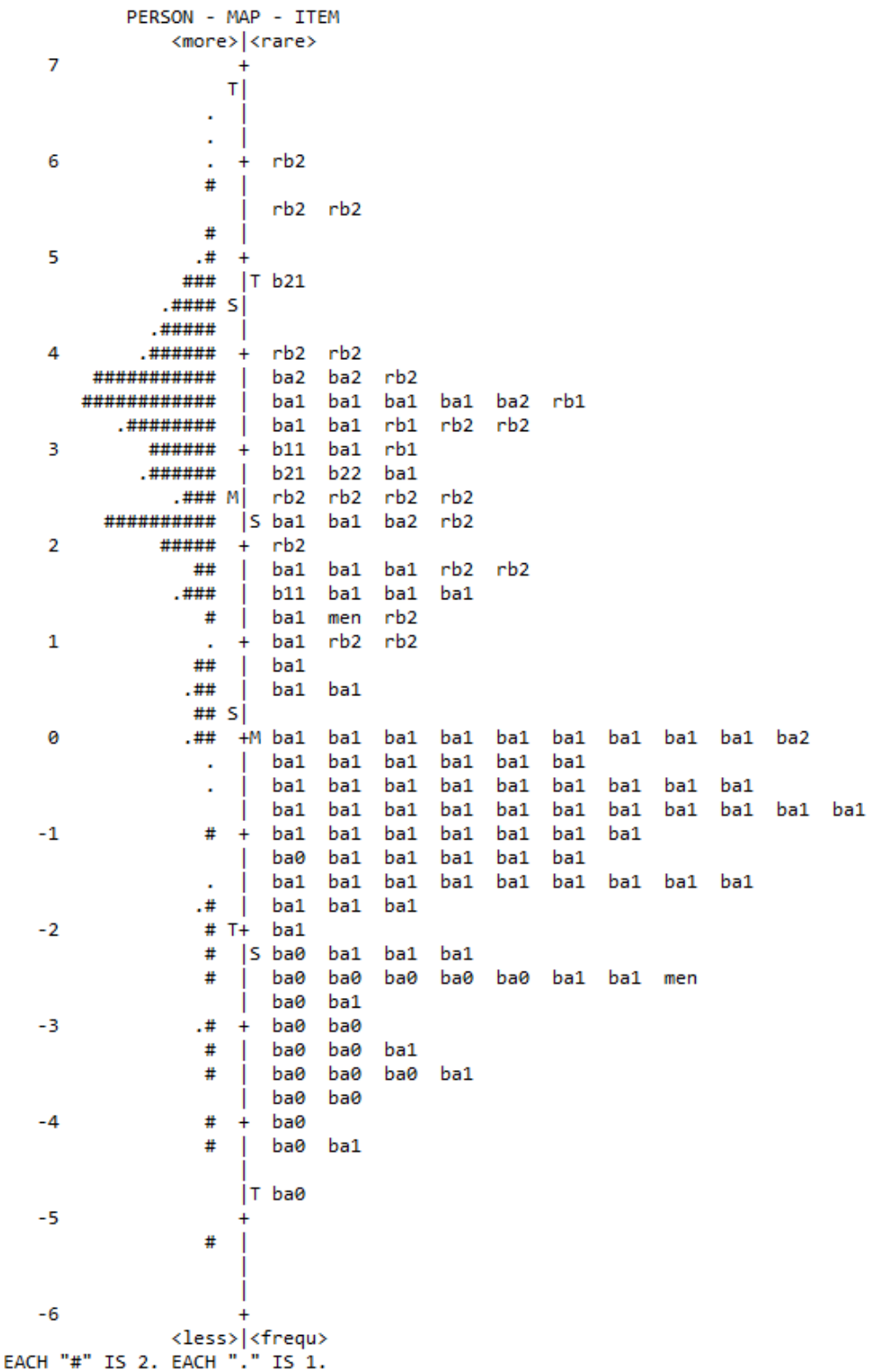
Math - Grade 8


PERSON - MAP - ITEM
<more>|<rare>
4
3
2
1
0
-1
-2
-3
-4
sc_rm82
sc_rm70
sc_rm80 sc_rm84
sc_rm81 sc_rm83
sc_rm80
sc_rm72
sc_rm82
sc_rm74
sc_rm82 sc_rm83
sc_rm80 sc_rm80 sc_rm83 sc_rm83
sc_rm81 sc_rm81 sc_rm82 sc_rm82 sc_rm84 sc_rm84
sc_rm81 sc_rm82
sc_rm72 sc_rm80 sc_rm81 sc_rm82 sc_rm82 sc_rm84
sc_rm81 sc_rm81 sc_rm83 sc_rm84
sc_rm75 sc_rm80 sc_rm81 sc_rm83 sc_rm83 sc_rm84
sc_rm80 sc_rm84
sc_rm84
sc_rm84
sc_rm74 sc_rm80 sc_rm83
sc_rm70
sc_rm82
sc_rm71 sc_rm73 sc_rm81
sc_rm80
sc_rm83
sc_rm81
sc_rm82 sc_rm85
sc_rm84
sc_rm83
<less>|<frequ>
EACH "#" IS 2. EACH "." IS 1.

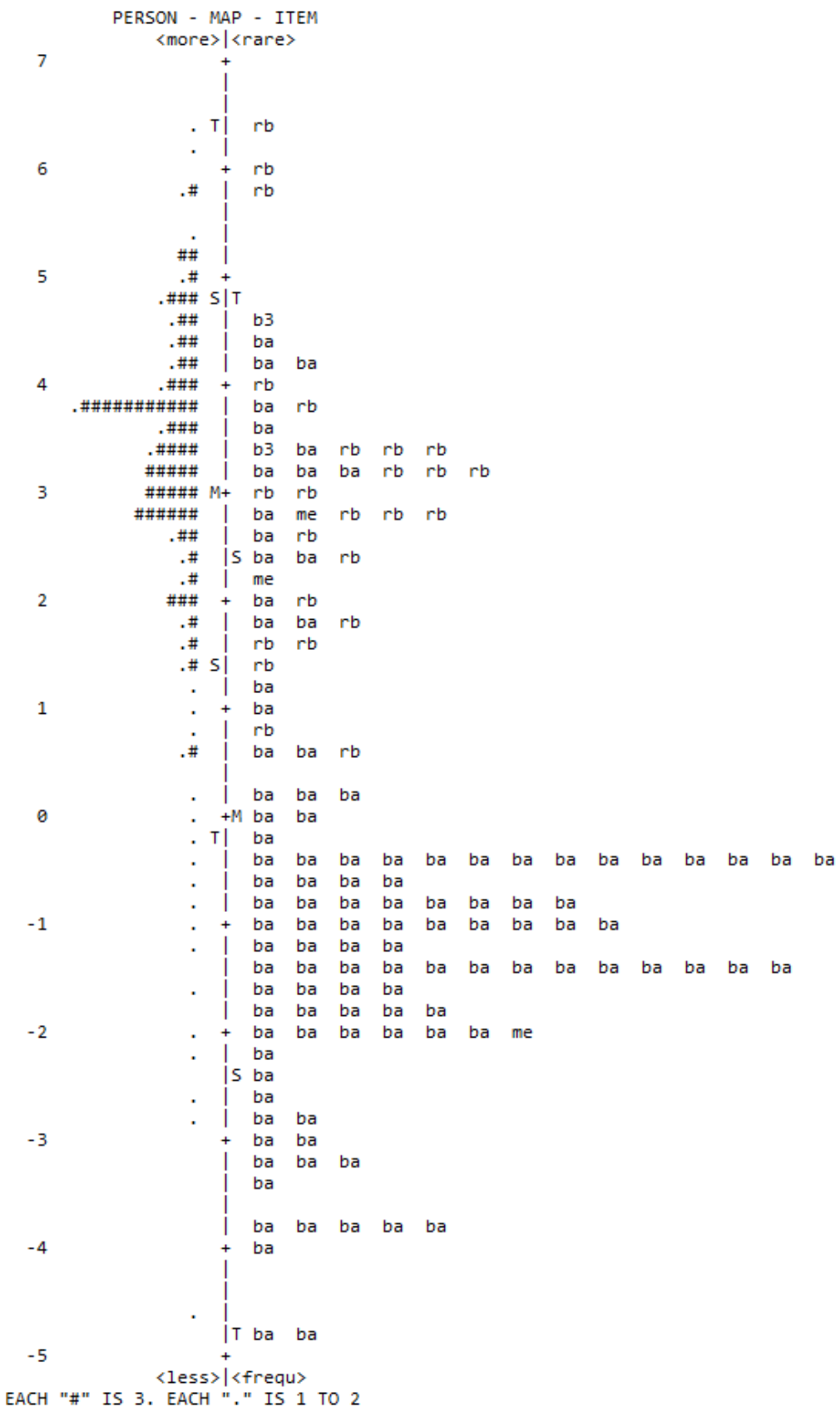
Indonesian - Grade 1

```
        PERSON - MAP - ITEM
            <more>|<rare>
 7                +
                X T|
                   |
                   |
 6                 +
             XXX   |  ba1  rb1
                   |
            XXXX   |
 5        XXXXXX   +
      XXXXXXXXXX   |T
           XXXXX   |  ba1  rb1
   XXXXXXXXXXXXX S|  rb1
 4     XXXXXXXXXX  +
      XXXXXXXXXX   |  ba1  rb1
    XXXXXXXXXXXX   |  b21  ba1  rb1
      XXXXXXXXXX   |  b11  ba1  rb1  rb1
 3   XXXXXXXXXXX   +  b11  ba1  ba1  ba1  men  rb1
        XXXXXXXX   |  ba1  ba1  ba1  ba1
         XXXXXXX   |  ba1  ba1  rb1  rb1  rb1  rb1
      XXXXXXXXXX   |S b11  ba1  rb1
 2       XXXXXXX   +  ba1  ba1  men  rb1
        XXXXXXXX M|  ba1
            XXXX   |  ba1
             XXX   |  ba1  ba1
 1         XXXXX   +  ba1  ba1
             XXX   |  ba1  ba1
          XXXXXXX  |  ba1  ba1  ba1  ba1  ba1  ba1  ba1  ba1
           XXXXXX  |  ba1  ba1  ba1  ba1  ba1  ba1
 0         XXXXXX  +M ba1  ba1  ba1  ba1  ba1  ba1
       XXXXXXXXX   |  ba1  ba1  ba1  ba1  ba1  ba1  ba1  ba1  ba1  ba1  ba1
    XXXXXXXXXXXX   |  ba1  ba1  ba1  ba1
             XXX S|  ba0  ba1  ba1  ba1  ba1
-1           XXX   +  ba1  ba1  ba1  ba1
            XXXX   |  ba1  ba1  ba1  ba1  ba1  ba1
             XXX   |  ba0  ba1  ba1  ba1  ba1  ba1
           XXXXX   |  ba0  ba1  ba1
-2           XXX   +  ba0  ba1  ba1  ba1  ba1
            XXXX   |S ba0  ba0  ba1  men
              XX   |  ba0  ba0  ba1  ba1  ba1
                   |  ba0  ba1  ba1
-3            XX   +  ba0  ba1
            XXXX   |  ba0  ba0
              XX T|  ba0  ba0  ba1  ba1
              XX   |
-4            XX   +  ba0
                   |  ba0  ba0
                   |  ba0
                  |T
-5              X  +  ba0
                X  |
                X  |
                   |  ba0
-6                 +
            <less>|<frequ>
```

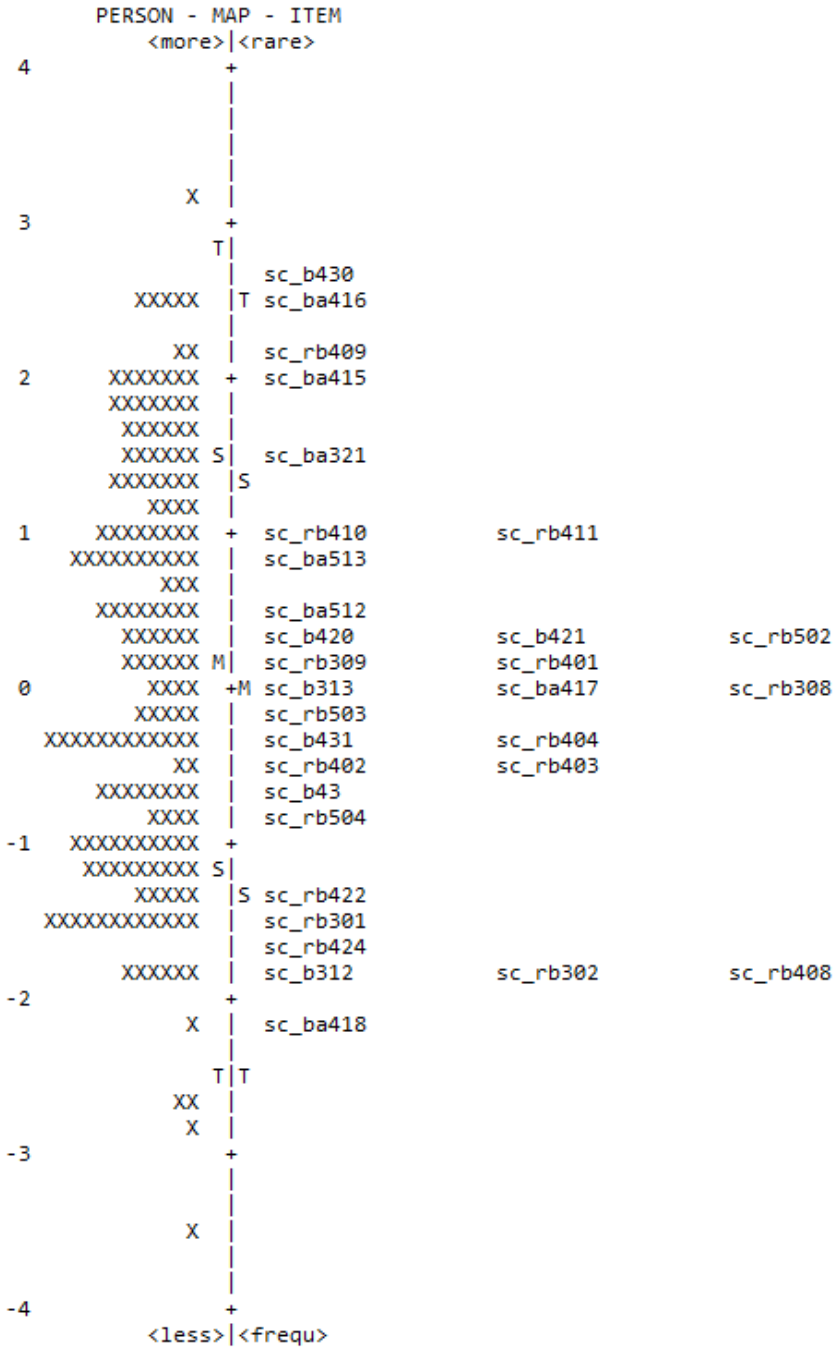
Indonesian - Grade 2



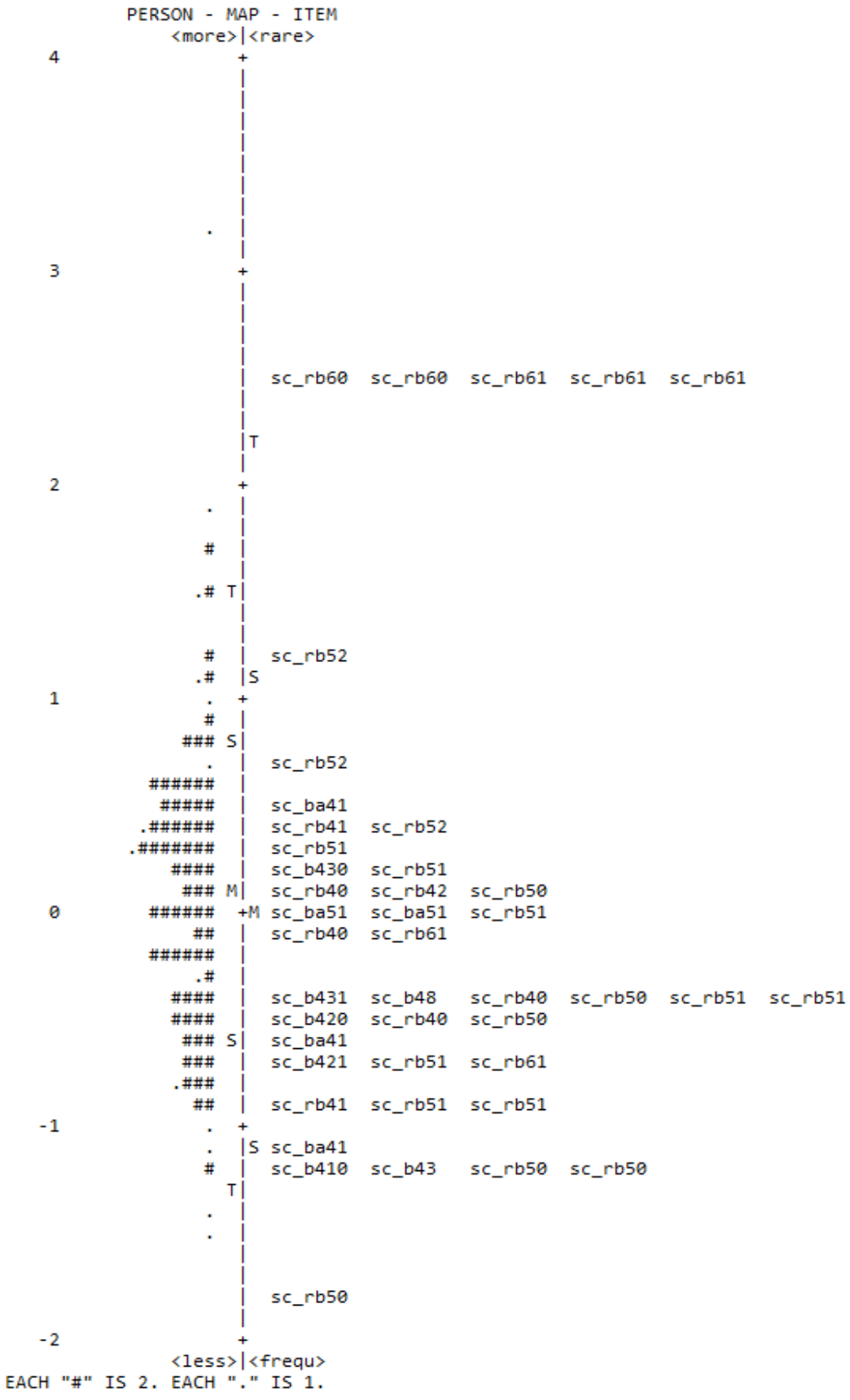
Indonesian - Grade 3



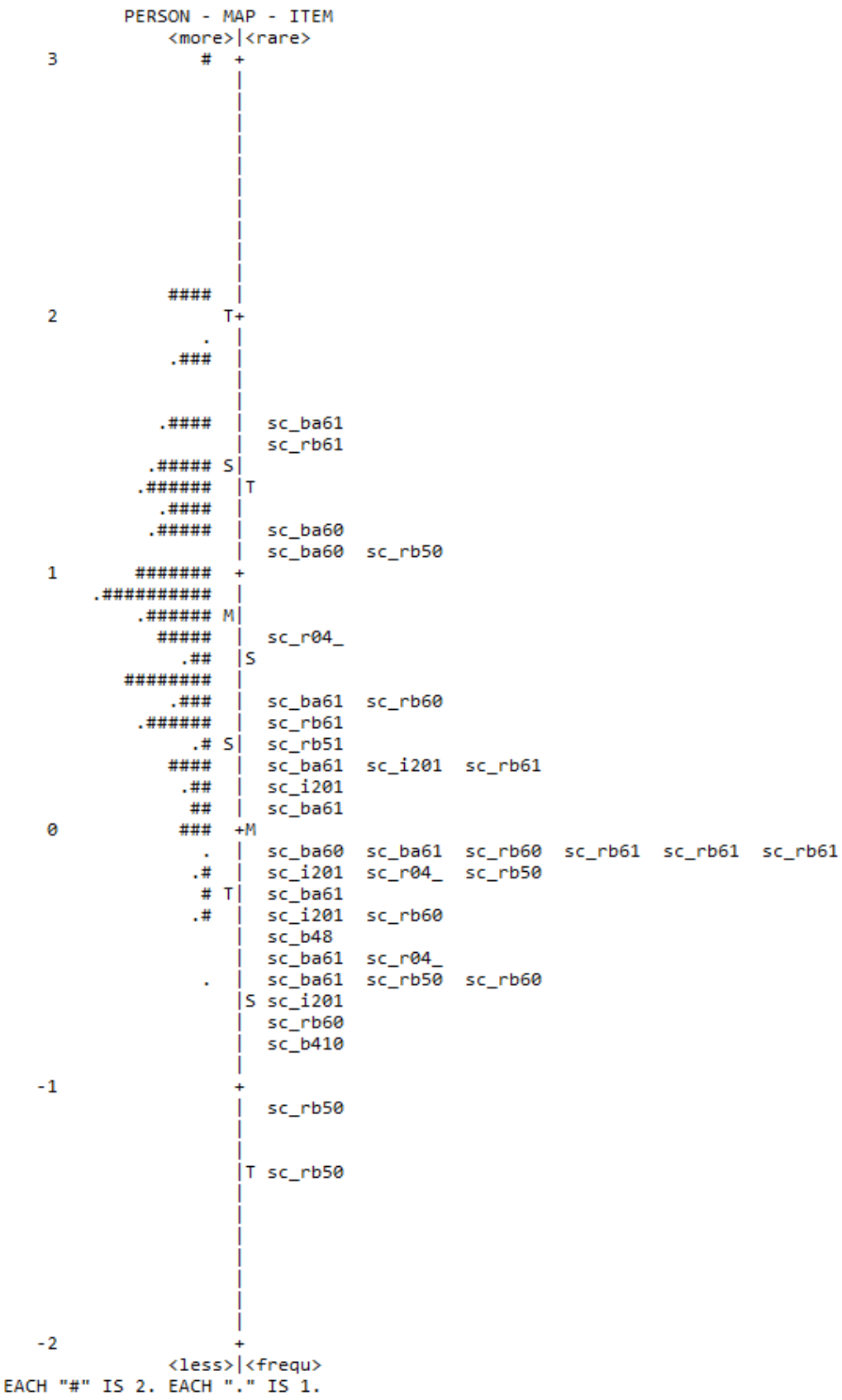
Indonesian - Grade 4



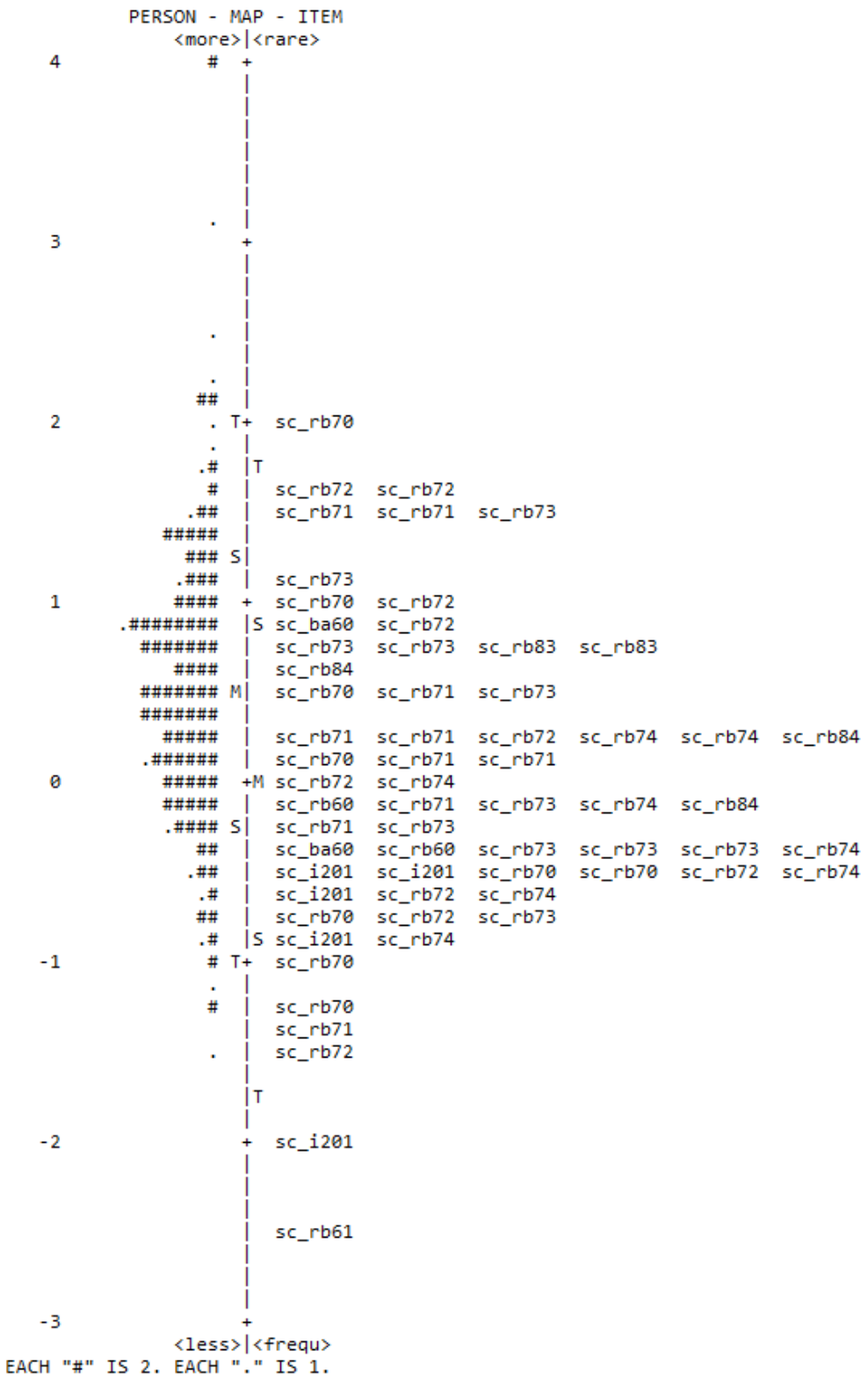
Indonesian - Grade 5



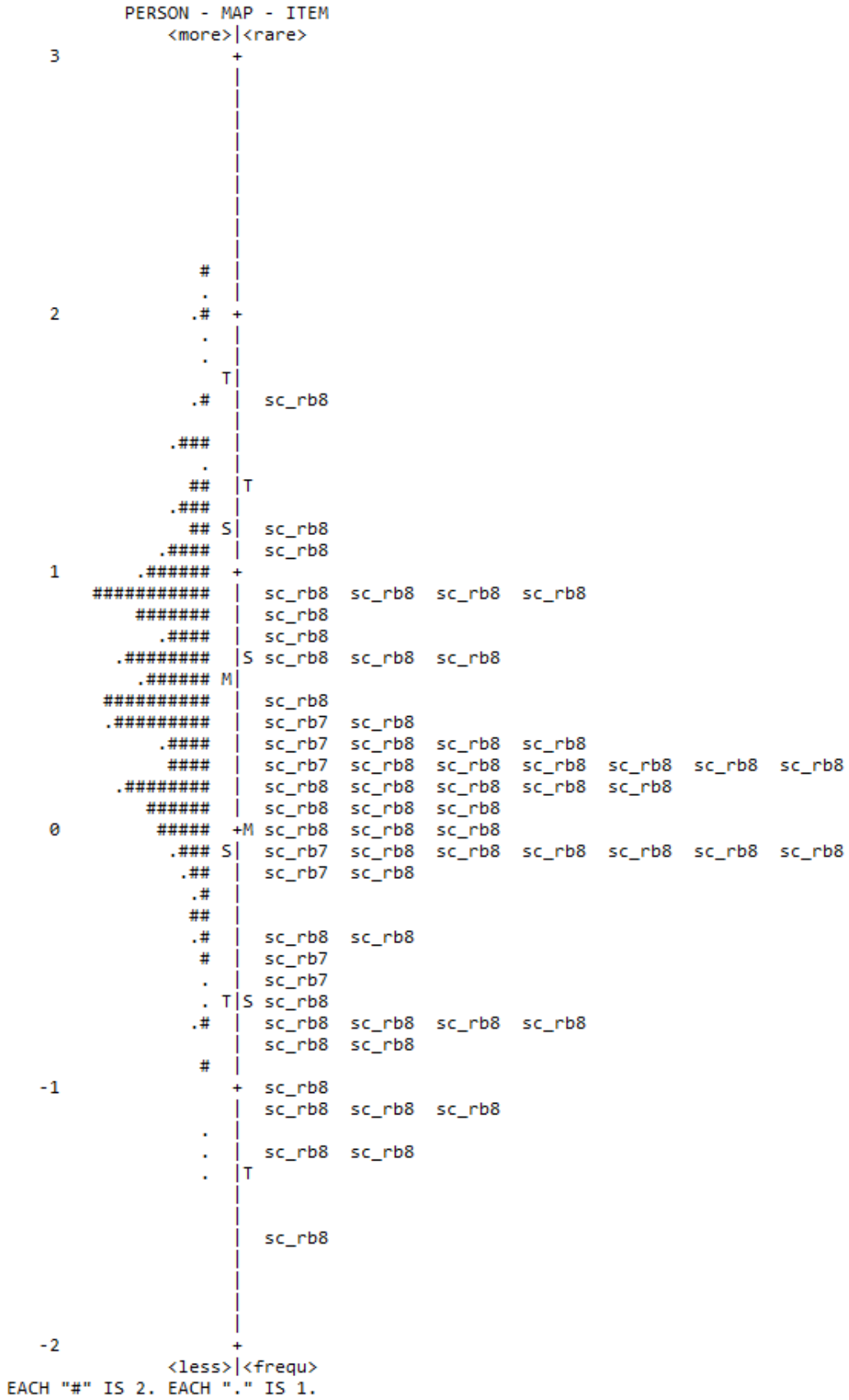
Indonesian - Grade 6



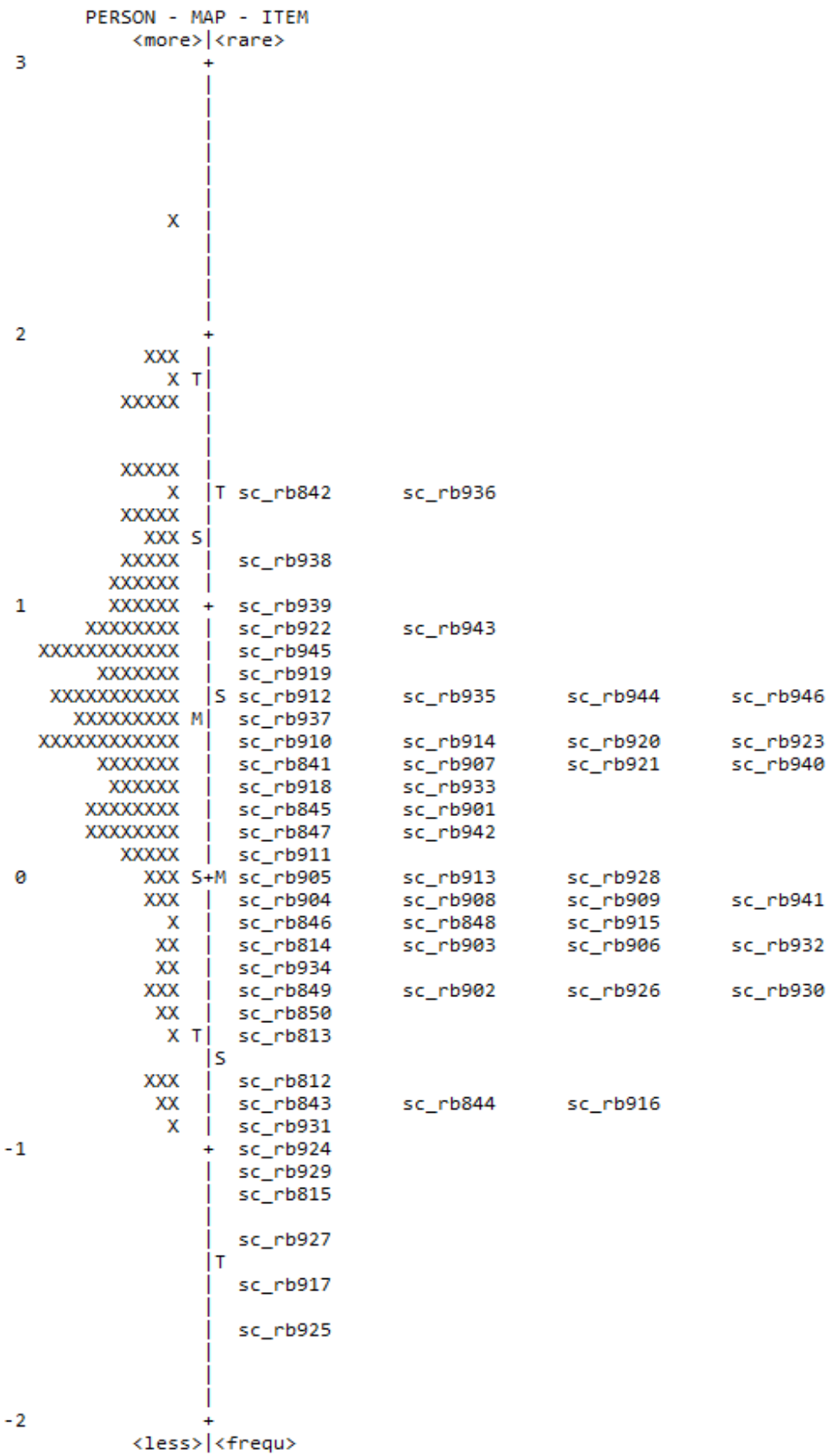
Indonesian - Grade 7



Indonesian - Grade 8



Indonesian - Grade 9



GLOSSARY

Anchor items	Test items administered in combination with two or more problem sets to connect the tests so that their scores are comparable
Assessment framework	A framework that explains the detailed content of the assessment
Bloom's Taxonomy	A set of hierarchical models used to classify educational learning objectives into levels of complexity and specificity
Classical test	A test administered in groups, commonly involving all students in one classroom
Closed constructed-response problem	A question similar to the more traditional fill-in-the-blank type of questions and having only one right answer
Cognitive domain	A domain that involves the development of students' cognitive skills and the acquisition of knowledge linked to the subject
Content domain	The body of knowledge, skills, or abilities being measured or examined by the student learning assessment tool
Graded response model (GRM)	A model to analyse whether a dichotomous response is correct
High-stakes test	A test that has major consequences for the test taker
Horizontal equating strategy	A strategy that refers to the equating of test scores administered to groups with similar abilities (e.g., different tests for students in the same grade or age)
Infit	Inlier-sensitive fit that shows how sensitive the pattern of the response to the item that detects the ability of a person
Latent ability	Individual's actual ability or true score
Low-stakes test	A test with less important consequences for the test taker
Lower-order thinking skills	A list of thinking skills which include remembering and understanding, and are generally attained by rote memorization
Middle-order thinking skills	A list of thinking skills which include applying and are generally attained after accomplishing lower-order thinking skills
Numeracy staircase	A set of numeracy abilities divided into several different levels
Open constructed-response problem	A test problem that requires the test taker to answer with a constructed response
Outfit	Outlier-sensitive fit where more difficult items are more sensitive to people with lower ability, and vice versa

Partial credit model (PCM)	A model used to analyse polytomous response, where each item has its own scale
Principal component analysis (PCA)	A statistical procedure that converts a set of observation of possibly correlated variables into a set of values of linearly uncorrelated variables
Procedural knowledge	The explicit knowledge that is required in performing a set of straightforward tasks
Psychometric properties	Properties that refer to the reliability and validity of the instrument (e.g., difficulty level, discrimination power, pseudo-guessing parameter)
Rasch model	A family of psychometric models for creating measurements from categorical data that uses difficulty level as the main parameter
Reliability	Quality of the test that shows how consistent it can assess a particular ability
Teaching to test	A practice of teaching the purpose of which is answering the test problems correctly
Text level gradient	A gradual change of difficulties of the text, including the number of difficult words, length, and genre
Validity	The quality of a test that shows how accurate the test can assess the intended ability
Vertical continuity	The continuous properties of a scale owned by different levels of tests to make them comparable
Vertical equating strategy	A strategy that refers to the equating of test scores administered to groups with similar abilities
Wright map (item-person map)	A map that juxtaposes the item's difficulty level with the person's ability level

The SMERU Research Institute

Telephone : +62 21 3193 6336
Fax : +62 21 3193 0850
E-mail : smeru@smeru.or.id
Website : www.smeru.or.id
Facebook : @SMERUInstitute
Twitter : @SMERUInstitute
YouTube : The SMERU Research Institute

9 786237 492283
9 786237 492276