KIAT Guru

BACKGROUND PAPER

Student Learning Assessment:

A Tool to Measure Primary Grade Student Learning Outcomes in Indonesia's Remote Areas

—

MARCH 2021

Australian Government

THE WORLD BANK
IBRD • IDA | WORLD BANK GROUP

USAID
FROM THE AMERICAN PEOPLE

BaKTI

BACKGROUND PAPER

Student Learning Assessment:

**A Tool to Measure Primary Grade Student
Learning Outcomes in Indonesia's Remote Areas**

Table of Contents

Figures

Graph

Tables

Acknowledgments

Abbreviations

ACDP	Analytical and Capacity Development Partnership
CREDO	Creative Education Indonesia Foundation
CPPS-UGM	Center for Population and Policy Studies – Gadjah Mada University
EGMA	Early Grade Mathematics Assessment
EGRA	Early Grade Reading Assessment
INAP	Indonesian National Assessment Programme (the Indonesian version is AKSI or *Asesmen Kompetensi Siswa Indonesia*)
INOVASI	Innovation for Indonesia's School Children
KIAT Guru	Teacher Performance and Accountability (*Kinerja dan Akuntabiltas Guru*)
KTSP	2006 Curriculum (*Kurikulum Tingkat Satuan Pendidikan*)
MoEC	Ministry of Education and Culture
NTT	Nusa Tenggara Province (Nusa Tenggara Timur)
OECD	Organisation for Economic Co-operation and Development
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
PUSPENDIK	*Pusat Penilaian Pendidikan* (Education Assessment Centre)
RISE	Research on Improving Systems of Education
SD	Primary school (*Sekolah Dasar*)
SLA	Student Learning Assessment
SMERU	Social Monitoring and Early Response Unit
TIMSS	Trends in International Mathematics and Science Study
TNP2K	The National Team for the Acceleration of Poverty Reduction (*Tim Nasional Percepatan Penanggulangan Kemiskinan*)
UN	National Examination (*Ujian Nasional*)
USAID	The United States Agency for International Development

01 Introduction

Learning outcomes of students in remote areas of Indonesia are behind their urban peers (ACDP, 2014; Stern & Nordstrum, 2014). KIAT Guru pilot aims to improve teacher performance and accountability in Indonesia's remote areas by means of community empowerment and performance-based teacher pay (Gaduh et al, 2020). Learning outcomes in remote schools of five KIAT Guru districts is lower by over four percent.[1]

To evaluate KIAT Guru's impact on learning outcomes we developed a Student Learning Assessment (SLA) tool. We conducted a review of existing learning assessment tools in basic reading and mathematics. The Ministry of Education and Culture (MoEC) has the national exam (Ujian Nasional), which covers grade 6, 9, and 12, and the Indonesian National Assessment Program (INAP) which covers grade 4. USAID PRIORITAS developed the Early Grade Reading Assessment (EGRA), and World Bank's Better Education through Reformed Management and Universal Teacher Upgrading (BERMUTU) developed the early childhood and primary grade assessments. While these tools are for urban contexts, the SLA is contextualized to remote areas' settings.

SLA assesses basic reading and math competencies in all six primary grades, based on the national curriculum (KTSP 2006) and international (INAP, PIRLS, TIMSS) framework for reading and math. The assessment was developed in collaboration with MoEC's Center for Education Assessment (*Pusat Penilaian Pendidikan/* Puspendik), The National Team for the Acceleration of Poverty Reduction (TNP2K) and Innovation for Indonesia's School Children (INOVASI); and was piloted with Center for Population and Policy Studies – Gadjah Mada University (CPPS-UGM), World Bank, and the Research on Improving Systems of Education (RISE) team at SMERU (Social Monitoring and Early Response Unit).

---

[1] Authors' analysis based on Grade 6 data from 2016 Indonesia's National Examination (*Ujian Nasional*) comparing remote and urban schools across 5 KIAT Guru districts: Sintang, Landak, Ketapang, West Manggarai, and East Manggarai.

Compared to aforementioned assessment tools, SLA offers some new features: (a) covers early and late primary grades, (b) vertical continuity across grades to inform whether a student's ability is within the range of their current grade, and (c) contextualized to remote area settings. The third feature is done by applying more test items that fall on the lower spectrum of difficulty levels while maintaining adherence to competency standards. This background paper aims to document the development, implementation, and scoring or analysis of the SLA for KIAT Guru pilot. The rest of the paper is organized as follows. The next section discusses the framework of the test, while section 3, 4, and 5 describes the development and finalization of test design and item statistics, and the paper concludes with section 6.

02 Framework

SLA refers to both national and international framework on primary grade reading and math. The tool is in coherent to the content and cognitive domain required by the 2006 national curriculum, PIRLS Literacy, and TIMSS Numeracy. Content in reading assess decoding and comprehension skills. Content in math assess skills in solving problems related to numbers, geometry and measurement, as well as data and statistics. The cognitive skills to understand, apply, and reason with these reading and math contents are also assessed. The quality of SLA test items, or item statistics, were developed based on criteria set by the INAP, which adapted the PIRLS and TIMSS. SLA also contains anchor items that allow test scores to be comparable across grades and time.

The SLA tool was developed to measure primary grade reading and mathematics, based on national and international standards. Both skills are fundamental prerequisite to support student's learning in school, as well as their overall success in today's increasingly information and technology-based society (OECD, 2019). To ensure its quality and global comparativeness while maintaining adherence to national standards, competencies measured in the SLA tool are based on standards set in the national curriculum as well as international standardized tests.

We chose the KTSP 2006 or Indonesia's 2006 national curriculum as it is still the primary pedagogical guideline implemented in remote schools of the five KIAT Guru's districts. We referred to the 2016 PIRLS Literacy and 2015 TIMSS Numeracy frameworks, as these tests mainly focus on fundamental reading and mathematical knowledge, procedures, and problem-solving strategies. These alternate versions of the PIRLS and TIMSS were developed as a response to findings that upper primary grade students in some countries are still developing fundamental reading and math skills, and are behind the global average (Mullis & Martin, 2013; Mullis & Martin, 2015). Such state of learning corresponds to the settings found Indonesia's remote areas. Following these frameworks, the SLA features problem sets that have easier passages or numbers and more straightforward procedures.

SLA tests reading and math competencies in content and cognitive domains. Building upon the TIMSS and PIRLS framework, two dimensions were used to design progression of difficulty levels in reading and math competencies: (1) content domain, and (2) cognitive domain. As the SLA was designed to be a time series testing tool, test "packages" across the two phases of the survey were ensured to have similar framework of measurement.

2.1   Content Domain

Content domain consists of topic areas or subject matters learned in primary grade reading and mathematics. Content domain in SLA was determined by both the national curriculum and international frameworks, with sub-sections below presenting details on reading and math content domains.

2.1.1. Reading

Reading literacy is the ability to understand and use written language forms required by society and/or valued by the individual. Readers can construct meaning from texts in a variety of forms. They read to learn, to participate in communities of readers in school and everyday life, and for enjoyment (PIRLS, 2016; OECD, 2019).

Reading literacy is further divided into two general skills: decoding and comprehension.

- **Decoding skill is the technical or foundational ability required to be able to read or understand a text.** This skill is fundamental to have before students can ever "read to learn." This involves competencies in understanding phonology and syllables, and in recognizing letters and words (RTI International, 2015). Although the national curriculum expects students to be able to read letters and simple words by the first grade, it does not break down in detail the progression in achieving this foundational skill. We then referred to framework developed by EGRA[2] (RTI International,

2015) in detailing content domains of decoding skills to be assessed in SLA.

- **Comprehension skill is the ability to construct meaning from written texts.** We reviewed KTSP 2006 and PIRLS Literacy to determine content domain for comprehension. Similar to decoding skill, the national curriculum is limited in its definition and structure of the comprehension skill. As mapped out in Figure 1a in Appendix 1, progression of comprehension in KTSP 2006 is simply determined by the speed of reading fluency, length and type of reading passages, and the ability to summarize and conclude main idea from a passage, including in applying information retrieved to a novel situation. Recognizing this limitation, the INAP expands its reading test framework by adapting PIRLS' definition and processes of comprehension. Correspondingly, we referred to PIRLS Literacy framework to build on KTSP 2006 reading standards by further breaking down content domain involved in comprehension skills. As seen in Figure 1c in Appendix 1, comprehension begins with the ability to retrieve explicit information from the text and progresses to achieve the ability to examine and evaluate implicit information.

Reading literacy is connected to the purpose in reading. Such purpose is determined by the type of passages being read, which is divided into two primary types of texts: literary and informational. The national curriculum does provide guidelines in choosing types of texts for each grade level.

- **Literary or fictional texts are primarily read by early grade students, for pleasure and personal interests** (Mullis & Martin, 2015). As they are catered for younger readers, these passages are typically written in larger fonts, with visual aids, sentences and words of shorter lengths and common usages.

- **Informational texts are primarily read by upper grade students, for learning and participation in society** (Mullis & Martin, 2015). These are passages that students would often find in their daily or direct surroundings, such as letters, announcement, and news articles.

---

[2] Early Grade Reading Assessment or EGRA is a simple instrument that could report on the foundation levels of student learning, including assessment of the first steps students take in learning to read: https://www.rti.org/rti-press-publication/early-grade-reading-assessment

Table 1.
Primary Grade Reading Content Domains Measured in SLA

Process		Competencies	Grade					
			1	2	3	4	5	6
Decoding		**Letter, Word, and Sentence Recognition**	x					
Comprehension	Focus on and retrieve explicitly stated information	Looking for specific ideas						x
		Identifying the setting of a story (e.g., time, place)	x	x		x	x	
		Finding the topic sentence or main idea (when explicitly stated)			x			
	Make straightforward inferences	Inferring that one event caused another event		x	x			
		Concluding what is the main point made by a series of arguments			x			x
		Determining the referent of a pronoun	x					x
		Identifying generalizations made in the text			x			
	Interpret and integrate ideas and information	Discerning the overall message or theme of a text			x	x		
		Considering an alternative to actions of characters				x		
		Interpreting a real-world application of text information			x		x	x
	Examine and evaluate content, language, and textual elements	Evaluating the likelihood that the events described could really happen					x	
		Judging the completeness or clarity of information in the text				x		x
		Determining an author's perspective on the central topic				x		
Type of Text		Literary	x	x	x	x		
		Informational				x	x	x

In SLA, both literary and informational reading passages used were contextually relevant to the learning environment of KIAT Guru's participating schools. This included stories that have rural setting or background, as well as elements, characters and plots that are neutral in terms of religion, gender and culture, especially in respect to cultural values embraced by eastern and western Indonesia.[3]

Based on the standards determined by KTSP 2006, EGRA, and PIRLS Literacy, we then mapped out reading content domains to be assessed for each grade level in the SLA (Table 1).

2.1.2. Math

In math, there are three general content domains: numbers, geometry and measurement, and data and statistics. "Mathematics is essential in daily life for such activities as counting, cooking, managing money, and building things" (Mullis & Martin, 2013). Unlike reading, the national curriculum does lay out in detail the progression of the three math content domains across grade levels (Appendix 2). This progression also corresponds to that measured by TIMSS Numeracy (Mullis & Martin, 2013).

Based on KTSP 2006 and TIMSS Numeracy, we mapped out math content domains to be assessed for each grade level in the SLA. However, some content domains are excluded from the SLA (Table 2). Following rounds of piloting the test items as explained in section 3.2, we had to drop some content domains in several grades as they were found to be too difficult (zero response rate) for students in our remote sample schools. The SLA focuses only on math concepts that are newly introduced in respective grade according to the curriculum standards. Earlier grades' math concepts and content that are repeated in later grades are also excluded from the test.

---

[3] Elements of test items were adapted to suit context of rural villages, including name of persons, places, and objects. Name of persons written in test items were either common names found in West Kalimantan, particularly the Melayu and "Dayak", or East Nusa Tenggara communities. Names from the Bible or the Al-Quran were avoided to reduce religious bias. Places that were used in test items were either those familiar to West Kalimantan and East Nusa Tenggara settings, or those of public knowledge such as the capital city and province of Indonesia. Objects in test items were also ensured to be relevant and present in rural settings. These include type of food, transportation, sports games, animals, and technology.

Table 2.
Primary Grade Math Content Domains measured in SLA

Process	Competencies	Grade					
		1	2	3	4	5	6
Numbers	Whole Numbers	x	x	x	x	x	x
	Fractions & Decimals			x	x	x	x
Geometry & Measurement	Measurement	x	x	x	x	x	x
	Geometry	x	x	x	x	x	x
Data & Statistics	Data & Statistics						x

x	Excluded from SLA

2.2.   Cognitive Domain

Items in the SLA also assesses a range of cognitive or thinking skills involved in solving a problem: knowing, applying, and reasoning. SLA refers to cognitive processes measured by TIMSS and PIRLS' frameworks, which were adapted from Bloom's Taxonomy (Mullis & Martin, 2013; Mullis & Martin, 2015; Anderson & Krathwol, 2001). There are three categories of cognitive domain measured in TIMSS and PIRLS, namely knowing, applying, and reasoning. As portrayed in Table 3 these domains follow the order of thinking processes laid out by Bloom's taxonomy. The order also indicates a continuum from Lower Order Thinking Skills (LOTS) to Higher Order Thinking Skills (HOTS). In measuring these domains, SLA items assess students' abilities in:

- Demonstrating learned knowledge in reading and math,
- Applying what they have learned in solving novel reading and math problems, and
- Reasoning their solution to reading and math problems through analysis and logical thinking.

Table 3 further specifies the type of thinking skills measured in reading and math competencies for each cognitive domain.

Table 3.
Bloom's Cognitive Processes as adapted PIRLS & TIMSS

	Bloom's Taxonomy	Definition	PIRLS & TIMSS	Reading	Math
Lower-Order Thinking	*Remember*	Retrieving relevant knowledge from long-term memory	*Knowing*	To locate and understand content of text that is relevant to the question posed. To move beyond the surface of texts and to resolve the gaps in meaning occurring in texts	Familiarity with mathematical concepts and fluency in mathematical skills
	Understand	Determining the meaning of instructional messages, including oral, written, and graphic communication			
	Apply	Carrying out or using a procedure in a given situation	*Applying*	To construct a more specific or complete understanding of the text by integrating personal knowledge and experience with meaning that resides within the text.	The application of mathematics in a range of contexts.
Higher-Order Thinking	*Analyse*	Breaking material into its constituent parts and detecting how the parts relate to one another and to an overall structure or purpose	*Reasoning*	To critically consider the text from a personal perspective or with an objective view.	Involves logical, systematic thinking; includes intuitive and inductive reasoning based on patterns and regularities to solve problems set in novel or unfamiliar situations
	Evaluate	Making judgments based on criteria and standards			

Table 4.
INAP's Range and Proportion of Item by Difficulty and Discrimination

Level of Item	Difficulty Index	Discrimination Index	% of Items in Test
Easy	(-2.75) - (-1.25)		20%
Medium	(-1.25) - 1.25	$a \geq 0.5$	70%
Difficult	1.25 - 2.75		10%

2.3. Item Statistics

As part of psychometric quality, item statistics is required to determine the test's ability to detect the differing levels of ability of its subjects, and to discriminate between high- and low- performing subjects. In adherence to this standard, SLA test items were designed to be sensitive in measuring the range of student abilities found in remote schools where KIAT Guru pilot is implemented. This would allow the test to detect granular changes in learning outcomes over the years of the intervention. In doing so, the assessment contains distribution of items with range of difficulty levels that can differentiate between high- and low- ability students.

In measuring the test's psychometric quality, we use the 2-Parameter Logistic (2-PL) Item Response Theory (IRT). The two parameters measured are item difficulty and item discrimination. Item difficulty measures the ability needed to pass an item. Item discrimination measures the ability of an item in differentiating low- and high- scoring test takers.

We adapted INAP[4]'s (Puspendik, 2012) criteria for appropriate range of item difficulty and item discrimination parameters, as well as distribution of difficulty level items within a test booklet (Table 4).

---

[4] Indonesian National Assessment Program or INAP is a sample-based national assessment program to monitor student learning outcomes in mathematics, reading and science nationally, as well as providing regional and international comparisons.

2.4. Anchor Items

The SLA tracked learning over time by including anchor items in the different tests. It was designed to track the learning progress of primary grade students before and after KIAT Guru was implemented. To do so, these tests needed to have a set of items in common, also known as anchor items. Anchor items ensure scores are comparable across grades and time by equating the psychometric properties of these tests. As they serve to link tests, anchor items need to represent both content and statistical characteristics of the overall tests (Sinharay & Holland, 2006). These items were selected from each content domain covered in the test, and with difficulty and discrimination levels that cover the range in Table 4.

Similar to other testing tools (Sinharay & Holland, 2006; Rarasati et al., 2020), the SLA contains about 50% anchor items that links across test booklets within the same time period or package. It also contains approximately 30% anchor items linking test booklets of corresponding grade level across time or packages. Although the percentage of anchor items across time series is less than within the same test package, the distribution of competencies measured remain the same. This was done to ensure student performance is still comparable across time without jeopardizing risk of memorization due to re-using the same items as previous test. Anchor items were selected based on their psychometric quality, as they have sufficient item difficulty and item discrimination.

03

Development of the Test

SLA test items were developed based on the national curriculum textbooks, with assistance of experts from the Ministry of Education, other development program, academic institutions, and a local non-governmental organization. These test items were piloted for 11 rounds in 54 remote schools across Indonesia to ensure its psychometric quality.

3.1.   Item Development

Reading and math items were adapted from KTSP 2006 textbooks, and further developed with support from INOVASI, curriculum experts at Yogyakarta State University (UNY), Sanata Dharma University, and CREDO[5]. Puspendik conducted final review of these items.

3.2.   Item Pre-Testing

Test items were piloted for a total of 11 rounds to 3,755 students in 54 remote schools. Prior to its administration in KIAT Guru's 270 pilot schools, the test must first be validated for its psychometric quality and coherence to the framework determined in section 2. The pre-testing covered both east and west of Indonesia, including East and West Java, Special Region of Yogyakarta, West Kalimantan, South Sulawesi, Papua and West Papua. In each round, every test item was tested to a minimum of 100 students per grade level. 2-PL IRT analysis was conducted to measure item statistics. Items with difficulty and discrimination quality that fell within the set range are kept for subsequent rounds, and ultimately as final test items. In contrary, items that were too easy or too difficult or cannot discriminate students' ability well, were revised and retested in subsequent rounds.

---

[5] Creative Education Indonesia Foundation or CREDO is a local NGO that works in promoting early grade literacy and numeracy.

04 Test Design

SLA contains reading and math test items in proportion to the standards set by the 2006 national curriculum, PIRLS Literacy and TIMSS Numeracy. Following multiple rounds of piloting and revisions, proportion of content and cognitive domains assessed in the SLA were adjusted to fit ability of targeted students in Indonesia's remote schools. The finalized version consists of 23 reading items and 30 math items for each grade level test booklet. SLA is presented in multiple-choice item format and is administered differently: oral and individual tests for grades 1 and 2 students, and classical and group tests for grades 3 through 6 students. Individual tests last for up to 40 minutes per student, with 25 minutes for reading and 15 minutes for math. Group tests last for up to 95 minutes, with 45 minutes for reading and 50 minutes for math. Student performance is scored by the percentage of items answered correctly, which can also be classified into its grade-level competency based on the national curriculum standard.

4.1. Test Items by Content Domain

4.1.1 Reading

The final reading tests measured foundational reading skills per national competencies and international standards. Content domain measured in SLA reading test was primarily adapted from KTSP 2006 and PIRLS Literacy. This combination ensures that the competencies being measured are those that are supposed to be taught in Indonesian classrooms, as opposed to having it measured against an ideal framework. Nevertheless, as portrayed in Table 5, finalized proportion of reading content domains measured in the SLA also corresponds to those measured in international tests, but focuses more foundational skills in its reading test. Table 6 maps out proportion of content domains measured per grade-level.

Table 5.
Proportion of Reading Content Domains Measured in PIRLS, PIRLS Literacy, and SLA

Grade	4	4-6	1-6
Competencies	PIRLS	PIRLS Literacy	SLA
Letter, Word, and Sentence Recognition (%)	-	-	26%
Focus on and retrieve explicitly stated information (%)	20%	50%	30%
Make straightforward inferences (%)	30%	25%	16%
Interpret and integrate ideas and information (%)	30%	25%	17%
Examine and evaluate content, language, and textual elements (%)	20%		11%

Table 6.
Proportion of Reading Content Domain Measured in SLA by Grade Level

Grade	Decoding	Comprehension			
	Letter, Word, and Sentence Recognition (%)	Focus on and retrieve explicitly stated information (%)	Make straightforward inferences (%)	Interpret and integrate ideas and information (%)	Examine and evaluate content, language, and textual elements (%)
1	77%	23%			
2		50%	30%	20%	
3		18%	21%	29%	32%
4		36%	18%	33%	12%
5		29%	21%	17%	42%
6		17%	50%	17%	17%

A grade-level reading test booklet covers three-grade levels, with a total of 23 reading items. As portrayed in Table 7, each booklet is composed of items that are either (i) two levels below, (ii) one level below, and (iii) at level the respective grade of test subject.

This proportion was determined based on findings from series of test piloting (Section 3.2) to ensure that each test booklet accommodates the range of student abilities found in remote schools.

Table 7.
Proportion of Grade-Level Reading Competencies In SLA Across Test Booklets

		Item by Standard					
		Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6
Test Booklet	Grade 1	100%					
	Grade 2	83%	17%				
	Grade 3	9%	65%	26%			
	Grade 4		30%	57%	13%		
	Grade 5			17%	65%	17%	
	Grade 6				30%	43%	26%

4.1.2 Math

The final math tests cover more of the number concept. Content domain measured in SLA math test was primarily adapted from KTSP 2006 and TIMSS Numeracy. As portrayed in Table 8 however, proportion of math content domains measured in the SLA is not aligned as much to those measured in international tests. This deviance corresponds to test piloting that found students in remote schools were still primarily learning about number concept. Table 9 further maps out proportion of content domains measured per grade-level.

Table 8.
Proportion of Math Content Domain Measured in TIMSS, TIMSS Numeracy, and SLA

Grade	4	4-6	1-6
Comprehension Skill	**TIMSS**	**TIMSS Numeracy**	**SLA**
Number (%)	50%	65%	92%
Geometric & Measurement (%)	35%	35%	7%
Data & Statistics (%)	15%	N/A	1%

Table 9.
Proportion of Math Content Domain Measured in SLA by Grade Level

Grade	Number (%)		Geometry & Measurement (%)		Data & Statistics (%)
	Whole Numbers	**Fractions & Decimals**	**Measurement**	**Geometry**	
1	100%				
2	91%			9%	
3	70%	20%		10%	
4	70%	25%		5%	
5	44%		11%	44%	
6	40%		40%		20%

A grade-level math test booklet also covers three-grade levels, with a total of 30 math items. As portrayed in Table 10, most test booklets are composed of items that are either (i) two levels below, (ii) one level below, and (iii) at level the respective grade of test subject. An exception is applied to grade 3 and 5 tests that do not contain at level items. Results from the pilots (Section 3.2) indicated that students found these test items too difficult to work on. Each test booklet is ensured to be able to accommodate the range of student abilities found in remote schools.

Table 10.
Proportion of Grade-Level Math Competencies in SLA Across Test Booklets

		Item by Standard					
		Grade 1	**Grade 2**	**Grade 3**	**Grade 4**	**Grade 5**	**Grade 6**
Test Booklet	Grade 1	100%					
	Grade 2	77%	23%				
	Grade 3	13%	87%				
	Grade 4		13%	67%	20%		
	Grade 5			40%	60%		
	Grade 6				53%	30%	17%

Table 11.
Proportion of Cognitive Processes in PIRLS Literacy, TIMSS Numeracy, and SLA

		Reading			Math		
	Grade	4-6		1-6	4-6		1-4
	Cognitive Processes	PIRLS	PIRLS Literacy	SLA	TIMSS	TIMSS for Numeracy	SLA
Lower-Order Thinking	Knowing	20%	50%	73%	40%	50%	8%
	Applying	60%	25%	7%	40%	35%	90%
Higher-Order Thinking	Reasoning	20%	25%	20%	20%	15%	2%

4.2. Test Items by Cognitive Domain

SLA test items primarily measures students' knowledge in reading, while measuring students' application in math. Following its main reference, the PIRLS Literacy and TIMSS Numeracy, the SLA focuses on measuring student's lower order thinking skills. Finalized cognitive domains assessed in SLA were determined following a series of test piloting and multiple revisions to adjust the test to the ability of students in Indonesia's remote schools. Table 11 presents proportion of cognitive domains assessed in SLA, compared to PIRLS Literacy and TIMSS Numeracy.

4.3. Test Format and Administration

4.3.1. Test Format

To ensure efficiency in test administration and objectivity in scoring, SLA test items were written as multiple-choice questions with three to four answer options. The answer options, both correct and incorrect, were constructed carefully to reduce any external factors aside to student's ability. For reading items, these included random ordering of the correct option across questions and applying similar length of sentence or paragraph across options. For math items, these included having similar numbers across options, and using results from applying incorrect concept or procedure in solving the problem[6].

4.3.2. Test Administration

The SLA is administered in two ways: (i) oral and individual test, and (ii) classical and group test.

- **Oral and individual tests are administered to grades 1 and 2 students to accommodate students who are unable to read or even recognize a letter.** Each student receives verbal instruction from an enumerator throughout the testing session. The enumerator is only allowed to read the prompts written in the manual as not to provide hints or aids to students. Students provide their answer verbally, through writing or by pointing the option they choose, which is recorded by the enumerator. Modelled after the EGRA (RTI International, 2015) and ASER[7] (Pratham, n.d.), the test is semi-adaptive in its administration. Students must first successfully answer a set of initial reading decoding items before continuing to work on the rest of the test items. The test normally takes about 40 minutes per student, starting with 25 minutes for reading, followed by 15 minutes for math. This unfortunately extends the time needed to administer the test to a whole class of grades 1 and 2.

- **Classical and group tests are administered to grade 3 through 6 students.** These tests are given to a group of students seated in a classroom, and written instructions are provided in the test booklets. Students who sit on the same desk are

---

[6] An example would be a subtraction problem with an answer option that applies addition instead of subtraction to the numbers presented in the problem. Another example would be a geometry problem that asks for an area of a shape, but instead provides an answer option that calculates the perimeter of the shape.

[7] Annual Status of Education Report or ASER is is a national citizen-led rapid assessment of children's ability to read simple text and do basic arithmetic. ASER is designed and facilitated by the Indian nongovernment organisation Pratham, and has been conducted every year since 2005 by partner organisations in every rural district of India

given test booklets of differing versions, i.e. version A and version B, as to prevent cheating. Students write their answers in the sheets provided. Similarly, testing begins with reading and then followed by math. The test normally lasts for 95 minutes, with 45 minutes for reading, and 50 minutes for math.

4.4. Test Scoring

SLA can be scored using percent or competency classification, which can be further interpreted against the national curriculum target. There are three ways of scoring students' results in the SLA: percent score, and competency classification.

- **Percent score** is calculated by the percentage of correctly answered items. A student earns a full one point for selecting the correct answer option for a test item, and get zero for selecting incorrect answer options.

- **Competency classification** is determined by students' percent score, calculated based on the composition of grade-level items in each test booklet. It indicates which grade-level competency the student's ability is classified into based on their results. Table 12-13 demonstrates range of percent score mapped to its equivalent competency classification.

- **Competency interpretation** against the national curriculum is identified by marking if student's classified grade-level ability is (i) at/ above level or corresponds to their current grade or higher, (ii) below level or behind their current grade, or (iii) at illiteracy state or unable to recognize a set of letters or numbers. The third interpretation is exclusive to grade 1 and 2 tests. Table 12-13 further maps out in colour these three ways of interpreting student's classified competency.

Table 12.
Reading Test Score and Grade-Level Competency Equivalence

Test	Score Range for Competency Level						
	Illiteracy	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6
Grade 1	<30%	>85%					
Grade 2	<13%	45-95%	>95%				
Grade 3		5-40%	40-87%	>87%			
Grade 4			15-60%	60-94%	>94%		
Grade 5				10-42%	42-82.5%	>82,5%	
Grade 6					17-56.5%	56,6-91%	>91%

Table 13.
Math Test Score and Grade-Level Competency Equivalence

Test	Score Range for Competency Level						
	Innumeracy	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6
Grade 1	<20%	>70%					
Grade 2	<10%	40-86.7%	>86.7%				
Grade 3		7-50%	>50%				
Grade 4			7-40%	40-90%	>90%		
Grade 5				20-50%	>50%		
Grade 6					10-53%	54-93%	>93%

Illiteracy/Innumeracy	Below Level	At/ Above Level

05 Psychometric Properties of the Test

Following multiple rounds of piloting and revisions, SLA test items also reached sound statistics. These test items cover a range of difficulty levels appropriate to the distribution of ability of targeted students in remote schools. Analysis were conducted using1 or 2-PL Item Response Theory and the Rasch Model.

5.1. Item Statistics

Using 1 or 2-PL IRT analysis, we found that 90% of SLA test items had sound item statistics based on the set criteria. Graph 1 maps proportion of test items that were categorized as easy, medium, and difficult.

Graph 1.
Proportion of SLA Reading and Math Test Items based on their Difficulty Level

Easy
3%

Difficult
21%

Medium
68%

Reading

Easy
4%

Difficult
16%

Medium
69%

Math
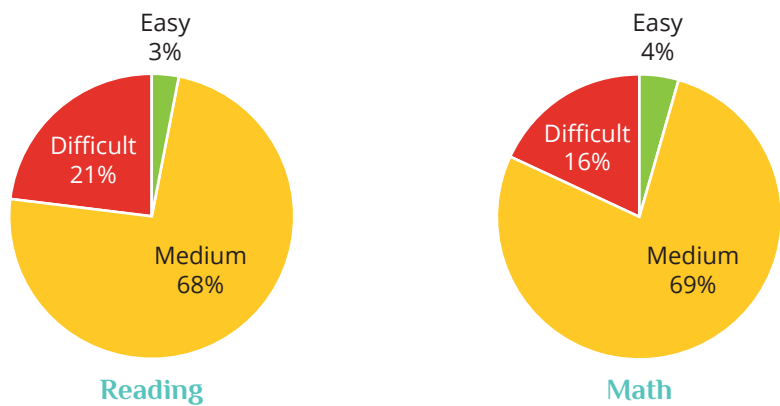
5.2. Rasch Model

We used the Rasch Model and found the appropriateness of SLA for remote schools in Indonesia's lagging regions. The Rasch Model evaluate the distribution of difficulty level and differentiate between low- and high- ability students in its targeted population, which An item-to-person map, or the Wright Map, is created to display the fitness of the distribution of SLA items' difficulty level with targeted students' ability. On the right side of Figure 1, we can see that distribution of grade 1 SLA reading items corresponded to the distribution of grade 1 student ability in KIAT Guru's 270 remote schools. Interestingly, this distribution differs for grade 1 students in remote schools of Yogyakarta, an advantaged region, where we piloted the test. Here, students only fell within the mid-to- high ability range, deeming the test to be easier for them relative to students in KIAT Guru's schools. This trend seemed to also apply to early grade students in math, as portrayed in Figure 2.

Figure 1.
Range of Grade 1 Student Reading Abilities in Remote Schools of Disadvantaged and Advantaged Regions
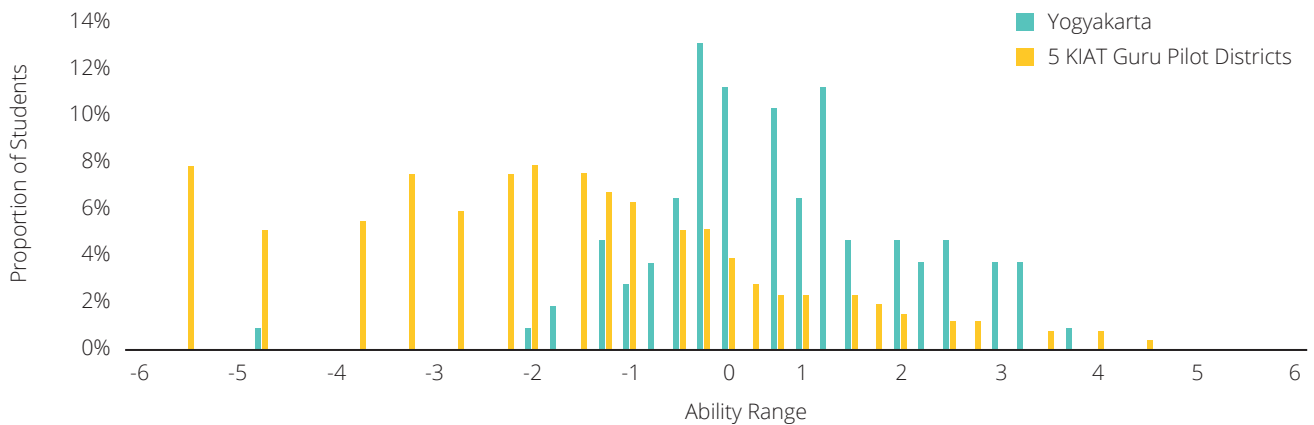
Data Source: SLA piloting in Yogyakarta and SLA Baseline in 5 KIAT Guru Pilot Districts, World Bank, 2016

Figure 2.
Range of Grade 2 Student Math Abilities in Remote Schools of Disadvantaged and Advantaged Regions
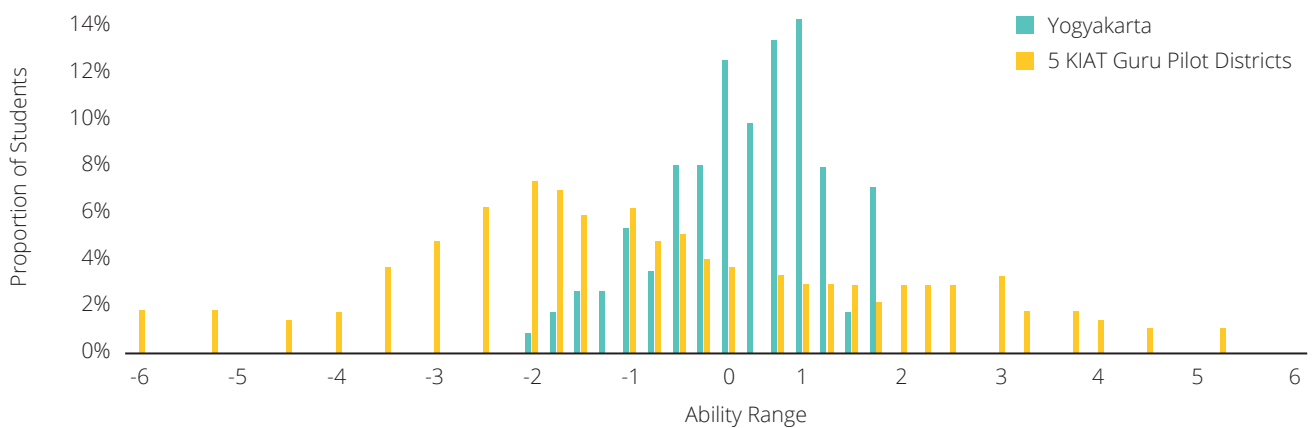
Data Source: SLA piloting in Yogyakarta and SLA Baseline in 5 KIAT Guru Pilot Districts, World Bank, 2016

This finding further demonstrates the differing state of learning in remote areas of Indonesia as compared to those in the country's more developed regions. This SLA tool is mostly appropriate for assessing student learning outcomes in remote schools of Indonesia's lagging regions.

06 Summary and Discussion

Building upon both national and international framework, the SLA is designed and developed to be able to measure primary grade student learning outcomes in reading and math in Indonesia's remote schools. This technical paper documents that:

SLA is a valid measure to assess learning outcomes in primary grade reading and math, covering a wide range of content and cognitive domains demanded by the national curriculum and international standardized tests;

SLA test items were developed and contextualized according to the ability of students in targeted schools;

SLA anchor items allowed test scores to be comparable across grades and time, and

SLA produces test scores can easily classify and interpret student's competencies against the national curriculum.

The SLA differs from its predecessors in Indonesia, as it covers all primary grade reading and math, and more appropriate for remote schools in Indonesia's lagging districts. Unlike EGRA, BERMUTU, and INAP, the SLA not only covers a more comprehensive range and domains of primary grade reading and math, it is also tailored to measure learning in remote area contexts of disadvantaged regions. It is critical for an assessment to be adjusted to the ability of its subjects. This allows the tool to detect granular changes in learning. The study found that students in remote schools of disadvantaged districts in Indonesia were behind even their peers in remote schools of advantaged districts. Consequently, an assessment that seeks to measure learning across Indonesia's diverse socioeconomic conditions will be limited in its capacity to recognize differences within the low-ability student population of the country's remote and disadvantaged areas. The SLA can address this limitation and provide thorough examination of changes or differences in student learning of disadvantaged remote area context.

Appendixes

Appendix 1. Progression of Primary Grade Reading Skills across National and International Standards

Figure 1a. Progression of Primary Grade Reading Skills in KTSP 2006

Standards	Decoding & Comprehension					
	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6
	Fluency (syllable, word, sentence)	Fluency (text)	Fluency (text)	Fluency (text)	Fluency (text)	Retrieve information from multiple texts
		Summarize text	Main idea of text	Main idea of text	Summarize text	Evaluate implicit messages from text
				Apply information from text	Retrieve information from multiple texts	
Type of Text	Literature	Literature	Literature	Literature Information	Information	Information
Length of Text (sentences)	2-4	10-25	20-40	30-40	N/A	N/A

Figure 1b. Progression of Primary Grade Reading Skills in EGRA

	Decoding						
Orientation to Print	Letter Name Identification	Letter Sound Identification	Initial Sound Identification	Segmentation (phoneme or syllables)	Syllable Identification	Familiar Word Reading	Non-word Reading

Figure 1c. Progression of Primary Grade Reading Skills in PIRLS Literacy

	Comprehension			
Processes of Comprehension	Focus on and retrieve explicitly stated information	Make straightforward inferences	Interpret and integrate ideas and information	Examine and evaluate content, language, and textual elements
Competencies	Identifying information that is relevant to the specific goal of reading	Inferring that one event caused another event	Discerning the overall message or theme of a text	Evaluating the likelihood that the events described could really happen
	Looking for specific ideas	Concluding what is the main point made by a series of arguments	Considering an alternative to actions of characters	Describing how the author devised a surprise ending
	Searching for definition of words or phrases	Determining the referent of a pronoun	Comparing and contrasting text information	Judging the completeness or clarity of information in the text
	Identifying the setting of a story (e.g., time, place)	Identifying generalizations made in the text	Inferring a story's mood or tone	Determining an author's perspective on the central topic
	Finding the topic sentence or main idea (when explicitly stated)	Describing the relationship between two characters	Interpreting a real-world application of text information	

Appendix 2. Progression of Primary Grade Math Skills across National and International Standards

Content Domain	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6
Numbers						
Whole Numbers	Represent, count and order numbers 1-20	Compare and order numbers 1-500	Compare and order numbers in number line	Order and compute with whole numbers	Compute and round whole numbers	Solve problems by computing whole numbers, GCF and LCM, multiples, and roots
	Place Value: ones, tens	Place Value: ones, tens, hundreds		Round whole numbers		
	Add and subtract numbers 1-20	Add and subtract numbers 1-500	Add and subtract numbers up to 3-digits	Solve problems involving money		
		Multiply and divide with numbers up to 2-digits	Solve problems involving money	Identify and solve problems with Greatest Common Factor (GCF) and Least Common Multiple (LCM)	Use prime factors to identify GCF and LCM	
					Solve problems by computing whole numbers, GCF and LCM	
				Identify Roman numerals	Identify multiples and roots of numbers	Identify cube roots of numbers
Fractions & Decimals			Identify and compare simple fractions	Identify, compare, and order fractions	Identify equivalent of fractions in percentages and decimals	Identify and order equivalent of fractions in simple fractions, decimals
				Add and subtract fractions	Add, subtract, multiply, and divide fractions	Compute with different formats of fractions
					Solve problems related to comparison and scales with fractions	Solve problems related to comparison and scales

Content Domain	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6
Geometry & Measurements						
Measurement	Measure and estimate time, lengths, and weights	Measure and estimate time, lengths, and weights with common measurement units	Identify and use measurement tool to measure time, length, and weight	Identify different types of angles	Identify and use military hours unit to measure time	Identify measurement units for volume flow rate
			Identify relationships between measurement units used in time, length, and weight	Identify relationships between measurement units used in time, length, and weight	Compute time	
					Compute angle	
					Identify measurement units for distance and speed	
					Solve problems related to time, distance, and speed	
Geometry	Describe and compare common two- and three-dimensional shapes	Compare common two-dimensional shapes	Identify common two-dimensional shapes by lines and angles	Identify properties and relationships across common two-dimensional shapes	Identify properties of common two- and three-dimensional shapes	Calculate areas of two-dimensional shapes with multiple lines of symmetry
			Calculate perimeter and area of squares and rectangles	Calculate perimeter and area of parallelogram and triangles	Calculate area of trapezium and quadrilateral	Calculate areas of circles
					Calculate and solve problems related to volume of cubes and cuboids	Calculate volume of prisms and cylinders
					Solve problems related to two- and three-dimensional shapes	Identify and use coordinate systems to locate a point
Data & Statistics						
Data & Statistics						Collect, read, and compare data
						Represent data in tables, diagrams, graphs, and pie charts
						Order and compute means, modes, maximum and minimum of data

Appendix 3. IRT Table

SLA Reading Baseline (Package 1)

Item	Grade 1		Grade 2		Grade 3		Grade 4		Grade 5	
	Discrimination	Difficulty	Discrimination	Difficulty	Discrimination	Difficulty	Discrimination	Difficulty	Discrimination	Difficulty
1	1.726082	-1.148702	1.399046	-2.952959	0.507956	2.60884	1.33826	-0.908697	1.666646	-0.454285
2	1.696719	-1.133678	1.14399	-2.90031	0.548841	1.764333	1.375389	0.0071913	0.534914	0.4384911
3	1.825638	-0.502765	1.285736	-2.035691	1.457256	-0.015289	0.336514	2.010362	1.239825	-0.068568
4	0.792211	1.343351	0.774957	0.3677987	0.820439	1.752396	1.189902	0.6788612	0.651593	0.4726871
5	1.939554	-0.653659	2.54136	-1.058647	0.953754	0.7204816	0.665786	0.9033396	0.593978	1.298892
6	1.654596	0.1334687	1.522618	-0.091557	0.724781	1.64537	1.109828	0.1532226	0.96972	0.3785087
7	1.470525	0.3262142	2.890909	-0.760393	1.708136	0.1372199	0.458575	2.037976	1.171444	-0.497438
8	2.32125	-0.538264	1.554807	0.7170989	1.170177	0.5014617	1.701716	-0.263989	0.417362	2.096051
9	4.176869	0.1941315	1.772324	-0.465269	1.166683	1.285835	1.081264	1.030985	1.40734	0.5327368
10	2.424922	1.097693	1.950722	0.0146553	1.048081	0.6641827	0.927444	0.707176	1.148915	-0.204168
11	2.60101	0.9398396	3.134325	-0.39892	0.388385	3.684493	1.665614	-0.308622	0.572812	1.401499
12	5.046417	0.4413528	1.298913	0.709576	1.343786	0.5755389	1.473242	-0.088808	0.222292	0.9241387
13	2.257278	1.460216	3.687311	0.7856178	1.12887	0.7259374	0.61448	1.198723	0.879319	1.170283
14	3.513505	0.5920905	2.762969	0.9748071	0.473325	1.847306	0.078554	17.47925	1.562218	-0.343554
15	2.624229	1.087814	1.594772	-0.010854	1.755125	0.4289408	0.611668	2.183892	0.57346	-0.108889
16	4.958155	0.664636	3.544031	0.8573737	0.502908	2.700921	1.137979	-0.585788	0.047526	33.35498
17	2.386752	1.399921	5.426226	0.7858123	1.212225	0.5443407	0.592595	1.828072	0.754175	1.292472
18	3.384257	0.8420553	5.535071	0.9386512	0.519692	2.258052	1.007507	0.7095546	0.512124	1.58888
19	2.728372	1.451376	3.263832	1.112546	1.001249	1.817369	0.805943	1.018006	0.580477	1.926745
20	4.382351	1.703075	5.37151	0.8227956	0.576716	2.147075	0.38124	3.14588	1.410154	-0.195211
21	5.193728	1.699695	3.35792	1.492437	1.269643	0.7407944	0.963831	0.1380707	0.384853	1.509153
22	5.316992	1.746375	3.217904	1.615482	0.901129	1.676785	0.322565	3.775203	1.1556	0.1124483
23	4.593729	1.716049	4.436557	1.150323	0.650642	1.955094	0.199233	2.930404	0.723992	0.3710293

SLA Math Baseline (Package 1)

Item	Grade 1		Grade 2		Grade 3		Grade 4		Grade 5	
	Discrimination	Difficulty	Discrimination	Difficulty	Discrimination	Difficulty	Discrimination	Difficulty	Discrimination	Difficulty
1	2.53135	-1.238289	2.088436	-2.99628	2.349847	-0.266395	0.633507	0.75068	1.419173	0.3337839
2	2.475974	-0.448777	1.839738	-2.088702	1.812478	-0.132339	1.934349	-0.339064	0.65432	1.012898
3	2.569052	-0.484728	1.667684	-1.857549	1.813929	0.3477874	2.154471	-0.425487	0.760909	0.1246353
4	2.235379	-0.35526	1.946031	-1.874666	1.805011	0.2752656	2.165457	-0.279046	1.123127	0.5424175
5	1.554726	0.4919961	2.140272	-2.048421	1.914074	0.0367742	0.650078	3.560769	0.237501	2.845082
6	2.815867	-0.275918	1.319105	-1.330723	1.276006	2.8500056	0.814901	0.9615421	0.808051	0.5509062
7	2.522269	-0.487209	1.20165	-0.931037	1.703279	0.2793539	0.471489	1.839897	0.77557	0.6014187
8	4.582659	-0.504271	1.152918	-0.47654	2.15066	-0.311548	1.433094	-0.73216	0.747539	1.015794
9	2.733152	-0.003826	1.513159	-0.036162	2.263477	0.1145736	1.02549	1.08656	0.491139	2.613124
10	2.948763	-0.01886	1.518888	-0.041774	1.716489	0.01755894	0.993566	1.070696	0.30302	2.954376
11	3.032406	0.0935444	1.525797	-0.18339	0.905004	0.6820449	0.436437	2.678861	0.990081	0.2606717
12	3.68178	0.0642398	3.062258	0.0927456	0.853948	1.367185	1.177392	-0.003887	1.408806	-0.319252
13	2.402454	0.656554	3.878936	0.037508	0.489004	2.427819	0.799108	0.4214697	1.006308	-0.76187
14	3.843054	0.0616186	4.749316	0.3176988	2.660355	0.0718763	0.822519	0.4144612	0.656377	1.612259
15	2.010597	0.8212416	6.363282	0.3143823	2.751431	0.0642512	0.698083	1.874505	1.007088	0.1342398
16	2.624795	0.8381935	6.330131	0.4285062	2.584947	0.0771308	0.763282	0.2863472	0.575134	2.088858
17	2.470231	0.8873087	4.180654	0.6169829	1.148705	1.020185	0.693961	1.727293	0.287106	6.22812
18	3.261759	0.9255081	3.828617	0.7495462	1.025696	1.079888	0.284867	3.300461	0.548746	2.400365
19	5.241226	0.8866924	4.119555	0.7757928	0.640103	2.59177	0.550415	1.453706	0.267142	1.458036
20	5.137071	1.098709	9.19978	0.7439134	1.262369	0.8073305	0.464686	3.00204	0.564847	1.265079
21	6.91919	1.045022	11.67996	0.7466294	2.433312	0.1818371	0.214958	5.676555	0.085022	8.475554
22	5.578325	1.229977	10.76973	0.7825806	1.98794	0.0449504	0.230056	4.546078	1.000409	-0.11093
23	6.467205	1.085797	10.07992	0.8550429	2.09353	0.2784479	0.936113	0.8175254	0.378395	2.310343
24	5.653019	1.283602	8.744648	0.8950822	1.427721	0.4596296	0.182099	6.343643	0.532298	1.322774
25	6.820591	1.537392	5.964198	1.146937	1.6843	0.3600213	0.799817	0.6320992	0.229523	4.517308
26	8.648647	1.510678	5.5453	1.206978	1.136455	0.872468	0.897107	0.2061405	0.564847	1.265079
27	8.281582	1.606504	4.353544	1.541147	0.631459	1.936271	0.424584	1.968102	0.611463	1.298977
28	7.11206	1.634447	5.626723	1.389316	0.76788	1.453	0.90017	0.8792654	0.671055	1.176517
29	6.892256	1.717722	4.215483	1.413863	1.613272	0.366846	0.275459	3.648467	0.392126	1.722477
30	6.63258	1.731589	4.32549	1.445767	1.59079	0.4077776	0.170536	9.003603	0.920828	0.2753411

SLA Reading End line (Package 2)

Item	Grade 1		Grade 2		Grade 3		Grade 4		Grade 5		Grade 6	
	Discrimination	Difficulty	Discrimination	Difficulty	Discrimination	Difficulty	Discrimination	Difficulty	Discrimination	Difficulty	Discrimination	Difficulty
1	0.833904	-1.532764	0.6162487	-2.799742	1.37519	-0.818677	1.633951	-0.677074	1.510678	-1.114951	1.282651	0.0578075
2	0.994297	-2.813318	0.4058514	-7.778756	1.67861	-0.097874	1.664499	-0.562226	0.7582155	-0.100539	1.236434	-0.806215
3	1.276882	-1.254579	0.9075249	-1.880666	1.742441	-0.128429	1.050764	0.3266511	1.174332	-0.242745	0.8252008	-0.142097
4	1.624237	-1.473206	1.035016	-1.774996	1.539841	-0.25655	1.546466	-0.202602	1.177311	-0.700482	0.984271	-0.69864
5	1.496301	-0.980163	1.237714	-1.009673	1.802928	-0.161483	0.6828684	0.4733046	0.6135073	-0.029177	0.5568585	-1.213205
6	0.906038	-0.747206	1.120071	-0.274971	1.135667	0.0714936	1.519795	-0.636958	1.344011	-0.741249	1.003763	0.1152721
7	1.014262	-0.572244	1.034306	-0.600256	1.796814	-0.113701	0.987738	-0.452085	1.368075	-0.997362	1.453542	-0.620366
8	1.38883	-1.576089	0.7912606	-0.832093	1.494885	-0.132124	1.690552	-0.53904	0.8397941	0.6337614	0.9082666	-0.145394
9	1.885474	0.2529645	1.025591	-0.172668	0.8165959	0.0298401	0.9975456	-0.511315	1.420169	-0.202104	1.0276	-0.744462
10	1.53056	-0.141633	1.148225	-0.134474	1.667049	-0.08724	0.871916	0.4015506	0.9443711	-0.115369	1.034344	-1.113038
11	1.560217	0.5014766	0.7512388	-0.214666	0.4337181	1.949173	1.4023	-0.455741	0.9271455	-0.236786	0.9604892	-0.778243
12	2.572899	0.050256	0.7210002	1.136703	1.165068	-0.018011	1.637369	-0.524074	1.477284	-0.388581	0.9343631	0.1576769
13	1.193696	0.2460793	0.4925165	2.416409	1.05114	0.45543	0.7643273	0.40108	1.007021	0.6690943	1.09002	-1.070896
14	1.384348	0.7469141	3.978685	0.4689317	0.3444733	4.921507	0.5573671	1.090566	1.559609	-0.240608	1.595532	-1.107776
15	1.733386	0.5909955	0.5770744	5.978484	1.465587	0.2573165	0.8647534	0.8098092	1.365648	-0.303473	0.7043667	-0.19988
16	1.132327	0.5907498	2.753894	1.091929	0.6939971	1.368382	1.401273	-0.281064	0.6587065	0.3121343	1.272945	-0.745601
17	0.876522	1.800397	4.025992	0.9063923	1.435129	0.2454346	0.7707307	0.639303	1.259404	0.0697375	1.386296	-0.376134
18	0.527901	3.154615	5.817271	0.7505379	0.8065621	1.16174	0.963354	0.9902582	0.701488	0.5064796	0.9055378	-0.692884
19	1.311387	0.2600354	0.320976	3.9005641	0.9880878	0.4329442	1.026352	0.2168748	0.8516309	0.5180018	0.645038	0.775244
20	1.483049	1.589202	7.064068	0.8579724	0.6530824	1.542648	0.9186006	0.4837157	1.383829	-0.897509	1.083238	-0.575654
21	2.158245	1.673952	5.746078	0.8827724	1.176072	0.2746286	1.105376	-0.513163	0.2620303	6.575894	0.55425	0.9402913
22	2.698539	1.94614	5.395918	0.8171604	0.9074641	1.34611	0.9659467	0.1761922	0.8529163	0.092485	0.7615626	-0.474541
23	1.999061	2.183489	5.402901	0.9754183	0.7595127	1.183291	1.401273	-0.281064	0.5492888	2.574139	0.077872	9.205555

SLA Math End line (Package 2)

Item	Grade 1		Grade 2		Grade 3		Grade 4		Grade 5		Grade 6	
	Discrimination	Difficulty	Discrimination	Difficulty	Discrimination	Difficulty	Discrimination	Difficulty	Discrimination	Difficulty	Discrimination	Difficulty
1	2.042642	-1.900745	0.7855323	-5.386684	2.082587	-0.701139	0.6877296	0.1667201	1.553771	-0.803973	1.126063	1.126063
2	2.360354	-1.509536	0.9313111	-3.710247	1.73594	-0.448182	1.84524	-0.46608	0.8189549	-0.283281	1.141456	1.141456
3	2.017997	-1.40555	0.9756763	-3.49297	1.463903	0.008951	0.997431	-2.965913	1.109218	-0.191009	1.612399	1.612399
4	1.844772	-1.022006	0.8047974	-4.315434	1.56233	-0.062389	2.027465	-0.47259	1.195919	-0.174253	0.9963299	0.9963299
5	0.890759	-0.56484	0.9536321	-3.579793	1.829509	-0.118001	1.379429	3.87591	-0.6105352	-3.598425	1.045397	1.045397
6	1.958429	-1.269148	0.6815197	-2.671911	1.25447	0.1359126	0.872019	3.86541	1.117318	0.018445	0.6288742	0.6288742
7	2.064287	-1.134668	-0.3758798	-7.288909	1.666829	-0.148997	0.8970154	0.2998873	0.9614638	-0.347058	0.9146634	0.9146634
8	2.041411	-1.242241	0.5363857	-2.095408	1.96205	-0.639014	1.718846	-0.966404	0.8946728	0.1351665	0.8881107	0.8881107
9	1.263447	-0.89402	0.7958659	-1.12394	1.678794	-0.175131	1.564291	4.88861	1.225008	-0.697185	1.875417	1.875417
10	1.435289	-0.810341	0.7120551	-0.402816	1.361516	-0.05531	0.9705069	0.7691544	0.7092479	1.224287	1.025863	1.025863
11	1.477731	-0.568066	1.177797	-0.213459	0.7885114	2.166542	0.874319	6.81721	1.63151	-0.448658	1.211493	1.211493
12	1.988402	-0.624722	1.99246	-0.443764	1.472983	0.6612677	1.276698	-0.269723	1.590349	-0.667327	1.383395	1.383395
13	1.499922	-0.093788	2.230628	-0.577486	1.679182	-0.064927	1.231214	-0.113512	1.100513	-0.939355	0.0251851	0.0251851
14	2.196904	-0.725192	2.701432	-0.428865	2.18181	-0.334862	0.9481405	0.367906	1.103533	0.4798102	0.5292331	0.5292331
15	0.9362694	1.351786	2.956135	-0.271385	2.02119	-0.265959	0.6438759	1.299462	0.7386974	-0.044091	0.6716359	0.6716359
16	1.316832	-0.168166	2.976146	-0.095977	1.989327	-0.248866	1.16197	-0.322528	0.9717749	0.1289318	0.8055911	0.8055911
17	1.011773	0.9003263	2.317997	0.2550298	2.195968	0.0416416	1.012796	0.5350371	0.5165197	0.3820773	0.7650541	0.7650541
18	1.554675	0.4056531	2.146619	0.3927998	1.305367	0.3604885	0.4658422	1.803055	0.0572554	1.19738	1.381376	1.381376
19	1.768897	0.5816198	2.505661	0.370739	1.117524	0.472973	0.6350774	0.5647212	0.1257055	0.6057673	0.9377368	0.9377368
20	1.828941	0.8174596	4.433222	0.2642808	0.4235345	2.581048	0.4651424	1.89728	1.33525	-0.569748	0.7442676	0.7442676
21	2.015649	0.549603	5.248918	0.2463295	1.621203	-0.411999	0.632928	0.8837623	0.2094887	1.192318	0.7934919	0.7934919
22	1.883999	1.144427	3.954407	0.271305	2.008393	-0.071326	0.349865	-4.765273	0.6965592	0.0384609	0.240596	3.875202
23	2.691537	0.5800011	4.589058	0.4280168	1.920986	-0.098869	1.657754	-0.731518	0.7885397	0.8072004	0.8063371	0.8063371
24	1.899737	1.155924	3.354958	0.5526276	1.745436	-0.182519	0.5648879	0.2177831	0.7234799	-0.446016	-0.0459637	-0.045964
25	1.732492	1.122509	2.512169	0.9375793	1.458485	-0.025181	0.8081241	-0.020549	0.3496982	1.535067	0.948255	0.948255
26	2.198837	1.256283	1.983538	1.013151	1.012222	0.4685539	0.8019485	0.0766484	1.027744	-0.226771	0.09865	2.943502
27	2.32217	1.495123	1.678153	1.814372	0.9496688	0.7562888	-0.3353393	-5.469203	0.7030986	0.8244066	0.1818646	0.1818646
28	1.539527	1.522275	2.556258	1.322562	0.7649633	1.051652	0.8783163	0.3169514	0.8818588	-0.029632	0.8935592	0.8935592
29	1.456849	1.753378	2.378755	1.346242	1.589777	0.1899712	0.2667942	4.666095	-0.6907756	-3.052388	0.6061188	0.6061188
30	1.75443	1.87919	1.866565	1.365129	1.534418	0.0448854	0.3459536	3.746009	0.9587477	-0.235476	1.286371	1.286371

References

ACDP. (2014). Study on Teacher Absenteeism in Indonesia 2014. Jakarta, Indonesia: Education Sector Analytical and Capacity Development Partnership.

Anderson, L.W., and D. Krathwohl (Eds.) (2001). A Taxonomy for Learning, Teaching and Assessing: a Revision of Bloom's Taxonomy of Educational Objectives. Longman, New York.

Gaduh, A., Pradhan, M., Priebe, J., and Susanti, D. (2020). Scores, Camera, Action? Incentivizing Teachers in Remote Areas. RISE Working Paper Series. 20/035. https://doi.org/10.35489/BSG-RISE-WP_2020/035.

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.

Mullis, Ina V.S. and Michael O. Martin (eds.) (2015) PIRLS 2016 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.

OECD. (2019). PISA 2018 Assessment and Analytical Framework. PISA, OECD Publishing, Paris, https://doi.org/10.1787/b25efab8-en.

Mullis, I. V. S., & Martin, M. O. (2000). Item analysis and review. In M. O. Martin, K. D. Gregory, & S. E. Stemler (Eds.), TIMSS 1999 technical report (pp. 225–234). Chestnut Hill, MA: Boston College, International Study Center.

Pratham (n.d.). ASER Assessment and Survey Framework. Mumbai: Pratham Resource Centre.

Pusat Penilaian Pendidikan (Puspendik). (2012). 2012 Indonesian National Assessment Program. *Kemampuan Membaca Siswa Kelas IV Sekolah Dasar: di Provinsi Kalimantan Timur dan D.I. Yogyakarta*. Kementerian Pendidikan dan Kebudayaan, Badan Penelitian dan Pengembangan, Pusat Penilaian Pendidikan.

Pusat Penilaian Pendidikan (Puspendik). (2012). 2012 Indonesian National Assessment Program. *Kemampuan Matematika Siswa Kelas IV Sekolah Dasar: di Provinsi Kalimantan Timur dan D.I. Yogyakarta*. Kementerian Pendidikan dan Kebudayaan, Badan Penelitian dan Pengembangan, Pusat Penilaian Pendidikan.

Rarasati, N., Dharmawan, G., Swarnata, A., Zulfa, A.H., and Lim, D. (2020). Comprehensive Reading and Mathematics Assessment Tool (CERMAT). SMERU Technical Report. Jakarta: SMERU Research Institute, 2020.

RTI International. (2015). Early Grade Reading Assessment (EGRA) Toolkit, Second Edition. Washington, DC: United States Agency for International Development.

Sinharay, S., & Holland, P. (2006). Choice of Anchor Test in Equating. Research Report. ETS RR-06-35. ETS Research Report Series.

Stern, J. & Nordstrum, L. (2014). Indonesia 2014: The National Early Grade Reading Assessment (EGRA) and Snapshot of School Management Effectiveness (SSME) Survey. North Carolina, USA: Research Triangle Park Press.