

Bridging Assessment and Learning: A Formative Assessment Instrument for Fraction Understanding in Elementary Mathematics

Elita Rachmayanti¹

¹ Research and Educational Evaluation, Postgraduate Program, Universitas Negeri Jakarta, Indonesia

¹Email: elita.rachmayanti@gmail.com

Proceedings of the Kalijaga International Conference on Education 2026

Abstract. This study aims to develop and analyze the quality of a formative test instrument on fraction material for fifth-grade elementary school students using the Research and Development (R&D) method. The development process comprised the stages of needs analysis, blueprint construction, item writing, expert validation, field testing, and item analysis using the Classical Test Theory (CTT) approach. Data were collected through a tryout involving 30 students using 20 multiple-choice items. The analysis was conducted to determine item difficulty, discrimination index, distractor function, item validity, and instrument reliability. The results indicate that all items were declared valid, with difficulty levels ranging from easy to moderate. The reliability coefficient of $r_{11} = 0.934$ indicates that the instrument possesses very high internal consistency. This instrument can be used as a formative assessment tool to measure students' understanding of fraction concepts, encompassing addition, subtraction, multiplication, division, and comparison of fractions. This study further underscores the importance of construct validity and effective distractor functioning in producing high-quality assessment instruments.

Keywords: formative test instrument; fractions; item analysis; reliability; validity

INTRODUCTION

Assessment is an important component of the learning process because it functions as a tool to determine the extent to which students have achieved the expected competencies. Assessment carries a specific terminology that gives meaning to and describes the activities carried out by teachers to obtain information about students' attitudes, knowledge, and skills [1]. In other words, assessment plays a role in observing the learning process and its level of success, enabling teachers to make decisions based on the results obtained — whether in the form of feedback, selection, placement, or other decisions. Assessment that is specifically designed to provide feedback to students is referred to as formative assessment.

Formative assessment is an assessment conducted during the learning process, oriented toward the process itself, and provides teachers with information regarding the quality of the instruction that has taken place [2], [3]. The effectiveness of formative assessment becomes apparent when the feedback provided is continuous, specific, timely, and actionable for students. Therefore, formative assessment must be designed to enhance students' understanding of the material currently being studied [4].

At the elementary school level, mathematics is often regarded by students as a difficult subject [5], [6]. The topic of fractions is one of the materials that frequently causes conceptual difficulties for students. Numerous studies have shown that students often experience misconceptions regarding fraction material because they find it difficult to understand the relationship between the numerator and denominator [7]. Therefore, an assessment instrument is needed that is capable of accurately measuring student understanding and providing diagnostic information beneficial to teachers. A good assessment instrument must satisfy the requirements of validity and reliability, and possess good item characteristics, including difficulty level, discrimination index, and distractor function [8], [9]. According to Tavakol and Dennick [10], instrument reliability reflects the internal consistency of items, indicating the extent to which an instrument produces stable results when used under similar conditions. Meanwhile, validity indicates the extent to which an instrument is capable of measuring the intended construct [11].

In practice, the development of quality learning evaluation instruments often refers to the Research and Development (R&D) approach, which allows researchers to systematically develop educational products and test them through several stages [12]. This model has been widely used in the development of learning tools, including test instruments [13], [14], [15], [16]. This study focuses on the development and quality analysis of a formative test instrument on the topic of fractions for fifth-grade elementary school students. Through the analysis of difficulty level, discrimination index, distractor function, validity, and reliability, this study is expected to contribute to improving the quality of formative assessment in elementary schools and to enrich the literature on instrument development based on classical test theory.

METHODS

Research Type and Design

This study employed a Research and Development (R&D) approach aimed at producing an educational product in the form of a valid and reliable formative test instrument on fraction material for fifth-grade elementary school students. The development model adopted was adapted from the Borg and Gall model [17], tailored to the context of instrument development according to Mardapi [9] comprised the stages of needs analysis, blueprint construction, item writing, expert validation, field testing, and item analysis.

Research Subjects and Location

The research subjects consisted of 30 fifth-grade students from SDN 200 Leuwipanjang, Bandung City, selected purposively. Sample selection took into account similarity of characteristics with the target population. The tryout was conducted during the odd semester of the 2025/2026 academic year.

Instrument Development Procedure

The instrument development procedure was carried out through several systematic stages. The first stage was needs analysis, conducted by reviewing curriculum documents and interviewing teachers to identify the difficulties experienced by students in fraction material. The second stage was the construction of the instrument blueprint, developed based on the Learning Outcomes (Capaian Pembelajaran/CP) for the numbers element in the Merdeka Curriculum. Through the identification of these learning outcomes, several indicators were formulated to assess students' abilities in addition, subtraction, multiplication, division, and comparison of equivalent fractions, non-equivalent fractions, and fractions with different denominators. The overall learning outcome underlying the blueprint states that students are able to compare and order various fractions, including mixed fractions; perform addition and subtraction of fractions; and perform multiplication and division of fractions with whole numbers. The blueprint developed is presented in Table 1.

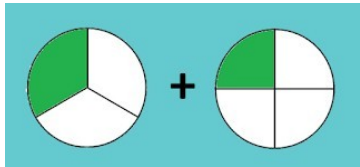
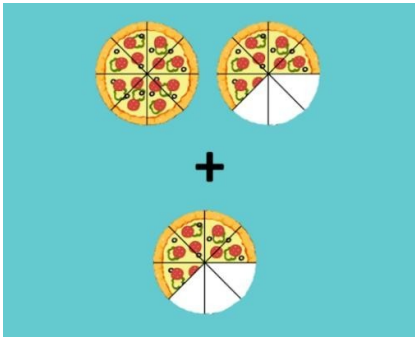
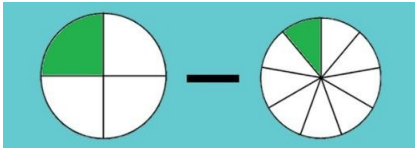
Table 1. Instrument Blueprint

Learning Objective (TP)	Indicator	Level	Item No.	Count
Perform addition of fractions with whole numbers	1. Operate (C3) addition of fractions	C3	1	1
	2. Analyze (C4) word/pictorial problems on addition of fractions	C4	2, 3, 4, 5	4
Perform subtraction of fractions with whole numbers	1. Operate (C3) subtraction of fractions from word problems	C3	7, 10	2
	2. Analyze (C4) pictorial problems on subtraction of fractions	C4	6, 8, 9	3
Analyze multiplication of fractions with whole numbers	1. Operate (C3) multiplication of fractions with whole numbers	C3	11, 12	2
	2. Analyze (C4) word problems on multiplication of fractions with whole numbers	C4	13, 14	2
Analyze division of fractions with whole numbers	1. Operate (C3) division of fractions with whole numbers	C3	15	1
	2. Analyze (C4) word problems on division of fractions with whole numbers	C4	16	1
Compare various fractions, including mixed fractions	1. Compare (C2) fractions with the same denominator	C2	17	1
	2. Compare (C2) fractions with different denominators	C2	18	1
	3. Compare (C2) simple fractions with mixed fractions	C2	19, 20	2

The third stage was item writing, which produced 20 multiple-choice items with five answer options. The researcher then conducted expert validation, carried out by one lecturer specializing in educational evaluation and one experienced fifth-grade elementary school teacher, to assess the appropriateness of the instrument's content, construction, and language. Following this, the instrument was revised based on input from the experts to ensure improved instrument quality.

Table 2. Items

Indicator	No.	Item
1. Operate (C3) addition of fractions. 2. Analyze (C4) word/pictorial problems on addition of fractions.	1	The result of $\frac{3}{15} + \frac{4}{15}$ is ...
	2	Mrs. Lia has $\frac{2}{9}$ kg of chicken meat. She then buys an additional $\frac{2}{3}$ kg of chicken meat from the supermarket. The amount of chicken meat Mrs. Lia now has is kg

Indicator	No.	Item
	3	<p>Look at the picture below! The sum of the two fractions shown in the picture is</p> 
	4	<p>Mia bought $2\frac{1}{8}$ m of batik fabric to have made into clothing at a tailor. However, after being measured, the tailor told Mia that she needed more fabric because it was insufficient. She then bought an additional $\frac{1}{2}$ m of the same batik fabric. What is the total length of batik fabric Mia has purchased?</p>
	5	<p>Mr. Riko has a rice supply of $5\frac{7}{12}$ kg in his shop. One afternoon, a regular customer bought $2\frac{1}{4}$ kg of rice from Mr. Riko. A few minutes later, Danu bought $1\frac{1}{12}$ kg of rice from Mr. Riko. The total amount of rice sold that day is</p>
<p>1. Operate (C3) subtraction of fractions from word problems. 2. Analyze (C4) pictorial problems on subtraction of fractions.</p>	6	<p>Look at the picture below! Dimas bought 2 boxes of pizza to give to Yudi, who was celebrating his birthday. Randi also bought 1 box of the same pizza for Yudi. Yudi received all three boxes and ate them together with friends. If each pizza is cut into 8 slices and, of all the slices, only 6 were eaten, how many slices of pizza does Yudi have left?</p> 
	7	<p>Look at the picture below! The result of subtracting the two fractions shown in the picture above is ...</p> 
	8	<p>Bisma made 4 jars of nastar cookies for Eid. He then gave $1\frac{1}{2}$ jars of the cookies to his neighbor. The amount of nastar cookies Bisma now has left is jars</p>
	9	<p>Dara wants to decorate a gift for her friend using ribbon. The ribbon she has is $\frac{1}{2}$ meter long. She has used $\frac{2}{5}$ meter of the ribbon. The remaining length of Dara's ribbon is ...</p>
	10	<p>The result of $1\frac{3}{4} - \frac{1}{4}$ is ...</p>
<p>1. Operate (C3) multiplication of fractions with whole numbers. 2. Analyze (C4) word problems on multiplication of fractions with whole</p>	11	<p>The result of the multiplication $\frac{1}{3} \times 2$ is</p>
	12	<p>The result of $\frac{5}{8} \times \frac{4}{5}$ is</p>
	13	<p>Dinda and Ami went to the market to buy oranges. At one stall, Dinda bought 3 kg of</p>

Indicator	No.	Item
numbers.		oranges. At another stall, Ami bought $\frac{1}{2}$ times the amount of oranges Dinda bought. How many kg of oranges did Ami buy?
	14	Toni's daily allowance is Rp 5,000. If Toni saves $\frac{1}{10}$ of his allowance every day, how much money does Toni save?
1. Operate (C3) division of fractions with whole numbers. 2. Analyze (C4) word problems on division of fractions with whole numbers.	15	The result of $\frac{1}{2} \div 2$ is ...
	16	Mother bought $5 \frac{1}{2}$ liters of cooking oil. The cooking oil will be divided into 10 plastic bags. How many liters of cooking oil are in each plastic bag?
1. Compare (C2) fractions with the same denominator. 2. Compare (C2) fractions with different denominators. 3. Compare (C2) simple fractions with mixed fractions.	17	Compare the following fractions using the correct symbol! $\frac{2}{7} \dots \frac{4}{7}$
	18	Compare the following fractions using the correct symbol! $\frac{4}{6} \dots \frac{2}{3}$
	19	Compare the following fractions using the correct symbol! $1 \frac{3}{5} \dots 1 \frac{4}{5}$
	20	Compare the following fractions using the correct symbol! $2 \frac{3}{6} \dots 2 \frac{1}{2}$

After the items were finalized, the fifth stage was field testing, conducted by the researcher on 30 fifth-grade students of SDN 200 Leuwipanjang. Student response data were then used to perform item analysis to assess the validity and reliability of the developed instrument.

Data Analysis Technique

Data from the instrument that had been tried out were analyzed by the researcher using the Classical Test Theory (CTT) approach, with the analysis steps comprising:

1. Testing the difficulty level (p), determined based on the proportion of students who answered correctly.

Difficulty level describes the extent to which a test item is easy or difficult for test-takers to answer. The difficulty index value ranges from 0–1. A good test item is one that is neither too easy nor too difficult [18]. If a test item is too easy, students will not be motivated to make a greater effort to solve it, whereas if a test item is too difficult, students will become discouraged and give up trying to solve it because it is beyond their ability. The difficulty level (P) is determined using the formula:

$$P = B / JS$$

Where: P = item difficulty index; B = number of test-takers who answered the item correctly; JS = total number of test-takers.

The larger the value of P, the easier the item; conversely, the smaller the value of P, the more difficult the item. According to Robert L. Thorndike and Elizabeth Hagen [18], the difficulty index criteria are interpreted as follows.

Table 3. Item Difficulty Level Categories

Value of p	Category
$p > 0.70$	Easy
$0.30 \leq p \leq 0.70$	Moderate
$p < 0.30$	Difficult

2. Testing the discrimination index (D) through the correlation between item scores and total scores using point-biserial correlation.

Discrimination index indicates an item's ability to differentiate between high-ability and low-ability students. The formula for testing discrimination index is as follows:

$$D = \frac{B_A - B_B}{N/2}$$

Where: BA = number of students in the upper group who answered correctly; BB = number of students in the lower group who answered correctly; N = total number of students in both groups (upper and lower).

In Kelley's theory [19], 27% of the total number of respondents is taken to represent the upper and lower groups. In this study, the number of student respondents was N = 30, so each of the upper and lower groups is represented by approximately 8 participants. The discrimination index categories are as follows [20].

Table 4. Discrimination Index Categories

Category	Value Range (D)
Very Good	≥ 0.40
Good	0.30–0.39
Fair	0.20–0.29
Poor	< 0.20

3. Testing distractor function, evaluated based on the proportion of participants who selected each answer option.

A distractor refers to all answer alternatives (options) other than the correct answer (key). The purpose of using distractors is to mislead students who are less able or do not know the answer, so that they can be distinguished from students who are able to complete the test. Therefore, a good distractor is one that is avoided by high-achieving students and selected by low-achieving students, and is considered to be functioning if selected by at least 5% of participants [18].

4. Testing item validity, calculated using the point-biserial correlation formula (rpbi).

A test is said to have high validity if the instrument can perform its measurement function accurately, producing measurement results that correspond to the intended purpose of the measurement [21]. This means that the measurement results represent a quantity that reflects the actual facts or conditions of what is being measured. The validity assessed in this study consisted of content validity and item validity. Content validity was carried out by the fifth-grade teacher of SDN 200 Leuwipanjang based on the blueprint and items prepared by the researcher. Item validity was determined using point-biserial correlation. The formula for item validity is as follows.

$$r_{pbis} = \frac{\bar{X}_i - \bar{X}}{S_x} \sqrt{\frac{p}{1-p}}$$

Where:

$\bar{X}_i$ = mean test score of subjects who answered correctly

$\bar{X}$ = mean test score of all subjects

S_x = standard deviation of scores of all subjects

p = proportion of subjects who answered correctly

5. Testing reliability, calculated using KR-20.

Reliability refers to the consistency demonstrated by a measurement instrument, even when the instrument is tested under different conditions. A value of $r_{11} \geq 0.70$ is considered to indicate high reliability [10]. The KR-20 reliability formula is as follows.

$$r_{11} = \left[\frac{n}{(n-1)} \right] \left[1 - \frac{\sum \sigma_i^2}{\sigma_t^2} \right]$$

Where:

r_{11} = instrument reliability

$\sum \sigma_i^2$ = sum of the item variances

n = total number of items

σ_t^2 = total variance

RESULTS AND DISCUSSION

Item Analysis Results

The developed formative test instrument consists of 20 multiple-choice items that were administered to 30 fifth-grade elementary school students. Item analysis was conducted to assess difficulty level, discrimination index, distractor function, validity, and reliability.

a. Difficulty Level

Based on the analysis using MS Excel, the difficulty levels of the items in the fraction formative assessment instrument are as follows.

Table 5. Item Difficulty Analysis Results

Item	JS	B	P	Category
Q1	30	25	0.833	Easy
Q2	30	24	0.800	Easy
Q3	30	23	0.767	Easy
Q4	30	23	0.767	Easy
Q5	30	21	0.700	Moderate
Q6	30	21	0.700	Moderate
Q7	30	25	0.833	Easy
Q8	30	20	0.667	Moderate
Q9	30	22	0.733	Easy
Q10	30	18	0.600	Moderate
Q11	30	18	0.600	Moderate
Q12	30	17	0.567	Moderate
Q13	30	18	0.600	Moderate
Q14	30	19	0.633	Moderate
Q15	30	18	0.600	Moderate
Q16	30	14	0.467	Moderate
Q17	30	13	0.433	Moderate
Q18	30	16	0.533	Moderate
Q19	30	17	0.567	Moderate
Q20	30	13	0.433	Moderate

From the MS Excel output above, the percentage distribution of item difficulty levels is summarized as follows.

Table 6. Summary of Item Difficulty Level Analysis

Category	P	No. of Items	Item No.	Percentage
Easy	> 0.70	6	1, 2, 3, 4, 7, 9	30%
Moderate	0.30–0.70	18	5, 6, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20	70%
Difficult	< 0.30	0	-	0%

The majority of items (70%) fall into the "moderate" category, indicating that the instrument has a good distribution of difficulty levels. This distribution is consistent with the principle that a good test should have a greater proportion of moderate items in order to measure student ability across a wide range [8]. However, no items fell into the "difficult" category. This can be understood because the operational verbs used by the researcher in constructing the blueprint ranged from C2 to C4 (low to early high-order thinking skills).

b. Discrimination Index

Testing item discrimination is necessary to observe differences in ability between high-ability and low-ability students. The results of the discrimination index analysis using MS Excel for the developed items are as follows.

Table 7. Item Discrimination Index Analysis Results

Item	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
D	0.313	0.4375	0.5	0.5	0.313	0.5	0.25	0.5	0.313	0.5
Item	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20
D	0.375	0.5	0.375	0.125	0.1	0.5	0.5	0.313	0.3125	0.438

Based on the discrimination index analysis above, the following conclusions can be drawn.

Table 8. Summary of Item Discrimination Index Analysis

Category	Value Range (D)	No. of Items	Item No.	Percentage
Very Good	≥ 0.40	10	2, 3, 4, 6, 8, 10, 12, 16, 17, 20	50%
Good	0.30–0.39	7	1, 5, 9, 11, 13, 18, 19	35%
Fair	0.20–0.29	1	7	5%
Poor	< 0.20	2	14, 15	10%

A total of 85% of items had a discrimination index in the "good" to "very good" category, indicating that most items are able to effectively differentiate between high-ability and low-ability students. Two items (numbers 14 and 15) had low discrimination indices and are recommended for revision or removal.

c. Distractor Function

The results of the distractor-function analysis using MS Excel are presented in the table below.

Table 9. Distractor Function Analysis Results

Item	Functioning Distractors	Non-functioning Distractors	Conclusion
Q1	3	1	Very Good
Q2	3	3	Very Good
Q3	4	0	Very Good
Q4	0	4	Poor
Q5	2	2	Good
Q6	2	2	Good
Q7	1	3	Fair
Q8	2	2	Good
Q9	2	2	Good
Q10	3	1	Very Good
Q11	1	3	Fair
Q12	1	3	Fair
Q13	1	3	Fair
Q14	2	2	Good
Q15	2	2	Good
Q16	4	0	Very Good
Q17	1	3	Fair
Q18	1	3	Fair
Q19	2	2	Good
Q20	2	2	Good

Total: 39 functioning / 43 non-functioning

Based on the analysis above, 5 items had "very good" distractor function, 8 items had "good" distractor function, 6 items had "fair" distractor function, and 1 item had "poor" distractor function.

d. Item Validity

Empirical validity was calculated using point-biserial correlation (rpbi), considering that the developed items were dichotomous. The results of the validity analysis using MS Excel are presented in the table below.

Table 10. Item Validity Test Results

Item	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
rpbi	0.5967	0.6888	0.7387	0.7387	0.5835	0.8075	0.5098	0.7487	0.5457	0.8822
r-table	0.361	0.361	0.361	0.361	0.361	0.361	0.361	0.361	0.361	0.361
Category	VALID	VALID	VALID	VALID	VALID	VALID	VALID	VALID	VALID	VALID
Item	Q11	Q12	Q13	Q14	Q15	Q16	Q17	Q18	Q19	Q20

Item	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
rpbi	0.5514	0.9139	0.6286	0.4167	0.3639	0.9132	0.8849	0.6136	0.5323	0.6450
r-table	0.361	0.361	0.361	0.361	0.361	0.361	0.361	0.361	0.361	0.361
Category	VALID	VALID	VALID	VALID	VALID	VALID	VALID	VALID	VALID	VALID

Table 11. Summary of Item Validity Test Results

Category	Criterion	No. of Items	Item No.	Percentage
Valid	rpbi > r-table (0.361)	20	1–20	100%

From the table above, it can be concluded that all items have a significant correlation with the total score, as evidenced by $rpbi > r\text{-table}$ (0.361). This indicates that all items are "valid" and measure the same construct as the overall test. This result reinforces the empirical validity of the instrument.

e. Instrument Reliability

Instrument reliability was measured using KR-20, with the analysis results in MS Excel as follows.

Table 12. Instrument Reliability Test Results

Test Type	R11	Criterion	Interpretation
KR-20	0.934	≥ 0.70	Very reliable

A value of $r_{11} = 0.934$ indicates that the instrument has very high internal consistency [21]. This means that the items within the instrument are homogeneous and consistently measure the same construct, so it can be concluded that the developed instrument meets the "very reliable" criterion, with $r_{11} > 0.7$.

Discussion

Based on the analysis results, it can be concluded that the developed formative test instrument on fraction material possesses good item characteristics and meets the criteria for a high-quality assessment instrument. The dominant proportion of items in the "moderate" difficulty category indicates an ideal balance in test construction. This composition is important because items with moderate difficulty generally have optimal discrimination power, meaning they can more accurately differentiate between high-ability and low-ability students. This is consistent with Fatimah and Alfath [18], who argue that an instrument with a balanced distribution of difficulty levels and high reliability will produce assessment results that are more valid and representative of students' abilities. Thus, the proportional difficulty structure of this instrument supports the diagnostic function of formative testing in comprehensively mapping student ability.

Furthermore, the reliability analysis results show that the r_{11} value of 0.934 indicates a very high level of internal consistency. This value far exceeds the minimum reliability threshold of 0.700 suggested by Creswell and Creswell [22] and is further reinforced by Pulungan [23], who states that reliability above 0.90 indicates the reliability of an instrument in consistently measuring the same construct across different measurement occasions. This means that if this instrument were used again under different conditions, the results would remain stable and trustworthy. Thus, this reliability aspect strengthens the conclusion that the instrument has met the quality criteria for measurement in the context of educational assessment.

Nevertheless, the analysis also identified several items requiring revision, namely items 7, 14, and 15, which had low discrimination indices, and item 4, which contained a non-functioning distractor. This condition indicates that the distractor options have not been able to function optimally in stimulating different responses between high-ability and low-ability students. Psychometrically, this indicates that these items have not been effective in sharply capturing variation in student ability. Therefore, improvements can be made through revising the item wording, adjusting the context, or replacing answer alternatives with more logical ones to increase discrimination power. This step is important to ensure that every item genuinely contributes to the overall validity of the test.

The item validity test results using point-biserial correlation also show that of the 20 items analyzed, all were declared valid, with $rpbi$ values greater than $r\text{-table}$ (0.361). This result indicates that each item has a significant relationship with the total score, so it can be concluded that these items measure the same construct — namely, students' ability to understand fraction concepts. According to Arif [2], Khumaedi [24], and Ndiung and Jediut [25], instrument validity reflects the extent to which a measurement tool is able to accurately and consistently measure what it is intended to measure. Thus, these results indicate that the developed formative assessment instrument has a high level of validity, as its items are able to accurately represent the competency indicators in the blueprint.

Overall, the results of the difficulty level, discrimination index, distractor function, item validity, and reliability tests demonstrate that this instrument is suitable for use as a formative assessment tool for fraction material in elementary schools. This instrument is not only able to provide a clear picture of students' mastery of fraction concepts, but also functions as a diagnostic tool that can help teachers identify learning difficulties and design more targeted and adaptive follow-up learning strategies for students' needs. Thus, this instrument holds strategic value

in supporting the creation of a learning process oriented toward improving the quality of students' conceptual understanding, in line with the principle of authentic assessment in the Merdeka Curriculum, which emphasizes formative feedback and continuous improvement of the learning process.

CONCLUSION

This study has produced a formative test instrument for fraction material for fifth-grade elementary school students through the stages of the Research and Development (R&D) model. Based on the tryout and item analysis results, it can be concluded that the fraction formative assessment instrument can be used as a valid and reliable formative assessment tool to measure students' understanding of fraction concepts. This is evidenced by all items being declared empirically valid, and a reliability value of $r_{11} = 0.934$ (> 0.7), indicating very high internal consistency of the instrument. The instrument has varying difficulty levels, with the majority of items falling into the "moderate" category. In addition, most items have discrimination indices in the "good" to "very good" category, indicating the items' ability to differentiate between high-ability and low-ability students. However, one item distractor has not yet been able to function optimally, requiring improvement through revising the item wording, adjusting the context, or replacing answer alternatives with more logical ones. Based on the results of this study, teachers can use the developed fraction formative assessment instrument to conduct formative assessments and continuously monitor students' learning progress. Future researchers are advised to conduct construct validity testing using Confirmatory Factor Analysis (CFA) to strengthen the empirical evidence of instrument validity, or to expand the tryout sample so that the results are more generalizable while accounting for differences in school context.

ACKNOWLEDGMENTS

The author would like to thank the fifth-grade teacher and students of SDN 200 Leuwipanjang, Bandung City, for their participation and support during the tryout of the instrument, as well as the expert validators for their valuable input during the instrument development process.

REFERENCES

- [1] D. Firmanzah and E. Sudibyo, "Implementasi Asesmen Diagnostik dalam Pembelajaran IPA pada Masa Pandemi Covid-19 di SMP/MTs Wilayah Menganti Gresik," *PENSA E-Jurnal: Pendidikan Sains*, vol. 9, no. 2, pp. 165–170, 2021.
- [2] M. Arif, "Pengembangan Instrumen Penilaian Mapel Sains melalui Pendekatan Keterampilan Proses Sains SD/MI," *Ta'allum: Jurnal Pendidikan Islam*, vol. 4, no. 1, pp. 123–148, 2016, doi: 10.21274/taalum.2016.4.1.123-148.
- [3] A. L. F. Tilaar, *Asesmen pembelajaran matematika*. Unima Press, 2018.
- [4] M. A. A. Ismail, A. Ahmad, J. A. M. Mohammad, N. M. R. M. Fakri, M. Z. M. Nor, and M. N. M. Pa, "Using Kahoot! as a formative assessment tool in medical education: A phenomenological study," *BMC Medical Education*, vol. 19, no. 1, pp. 1–8, 2019, doi: 10.1186/s12909-019-1658-z.
- [5] D. A. Rahmah and A. P. Abadi, "Kesulitan Belajar Siswa Pada Proses Pembelajaran Matematika," in *Prosiding Seminar Nasional Matematika Dan Pendidikan Matematika Sesiomadika 2019*, pp. 945–949, 2019.
- [6] A. Wijaya, H. Retnawati, W. Setyaningrum, K. Aoyama, and S. Sugiman, "Diagnosing Students' Learning Difficulties in the Eyes of Indonesian Mathematics Teachers," *Journal on Mathematics Education*, vol. 10, no. 3, pp. 357–364, 2019.
- [7] S. N. Ariyani and Y. Yuhana, "Analisis Miskonsepsi Peserta didik Sekolah Dasar tentang Materi Pecahan," *Jurnal Fibonacci: Jurnal Pendidikan Matematika*, vol. 4, no. 2, p. 40, 2023, doi: 10.24114/jfi.v4i2.54085.
- [8] S. Arikunto, *Dasar-Dasar Evaluasi Pendidikan (Revisi)*. PT Bumi Aksara, 2008.
- [9] D. Mardapi, *Pengukuran, Penilaian, dan Evaluasi Pendidikan*. Parama Publishing, 2017.
- [10] M. Tavakol and R. Dennick, "Making sense of Cronbach's alpha," *International Journal of Medical Education*, vol. 2, pp. 53–55, 2011, doi: 10.5116/ijme.4dfb.8dfd.
- [11] J. W. Creswell, *Research design: Qualitative, quantitative, and mixed methods approaches*, 1999.
- [12] J. Creswell, *Research Design*, 4th ed. Pustaka Pelajar, 2016.
- [13] A. A. Nugraha and Y. Supriyati, "Developing Attitudes on Global Awareness Instrument of Senior High School Students," *JISAE: Journal of Indonesian Student Assessment and Evaluation*, vol. 4, no. 2, pp. 2442–4919, 2018, doi: 10.21009/jisae.v4i2.7345.
- [14] V. Pandra and Bellinda, "Systematic Literature Review: Model Pengembangan Instrumen Tes," *Jurnal Cendekia: Jurnal Pendidikan Matematika*, vol. 8, no. 2, pp. 1711–1724, 2024.
- [15] I. Sarifah and E. M. Kurnianti, "Development of Performance Appraisals As an Alternative To Measuring Critical Thinking Skills in Elementary Mathematics Learning," *International Journal of Business, Law, and Education*, vol. 4, no. 2, pp. 717–729, 2023, doi: 10.56442/ijble.v4i2.239.
- [16] A. I. Sofiyat, A. Tjalla, and I. Sarifah, "Pengembangan Instrumen Tes Kemampuan Penalaran Statistika Mahasiswa Pai Universitas Islam As-Syafi'iyah," *Matematika Sains*, vol. 1, no. 1, pp. 12–18, 2023.
- [17] M. D. Gall, J. P. Gall, and W. R. Borg, *Educational Research: An Introduction*, 8th ed. Pearson, 2015.
- [18] U. Fatimah and K. Alfath, "Analisis Kesukaran Soal, Daya Pembeda dan Fungsi Distraktor," *Jurnal Komunikasi Dan Pendidikan Islam*, vol. 8, no. 2, pp. 1–9, 2019.
- [19] T. L. Kelley, "The selection of upper and lower groups for the validation of test items," *Journal of Educational Psychology*, vol. 30, no. 1, pp. 17–24, 1939, doi: 10.1037/h0057123.

- [20] M. Duskri, K. Kumaidi, and S. Suryanto, "Pengembangan Tes Diagnostik Kesulitan Belajar Matematika di SD," *Jurnal Penelitian Dan Evaluasi Pendidikan*, vol. 18, no. 1, pp. 44–56, 2014, doi: 10.21831/pep.v18i1.2123.
- [21] Z. Matondang, "Validitas dan reliabilitas suatu instrumen penelitian," *Jurnal Tabularasa PPS UNIMED*, vol. 6, no. 1, pp. 87–97, 2009.
- [22] J. W. Creswell and J. D. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, 5th ed. SAGE Publications Inc., 2018.
- [23] D. A. Pulungan, "Pengembangan Instrumen Tes Literasi Matematika Model Pisa," *Journal of Educational Research and Evaluation*, vol. 3, no. 2, 2014, doi: 10.15294/jere.
- [24] M. Khumaedi, "Reliabilitas Instrumen Penelitian Pendidikan," *Jurnal Pendidikan Teknik Mesin Unnes*, vol. 12, no. 1, 2012.
- [25] S. Ndiung and M. Jediut, "Pengembangan instrumen tes hasil belajar matematika peserta didik sekolah dasar berorientasi pada berpikir tingkat tinggi," *Premiere Educandum: Jurnal Pendidikan Dasar Dan Pembelajaran*, vol. 10, no. 1, pp. 94–111, 2020, doi: 10.25273/pe.v10i1.6274.