

Validation of Achievement Items for Fifth-Grade Elementary School Students on Multiplication and Division of Fractions

Meylisa¹, Zetra Hainul Putra¹, Guslinda¹

¹ Primary School Teacher Education, Faculty of Teacher Training and Education, Universitas Riau, Indonesia

Article Info

Article history:

Received: February 15th, 2026

Revised: March 17th, 2026

Accepted: April 27th, 2026

Keywords:

Achievement Test
Elementary School Students
Fraction Division
Fraction Multiplication
Item Validation

ABSTRACT

Achievement tests are essential for accurately assessing elementary school students' understanding of mathematical concepts, particularly in fraction multiplication and division. However, assessment instruments specifically designed to measure fifth-grade students' achievement in these topics need to be carefully examined for their psychometric quality. This study aimed to develop and conduct a preliminary psychometric evaluation of an achievement test instrument for fifth-grade elementary school students on multiplication and division of fractions. Using a quantitative approach with an instrument development and psychometric testing design, the study involved 30 fifth-grade students in the preliminary instrument testing. The instrument consisted of 30 multiple-choice items covering multiplication and division of fractions. Data were analyzed using item validity, Cronbach's Alpha, item difficulty, and item discrimination analysis. The results showed that the instrument demonstrated high internal consistency, with a Cronbach's Alpha coefficient of 0.920. The item analysis also provided information regarding the validity, difficulty level, and discriminating power of each item. These findings indicate that the developed instrument has promising initial psychometric quality for assessing fifth-grade students' achievement in multiplication and division of fractions, although items with weaker characteristics require further revision before broader application.

This is an open access article under the [CC BY-NC-SA](#) license.



Corresponding Author:

Meylisa
Universitas Riau, Pekanbaru, Indonesia
Email: meylisa03072@student.unri.ac.id

INTRODUCTION

Fractions are an essential component of elementary mathematics because students' understanding of fractions provides an important foundation for subsequent mathematical learning. Fraction knowledge is not limited to performing arithmetic procedures but also involves understanding the relationships between quantities, representations, and numerical magnitudes. Previous research has shown that fraction understanding is closely associated with students' broader mathematical achievement. In a longitudinal study involving students from fourth to sixth grade, fraction magnitude knowledge and mathematics achievement were found to develop in relation to each other, indicating the important role of fraction understanding in students' mathematical development (Hansen et al., 2017). Similarly, research involving elementary school

students has demonstrated that relational reasoning is associated with fraction knowledge, suggesting that understanding the relational structure of fractions is important for developing students' mathematical competence (Kalra et al., 2020).

Fraction multiplication and division are particularly important because these operations require students to coordinate several mathematical concepts and procedures. Unlike whole-number operations, fraction operations require students to understand the relationships between numerators and denominators and to apply appropriate procedures in different situations. Difficulties in fraction learning may persist across elementary school years when students do not develop adequate conceptual and procedural knowledge. A longitudinal study of students from third through sixth grade found substantial differences in the development of fraction knowledge and identified persistent difficulties in several fraction concepts and procedures (Hansen et al., 2017). More recent research has also shown that fraction understanding consists of multiple interconnected aspects and that early differences in fraction knowledge are related to later mathematics achievement.

Because fraction knowledge is an important component of mathematics learning, appropriate assessment is needed to determine the extent to which students have achieved the expected learning outcomes. Assessment instruments provide information about students' knowledge and performance and can support teachers in identifying areas of learning that require further attention. A well-designed mathematics assessment should not only produce scores but should also provide valid and reliable evidence regarding the mathematical knowledge being measured. Research on the development of mathematics assessments has demonstrated the importance of establishing psychometric evidence when constructing instruments for measuring students' mathematical abilities. For example, Dündar (2016) developed a mathematical ability test and examined its validity and reliability as part of a systematic test development process.

The quality of an achievement test depends not only on the content selected but also on the empirical characteristics of its items. Therefore, item analysis is an important stage in the development and evaluation of an assessment instrument. Item analysis provides information about the validity and reliability of test items, as well as their difficulty and discrimination characteristics. Difficulty analysis indicates the proportion of students who answer an item correctly, while discrimination analysis indicates the ability of an item to distinguish between students with higher and lower levels of achievement. In addition, item analysis can help identify items that function appropriately and those that require revision or removal (Kumar et al., 2021).

In mathematics education, the development of assessment instruments for elementary school students requires particular attention because mathematical knowledge develops progressively and assessment items need to correspond with students' conceptual and procedural development. The development of the Research-Based Early Maths Assessment, for example, demonstrated the importance of developing a theoretically grounded and empirically tested mathematics assessment instrument for young learners. The study used psychometric analysis to refine the instrument and provided evidence of reliability and developmental appropriateness (Clements et al., 2008).

For fraction learning specifically, assessment needs to capture students' ability to perform operations as well as their understanding of fraction concepts. Studies have shown that students can experience difficulties with fractions even when they have received formal mathematical instruction. Research involving fifth-grade students has demonstrated that interventions designed to strengthen fraction conceptual understanding can improve students' performance, highlighting the importance of accurately identifying students' level of fraction knowledge through appropriate assessment (Björkhammer et al., 2024). Furthermore, research on fraction learning has shown that

understanding fraction magnitude and other conceptual aspects is closely related to later mathematics achievement, emphasizing the need for assessment instruments that can provide meaningful information about students' mathematical performance (Hansen et al., 2017).

Although research on fraction learning and assessment has continued to develop internationally, there remains a need for classroom-based achievement instruments that are specifically designed and empirically examined for fifth-grade students learning multiplication and division of fractions. In particular, an instrument used in an experimental study should demonstrate adequate item characteristics before it is used to measure the effects of an instructional intervention. Without adequate evidence of item validity, reliability, difficulty, and discrimination, differences in students' test scores may be difficult to interpret as evidence of their actual learning achievement. Therefore, systematic analysis of the instrument is necessary before it is used as a measure of learning outcomes.

The present study addresses this need by developing and evaluating a mathematics achievement test for fifth-grade elementary school students on multiplication and division of fractions. The instrument was designed based on the learning indicators and mathematical content covered in the fifth-grade curriculum and consisted of multiple-choice items representing the targeted fraction operations. The study focuses on examining the psychometric characteristics of the items, including item validity, internal consistency reliability, item difficulty, and item discrimination. These analyses are intended to identify items that function adequately and items that require revision before the instrument is used in further research.

The development of this instrument is particularly relevant because the test is intended to provide empirical evidence of students' achievement in multiplication and division of fractions. A valid and reliable instrument can provide more accurate information about students' learning outcomes and strengthen the interpretation of experimental findings in mathematics education. The results can also provide a basis for refining fraction achievement items and developing more appropriate assessment instruments for elementary school mathematics.

Accordingly, this study aims to develop and conduct a preliminary psychometric evaluation of an achievement test instrument for fifth-grade elementary school students on multiplication and division of fractions. Specifically, the study examines the validity, reliability, difficulty level, and discriminating power of the test items. The findings are expected to contribute to the availability of a psychometrically examined instrument that can be used to assess fifth-grade students' achievement in multiplication and division of fractions and to support future research and classroom assessment in elementary mathematics education.

METHOD

This study employed a quantitative approach using a preliminary instrument development and psychometric testing design. The study was intended to develop an initial set of mathematics achievement test items and examine their preliminary psychometric quality for assessing sixth-grade students' achievement in multiplication and division of fractions. The instrument was designed to measure students' understanding and application of fraction operations through conceptual, procedural, and problem-solving items. The empirical evaluation focused on item validity, internal consistency reliability, item difficulty, and item discrimination, which are important components in the evaluation of multiple-choice assessment instruments (Dündar et al., 2016; DeCarlo, 2023).

The study was conducted at SD Negeri 184 Pekanbaru, Indonesia, during the 2025/2026 academic year. The participants were 30 sixth-grade students who had studied multiplication and division of fractions relevant to the content domain of the instrument. The participants were

involved in the preliminary field testing of the developed items. The study used the students available in the selected class as participants because the purpose of this stage was to examine the initial empirical functioning of the test items rather than to provide final large-scale validation or population-level generalization. The students' responses were recorded for analysis, and the data were used to evaluate the characteristics of the developed instrument.

The instrument was developed based on the learning indicators of multiplication and division of fractions. The content domain covered multiplication of fractions by whole numbers, multiplication of whole numbers by fractions, multiplication of fractions by fractions, division of fractions by whole numbers, division of whole numbers by fractions, and division of fractions by fractions. The items were also designed to represent three types of assessment tasks: conceptual, procedural, and problem-solving items. Conceptual items assessed students' understanding of basic fraction concepts and representations, including identifying fractions, interpreting the numerator and denominator, and determining fractions represented by visual models. Procedural items assessed students' ability to perform multiplication and division involving fractions and whole numbers. Problem-solving items presented fraction operations in contextual situations that required students to identify and apply the appropriate mathematical operation. The inclusion of conceptual and representational aspects is consistent with research on fraction assessment, which emphasizes the importance of examining different aspects of students' fraction knowledge, including graphical representations, numerical symbols, and contextual information (Schadl & Lindmeier, 2025). Research has also shown that fraction magnitude knowledge is associated with subsequent mathematics achievement, supporting the importance of assessing students' fraction understanding through structured mathematical tasks (Bal-Incebacak & Ersoy, 2017).

The initial instrument consisted of 30 multiple-choice items with four response alternatives (A–D). The items were developed across cognitive levels C2 to C4. The first eight items focused on conceptual understanding of fractions and fraction representations. The procedural items addressed multiplication and division involving whole numbers and fractions, including direct computation and visual representation. The remaining items consisted of contextual problem-solving tasks involving multiplication and division of fractions. Several items incorporated visual representations to assess students' ability to connect symbolic fraction operations with their graphical representations. For example, one item asked students to identify the correct fraction representation for $\frac{1}{4} \times \frac{2}{3}$, while another required students to determine the correct representation of $2 \div \frac{1}{2}$. These items were included to assess students' understanding of fraction operations beyond direct symbolic calculation.

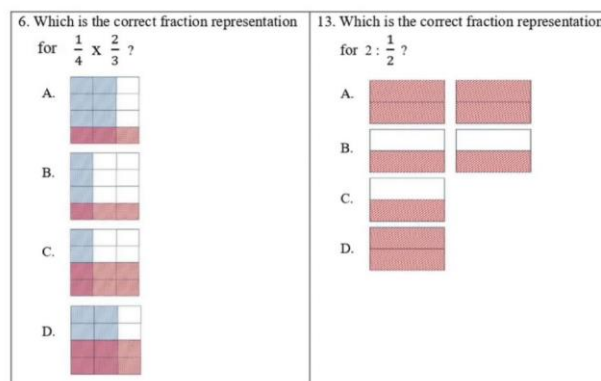


Figure 1. Examples of Visual Representation Items for Fraction Multiplication and Division

The instrument also included procedural items requiring students to calculate the results of fraction operations directly, as well as problem-solving items presented in everyday contexts. The combination of conceptual, procedural, visual-representation, and problem-solving items was intended to provide broader coverage of students' achievement in multiplication and division of fractions.

The problem-solving items presented everyday situations involving multiplication and division of fractions. For example, students were asked to determine the amount of flour required when several cakes used $\frac{3}{4}$ kg of flour each, or to determine the number of containers that could be filled using a given amount of water when each container had a fractional capacity. These contextual tasks were designed to assess students' ability to identify the mathematical operation represented by a situation and apply the appropriate fraction procedure. The instrument therefore included both direct mathematical tasks and contextual problems to represent different forms of students' achievement in multiplication and division of fractions.

Each item was scored dichotomously, with a score of 1 for a correct answer and 0 for an incorrect answer. Thus, the initial 30-item instrument provided a maximum possible score of 30. The test was administered in the school setting in one testing session lasting approximately 1.5 hours. Before the test began, students were given instructions on how to complete the items. Students then worked individually on the test, and their answer sheets were collected after the testing session. The responses were subsequently scored using the predetermined answer key and entered into IBM SPSS Statistics version 26 for analysis.

Data were analyzed using IBM SPSS Statistics 26, with the preliminary psychometric evaluation focusing on four main aspects of the instrument, as summarized in Table 1. Item validity was examined by comparing the calculated correlation coefficient $r_{\text{calculated}}$ with the critical value r_{table} . With 30 participants, the r_{table} value at the 5% significance level was 0.361. Items with $r_{\text{calculated}} > r_{\text{table}}$ were classified as valid. The validity analysis resulted in 24 valid items and six invalid items, namely Items 1, 2, 4, 5, 6, and 10.

Internal consistency reliability was examined using Cronbach's Alpha. The reliability analysis was conducted on the 24 valid items obtained after the item validity analysis. The analysis included 30 valid cases and produced a Cronbach's Alpha coefficient of 0.920, which was interpreted as very good reliability according to the reliability criterion used in the study.

Item difficulty was analyzed based on the proportion of students who answered each item correctly. The analysis of the retained items resulted in 10 items classified as moderate and 14 items classified as difficult according to the classification used in the study. The reported item difficulty values ranged from 0.53 to 0.67. Item discrimination was examined to determine the extent to which each item differentiated students according to their overall achievement. Among the 24 retained items, 22 items were classified as good and two items as fair. The two items classified as fair were Item 7, with a discrimination value of 0.396, and Item 20, with a value of 0.369.

The preliminary analysis resulted in the retention of 24 items from the initial 30-item instrument. The six items that did not meet the validity criterion were Items 1, 2, 4, 5, 6, and 10. The 24 retained items demonstrated high internal consistency, with a Cronbach's Alpha coefficient of 0.920. The subsequent item analysis showed that the retained items consisted of 10 items classified as moderate and 14 as difficult in terms of difficulty, while 22 items demonstrated good discrimination and two demonstrated fair discrimination.

Table 1. The Instrument's Preliminary Psychometric Quality

Analysis Aspect	Statistical Technique	Decision Rule / Classification	Purpose
Item validity	Item-total correlation	$r_{\text{calculated}} > r_{\text{table}}, r_{\text{table}} = 0.361$	To examine whether each item was related to the overall test score
Internal consistency reliability	Cronbach's Alpha	$\alpha = 0.920$	To assess the internal consistency of the 24 valid items
Item difficulty	Proportion of correct responses	Moderate and difficult	To determine the difficulty level of each item
Item discrimination	Item discrimination coefficient	Fair and good	To determine how well each item distinguished students with different levels of achievement

RESULT AND DISCUSSION

The study produced four main findings regarding the quality of the mathematics achievement test instrument: item validity, reliability, item difficulty, and item discrimination. The initial instrument consisted of 30 multiple-choice items developed to assess sixth-grade students' achievement in multiplication and division of fractions. After empirical testing with 30 students, 24 items met the validity criterion, while six items were classified as invalid. The reliability analysis of the 24 valid items produced a Cronbach's Alpha coefficient of 0.920. The item analysis further showed that the valid items consisted of moderate and difficult items, while most of the items demonstrated good discrimination.

Table 2. Summary of Mathematics Achievement Test Instrument Testing

Analysis	Result
Initial number of items	30
Valid items	24
Invalid items	6
Invalid item numbers	1, 2, 4, 5, 6, 10
Cronbach's Alpha	0.920
Moderate difficulty items	10
Difficult items	14
Good discrimination	22
Fair discrimination	2

The validity analysis showed that 24 of the 30 items met the validity criterion based on a comparison between $r_{\text{calculated}}$ dengan r_{table} of 0.361. Items 1, 2, 4, 5, 6, and 10 memiliki nilai $r_{\text{calculated}}$ below 0.361 and were therefore classified as invalid. The remaining 24 items had correlation coefficients ranging from 0.423 to 0.713 and met the validity criterion.

The finding indicates that most of the developed items demonstrated a sufficient empirical relationship with the overall test score. Item validity analysis is important because it provides information about whether individual multiple-choice items function appropriately within an assessment. (Saat, 2020) explains that item analysis is used to examine the quality of individual test items and to determine whether items should be retained, modified, or discarded. In the present study, the six invalid items require further examination before being considered for use in the final instrument.

The reliability analysis of the 24 valid items produced a Cronbach's Alpha coefficient of 0.920. The analysis was based on 30 student responses, with all cases included in the analysis. This

coefficient indicates a high level of internal consistency among the retained items under the reliability criterion used in the study. However, Cronbach's Alpha should be interpreted carefully because the coefficient provides evidence of internal consistency rather than by itself establishing the complete validity of an instrument (Taber, 2018).

The item difficulty analysis showed that the retained items consisted of 10 moderate items and 14 difficult items according to the classification reported in the study. The item-level difficulty values presented in the analysis ranged from 0.53 to 0.67. The distribution indicates that the test was not dominated by items classified as easy. This pattern provides variation in the level of challenge presented to students when answering questions on multiplication and division of fractions.

The difficulty characteristics are relevant because an achievement test should contain items that can provide variation in students' responses rather than producing uniformly similar response patterns. (Quaigrain & Arhin, 2017) explain that the difficulty index indicates the proportion of students who answer an item correctly and can be used to evaluate the quality of multiple-choice test items. They also note that items with moderate difficulty can provide useful discrimination among students. In the present instrument, the combination of moderate and difficult items provides information about students' performance across different levels of challenge.

The discrimination analysis showed that 22 of the 24 retained items had good discrimination, while two items had fair discrimination. The two fair items were Item 7 with a discrimination value of 0.396 and Item 20 with a value of 0.369; the remaining 22 items were classified as good. This result indicates that most retained items were able to distinguish students according to differences in their overall achievement on the test.

The predominance of items with good discrimination is relevant to the purpose of the instrument because an achievement test should be capable of differentiating students who demonstrate different levels of mastery of the assessed content. Item analysis research indicates that the discrimination index provides information about an item's ability to distinguish between students with different levels of knowledge and can be considered together with item difficulty when evaluating multiple-choice items (Amo-Salas et al., 2014). The two items with fair discrimination may therefore warrant additional examination in subsequent instrument refinement.

Overall, the findings provide preliminary empirical evidence regarding the quality of the developed mathematics achievement test for multiplication and division of fractions. Of the 30 initial items, 24 met the validity criterion, and these retained items demonstrated a Cronbach's Alpha coefficient of 0.920. The retained items also showed variation in difficulty and predominantly good discrimination. These findings suggest that the instrument has useful initial psychometric characteristics for assessing sixth-grade students' achievement in multiplication and division of fractions. Nevertheless, because the preliminary testing involved only 30 students from one school, the findings should be interpreted as initial evidence rather than final large-scale validation. Further testing with larger and more diverse samples would be useful to examine whether the item characteristics remain stable across different student groups.

CONCLUSION

This study indicates that the developed achievement test instrument provides promising initial psychometric evidence for assessing fifth-grade students' achievement in multiplication and division of fractions. Of the 30 initial items, 24 were identified as valid, with a Cronbach's Alpha coefficient of 0.920, indicating high internal consistency. The item analysis showed that 10 items were classified as moderate and 14 as difficult, while 22 items demonstrated good discrimination and 2 items demonstrated fair discrimination. These findings suggest that the instrument has

promising initial characteristics for measuring students' achievement through conceptual, procedural, and problem-solving tasks involving multiplication and division of fractions. However, the six invalid items indicate that further revision is necessary before the instrument can be applied more broadly.

This study has several limitations. The preliminary testing involved only 30 students from a single elementary school, limiting the generalizability of the findings to a broader population. In addition, the multiple-choice format with dichotomous scoring primarily captures students' selected answers and does not directly reveal their reasoning or problem-solving processes. The study also represents an initial stage of instrument development and did not include expert judgment or broader validation procedures. Future studies should therefore involve larger and more diverse samples, revise and retest the invalid items, incorporate expert review, and apply additional validation procedures to obtain stronger evidence of the instrument's psychometric quality for assessing students' achievement in multiplication and division of fractions.

REFERENCE

- Amo-Salas, M., Arroyo-Jimenez, M. del M., Bustos-Escribano, D., Fairén-Jiménez, E., & López-Fidalgo, J. (2014). New Indices for Refining Multiple Choice Questions. *Journal of Probability and Statistics*, 2014, 1–8. <https://doi.org/10.1155/2014/240263>
- Bal-Incebacak, B., & Ersoy, E. (2017). Developing An Achievement Test for Fraction Teaching: Validity and Reliability Analysis. *ITM Web of Conferences*, 13, 01003. <https://doi.org/10.1051/itmconf/20171301003>
- Björkhammar, C., Samuelsson, J., Träff, U., & Östergren, R. (2024). The effects of a whole-class mathematics intervention on students' fraction knowledge in primary school. *Scandinavian Journal of Educational Research*, 68(6), 1275–1289. <https://doi.org/10.1080/00313831.2023.2228818>
- Clements, D. H., Sarama, J. H., & Liu, X. H. (2008). Development of a measure of early mathematics achievement using the Rasch model: the Research-Based Early Maths Assessment. *Educational Psychology*, 28(4), 457–482. <https://doi.org/10.1080/01443410701777272>
- DeCarlo, L. T. (2023). Classical Item Analysis from a Signal Detection Perspective. *Journal of Educational Measurement*, 60(3), 520–547. <https://doi.org/10.1111/jedm.12358>
- Dündar, S., Temel, H., & Gündüz, N. (2016). Development of a mathematical ability test: a validity and reliability study. *International Journal of Mathematical Education in Science and Technology*, 47(7), 1061–1075. <https://doi.org/10.1080/0020739X.2016.1153734>
- Hansen, N., Rinne, L., Jordan, N. C., Ye, A., Resnick, I., & Rodrigues, J. (2017). Co-development of fraction magnitude knowledge and mathematics achievement from fourth through sixth grade. *Learning and Individual Differences*, 60, 18–32. <https://doi.org/10.1016/J.LINDIF.2017.10.005>
- Kalra, P. B., Hubbard, E. M., & Matthews, P. G. (2020). Taking the relational structure of fractions seriously: Relational reasoning predicts fraction knowledge in elementary school children. *Contemporary Educational Psychology*, 62, 101896. <https://doi.org/10.1016/J.CEDPSYCH.2020.101896>
- Kumar, D., Jaipurkar, R., Shekhar, A., Sikri, G., & Srinivas, V. (2021). Item analysis of multiple choice questions: A quality assurance test for an assessment tool. *Medical Journal Armed Forces India*, 77, S85–S89. <https://doi.org/10.1016/J.MJAFI.2020.11.007>

- Quaigrain, K., & Arhin, A. K. (2017). Using reliability and item analysis to evaluate a teacher-developed test in educational measurement and evaluation. *Cogent Education*, 4(1), 1301013. <https://doi.org/10.1080/2331186X.2017.1301013>
- Saat, N. A. (2020). Summative Test Items Analysis Using Classical Test Theory (CTT)/ Analisis Item Kertas Peperiksaan Sumatif Menggunakan Teori Ujian Klasik (TUK). *Sains Humanika*, 12(2–2). <https://doi.org/10.11113/sh.v12n2-2.1788>
- Schadl, C., & Lindmeier, A. (2025). Searching for efficient and informative instruments to digitally monitor fraction learning: putting tests for fraction subconstruct knowledge and informal fraction knowledge on the test bench. *Frontiers in Education*, 9. <https://doi.org/10.3389/feduc.2024.1367942>
- Taber, K. S. (2018). The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Research in Science Education*, 48(6), 1273–1296. <https://doi.org/10.1007/s11165-016-9602-2>