



Journal of Education, Teaching, and Learning is licensed under
A [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).

DEVELOPMENT OF THREE-TIER DIAGNOSTIC TEST TO MEASURE MISCONCEPTIONS OF GRADE 5 STUDENTS ABOUT FRACTIONS

Ariza Fitriani Safira¹⁾, Harun²⁾

¹⁾ Universitas Negeri Yogyakarta, Yogyakarta, Indonesia
E-mail: arizafitriani.2020@student.uny.ac.id

²⁾ Universitas Negeri Yogyakarta, Yogyakarta, Indonesia
E-mail: harun@uny.ac.id

Abstract. This research developed a measuring instrument in the form of a three-tier diagnostic test to measure students' misconceptions in fraction material in grade V. This research aims to develop three-tier diagnostic test that has good validity and reliability and to know the level of misconceptions of grade V elementary school students in Pontianak. This research used quantitative approach and test development model by Mardapi, that has nine steps. The subjects were fifth-grade students in Pontianak, and the developed test instruments focused on addition-subtraction of fractions, each with two sets of questions A and B. The instrument were analyzed for content and construct validity, reliability, and item characteristics based on IRT using Partial Credit Model (PCM). The results show that the developed test instrument demonstrated good content and construct validity. Reliability analysis shows high reliability, with Cronbach's Alpha values $>0,7$. Item fit analysis reveal that 91% and 73% were fit, therefore the unfit items were fixed or eliminated. The analysis of misconceptions shows that 39.52% of students do not understand the concept of fractions' addition and subtraction, and 29.09% experienced misconceptions. Additionally, 12.61% have false positives, and 8.97% have false negatives. Only 6.30% understood the concept correctly, and 3.52% guessed correctly or understood but had doubts. Generally, the final test instrument demonstrated good validity, reliability and item qualities.

Keywords: diagnostic test; three-tier test; misconceptions; fraction

I. INTRODUCTION

Basic mathematical knowledge at the elementary school level becomes the starting point for students to build mathematical abilities, logical thinking, analytical skills, and systematic approach, not only at school, but also in daily activities (Fauzan & Sari, 2017; Theis et al., 2022). A strong foundation in basic concepts must be built first so students can understand and solve subsequent mathematical problems (Zulkardi in Rismawati & Hutagaol, 2018).

In the process of students constructing their conceptual understanding, errors are prone to occur due to cognitive conflicts that arise during the process of assimilating existing knowledge with new knowledge being learned. These errors are known as misconceptions (Jefi et al., 2020; Moosapoor, 2023). Misconception is discrepancy between the concept that students have, and the correct concept as defined by experts (McAfee & Hoffman, 2021; Yuliati, 2017). Misconceptions are cognitive structures that stably prevent change or limiting knowledge's reconstructions and negatively affect students' understanding (Taslidere, 2016) Weak or inadequate understanding of mathematical concepts can lead to various misconceptions in students, and likewise, misconceptions will make it difficult for students to

understand concepts in learning (Makonye & Fakude, 2016; Theis et al., 2022). Considering the disadvantages, misconceptions need to be addressed properly.

One of the mathematical topics that experiences the highest misconceptions or considered difficult to understand by elementary students is the topic of fractions (Deringöl, 2019; Eichhorn, 2018; Ermayani et al., 2019; Makhubele, 2021; Safitri, 2018; Yanto et al., 2021). Misconceptions in fraction experienced by elementary students are important to be overcome. If not identified and resolved, these misconceptions will continue to persist and causing repeated and continuous errors (Eichhorn, 2018; Im & Jitendra, 2020; Kor et al., 2018; Morris et al., 2022; Yuliati, 2017). This is because math is fundamentally hierarchical, and systematic, progressing from simple to complex, so foundational concepts like fraction will affecting other concepts' understanding, as the concepts are interconnected (Purwaningrum & Bintoro, 2018; Rismawati & Hutagaol, 2018). Students need the understanding of calculating fractions with same or different denominators, mixed fraction numbers, or calculating decimal numbers in order to do algebraic thinking, geometric analysis, proportional reasoning or other mathematic analysis (N. A. Abbas et al., 2022; Kor et al., 2018; Malone et al., 2019; Purnomo et al.,

2022). Not only in mathematic, fraction as basic concept also needed in other subjects like chemistry, physics, biology, economy, medics, engineering and every day activities (Lortie-Forgues et al., 2015). This shows that misconceptions about fraction is important to be analysed and need more attention in learning.

Ironically, fractions not only hard to learn for elementary students, but also considered hard to teach for teacher (Agbozo, 2020; Kor et al., 2018; Ubah, 2021). One of the main causes is due to the transition of concepts from whole numbers to fractions which have different concepts and rules (Jarrah et al., 2022; Yu et al., 2024). Many students failed to differentiate and understand this transition of concepts that caused them to have misconceptions such as adding numerator and denominator separately (Saputri, 2019). Misconceptions also tend to arise when students encounter problems that that require them to calculate fractions in different forms, such as calculation between ordinary fractions with mixed fractions or whole numbers with ordinary fractions (Aksoy & Yazlik, 2017; Üzel, 2018).

Misconceptions usually addressed by remedial programs (Hidayat & Suyono, 2016). When students are struggling in math, what they need is targeted instruction and right constructive feedback. This instructions and feedback should be able to explains their concept misunderstanding, and explanations of the correct concept (Arifin et al., 2020; Eichhorn, 2018). However, based on the pre-research's result, teachers in Pontianak tend to give re-formative tests or additional assignments if they find students whose scores do not meet the KKM. Teachers said that it can be pretty confusing to design the right effective remedial activities because some misconceptions couldn't be identified in daily assignments, or there are cases when students gives the right answers but they actually haven't grasp the concepts completely. Furthermore, teacher also said that elementary students often got confused when they were asked about parts that they still don't understand. This pre-research founding is in line with Caleon & Subramaniam (2010), who found that "most students don't know what they do not know". They could have "illusion of knowing", that tend to increase when confronted unfamiliar topics. It can be difficult for teachers to give the right feedbacks when they don't have specific and accurate information of students' misconceptions while repeating the topic from beginning will take too much time.

This is why the identification process is a fundamental step for teachers to know what misconceptions were experienced by students in order to design suitable learning (Kshetree et al., 2021; Moosapoor, 2023). This is where the important role of diagnostic assessment needed. Diagnostic assessment is an assessment carried out specifically to identify the competencies, strengths and weaknesses of students, so that learning can be designed in accordance with the conditions of students (Kepmendikbud, 2020).

Several methods have been used to do diagnostic assessment in math, such as interviews, open ended questions, and concept maps (Aliustaoğlu et al., 2018; Deringöl, 2019; Liu & Jacobson, 2022; Ratnasari, 2018;

Turidho et al., 2021). Interview was the most popular method, but this method take too much time to gather data and analyse them (Kirbulut & Geban, 2014). Furthermore, obtaining good data requires skilful interviewer, therefore it needs training for unexperienced people. Beside, interviewer could be biased or students could give unserious response if they don't comfortable with the interviewer (Gurel et al., 2015).

In order to get accurate but simpler analysis, open ended test became commonly used in analysing students' understanding. This help teachers to get proper information, but results analysis is tiring and could only used for small number of students, so the generality of result could be questioned (Milenković et al., 2016). It's also found that students tend to give little response when they couldn't do the question, and don't write their reasoning completely (Gurel et al., 2015). Beside, essay or open ended test's scoring tends to be subjective, because students' interpretations of the questions can be various as well as the teachers who assess the answers, especially if the teacher lacks the ability of mathematical concepts to assess objectively (Asnawi et al., 2023). This can be tricky, while not all elementary school teachers have mathematic background of education. On the other hand, multiple choice diagnostic test provide teachers with a user-friendly and efficient execution that is both valid and reliable for assessing student understanding. These test are simple to administer and easy to grade for assessing students' comprehension. Interestingly, even some researches that initially against this multiple choice method also draw backed (Kanwal & Farooq, 2021). This highlight the multiple choice diagnostic test as impactful, and efficient assessment tool, while also user-friendly for teachers.

Diagnostic test is considered as the most preferred one to identify students' misconceptions. The implementation of diagnostic test is easier and the application of the test can be done on a larger scale (Anggrayni & Ermawati, 2019; Taslidere, 2016). By using diagnostic test, the identification of weaknesses and the follow-up that teachers provide will be more precise, that's why it's considered as one of the effective and recommended way to overcome misconceptions, (Kanwal & Farooq, 2021; Nurharyanto & Retnawati, 2020).

Diagnostic tests have several forms, such as diagnostic tests in the form of ordinary multiple choice (one-tier), two-tier diagnostic tests supported with reasoning for answers, and three-tier diagnostic tests supported with reasoning and the level of confidence of participants in their answers and reasoning (Diani et al., 2019). One-tier test or ordinary multiple choices can only get right-wrong response, so it is possible for learners to give correct answers but wrong reasons, or wrong answers but correct reasons. Two-tier diagnostic test is superior to ordinary multiple-choice questions because they are equipped with reasons for the answers of students. But this type of diagnostic test has a weakness in distinguishing students who had misconceptions, lack of concepts, or lucky quessing, because it's unknown

whether students answer confidently, or just by guess (Dewi et al., 2018; Zaleha et al., 2017).

Three tier test is a type of diagnostic test that have three levels of response. The first level is the answer choice, followed by the second level in the form of the reason for the answer of the first level and the third level is the level of confidence in the response that has been given in the first and second levels. This is based on the finding that learners' concept understanding can be evaluated by looking at their level of confidence in answering questions (Chewoh & Marmoah, 2021; Kirbulut & Geban, 2014). The Three-Tier Diagnostic Test can distinguish students who are classified as already understanding the concept, experiencing pure misconceptions, only guessing, understanding the concept but not sure, not understanding the concept, and those who experience false positive misconceptions or false negative misconceptions (M. L. H. Abbas, 2016; Gurel et al., 2015). Therefore, it can be concluded that three-tier is better than ordinary multiple choice tests and two-tier diagnostic tests (Kirbulut & Geban, 2014; Taslidere, 2016). Three-tier diagnostic test can help teachers to diagnose in-depth understanding of students' misconceptions to determine parts that need more emphasis in class (Theis et al., 2022).

However, contradictory to the needs, not many diagnostic tests have been developed in elementary school level. Previous research showed that only 6,67% elementary school teachers use diagnostic test regularly, and 100% teachers want to have a standard diagnostic test instrument (Asnawi et al., 2023). These results are in line with previous researches, that recommend to use standardized diagnostic tests that have been through proper development steps and analysis (Eichhorn, 2018; Nurharyanto & Retnawati, 2020; Takker & Subramaniam, 2019). Using qualified test that's valid and reliable instead of just using test screener that's commonly used will give proper information needed.

Therefore, to address the issues and needs in the background, it appears necessary to conduct research to develop a three-tier diagnostic test instrument to detect misconceptions in fraction material that is ready for use. This developed three-tier diagnostic test instrument is intended to be used to measure mathematical misconceptions of students at the elementary school level.

II. METHODS

This research falls under the category of Research and Development (R&D) aimed to develop a diagnostic test instrument in order to measure students' misconceptions in fraction material for 5th grade elementary school. The development stages of the diagnostic test instrument in Mardapi's (2017) development model include nine steps. The model was chosen because it's recommended by Suwanto (2017) for creating a good and systematic diagnostic test. These nine steps are:

1) Compiling test specifications. In this step the purpose of the test was determined, which is to measure the misconceptions of grade 5 about fraction calculation operations in Pontianak. In determining it, theoretical and practical research was carried out by literature review and

pre-research at elementary schools in Pontianak. After some research, it was concluded that product should be a diagnostic test to measure students' misconceptions about fraction. Then, designing outline about computing addition and subtraction of fraction in grade 5 SD was carried out. The outline was constructed according to the forms of misconceptions related to fraction counting operations that have been studied in previous studies. Lastly, in this step the form of the test was decided, which is three-tier diagnostic test that has two packages of questions A and B.

2) Writing the test. The next step was to write the draft question items. The questions in each item refer to the test outline that has been designed in step 1. Question items was written by following the rules for writing multiple-choice tests (Mardapi, 2017). As for each answer choices were designed based on the forms of misconceptions in previous studies. Each answer choice in the first tier was associated with a choice of reasons in the second tier that also based on the misconceptions found in previous researches.

3) Reviewing the test. The review of items on the developed test was on the test outline, test instrument and answer key through a validation process by experts (validators). Validators assess the feasibility of the instrument based on criteria relevant to the characteristics of the preparation of multiple choice test items (Mardapi, 2017). The review was also about mathematical concepts, item's arrangement, images and the form of misconception referenced. Feedbacks by validators became the reference to revised the items.

4) Conducting trials, the items that have been written and reviewed were tested first in order to know the validities and reliabilities. Item trials were conducted to determine the quality and characteristics of the items by analysing data from the responses of the trial participants' answers. The trial process was carried out on 298 fifth-grade elementary school students in Pontianak using random sampling. The results of the trial analysis were used as a reference to keep the items, or eliminate the unfit items (Mardapi, 2017).

5) Analysing test items, Item analysis was conducted after obtaining data from the test results. The analysis was carried out by testing content validity, construct validity and reliability. Content validity was analysed by using Aiken's V, while construct validity was analyzed with the EFA and IRT. The reliability was conducted by using Alpha Cronbach. In item response theory, this research used PCM model which is a development of the Rasch model for polytomous scoring. Item analysis was using excel, SPSS, jmetrik and R studio softwares. In this analysis, the sample adequacy test, assumption test, and item fit per item were carried out, which are the MNSQ (Outfit Mean Square) and ZTSD (Outfit Z Standard) values, information functions or TIF (Test Information Function) and measurement errors (Standard Error Measured).

6) Revising the test. The question items that have been analysed and received feedback were improved. Question items that did not meet the criteria or were classified as poor were corrected or dropped. Good question items were retained.


7) Assembling the test. The question items that have been reviewed, tested, analysed and improved were assembled into the final form, that's ready to use. The final test has already has revised instruction for working on the questions and valid and reliable items only.

8) Implementing the test. The final test that is valid and reliable was used to measure student misconception. The test was implemented on 76 students in fifth grade who have already learnt about addition and subtraction of fractions.

9) Interpreting test results. After getting data, the answers or responses of students were analysed. Learners' answers were converted into categories of misconceptions, and profile of misconceptions also analysed in order to know the most problematic concepts.

Three-tier diagnostic test is the refinement of two-tier test that was developed by Treagust. In the construct point of view, three-tier diagnostic can be interpreted as two-tier test that is complemented by CRI (Certainty of Response Index) (Taslidere, 2016). The first tier serve questions about addition and subtraction of fractions while the second tier asked students about the reason of why they chose the answer in the first tier. Here's the design of three-tier diagnostic item

TABLE I
 DESIGN OF THREE-TIER DIAGNOSTIC ITEM

First Tier: Question about addition and subtraction of fraction... a. ... b. ... c. ...
Second Tier: Alasan: a. b. c.
Third Tier: Tingkat Keyakinan: 

In third tier, certainty level was categorised into "Sure/confident" and "Not sure/ Not confident" according to the table II below:

TABLE II
 CRI CATEGORIES

CRI	Validity Index	Category
5	Certain	High/ Confident/ Sure
4	Almost certain	Sure
3	Sure	
2	Not sure	Low/
1	Amost Guess	Unfonfident/ Not Sure
0	Totally guessed answer	Sure

(Diani et al., 2019)

Not all wrong answer and reason should be named "misconception", and not all correct answer and reason should be considered as understanding the concept. By adding certainty level, students' concept diagnosis expand into more categories which are "lack of knowledge", and "lucky guess" or "understand but not confident".

In this research, the test was constructed according to the outline designed as follows:

TABLE III
 TEST OUTLINE

Sub topic	Indicators	Item
Addition and subtraction of fractions with different denominator s	State and differentiate how to calculate fractions with different denominators	1A, 1B
	Gives the right example of fractions' calculation with different denominators	2A, 2B, 3A, 3B
	Calculate simple fractions, mixed fractions, and whole numbers correctly	5A, 5B, 6A, 6B, 7A, 7B
	Solve story problems about fraction operations	8A, 8B
Addition and subtraction of decimal fraction	Formulate the correct calculation operation for the given fractions	11A, 11B
	Gives the right example the calculation operation of decimal fractions correctly	4A, 4B
	Solve story problems about the calculation operation of decimal fractions	9A, 9B
	Comparing counting operations of decimal fractions	10A, 10B

Meanwhile, the analyses included content validity analysis, construct validity analysis and reliability analysis. Content validity analysis was carried out by expert judgment with Aiken's V. The developed diagnostic test instrument was reviewed and validated by two experts lecturers specialised in elementary school education with a concentration in mathematics, and one mathematics lecturer. The scores by validators then computed according to the equation below:

$$V = \sum \frac{s}{n(c-1)} \quad (1)$$

s : r-lo
 lo : lowest validity score
 c : highest validity score
 r : validity score given by validators
 n : number of validators (Azwar, 2016)

Then, to know the category of each item's content validity, the Aiken V value was analysed according to the tabel 1 below:

TABLE IV
 SCORING OF THREE-TIER DIAGNOSTIC TEST

Validity Index	Category
$0 \leq V \leq 0,4$	Invalid
$0,4 < V \leq 0,8$	Medium validity
$0,8 < V \leq 1$	Very Valid

In addition, test reliability analysis was also conducted using the Alpha Cronbach reliability test by using SPSS following the equation below:

$$r_{11} = \left[\frac{k}{k-1} \right] \left[1 - \frac{\sum a_b^2}{a_1^2} \right] \quad (2)$$

r_{11} = Reliabilitas Instrumen

K = Banyaknya item pertanyaan atau pertanyaan

$\sum a_b^2$ = jumlah varian butir

a_1^2 = varian total (Sumintono & Widhiarso, 2015)

To know the category of each item's content validity, the Alpha Cronbach value was analysed according to the table 1 below:

TABLE V
 ALPHA CRONBACH ANALYSIS

Reliability Index	Category
0,00 – 0,199	Very low reliability
0,20- 0,399	Low reliability
0,40 – 0,599	Mediam reliability
0,60 – 0,799	Strong reliability
0,80 – 1,000	Very strong reliability

(Sumintono & Widhiarso, 2015)

Beside Alpha Cronbach, reliability test also conducted by analysing person dan item reliability index by Rasch model using Jmetrik. If Alpha Cronbach value indicates the person and item interaction, person reliability reflects the consistency of responden in answering the test. On the other hand, item reliability in Rasch model shows the consistency of the item (Sumintono & Widhiarso, 2015).

In Rasch analysis, the separation index also executed using Jmetrik. The separation index refers to the differentiation between items and individuals. The separation index for individuals indicates differences based on their ability levels, while the separation index for items reflects the varying levels of difficulty of the items (Kamis et al., 2018). The category analysis followed the Table VI below:

TABLE VI
 RASCH REALIBILITY ANALYSIS

Person/ Item Reliability Index	Separation Index	Category
<0,67	≤ 2	Poor
0,67 – 0,80	2-3	Fair
0,8 – 0,9	3-4	Good

0,91-0,94	4-5	Very good
>0,94	>5	Excellent

(Fisher in Farozin et al., 2022)

The construct validity analysis of three-tier diagnostic test items was using EFA and PCM (Partial Credit Model) in item response theory. PCM is an extended branch of the Rasch model in item response theory aimed at instruments that have a polytomous answer response (not only 1 and 0). The scoring technique for three-tier diagnostic test in this study follows Istiyono's scoring technique (2020) as follows:

TABLE VII
 SCORING OF THREE-TIER DIAGNOSTIC TEST

First Tier	Second Tier	Third Tier	Score
Correct	Correct	Sure	8
Correct	Correct	Not Sure	7
Wrong	Correct	Sure	6
Wrong	Correct	Not Sure	5
Correct	Wrong	Sure	4
Correct	Wrong	Not Sure	3
Wrong	Wrong	Sure	2
Wrong	Wrong	Not Sure	1

Because the scoring in this research has interval of 1-8 and not only 1 and 0, so the scoring's categorized as polytomous. PCM that's an extension of the Rasch model, have same assumption and characteristics as Rasch model (Retnawati, 2017). So, just like Rasch model, PCM also help researchers to conduct analysis that is invariant between item and person, and not sample or test dependent (Hambleton et al., 1992; Planinic et al., 2019)

PCM was originally developed to analyse test items that require solving steps that need to have several steps of work, but each steps doesn't have to be sequential and do not have to have the same difficulty (Retnawati, 2017; Isgiyanto, 2013). So PCM is suitable for analysing three-tier test, because students can get wrong answer in first tier, but get the right answer in second tier tier while also giving low certainty in third tier. This characteristic makes PCM suitable for evaluating learners' responses when partially correct and partially incorrect answers are possible unorderedly.

When using IRT analysis, assumption test must be conducted as a prerequisite test for analysing item quality. Those assumption tests was unidimensionality, local independence, and parameter invariance (Hambleton et al., 1992). Unidimensionality test was done by looking at the Total Variance Explained obtained from analysis using SPSS. This unidimensional assumption test was carried out, to find out whether the results of student work, whether correct or incorrect, are indeed only influenced by one main factor. The unidimensionality assumption can be met when the dominant raw variance explained value is bigger than 20%, or the scree plot produced only one steepness (Hambleton et al., 1992). A factor can be said to be dominant if it has an eigenvalue at least 2 times greater than the second factor (Retnawati, 2017). Local independence essentially means that items become uncorrelated when variability can be isolated or managed. The concepts of

unidimensionality and local independence are fundamentally equivalent. As a result, satisfying the unidimensionality assumption automatically fulfills the local independence assumption as well (Hambleton et al., 1992; Wells & Faulkner-Bond, 2016).

The assumption of parameter invariance in the assumption test analysis means that the characteristics of the item do not depend on the distribution of participants' abilities, and the parameters that characterize the characteristics of the test takers have nothing to do with the characteristics of the items being tested (Wells & Faulkner-Bond, 2016). Invariance of item parameters can be known through the division of test participants into two groups odd and even or male and female then estimated separately and formed a scree plot between the two. Scree plot between the two. If the parameter distribution is close to the diagonal line ($y=x$) then it can be said that the item parameters have met the assumption of invariance. It can be said that the item parameters have met the assumption of parameter invariance (Retnawati, 2014). After the assumption tests were satisfied, analysis of item quality can be conducted using PCM.

Based on this item quality analysis, it can be determined whether the developed test items are already suitable or not. Items that didn't not meet the requirements in the IRT analysis were revised or annulled to obtain the final product with questions of good quality. The item analysis conducted were item difficulties, fit item (goodness of fit), and information function. For item difficulties, good questions should not have a level of difficulty that is too high, but also not too easy. Items can be said to be good if they have a bi value ranging between -2 and +2 (Hambleton et al., 1992).

Goodness of fit is determined using the MNSQ Outfit value ($0.5 < \text{MNSQ} < 1.5$) and ZSTD Outfit ($-2.0 < \text{ZSTD} < + 2.0$) (Sumintono and Widhiarso, 2015). A rule of thumb can be used to assess the implications of model fit for measurement, namely $\text{MNSQ} > 2.0$ means damaging the measurement system; $1.5 < \text{MNSQ} \leq 2.0$ means it has no meaning for measurement; $0.5 \leq \text{MNSQ} \leq 1.5$ means it is useful for measurement; and $\text{MNSQ} < 0.5$ means it is not useful for measurement even though it does not damage the measurement system (Sumintono and Widhiarso, 2015).

Information function can be defined as a function that provides information on the relationship between the characteristics of the items on the test, and the ability of test takers (Hambleton et al., 1992). The test information function is the total accumulation of the information function of each item that makes up the question. This information function can describe the accuracy level of test measurements on test takers. This is because the information function and measurement error have an inversely proportional relationship, meaning that the greater the value of the information function, the smaller the measurement error, and vice versa (Hambleton et al., 1992). This is represented mathematically through the measurement error formula against the information function as follows:

$$SE(\theta) = \frac{1}{\sqrt{1(\theta)}} \quad (3)$$

(Hambleton et al., 1992).

After the test items were analyzed and improved, misconception measurement was carried out using the final product of three-tier diagnostic test. The categorization of three-tier diagnostic tests was referring to three-tier test analysis category by (Arslan et al., 2012) as follows:

Table VIII
 Analyze of Three-Tier Diagnostic Test

First Tier	Second Tier	Third Tier	Category
Correct	Correct	Sure	Understand Concept
Correct	Wrong	Sure	False Negative Misconception
Wrong	Correct	Sure	False Negative Misconceptions
Wrong	Wrong	Sure	Misconceptions
Correct	Correct	Not Sure	Lucky Guess, understand but lack confidence
Correct	Wrong	Not Sure	Lack of Knowledge
Wrong	Correct	Not Sure	Lack of Knowledge
Wrong	Wrong	Not Sure	Lack of Knowledge

The data was examined to obtain misconception's results experienced by students in fraction. When a student gives a wrong answer, followed by a reason that is also wrong but the level of confidence is high, then the student has a pure misconception. False positive misconceptions occur when students are able to answer correctly, but give wrong reasons /have wrong concepts with confidence (Kamilah & Suwarna, 2019). In pedagogical aspects, "false positive" refers to a condition in which an evaluation instrument identifies an answer that appears to be correct or positive, but is actually inaccurate or negative (Rozycki, 2008). Students who experience false positives respond with answers as if there are no misconceptions, that is, answers that are correct and confident, when in fact they have false confidence because the basis of the given concept is wrong (false positive).

False negative misconceptions are identified when students give the wrong answer response but the reasoning is correct and answer with confidence (Maison et al., 2020). False negative refers to a situation where students appear to give the wrong answer during an assessment, when in fact they understand the material provided. Lack of knowledge or lack of understanding is a condition where students are less confident with the answers they give. Researchers before Arslan et al. (2012) classified all answers that had low confidence as lack of knowledge (I. Caleon & Subramaniam, 2010; Cetin-Dindar & Geban, 2011; Peşman & Eryilmaz, 2010). But Arslan (2012) stated that students who answered correctly followed by the correct reasoning, but were not sure of the answer were included in the lack of confidence or lucky guess .

In addition, the profiles of misconceptions experienced by students were also analysed. After that, a questionnaire

analysis was conducted to find out the responses of students and teachers to the three-tier diagnostic test developed.

III. RESULT AND DISCUSSION

This research on the development of a three-tier diagnostic test has resulted in a diagnostic test product about addition and subtraction of fractions with different denominators, that consisting of package A and package B. This instrument has 10 questions in each question package, and have been examined thoroughly.


A. Content Validity


After some revisions and consultation with the experts, the content validity of three-tier test based on the calculation of Aiken's V resulting that all items developed for both packages have good validity. Some feedbacks were given to improve the questions quality such as follows:

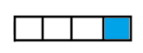
1. Some questions are too simple and weren't according to topic fraction in fifth grade.
2. Pre-requisite questions such as basic concepts of fractions that's still not about addition and subtraction should be revised or eliminated.
3. Double check the use of capital letters.
4. Make the questions between package A and B to be more variate.
5. The picture used should be useful and give mathematical information, and not as decoration only.
6. Fix the instructions to be clearer.

Based on the feedbacks, some revisions were done to the test prototype, like annulling pre-requisite questions about basic fraction concepts that were too simple. Although the questions are about fractions, validators advised to focus the test only for computing addition and subtractions. The eliminated items were as follows:


Dari gambar-gambar berikut, tentukan gambar yang menunjukkan bentuk yang SENILAI DENGAN 1/3

a. 

b. 

c. 

Berapakah nilai pecahan dari gambar berikut?



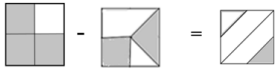
a. 2/4
 b. 4/2
 c. 2/2


Fig. 1 Examples of eliminated pre-requisite item

These types of items were replaced with items that focused on fraction's addition and subtraction. The example of revised item can be seen in image below:

Manakah dari gambar berikut yang menunjukkan contoh pengurangan pecahan

$$\frac{3}{4} - \frac{2}{4} = \frac{1}{4}$$

a. 

b. 


c. 

Fig. 2 Example of revised item 1

Some items that were too similar in package A and package B were changed to make them more varied between package A and package B. The questions were changed, but if question in A package was about addition, then in the B package questions would be about subtraction. The example of revised item displayed as follows:

3A Tentukan contoh penjumlahan pecahan yang tepat dari perhitungan di bawah ini!

a. $\frac{1}{2} + \frac{1}{4} = \frac{3}{4}$

b. $\frac{1}{2} + \frac{2}{3} = \frac{3}{5}$

c. $\frac{1}{6} + \frac{1}{3} = \frac{3}{8}$

3B Tentukan pengurangan yang tepat dari perhitungan di bawah ini!

a. $\frac{6}{8} - \frac{1}{2} = \frac{2}{4}$


b. $\frac{1}{2} - \frac{1}{4} = \frac{1}{3}$

c. $\frac{3}{5} - \frac{1}{2} = \frac{2}{3}$

Fig. 3 Examples of Revised Item 2

Feedback on the use of images was done by changing the initial images that were less representative to be more informative. The images used previously were as follows:

Ibu sedang menimbang menggunakan neraca seperti pada gambar di bawah ini:



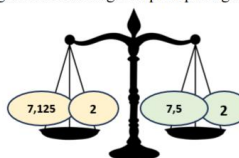
Jika Ibu meletakkan kantong tepung seberat $\frac{2}{3}$ Kg dan $\frac{1}{4}$ Kg di sebelah kanan timbangan. Kemudian ibu meletakkan $\frac{6}{12}$ Kg dan $\frac{5}{12}$ Kg kantong tepung di sebelah kiri timbangan, maka apakah yang akan terjadi?

a. Timbangan sebelah kiri lebih berat dari sebelah kanan
 b. Timbangan sebelah kanan lebih berat dari sebelah kiri
 c. Timbangan menjadi seimbang.

Fig. 4 Example of Uninformative Image

The picture was revised to give more information about the problem asked instead of just to make it looked good for students. The revised picture can be seen as follow:

Ibu sedang menimbang menggunakan timbangan seperti pada gambar di bawah ini:



Jika Ibu meletakkan kantong tepung seberat 7,5 Kg dan 2 Kg di sebelah kanan timbangan. Kemudian Ibu meletakkan 7,125 Kg dan 2 Kg kantong tepung di sebelah kiri timbangan, maka apakah yang akan terjadi?

- Timbangan menjadi miring ke kiri
- Timbangan menjadi miring ke kanan
- Timbangan menjadi seimbang.

Fig. 5 Example of Revised Item 3

Following a process of consultation and subsequent revisions, the developed three-tier test underwent evaluation using the Aiken V method. The results of this assessment are presented below.

TABLE IX
 VALIDITY RESULT

Item Number	A Package	Validity	B Package	Validity
1	0,78	Medium validity	0,89	Very valid
2	0,78	Medium validity	0,78	Medium validity
3	1,00	Very valid	1,00	Very valid
4	0,89	Very valid	0,89	Very valid
5	0,89	Very valid	0,78	Medium validity
6	0,78	Medium validity	0,78	Medium validity
7	0,89	Very valid	1,00	Very valid
8	1,00	Very valid	1,00	Very valid
9	0,89	Very valid	0,89	Very valid
10	1,00	Very valid	0,78	Medium validity
11	0,78	Medium validity	0,89	Very valid

B. Reliability

A reliable instrument is an instrument that's able to consistently measure a variable even when tested repeatedly at different times (Retnawati, 2017). The reliability was analysed using Alpha Cronbach, person reliability and item reliability in Rasch model. The outcomes of these analysis are presented in the following table:

TABLE X
 ANALYSIS OF RELIABILITY INDEX

Package	Separation		Reliability		Alpha Cronbatch
	Person	Item	Person	Item	
A	2,949	5,753	0,896	0,970	0,937
B	2,782	4,371	0,885	0,950	0,931

The analysis results show that the reliability of tests in package A and package B has reliability value at more than 0.7, which are 0,937 and 0,931 that's classified as having very strong reliability. The alpha cronbach value interpret the internal consistency which means it analyses the interaction between person and items, reliability of items and person in

Rasch represent the consistencies of items and people (Sumintono & Widhiarso, 2015). The person separation value was 2,949 and 2,782 that produces number of strata 4,2266 and 4,0430 which are rounded to 4. It means that the test can identify 4 groups of students by their abilities and classified very good. The item separation value was 4,371 and 5,753 that produce strata's number of 8,005 and 6,161 that's rounded into 8 and 6 groups of items. This number of strata categorized as excellent. The higher the separation value, the better the overall items and respondent (Mulyasari et al., 2024). Therefore, it can be concluded that all developed tests are reliable.

C. Construct Validity

To examine construct validity, prerequisite analysis was carried out by Exploratory Factor Analysis (EFA) using SPSS. This analysis was conducted by looking at the results of the KMO and Barlett Test of Sphericity tests. The KMO test is a test intended to measure the suitability of data for factor analysis, in other words, this test was carried out to see the adequacy of the sample (Shrestha, 2021). In SPSS, this test is usually paired with the Barlett Test of Sphericity which is used to identify whether the data set being tested is suitable for the factor analysis to be carried out. If it is adequate, then item quality analysis with IRT can be conducted.

Furthermore, based on the EFA analysis, the KMO value for packages A and B of has shown a number > 0.5, which are 0.935 for package A, and 0.941 for package B. Meanwhile, the results of the Barlett's Test of Sphericity analysis showed values that also met the requirements <0.05. Then, the two packages of questions designed gave a significance value of 0.000 in Barlett test, therefore the requirement has been met (Prihatni et al., 2019; Shrestha, 2021). Analysis result that's meet the requirement indicate that assumption and IRT analysis can be performed.

To do IRT analysis, assumption tests have to be executed first. Those assumption test are unidimensionality, local independence, and parameter invariance. The followings are the results of the unidimensionality test analysis on the developed diagnostic test instrument:

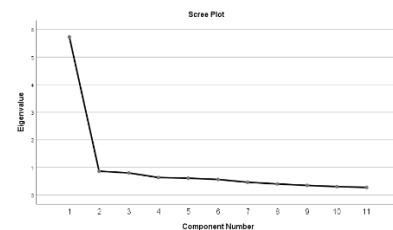


Fig. 6 Scree Plot of A Package

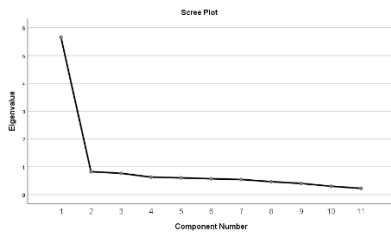


Fig. 7 Scree Plot of B Package

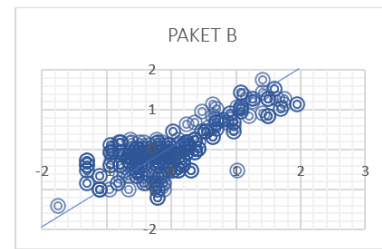


Fig. 9 Scatter plot of B Package

The scree plot shows that there is only one steepness, which indicates that this diagnostic test only measures one dimension (unidimensional), so the unidimensionality requirement has been met. The variance value extracted from the dominant factor analysis with an eigenvalue greater than one is 61.682% for A package, and 59.216% for B package. So it can be seen that there is only one dominant factor in the test instrument, which is indicated by a steepness in the graph that forms an “elbow”. In addition, the resulting variance value also has a value of more than 20% in both packages (Hambleton et al., 1992; Shrestha, 2021). Therefore, it can be concluded that the package A and package B test instruments have met the dimensional test requirements. The first eigenvalue has a value of more than 2.5 times compared to the second eigen (Retnawati, 2017). When eigenvalues drop significantly in size, then additional factors will add relatively little to the information already extracted. Therefore, the test instrument has met the assumption of measuring one dimension.

The assumption of local independence can be met if a test taker's response to one item does not affect their response to other items. It can be concluded that when the unidimensionality rules has been met, then the assumption of local independence is also proven (Hambleton et al., 1992). Since the unidimensionality assumption test has already been fulfilled, the assumption of local independence has also been met. Therefore no extra analysis was conducted to check the local independence assumption.

Parameter invariance analysis was conducted by measuring students' abilities in odd and even numbered items of the developed three-tier diagnostic test instrument. The results of the ability analysis were plotted into a scatterplot. If the scattered data approached a diagonal line, then the developed test met the assumption of parameter invariance (Sari & Suyanta, 2021). The results of the analysis for each developed question package were as follows:

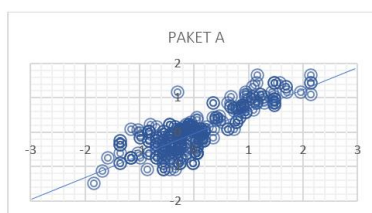


Fig. 8 Scatter plot of A Package

the parameter distribution gathered around the diagonal line $x=y$ in the scatter plot indicate that parameter invariance requirement has been fulfilled. After obtaining results that meet the assumption test requirements, the analysis can be continued on item quality which includes difficulty level, item fit, and information function and SEM.

1) *Item Difficulties*: Item difficulty analysis was carried out by looking at the measure values resulting from the PCM analysis. Based on the analysis carried out, all the questions developed in the package A and B have good level of difficulty because they have a difficulty value between -2 to +2. The results of the analysis show that all the questions developed have good item difficulty, so they are not too difficult, but also not too easy.

2) *Item Fit*: Item fit analysis was carried out with R studio based on the Partial Credit Model (PCM) by looking at the MNSQ (mean square) and ZTSD (Standardized as z-score) outfit values. As for the results of the item fit analysis obtained from all questions are as follows:

Table XI
 Item Analysis

Item	A Package		B Package	
	Oufit	ZTSD	Oufit	ZTSD
S1	0.736	-1.232	0.853	-2.235*
S2	1.052	-1.800	1.100	-0.151
S3	0.968	-1.311	1.306	0.999
S4	1.238	-0.708	1.201	0.549
S5	0.929	-3.288*	0.817	-2.751*
S6	1.347	-1.479	0.904	-2.154*
S7	0.989	-1.470	1.034	-0.780
S8	0.736	0.562	1.087	-0.561
S9	1.052	-1.945	0.959	-1.664
S10	0.968	1.832	1.082	-0.553
S11	1.238	-0.921	1.031	-1.589

Based on the analysis, the results obtained were that the test instrument was fit, because the mnsq outfit value for each question item in package A and package B was in the value range of 0.5-1.5. Meanwhile, when conducting analysis of the ZTSD outfit scores with the R program, it was found that several question items that are not suitable, which is number 5 in package A. Question items that are not suitable in package B were numbers 1, 5 and 6.

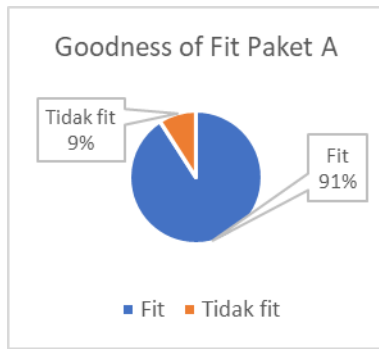


Fig. 10 Goodness of Fit A Package

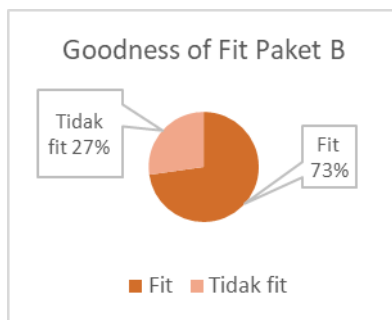


Fig. 11 Goodness of Fit B Package

According to the result, item 5 in A package that was unfit was eliminated. However, in B package although the unfit item by ZTSD was item 1, 5 and 6, the item eliminated was only item number 5. According to (Sumintono & Widhiarso, 2015), item can be used even if there's only one characteristics that fit. So actually all of the item can be used and there's no item eliminated because the MNSQ value showed fit result for all items. But, item number 5 was found unfit in both A and B package, while those items had similar question structure. Furthermore, it also has farthest value of ZTSD from the required value of -2 until +2, when being compared to other unfit items, it was decided that item number 5 better be eliminated. Item number 1 and 6 was re-analysed and revised to produce better items. Item number 1 and 6 already has correct mathematical concepts and solid misconception reference, but the question's description could be simpler so it can be understood more easily.

3) Information Function: The next analysis carried out in this research is the analysis of the test information function and SEM (Standard Error Measurement) with the help of the R console. Based on the results, the information and SEM function plots were obtained as follows:

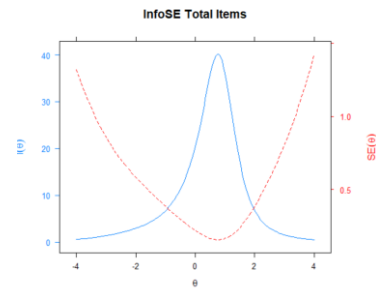


Fig. 12 InfoSE Graph of A Package

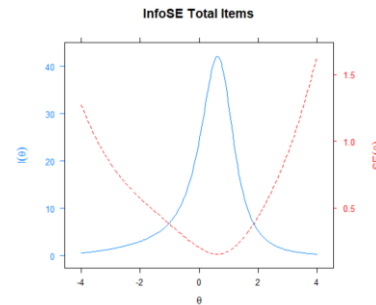


Fig. 13 InfoSE Graph of B Package

Based on the results of the analysis of the information function and SEM, it was found that there was an intersection of the curves at points -0.8 and 1.8 in the package A and B curves. So it can be concluded that the instrument developed is valid for use on students with abilities of -0.8 to 1.8 for both instruments of package A and package B.

4) Questionnaire analysis: The instrument was analyzed using a questionnaire given to students and teachers. The results of the questionnaire showed that 77.53% (high) of the material provided was suitable for students. Then the designed test display also received a good response from students, namely 83.08% felt that the test display was attractive and easy to understand. Furthermore, the results of the analysis regarding the systematics of the test gave a result of 79.17% which was included in the high category. This shows that the systematic test designed is also suitable for fifth grade elementary school students. Finally, student responses regarding the usefulness of the test are generally included in the very high category, which is 81.39%, so it can be concluded that students feel that this test is useful in helping them understand their conceptual understanding. Questionnaire analysis was also given to teachers to find out teachers' responses regarding the effectiveness of the tests that had been developed. The results of the analysis show that teachers responded well to the tests developed, proven by the data that all aspects analysed were included in the very high category. The results respectively were the material aspect at 89.42%, the appearance aspect at 91.67%, the systematic aspect at 86.92% and the usability aspect at 92.83%. Based on the results of the questionnaire analysis, it can be concluded that the three-tier diagnostic test developed received a good response from teachers and students.

5) Measurement results

Based on the results, it was found that 5th grade students who already understood the concept were only 6.30%; while 29.09% experienced pure misconceptions; 12.61% experienced false positive misconceptions; 8.97% experienced false negative misconceptions, 39.52% did not understand the concept, and 3.52% guessed right, or already understood but were still in doubt. Among the students that have misconceptions, the types of misconceptions and its distribution was broken down into as displayed below :

Types of misconceptions	A Package		B Package	
	No.	%	No.	%
Concept	1	17,94%	1	20,51%
	2	7,69%	2	11,28%
	3	42,15%	3	53,84%*
Procedural problem of simple fractions and decimal	4	20,76%	4	22,89%
	8	23,58%	8	23,07%
	9	35,89%	9	33,07%
Procedural problem of mixed numbers	6	43,15%	6	41,02%
	7	56,41%*	7	51,28%*
	11	41,58%	11	46,71%
Comparing decimal	10	61,53%*	10	48,38%

The most common forms of misconceptions in A package was decimal number equivalency in number 10, followed by procedural error mixed fractions in number 7. The question was a story problem about the comparison of fraction in the form of decimal numbers that have been calculated as listed in the following picture:

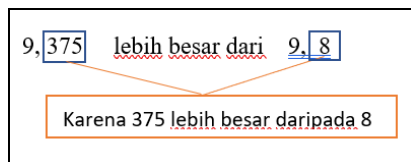


Fig. 1 Highest Misconception in A Package

In this misconception, students assume that decimal fractions have the same characteristics as whole numbers. They thought that when a number appears to have more numbers, it will have a greater value (Durkin & Johnson, 2015). This misconception was due to students mistake of still applying the concept of whole numbers in fractional numbers. When they see the comparison between 375 and 8, they immediately think that 9.375 must be greater than 9.8. This is in line with previous researches that students have tendency of "whole number bias", that is the application of the concept of whole numbers in fractional numbers (DeWolf & Vosniadou, 2015; Færch & Hodgen, 2023). In addition, fraction story problem is one of the most difficult topics to understand, and students tend to do better on problems that are symbolic than sentences (Agbozo, 2020; Nurharyanto & Retnawati, 2020; Üzel, 2018).

The second highest misconception was about procedural error of mixed numbers. The question was about addition of mixed fraction and whole number as follow:

$$2\frac{2}{3} + 4 = \dots$$

Many student answered that the right answer is $2\frac{6}{3}$, and choose the reason of simply adding the numerator with the whole number (Ratnasari, 2018; Üzel, 2018):

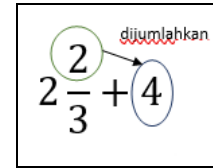


Fig. 14 Second Highest Misconception

In this question, students couldn't convert mixed fraction as simple fraction, as well as to convert whole number into fraction. Many students often find it hard to understand mixed fraction that represent a combination of whole number plus simple fraction. It appears that students tend to consider the whole number part and fraction part in mixed fraction separately (Aksoy & Yazlik, 2017; Siti Nurhani & Siti Mistima, 2018).

However, in contrast to question package A, the highest misconception in B package occurs in question number 3 about choosing correct subtraction problem along with its visual representation. Students chose the bar representation but have a wrong meaning (in this problem $3/5 - 1/2 = 2/3$). They thought that fraction can have the same size and shape, only the number of blocks on the bar is different without considering the whole part proportion (Isoda, 2013) as follows:

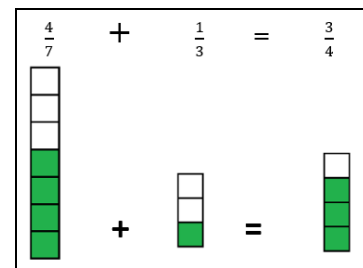


Fig. 15 Highest Misconception in B Package

This image portrayed the visual representation of common misconception about computing numerator and denominator separately (numerator with numerator, denominator with denominator).

Similar to A package, the second highest misconception in B package was also about calculating mixed numbers that need conversion as follows:

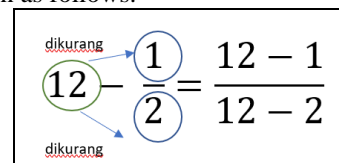


Fig. 16 Second Highest Misconception in B Package

The question shows that students simply subtract the whole number with the numerator and denominator of simple fraction. Students failed to convert whole number

into fraction and resulting in computational error (Aksoy & Yazlik, 2017). In compare with A package, both questions has whole number in the calculation. Although teacher said that this kind of calculation has been taught, it seems not having denominator in computing fraction has made students to get confused and chose the simplest way. Some students chose the reasoning option of converting whole number with dominator of "0" instead of "1" (Aksoy & Yazlik, 2017), but not many students chose this option. This result shows that students need more comprehensive understanding of whole numbers in fraction operations.

D. Discussion

Quality testing of the three tier diagnostic test includes content validity, construct validity, reliability, and item analysis with the PCM model. From the calculation of the Aiken V formula on the assessment given by experts, all question items (numbers 1 to 11) show a very high level of validity. Validity analysis of the test as a whole test also resulted in a validity value that is categorized very high. In general, both package A and package B have shown very high validity values. This shows that the diagnostic test instruments developed already have valid content.

Furthermore, construct validity analysis was conducted with EFA (Exploratory Factor Analysis) which gave the results of KMO values of 0.935 and 0.941 in package A and package B, respectively. This shows that the KMO value of the diagnostic test developed has met the requirements because those tests have value >0.5 . Likewise, the significance value of the Barlett Test of Sphericity shows a value of 0.000 for all question packages, so it has met the requirements because it is worth <0.05 (Prihatni et al., 2019; Shrestha, 2021). In addition, a reliability analysis was carried out using the Cronbatch Alpha formula which gave the result that the entire package of questions developed was reliable because it had an alpha value > 0.7 (Sugiyono, 2016), which are 0.937 and 0.931.

The item characteristics were analyzed with Item Response Theory (IRT) using the PCM model. Analysis with IRT must go through an assumption test which includes unidimensionality test, local independence test, and parameter invariance. The three-tier diagnostic test developed has fulfilled the three assumption tests, so the analysis can proceed to item analysis with PCM. Based on the PCM analysis, it was found that all questions developed about addition and subtraction in both package A and package B were classified as good because they were between -2 to +2. (Hambleton et al., 1992). This shows that the questions developed are not too difficult, but also not too easy, so they are good to use for students. As for the information function and SEM analysis, the results shows that the three-tier diagnostic test developed will provide maximum information if used on students with abilities of -0,8 to 1,8. It means this test will give optimum information when being used for students with range ability of -0,8 until 1,8. Furthermore, based on the item fit analysis, the MNSQ outfit value for all items in all question packages has met the fit requirements because it is between 0.5-1.5 (Mulyasari et

al., 2024). However, when analyzing the ZTSD outfit value, several items were found that did not fit, which were number 5 in package A, and numbers 1, 5, and 6 in package B. The instruments that have been tested and analyzed were revised and sorted to produce the final Three-Tier Diagnostic Test product. Therefore, the final product of the three-tier diagnostic test instrument in addition and subtraction material was obtained, each package consisting of 10 questions.

After the instrument was revised until the final product of the three-tier diagnostic test was obtained, the instrument was used to determine the misconception profile and forms of misconception of students. The study's findings revealed a diverse range of understanding among 5th grade students. Only a small fraction (6.30%) demonstrated a solid grasp of the concept. A significant portion (29.09%) held pure misconceptions. Some students (12.61%) exhibited false positive misconceptions, while others (8.97%) showed false negative misconceptions. The largest group (39.52%) lacked understanding of the concept altogether. A small percentage (3.52%) either guessed correctly or understood the concept but lacked confidence in their knowledge. This breakdown illustrates the varied levels of conceptual understanding and misconceptions present. This shows that students still do not master the topic of calculating fractions, because the results of the analysis are dominated by the percentage of misconceptions of 29.09%, and do not understand the concept of 39.52%, while only 6.30% are identified as understanding the concept. This shows that more than half of the students still have difficulty understanding or solving fraction calculation operation problems.

Based on the result, it can be seen that students conception about fraction addition and subtraction was dominated misconception and lack of knowledge. This could be useful information, because addressing misconception and lack of knowledge need different instructional methods (Peşman & Eryilmaz, 2010). Addressing misconceptions is harder than lack of knowledge, because the concepts have embedded in students' mind. Remediation of misconception requires conceptual change approach that stimulate cognitive conflict that confront incorrect ideas (McAfee & Hoffman, 2021). While lack of knowledge need direct instruction about students' missing content, and activities that offer scaffolding learning experiences (Merliza et al., 2020). Diagnostic test result provides information for teacher of the right action choice.

According to the measurement result, it was found that highest misconception surfaced in comparing decimal and procedural error of mixed fraction in A package. The misconception of comparing decimals in number 10 shows that the highest misconceptions have a tendency of "whole number bias". It was found that when students working on problems that require comparing and computing rational numbers like decimals, they tend to get the wrong idea of misapplying their understanding of whole number into decimal numbers (DeWolf & Vosniadou, 2015; Færch & Hodgen, 2023). Resnick (in Lai & Wong, 2017) named students perspective of comparing into three ideas that was

called Sackur-Grisvard and Leonard Rules 1,2 and 3. Rule 1 is about the misconception found in this research of “Longer-is-Larger” rule which is whole number bias of considering decimals as the another whole number on the right side of comma. The second rule was about “Shorter-is-Larger” rule which assume that decimal with smaller number will be larger because fraction with greater denominator is considered smaller, while the third rule is about the right concept of adding zeros to the decimal with fewer numbers. Actually, this test also contains the “Shorter-is-larger” reasoning option, but not many students answered them. It’s mostly because thinking that $\frac{1}{375}$ is larger than $\frac{1}{8}$ seems rather complicated for students than thinking that 375 is larger than 8.

In order to counter whole number bias, salient benchmark, especially 0 and 1, as it’s found to be effective to overcome the bias (Obersteiner et al., 2020). Teaching methods that emphasize understanding fractions through their representation on number lines can assist students to prevent whole number bias. It helps them connect whole numbers and fractions, and it broadens their understanding of numbers as part of a continuous spectrum. Furthermore, self regulatory activities also help students regarding whole number bias. When students were given assistance in explaining concepts to themselves, it led to two positive outcomes. They became more precise in comparing the size of fractions and decimals, and they were able to offer better-quality justifications for why one fraction is larger or smaller than another (Namkung et al., 2018).

The B package has highest misconceptions in number 3 about visual representation of fraction calculation with different denominators. In this question, it shows that students have begun to understand that fraction representation should be equally partitioned and have same size parts (Aliustaoğlu et al., 2018; Xu et al., 2024). But, they forgot that in computation of fractions with different denominators the representation of the “whole” in each fraction also should be proportioned with each other (Isoda, 2013).

Interestingly, although A package has similar question about visual representation of fraction addition, B package that’s about fraction subtraction has higher misconception. Previous researches have stated that generally addition in whole number was always easier than subtraction (Kamii et al., 2001; Makonye & Fakude, 2016), but no research has discussed deeply the comparison between those operations in fraction while addition and subtraction basically related to each other. It would be interesting to compare students’ performance in addition and subtraction separately. It’s only commonly known that working on fractions with same denominators is easier than different denominators (Liu & Jacobson, 2022), but no study has discussed students’ performance of fraction in different arithmetic operation. Regarding visual representation problems, digital tools is recommended to remediate students’ misconceptions (Yilmaz et al., 2018). When designed properly, digital tools can concretize the abstract mathematics concepts (Yilmaz et

al., 2018). This would help students to conceptualize the right images of fraction’s computations.

In both packages, the second highest misconception was procedural error of calculating mixed numbers in number 7. The fact that the misconception rate increasing significantly from calculating simple fractions only to mixed numbers shows that students really have problem in converting fractions and whole numbers. This is in line with previous misconceptions that when students adding mixed numbers, they tend to convert the fraction improperly before calculate them (Africa et al., 2020; Ratnasari, 2018). This is because there are many steps that students must follow to get the right answer, so it’s possible that they forgot the right steps and got confused (Nurharyanto & Retnawati, 2020; Ratnasari, 2018). However, most elementary school students did the conversion not because they identified different forms fractions that need to be adjusted, but because that’s how they’re taught (Africa et al., 2020). This measurement results showed that teachers should provide more explanation and better task sequence in explaining calculation of mixed fractions.

In order to remediate this misconception, Riyantini et al. (2015) suggest to use clinical interview that has cognitive conflict, as it’s effective to remediate students’ misconceptions about fractions’ procedurals. Clinical interviewing allows unique access to individual children as it effectively opens up the child’s way of thinking. Furthermore, teacher could also use interactive technological learning activities as effective method to overcome procedural misconceptions (Alkhateeb, 2020). Moreover, some research designed suitable learning or remedial activities based on diagnostic test results, and the combination of those two was proven to be effective against fraction misconceptions (Alkhateeb, 2020; Cut et al., 2023).

Overall, some remediation researches of fraction misconceptions stated the importance of using diagnostic test and how it helped to design effective remediation to overcome misconceptions in fractions (Alkhateeb, 2020; Cut et al., 2023). The information provided by the test can lead teachers to design better learning activities in fraction that can lead cognitive conflict in order to help conceptual change. Beside researches that’s proven effective in remediating fractions’ misconceptions, methods like refutational text, dialog, argumentation based or other self-reflection based activities also recommended to address misconception generally (McAfee & Hoffman, 2021; Rodrigues & Thacker, 2019). These methods stimulate cognitive conflict to make students realised about their wrong way of thinking to help them reconstruct their knowledge.

According to the Makhubele’s study (2021), there are four reasons students experience misconceptions or do not understand fraction. The first reason is that students tend not to understand the meaning of fractions, so students tend to memorize formulas and calculation patterns taught by teachers (Aliustaoğlu et al., 2018; Qomario et al., 2020). The second reason is “whole number bias”, that happens when students who have learned the concept of whole numbers

first struggling to understand new concepts and rules in fractions. Students who have not absorbed the concept of fractions will tend to assume that the characteristics of whole numbers can also be used in fractions (Yu et al., 2024). The next reason is that fractional numbers have many constructs and meanings. The concept of fractional numbers can be interpreted as a part of the whole, ratio and quotient (Kor et al., 2018). If students can understand the meaning of fractional numbers, they will be able to better understand why the computation and meaning of values in fractional numbers are different from whole numbers. This makes the concept of fractions more “reasonable” and less abstract, and prevents “whole number bias” (Eichhorn, 2018). In addition, fractional numbers also have the forms of pure, mixed, decimal, percent and permil fractions. Therefore, students have difficulty in understanding the concept and calculation of various fractional numbers. This is shown by the results of this study that there are students who can do calculation problems between ordinary fractions, but cannot do calculations of mixed fractions, or fractions with whole numbers. The next reason is that the textbook used emphasizes more on the calculation procedure of how to do the problem, and not on the meaning of the concept of fractional numbers that can be digested by elementary school children.

Regarding lack of knowledge, many researches have recommended various strategies to improve students understanding about fractions, such as daily life examples, repetition of preliminary knowledges, area models, number line model, asking leading examples, etc (Deringöl, 2019; Lai & Wong, 2017; Yu et al., 2024). It's also recommended to use Concrete-Picture-Abstract method in explaining fraction to elementary students (Africa et al., 2020; Ubah, 2021; Üzel, 2018). Combining multiple use of representations also could bring better result of students (Mohamed et al., 2021; Ubah, 2021).

Some researches advised to use media and tools, especially those that contain rectangular and circular forms to understand fraction values (Indrawati et al., 2018; Jarrah et al., 2022; Turidho et al., 2021; Unaenah et al., 2020). This result is in line with (Liu & Jacobson, 2022) who've found that rectangle strategy was the most popular among students, and was used the most to solve comparison questions, followed by circle strategy. The use of images could be executed by using digital pictures and animations to design more interesting and explanatory learning activities (Yilmaz et al., 2018).

Meanwhile, using pictures and interesting media to conduct fun learning is good for elementary school students, but it's also crucial to carefully explain the properties of fraction images. Teacher need to clearly explain that each parts of the picture has to be equal, because it's difficult for students to figure out the right picture that represent standardized whole when analysing fractions with different denominators (Aliustaoğlu et al., 2018; Liu & Jacobson, 2022), or else, students would have misconceptions just like question number 3 about visual representation that has highest misconception percentage of B package in this

research. Knowing the difficulties, it's recommended for teachers to allocate more time in teaching fractions. According to Jarrah et al. (2022), surprisingly students gave better result in addition and subtraction of fractions with different denominators instead of common denominators because teachers spend more time in explaining them.

However, this research has some limitations. This test is only about addition and subtraction of different denominators, so other topics couldn't be analysed. In content-wise, the three-tier test was designed to measure misconceptions according to the previous founds of misconceptions, so it couldn't discover new types of misconceptions. The research also had limited sample, so it would be better to use bigger sample size in the future. Moreover, this diagnostic test is able to measure the misconceptions only, so it couldn't analyse the causes and factors that influence students misconceptions.

IV. CONCLUSIONS

In conclusion, the three-tier diagnostic test developed was valid and reliable. It has fulfilled all the tests requirement of valid and reliable diagnostic test although some items should be eliminated and revised. The measurement result of this three-tier diagnostic test was that fifth-grade students still dominated by misconception and lack of knowledge. The highest misconceptions was about whole number bias and wrong concept of visual representation, followed by procedural error of mixed numbers. This result should become references for teachers to design more self-reflected activities to address misconceptions and use multi-representation media and tools to increase students knowledge about fraction's addition and subtraction. As McAfee & Hoffman, (2021) claimed, without proper measurement, an explicit misconception remediation strategy remains untenable.

As for suggestions for further research, researchers recommend developing diagnostic test products that are computer-based (Computer Based Test) and can be accessed online, so that the correction process can run automatically and results can be obtained instantly with easy access. The exploration of three-tier diagnostic test in other topics such as multiplication and division of fraction also would be very good to assist teachers. It' also would be better to use bigger sample size in order to expand the generality of research findings. In addition, conducting further research about students problem source and other affecting factors to students also recommended.

REFERENCES

- Abbas, M. L. H. (2016). Pengembangan instrumen three tier diagnostic test miskonsepsi suhu dan kalor. *Ed-Humanistics.*, 1(2), 83–92.
- Abbas, N. A., Abdullah, N. A., Shahrill, M., & Tengah, K. A. (2022). Primary school pupils' performance on the addition of fractions: conceptual and procedural knowledge. *Jurnal Pendidikan Matematika*, 16(2), 227–238.

- <https://doi.org/10.22342/jpm.16.2.17811.227-238>
- Africa, M., Borboran, A. M., Guilleno, M.-A., Mendiola, M., Portento, K. M., Rodrigues, R., Torre Franca, J., Viernes, R. M., & Conde, R. (2020). A lesson study on using the Concrete-Pictorial-Abstract (CPA): Approach in addressing misconceptions in learning fractions. *SABTON: Multidisciplinary Research Journal*, June, 1–8.
- Agbozo, K. K. (2020). Pre-service teachers' expressed attitude toward teaching of fractions and some concepts of fraction. *European Journal of Research and Reflection in Educational Sciences*, 8(3), 224–235. www.idpublications.org
- Aksoy, N. C., & Yazlik, D. O. (2017). Student errors in fractions and possible causes of these errors. *Journal of Education and Training Studies*, 5(11), 219. <https://doi.org/10.11114/jets.v5i11.2679>
- Aliustaoğlu, F., Tuna, A., & Biber, A. Ç. (2018). Misconceptions of sixth grade secondary school students on fractions. *International Electronic Journal of Elementary Education*, 10(5), 591–599. <https://doi.org/10.26822/iejee.2018541308>
- Alkhateeb, M. A. (2020). Correcting misconceptions in fractions using interactive technological learning activities. *International Journal of Learning, Teaching and Educational Research*, 19(4), 291–308. <https://doi.org/10.26803/ijlter.19.4.17>
- Anggrayni, S., & Ermawati, F. U. (2019). The validity of four-tier's misconception diagnostic test for work and energy concepts. *Journal of Physics: Conference Series*, 1171(1). <https://doi.org/10.1088/1742-6596/1171/1/012037>
- Arifin, S., Wahyudin, W., & Herman, T. (2020). The effects of contextual group guided discovery learning on students' mathematical understanding and reasoning. *Jurnal Prima Edukasia*, 8(2), 106–114. <https://doi.org/10.21831/jpe.v8i2.33059>
- Arslan, H. O., Cigdemoglu, C., & Moseley, C. (2012). A three-tier diagnostic test to assess pre-service teachers' misconceptions about global warming, greenhouse effect, ozone layer depletion, and acid rain. *International Journal of Science Education*, 34(11), 1667–1686. <https://doi.org/10.1080/09500693.2012.680618>
- Asnawi, A., Kenedi, A. K., Ramadhani, D., Sahudra, T. M., Angga, I., & Nasution, S. A. (2023). Analysis of needs for elementary school students' diagnostic instruments development. *Proceedings of the 1st International Conference on Early Childhood Education in Multiperspective, ICECEM 2022*. <https://doi.org/10.4108/eai.26-11-2022.2339525>
- Caleon, I. S., & Subramaniam, R. (2010). Do students know What they know and what they don't know? Using a four-tier diagnostic test to assess the nature of students' alternative conceptions. *Research in Science Education*, 40(3), 313–337. <https://doi.org/10.1007/s11165-009-9122-4>
- Caleon, I., & Subramaniam, R. (2010). Development and application of a three-tier diagnostic test to assess secondary students' understanding of waves. *International Journal of Science Education*, 32(7), 939–961. <https://doi.org/10.1080/09500690902890130>
- Cetin-Dindar, A., & Geban, O. (2011). Development of a three-tier test to assess high school students' understanding of acids and bases. *Procedia - Social and Behavioral Sciences*, 15, 600–604. <https://doi.org/10.1016/j.sbspro.2011.03.147>
- Chewoh, F., & Marmoah, S. (2021). Diagnostic test misconception of the sun in elementary school. *East African Scholars Journal of Education, Humanities and Literature*, 4(6), 237–241. <https://doi.org/10.36349/easjehl.2021.v04i06.001>
- Cut, S. S., Johar, R., & Mailizar, M. (2023). The effect of an e-learning-based remedial program on students' learning outcomes: The case of Fraction. *International Journal of Trends in Mathematics Education Research*, 6(2), 118–124. <https://doi.org/10.33122/ijtmer.v6i2.209>
- Deringöl, Y. (2019). Misconceptions of primary school students about the subject of fractions. *International Journal of Evaluation and Research in Education*, 8(1), 29–38. <https://doi.org/10.11591/ijere.v8i1.16290>
- Dewi, S. A., Susilaningih, E., & Sulistyarningsih, T. (2018). Conceptual understanding analysis through two-tier model diagnostic tests on acid-base materials. *JKPK (Jurnal Kimia Dan Pendidikan Kimia)*, 3(3), 160. <https://doi.org/10.20961/jkpk.v3i3.21904>
- DeWolf, M., & Vosniadou, S. (2015). The representation of fraction magnitudes and the whole number bias reconsidered. *Learning and Instruction*, 37, 39–49. <https://doi.org/10.1016/j.learninstruc.2014.07.002>
- Diani, R., Alfin, J., Anggraeni, Y. M., Mustari, M., & Fujiani, D. (2019). Four-tier diagnostic test with certainty of response index on the concepts of fluid. *Journal of Physics: Conference Series*, 1155(1). <https://doi.org/10.1088/1742-6596/1155/1/012078>
- Durkin, K., & Rittle-Johnson, B. (2015). Diagnosing misconceptions: Revealing changing decimal fraction knowledge. *Learning and Instruction*, 37, 21–29. <https://doi.org/10.1016/j.learninstruc.2014.08.003>
- Eichhorn, M. S. (2018). When the fractional cookie begins to crumble: Conceptual understanding of fractions in the fifth grade. *International Journal of Research in Education and Science*, 4(1), 39–54. <https://doi.org/10.21890/ijres.382933>
- Ermayani, L., Suarjana, I. M., & Parmiti, D. P. (2019). Analisis kemampuan siswa dalam menyelesaikan soal pecahan sederhana. *Jurnal Pedagogi Dan Pembelajaran*, 1(1), 9. <https://doi.org/10.23887/jp2.v1i1.19325>
- Færch, J. V., & Hodgen, J. (2023). Danish students' understanding of fractions: a replication study. *Implementation and Replication Studies in Mathematics Education*, 3(2), 200–242. <https://doi.org/10.1163/26670127-bja10014>
- Farozin, M., Astuti, B., Sanyata, S., Titi, D., Nurmalasari, E.,

- & Sutanti, N. (2022). Rasch model for the need assessment instrument of academic guidance and counseling program in junior high school. *Jurnal Penelitian Dan Evaluasi Pendidikan*, 26(2), 177–185. <https://doi.org/10.21831/pep.v26i2.51816>
- Fauzan, A., & Sari, O. Y. (2017). Pengembangan alur belajar pecahan berbasis realistic mathematics education. *Prosiding Seminar Nasional Pascasarjana Unsyiah Aceh*, 55–63. <https://jurnal.usk.ac.id/SNP-Unsyiah/article/view/6886/5619>
- Gurel, D. K., Eryilmaz, A., & McDermott, L. C. (2015). A review and comparison of diagnostic instruments to identify students' misconceptions in science. *Eurasia Journal of Mathematics, Science and Technology Education*, 11(5), 989–1008. <https://doi.org/10.12973/eurasia.2015.1369a>
- Hambleton, R. K., Swaminathan, H., & Rogers, H. H. (1992). Fundamentals of item response theory. In *Contemporary Sociology* (Vol. 21, Issue 2). Sage Publications. <https://doi.org/10.2307/2075521>
- Hidayat, C. M. P., & Suyono. (2016). Meremediasi siswa yang memiliki beban miskonsepsi tinggi pada ikatan kimia dan persepsi rendah menggunakan strategi analogi (students who have loads remediate misconceptions high on chemical bonding and perception of low strategy using analogy). *Unesa Journal of Chemical Education*, 5(3), 596–605. <https://doi.org/https://doi.org/10.26740/ujced.v5n3.p%25p>
- Im, S. H., & Jitendra, A. K. (2020). Analysis of proportional reasoning and misconceptions among students with mathematical learning disabilities. *Journal of Mathematical Behavior*, 57(November 2019). <https://doi.org/10.1016/j.jmathb.2019.100753>
- Indrawati, D., Cahyanti, D. N., Surabaya, U. N., Matematika, P., & Gerlach, M. (2018). Alternatif pembelajaran matematika menggunakan media papan arsir pecahan. *Jurnal Inventa*, II(2).
- Isgiyanto, A. (2013). Perbandingan penyekoran model rasch dan model partial credit pada matematika. *Jurnal Kependidikan*, 43(1), 9–18.
- Masami Isoda (2013). *Fraction for Teachers: Knowing What before Planning How to Teach*. Tokyo: CRICED, University of Tsukuba
- Istiyono, Edi. (2018). *Pengembangan instrumen penilaian dan analisis hasil belajar fisika dengan Teori Tes Klasik dan Modern*. Yogyakarta: UNY Press.
- Jarrah, A. M., Wardat, Y., & Gningue, S. (2022). Misconception on addition and subtraction of fractions in seventh-grade middle school students. *Eurasia Journal of Mathematics, Science and Technology Education*, 18(6). <https://doi.org/10.29333/ejmste/12070>
- Jefi, Muflihah, & Nurhadi, M. (2020). Penurunan miskonsepsi siswa melalui pembelajaran remedial dengan menggunakan model pembelajaran induktif pada materi reaksi reduksi dan oksidasi. *Bivalent: Chemical Studies Journal*, 3(1), 4–8. <https://doi.org/10.30872/bcsj.v3i1.321>
- Kamii, C., Lewis, B. A., & Kirkland, L. D. (2001). Fluency in subtraction compared with addition. *Journal of Mathematical Behavior*, 20(1), 33–42. [https://doi.org/10.1016/S0732-3123\(01\)00060-8](https://doi.org/10.1016/S0732-3123(01)00060-8)
- Kamilah, D. S., & Suwarna, I. P. (2019). Pengembangan three-tier test digital untuk mengidentifikasi miskonsepsi pada konsep fluida statis. *Edusains*, 8(2), 212–220. <https://doi.org/10.15408/es.v8i2.5192>
- Kamis, A., Rahim, M. B., Hussein, T., & Suhairom, N. (2018). Construct Validation for an Instrument to Measure Clothing Fashion Design Competency Knowledge (CFaDCK) using Rasch Model. *Journal of Asian Vocational Education and Training*, 11, 13–21.
- Kanwal, S., & Farooq, M. S. (2021). Development of Diagnostic Assessment Test for Misconceptions (ATM) in Mathematics at Elementary School Level. *Global Educational Studies Review*, VI(I), 94–102. [https://doi.org/10.31703/gesr.2021\(vi-i\).10](https://doi.org/10.31703/gesr.2021(vi-i).10)
- Keppmendikbud. (2020). Keppmendikbud Nomor 719/P/2020 tentang pedoman pelaksanaan kurikulum pada satuan pendidikan dalam kondisi khusus. *Www.Kemdikbud.Go.Id*, 022651, 9. <https://www.kemdikbud.go.id/main/blog/2020/08/kemendikbud-terbitkan-kurikulum-darurat-pada-satuan-pendidikan-dalam-kondisi-khusus>
- Kirbulut, Z. D., & Geban, O. (2014). Using three-tier diagnostic test to assess students' misconceptions of states of matter. *Eurasia Journal of Mathematics, Science and Technology Education*, 10(5), 509–521. <https://doi.org/10.12973/eurasia.2014.1128a>
- Kor, L.-K., Teoh, S.-H., Binti Mohamed, S. S. E., & Singh, P. (2018). Learning to make sense of fractions: some insights from the Malaysian primary 4 pupils. *International Electronic Journal of Mathematics Education*, 14(1), 169–182. <https://doi.org/10.29333/iejme/3985>
- Kshetree, M. P., Acharya, B. R., & Khanal, B. (2021). Eight grade students misconceptions and errors in mathematics learning in Nepal. *European Journal of Educational Research*, 10(3), 1199–1213. https://www.researchgate.net/profile/Ebrun-Eren/publication/348382981_Education_Policies_in_the_Context_of_Political_Communication_in_Turkey/links/5ffc2aeba6fdccdc846cc03/Education-Policies-in-the-Context-of-Political-Communication-in-Turkey.pdf
- Lai, M. Y., & Wong, J. P. (2017). Revisiting decimal misconceptions from a new perspective: The significance of whole number bias in the Chinese culture. *Journal of Mathematical Behavior*, 47(January), 96–108. <https://doi.org/10.1016/j.jmathb.2017.07.006>
- Lemonidis, C., & Pilianidis, N. (2020). The 8th grade students' competencies in alternating different symbolic representations of rational numbers. *International Electronic Journal of Mathematics*

- Education, 15(3). <https://doi.org/10.29333/iejme/7865>
- Liu, J., & Jacobson, E. (2022). Examining US elementary students' strategies for comparing fractions after the adoption of the common core state standards for mathematics. *Journal of Mathematical Behavior*, 67(May), 1–20. <https://doi.org/10.1016/j.jmathb.2022.100985>
- Lortie-Forgues, H., Tian, J., & Siegler, R. S. (2015). Why is learning fraction and decimal arithmetic so difficult? *Developmental Review*, 38, 201–221. <https://doi.org/10.1016/j.dr.2015.07.008>
- Maison, M., Lestari, N., & Widaningtyas, A. (2020). Identifikasi miskonsepsi siswa pada materi usaha dan energi. *Jurnal Penelitian Pendidikan IPA*, 6(1), 32–39. <https://doi.org/10.29303/jppipa.v6i1.314>
- Makhubele, Y. E. (2021). The analysis of grade 8 fractions errors displayed by learners due to deficient mastery of prerequisite concepts. *International Electronic Journal of Mathematics Education*, 16(3). <https://doi.org/10.29333/iejme/11004>
- Makonye, J. P., & Fakude, J. (2016). A study of errors and misconceptions in the learning of addition and subtraction of directed numbers in grade 8. *SAGE Open*, 6(4). <https://doi.org/10.1177/2158244016671375>
- Malone, A. S., Fuchs, L. S., Sterba, S. K., Fuchs, D., & Foreman-Murray, L. (2019). Does an integrated focus on fractions and decimals improve at-risk students' rational number magnitude performance? *Contemporary Educational Psychology*, 59(June), 101782. <https://doi.org/10.1016/j.cedpsych.2019.101782>
- Mardapi, Djemari.(2017). *Pengukuran penilaian dan evaluasi pendidikan*. Yogyakarta: Nuha Medika
- McAfee, M., & Hoffman, B. (2021). The Morass of Misconceptions: How Unjustified Beliefs Influence Pedagogy and Learning. *International Journal for the Scholarship of Teaching and Learning*, 15(1). <https://doi.org/10.20429/ijstl.2021.150104>
- Merliza, P., Ralmugiz, U., & Waritsman, A. (2020). Role Of Scaffolding Toward Enhancing Understanding Of Low-Achieving Students (LAS) In Mathematics Learning. *Jurnal International Conference On Research*, 30(May), 16–17.
- Milenković, D. D., Hrin, T. N., Segedinac, M. D., & Horvat, S. (2016). Development of a three-tier test as a valid diagnostic tool for identification of misconceptions related to carbohydrates. *Journal of Chemical Education*, 93(9), 1514–1520. <https://doi.org/10.1021/acs.jchemed.6b00261>
- Mohamed, R., Ghazali, M., & Samsudin, M. A. (2021). A systematic review on teaching fraction for understanding through representation on Web of Science database using PRISMA. *Lumat*, 9(1), 100–125. <https://doi.org/10.31129/LUMAT.9.1.1449>
- Moosapoor, M. (2023). New teachers' awareness of mathematical misconceptions in elementary students and their solution provision capabilities. *Education Research International*, 2023. <https://doi.org/10.1155/2023/4475027>
- Morris, E., Bossé, M. J., & Sevier, J. (2022). College developmental mathematics students' understanding of fraction operations. *International Journal for Mathematics Teaching And Learning*, 23(2), 22–39.
- Mulyasari, E., Ali, M., Sukirman, D., & Hamzah, R. M. (2024). Distribution map analysis of independent living skills through rasch model in primary school students. *Jurnal Prima Edukasia*, 12(1), 165–176.
- Namkung, J. M., Fuchs, L. S., & Koziol, N. (2018). Does Initial Learning about the Meaning of Fractions Present Similar Challenges for Students with and without Adequate Whole-Number Skill? *Physiology & Behavior*, 176(5), 139–148. <https://doi.org/10.4049/jimmunol.1801473>
- Nurharyanto, D. W., & Retnawati, H. (2020). The difficulties of the elementary school students in solving the mathematical narrative test items. *Jurnal Prima Edukasia*, 8(1), 29–39. <https://doi.org/10.21831/jpe.v8i1.29969>
- Obersteiner, A., Alibali, M. W., & Marupudi, V. (2020). Complex fraction comparisons and the natural number bias: The role of benchmarks. *Learning and Instruction*, 67(March), 101307. <https://doi.org/10.1016/j.learninstruc.2020.101307>
- Peşman, H., & Eryilmaz, A. (2010). Development of a three-tier test to assess misconceptions about simple electric circuits. *Journal of Educational Research*, 103(3), 208–222. <https://doi.org/10.1080/00220670903383002>
- Planinic, M., Boone, W. J., Susac, A., & Ivanjek, L. (2019). Rasch analysis in physics education research: Why measurement matters. *Physical Review Physics Education Research*, 15(2), 20111. <https://doi.org/10.1103/PhysRevPhysEducRes.15.020111>
- Prihatni, Y., Supriyoko, S., & Rahmatang, R. (2019). Development of attitude competency assessment test based on teaching of Ki Hadjar Dewantara in elementary and secondary school. *Jurnal Prima Edukasia*, 7(1), 1–8. <https://doi.org/10.21831/jpe.v7i1.21517>
- Purnomo, Y. W., Pasri, Aziz, T. A., Shahrill, M., & Prananto, I. W. (2022). Students' failure to understand fraction multiplication as part of a quantity. *Journal on Mathematics Education*, 13(4), 681–702. <https://doi.org/10.22342/jme.v13i4.pp681-702>
- Purwaningrum, J. P., & Bintoro, H. S. (2018). Miskonsepsi matematika materi bilangan pada mahasiswa calon guru sekolah dasar. *Prosiding Seminar Nasional MIPA 2018, November 2018*, 173–180.
- Retnawati, Heri. (2014). *Teori Respons Butir dan Penerapannya untuk Peneliti, Praktisi Pengukuran dan Pengujian, Mahasiswa Pascasarjana*. Yogyakarta: Parama Publishing
- Retnawati, H. (2017). *Validitas, reliabilitas & karakteristik butir*. Yoyakarta : Parama Publishing.
- Qomario, Q., Tohir, A., & Mashari, A. (2020). The effect of

- realistic mathematical approaches towards the students' math learning outcomes. *Jurnal Prima Edukasia*, 8(1), 78–85. <https://doi.org/10.21831/jpe.v8i1.32577>
- Ratnasari, R. (2018). Students' errors and misconceptions about operations of fractions in an Indonesian Primary School. *Southeast Asian Mathematics Education Journal*, 8(1), 83–98. <https://doi.org/10.46517/seamej.v8i1.66>
- Rismawati, M., & Hutagaol, A. S. R. (2018). Analisis kemampuan pemahaman konsep matematika mahasiswa PGSD STKIP PERSADA khatulistiwa. *Jurnal Pendidikan Dasar PerKhasa*, 4(1), 91–105. <https://doi.org/https://doi.org/10.31932/jpdp.v4i1.17>
- Riyantini, D., Yusmin, E., & Hamdani. (2015). Wawancara Klinis Berbasis Konflik Kognitif Untuk Mengatasi Miskonsepsi Siswa Pada Operasi Pecahan Bentuk Aljabar. *Jurnal Pendidikan Dan Pembelajaran*, 1–11.
- Rodrigues, J., & Thacker, I. (2019). Refuting a fraction misconception: a brief intervention promotes teachers' conceptual change. *Psychology of Mathematics & Education of North America*, 781–785.
- Rozycki, E. G. (2008). Classification error in schooling practice: the impact of the “false positive.” *Educational Horizons*, 86(4), 193–197. <https://www.jstor.org/stable/42923730>
- Safitri, I. M. S. D. P. P. P. E. A. (2018). Analisis kesulitan siswa dalam menyelesaikan operasi hitung pecahan siswa sekolah dasar. *International Journal of Elementary Education*, 2(2), 144–155.
- Saputri, M. E. E. (2019). Analisis miskonsepsi siswa kelas VI SD Negeri Gunung Pasir Jaya pada materi pecahan. *Jurnal Pendidikan Matematika*, 5(2), 61–70. <https://doi.org/http://dx.doi.org/10.23960/mtk/v9i2.pp211-222>
- Sari, A. R. P., & Suyanta, S. (2021). Measurement in Educational Research Critical Thinking Skills dan Science Process Skills. *Measurement in Educational Research*, 1(1), 26–38.
- Shrestha, N. (2021). Factor analysis as a tool for survey analysis. *American Journal of Applied Mathematics and Statistics*, 9(1), 4–11. <https://doi.org/10.12691/ajams-9-1-2>
- Siti Nurhani, A. G., & Siti Mistima, M. (2018). Misconception of fraction among middle grade year four pupils at primary school. *Research on Education and Psychology (REP)*, 2(1), 111–125. <http://journalrep.com>
- Sumintono & Widhiarso. (2015). Aplikasi pemodelan rasch pada assessment pendidikan. Cimahi: Trim Komunikata
- Suwarto. (2017). *Pengembangan tes diagnostik dalam pembelajaran*. Yogyakarta: Pustaka Pelajar
- Takker, S., & Subramaniam, K. (2019). Knowledge demands in teaching decimal numbers. *Journal of Mathematics Teacher Education*, 22(3), 257–280. <https://doi.org/10.1007/s10857-017-9393-z>
- Taslidere, E. (2016). Development and use of a three-tier diagnostic test to assess high school students' misconceptions about the photoelectric effect. *Research in Science and Technological Education*, 34(2), 164–186. <https://doi.org/10.1080/02635143.2015.1124409>
- Theis, R., Junita, R., Kamid, & Iriani, D. (2022). Diagnostic tests based on three level tests on fixed materials for students with misconceptions. *JPI (Jurnal Pendidikan Indonesia)*, 11(4), 638–646. <https://doi.org/10.23887/jpiundiksha.v11i4.50452>
- Turidho, A., Astuti, I. W., Islamirta, K. A., Sitio, S. D. U., Maharani, Y., Darmawijoyo, & Somakim. (2021). The impact of using a single image in a representation on a misconception of fraction concept. *Proceedings of the 1st International Conference on Mathematics and Mathematics Education (ICMMEd 2020)*, 550(Icmmmed 2020), 352–356. <https://doi.org/10.2991/assehr.k.210508.087>
- Ubah, I. J. A. (2021). The impact of different approaches to the teaching of grade 5 fraction by three experienced teachers. *South African Journal of Childhood Education*, 11(1), 1–14. <https://doi.org/10.4102/sajce.v11i1.854>
- Unaenah, E., Fazriandina, A., Damiyah, S. R. Al, Ningsih, Y. A., Alia, F., Berliana, N., & Gunawan, A. N. (2020). Analisis miskonsepsi pengenalan pada pecahan menggunakan media pembelajaran. *EDUCATIF: Journal of Education Research*, 2(2), 49–55.
- Üzel, D. (2018). Investigation of misconceptions and errors about division operation in fractions. *Universal Journal of Educational Research*, 6(11), 2656–2662. <https://doi.org/10.13189/ujer.2018.061131>
- Wells, C. S., & Faulkner-Bond, M. (2016). *Educational Measurement from Foundation to Future*. The Guilford Press.
- Xu, C., Di Lonardo Burr, S., Li, H., Liu, C., & Si, J. (2024). From whole numbers to fractions to word problems: Hierarchical relations in mathematics knowledge for Chinese Grade 6 students. *Journal of Experimental Child Psychology*, 242, 105884. <https://doi.org/10.1016/j.jecp.2024.105884>
- Yanto, E., T, A. Y., & Suratman, D. (2021). Miskonsepsi siswa pada bilangan pecahan dengan metode certainty of response index di kelas VII. *Jurnal Pendidikan Dan Pembelajaran Khatulistiwa*, 10(4), 1–11. <https://jurnal.untan.ac.id/index.php/jpdpb/article/view/File/46219/75676588996>
- Yilmaz, K., Gizem, F., Özdemir, Gökkurt, B., & Yasar, Z. (2018). Using digital stories to reduce misconceptions and mistakes about fractions: an action study. *International Journal of Mathematical Education in Science and Technology*, 49(6), 867–898. <https://doi.org/10.1080/0020739X.2017.1418919>
- Yu, S., Sidney, P., Kim, D., Thompson, C. A., & Opfer, J. E. (2024). From integers to fractions: The role of analogy in transfer and long-term learning. *Journal of Experimental Child Psychology*, 243, 105918. <https://doi.org/10.1016/j.jecp.2024.105918>

Yuliati, Y. (2017). Miskonsepsi siswa pada pembelajaran IPA serta remediasinya. *Jurnal Bio Educatio*, 2(2), 50–58.

Zaleha, Z., Samsudin, A., & Nugraha, M. G. (2017). Pengembangan instrumen tes diagnostik VCCI bentuk

four-tier test pada konsep getaran. *Jurnal Pendidikan Fisika Dan Keilmuan (JPFK)*, 3(1), 36.
<https://doi.org/10.25273/jpfk.v3i1.980>