International Journal of Science and Mathematics Education
https://doi.org/10.1007/s10763-020-10051-3

Sixth Grade Students' Performance, Misconception, and Confidence on a Three-Tier Number Sense Test

Der-Ching Yang[1] · Iwan Andi Jonri Sianturi[1,2]

Received: 25 March 2019 / Accepted: 4 January 2020/Published online: 01 February 2020

Abstract
Numbers and operations is widely considered the most essential domain in elementary school grades. Nevertheless, elementary students in Indonesia have consistently performed unsatisfactorily on this domain in several international mathematics assessments. This study examined 308 Indonesian sixth grade students' current performance(s), misconception(s), and confidence level(s) on a three-tier test containing 40 number sense-related questions. The results showed that most of the students performed unsatisfactorily (indicating low number sense) with a relatively high level of confidence and exhibited considerable misconceptions. The percentage of the correct responses (i.e. answers and reasons) was approximately 46.48% for the 40 questions. Specifically, only 25 students (8.12%) had good number sense, while 60.71% of the students held either strong or moderate misconceptions. The students' strong misconceptions were identified. In addition, other remarkable findings are presented and discussed. Furthermore, the implications of the results and the suggestions for future research are described.

Keywords Confidence rating · Number sense · Performance · Strong misconceptions · Sixth graders

✉ Iwan Andi Jonri Sianturi
iwansianturi75@gmail.com

Der-Ching Yang
dcyang@mail.ncyu.edu.tw

1 Graduate Institute of Mathematics and Science Education, National Chiayi University, 85, Wen Lung, Ming-Hsiung, 621 Chiayi, Taiwan

2 Graduate Institute of Curriculum and Instruction, National Taiwan Normal University, 162, Section 1, Heping E. Rd., 106 Taipei City, Taiwan

Published online: 01 February 2020

Introduction

Numbers and operations is the most notable and essential mathematics domain in elementary school grades to enable students to solve numerical problems in their daily lives (National Council of Teachers of Mathematics, [NCTM] 2000; Şengül & Gülbağcı, 2012; Yang & Sianturi, 2019a, 2019b). National Research Council (2001) asserted that the sufficient understanding of numbers and operations is extremely required to comprehend the concepts of geometry and measurement, algebra, and data analysis and probability, either in elementary schools or in more advanced levels of education. In numbers and operations domain, developing number sense is globally a trend and considered to be the key ingredient and most important mathematical skill in the twenty-first century for the K-12 mathematics education (Devlin, 2017; NCTM, 2000). Therefore, there is a growing amount of attention and research that examine students' number sense and their understanding of numbers and operations (Şengül & Gülbağcı, 2012; Yang & Sianturi, 2019a, 2019b). Number sense is widely regarded as an individual's general understanding of numbers and operations, which enable the individual to demonstrate some useful, flexible, and efficient strategies when solving numbers and operations-related problems (Markovits & Sowder, 1994; McIntosh, Reys, & Reys, 1992; Yang & Sianturi, 2019a, 2019b). An individual with good number sense tends to exhibit "sense-making approach, planning and control, and flexibility and appropriateness sense of reasonableness" when solving problems involving numbers and operations (Mohamed & Johnny, 2010, p. 318).

Many previous studies have revealed that most of elementary students in several countries (e.g. German, Indonesia, Philippines, Turkey, Taiwan, and Hong Kong) had low number sense—detailed in the "Theoretical Background" section. Particularly, elementary students in Indonesia consistently performed poorly on numbers and operations domain in international mathematics assessments, such as the International Mathematics and Science Study [TIMSS] and the Programme for International Student Assessment [PISA] (Mullis, Martin, Foy, & Hooper, 2016; Organisation for Economic Co-operation and Development [OECD], 2010, 2016). In addition, earlier studies (e.g. Herman, 2001; Purnomo, Kowiyah, Alyani, & Assiti, 2014) also reported that elementary students in Indonesia had low performance when solving numbers and operations-related problems and exhibited poor number sense. Nevertheless, the PISA and TIMSS studies and the number sense studies involving Indonesian students did not specifically report students' misconceptions and their confidence levels when solving the test items. To shed the light on these issues, this study examined the current performance, misconception, and confidence level of elementary students in Indonesia when responding to questions involving numbers and operations.

This study used a three-tier number sense test [TTNST] comprising three sequential tests. The first-tier evaluated descriptive knowledge of the students (Caleon & Subramaniam, 2010; Tsai & Chou, 2002). Descriptive knowledge, which is well known as declarative knowledge, can be thought of as "knowledge about" or answers to "WH- questions (What, When, Where, Who, Whom, Which, Whose, Why and How)" using the facts, events, processes, and their relation to each other (Sahdra & Thagard, 2003; ten Berge & van Hezewik, 1999; Yilmaz & Yalcin, 2012). Thus, descriptive knowledge can be regarded as suggestive or real knowledge (Sahdra & Thagard, 2003; Yilmaz & Yalcin, 2012). The real knowledge about numbers and operations acquired by elementary students in Indonesia, which were retrieved by the questions in the TTNST, is imperative to provide information and insights about their understanding of numbers and operations. The second-tier evaluated explanatory knowledge or so-called mental models of the students (Caleon & Subramaniam, 2010; Tsai & Chou, 2002). Mental models are internal constructs that comprise an individual's conceptual systems and are used to reason about the systems (Gogus, 2013). Moseley, Desjean-Perrotta, and Utley (2010) defined mental models as cognitive structures that we use to interpret and explain events in our world; the cognitive structures are developed by new understandings, prior knowledge, existing ideas, and past experiences. Assessing mental models of students is also imperative as mental models are regarded as a potential way to conceptualize student thinking (Lodge-Scharf, 2017). Assessing students' mental models could help us observe the bigger pictures of the application of their mathematical knowledge when solving a mathematical problem (Chinnappan, 1998; Gogus, 2013; Lodge-Scharf, 2017). In this study, mental models represent the reasons used to select an answer in the first-tier (Cheung & Yang, 2020; Yang & Sianturi, 2019a, 2019b). The third-tier examined students' confidence levels about their responses to the first two tiers (Caleon & Subramaniam, 2010; Cheung & Yang, 2020; Pesman & Eryilmaz, 2010; Yang & Sianturi, 2019a, 2019b).

Previous studies have highlighted that the strength of students' conceptual understanding can be examined on the basis of their confidence levels about their answers and reasons when solving a problem (Caleon & Subramaniam, 2010; Cetin-Dindar & Geban, 2011; Cheung & Yang, 2020; Pesman & Eryilmaz, 2010; Yang & Sianturi, 2019a, 2019b). In the field of psychology, confidence rating is essentially used to evaluate the accuracy of an individual's performance (Stankov & Crawford, 1997) and regarded as an "internal, estimated belief" of the individual's accuracy (Renner & Renner, 2001, p. 23). For instance, when solving a problem, individuals "who have low certainty in their answer combinations were possibly guessing; and, therefore had no understanding, or were confused about their understanding" (Odom & Barrow, 2007, p. 97). In mathematics and science education, some studies have used confidence ratings in a three-tier test to examine students' understanding of a topic and their misconceptions related to this topic (e.g. Caleon & Subramaniam, 2010; Cetin-Dindar & Geban, 2011; Cheung & Yang, 2020; Yang & Sianturi, 2019a, 2019b). These studies have shown that a three-tier test (question and answer, reason, and confidence rating) is more effective in assessing students' understanding of a topic rather than conventional multiple choice and/or essay tests because the three-tier test could differentiate misconceptions due to lack of knowledge, personality style, or merely guessing. In this study, the TTNST is beneficial for examining students' understanding of numbers and operations and for identifying their related misconceptions.

This study examined Indonesian sixth grade students' performance and confidence levels and identified their misconceptions when solving number sense-related questions on the basis of their performance and confidence levels (detailed in the "Methods" section). The methodology used for identifying the misconceptions offered in this study is beneficial for and more effective in detecting students' strong misconceptions when dealing with larger groups of students. Specifically, this study addressed the following research questions:

1. How do sixth grade students in Indonesia perform on the TTNST and how confident are they in deciding their responses to the TTNST?
2. How is the distribution of the sixth graders across their number sense and confidence levels based on their responses to the TTNST?
3. What are the strong misconceptions exhibited by the students when solving the test items in the TTNST?

Theoretical Background

Number Sense and Components of Number Sense

Number sense is demonstrated through the ability to handle numerical problems using flexible and efficient strategies (McIntosh et al., 1992; Şengül & Gülbağcı, 2014; Yang & Sianturi, 2019a, 2019b). Nevertheless, it should be noted that no two researchers defined number sense in exactly the same way (Berch, 2005; Gersten, Jordan, & Flojo, 2005). Thus, there are different components of number sense reported in the literature. Berch (2005) made a list of 30 alleged definitions or features of number sense collected from some related studies. Nevertheless, we recognized that there are five crucial components in the list to be the focus of this study [C1–C5]. The five components are pivotal and highlighted in several number sense-related studies (e.g. Berch, 2005; Gersten et al., 2005; Markovits & Sowder, 1994; McIntosh et al., 1992; NCTM, 2000) and in our studies on number sense involving students in Hong Kong and Taiwan (e.g. Cheung & Yang, 2020; Reys & Yang, 1998; Yang, 2019; Yang & Huang, 2004; Yang & Hsu, 2009; Yang, Li, & Lin, 2008; Yang & Lin, 2015; Yang & Sianturi, 2019a, 2019b). Most importantly, the five components are more relevant with the goals of the mathematics curriculum in numbers and operations domain for sixth grade students in Indonesia (Ministry of Education and Culture of Indonesia, 2013). Moreover, we intended to tie our decision on following the five components with the purpose of this study.

Being Able to Understand the Basic Meaning of Numbers and Operations [C-1]. This component overarches a profound understanding of basic meaning of numbers and operations, comprising the ability to make sense of number system (i.e. whole numbers, fractions, decimals, place value, and four basic operations). For example, students should understand that there are many different decimals between 0.27 and 0.28, and should be able to state these decimals (McIntosh et al., 1992; Yang & Sianturi, 2019a, 2019b).

Being Able to Recognize the Size of Numbers [C-2]. This component demonstrates the ability to compare and order some numbers flexibly. For instance, students should be able to compare two fractions without relying on standard written algorithms (Markovits & Sowder, 1994; McIntosh et al., 1992). To exemplify, when comparing $\frac{4}{5}$ and $\frac{3}{7}$, students do not need to find the least common denominator; instead, they should be able to recognize that $\frac{4}{5}$ is greater than $\frac{1}{2}$ and $\frac{3}{7}$ is less than $\frac{1}{2}$. Therefore, $\frac{4}{5}$ is larger than $\frac{3}{7}$.

Being Able to Use Multiple Representations of Numbers and Operations [C-3]. This component demonstrates an ability to use multiple representations (e.g. pictorial representation, symbolic representation, etc.) to solve numerical problems flexibly and requires an in-depth understanding of related concepts involved in the representations (McIntosh et al., 1992; Yang et al., 2008). For instance, when students are asked to determine which of the following alphabet the best to represent $\frac{3}{5} + \frac{1}{7}$ on a number line, without using paper-and-pencil, they should be able to decide the best answer (A, B, C, or D) using an estimation (e.g. (1) the students may think that $\frac{3}{5}$ is larger than $\frac{1}{2}$. Hence, the answer cannot be A; (2) the students may think that $\frac{3}{5} + \frac{1}{7}$ should be less than 1 because $\frac{3}{5} + \frac{1}{5} < 1$ and $\frac{1}{7} < \frac{1}{5}$. Thus, B should be the best answer to represent $\frac{3}{5} + \frac{1}{7}$ on the number line).

Being Able to Recognize the Relative Effect of Operations on Numbers [C-4]. This component indicates an ability to recognize how four basic operations affect the computational results in numbers representations and associates quantities (McIntosh et al., 1992; Yang et al., 2008). For instance, when students were asked to find out the best estimation for 201×0.95 or $201 \div 0.95$, they would not need to rely on written methods and algorithms to obtain the answers; instead, they should be able to decide that 201×0.95 results in a smaller number and $201 \div 0.95$ results in a larger number, indicating that they have sufficient understanding to recognize the relative effect of operations on numbers. In addition, C-4 indicates that the students should be able to grasp the meaning of operations. For instance, multiplication does not always produce a larger number, or division makes a number smaller.

Being Able to Judge the Reasonableness of a Computational Result [C-5]. This component demonstrates an ability to apply mental computation and estimation strategies to solve computational problems without using paper and pencil and to judge the reasonableness of a result (McIntosh et al., 1992). For instance, when students were asked to answer a question, "Please place the decimal point correctly using estimation(s)": $938.5 \times 0.496 = 465.496$; (a) 46.5496, (b) 465.496, (c) 4654.96, (d) Answer could not be found, they do not need to rely on using paper and pencil or recalling mathematical rules; instead, they should be able to think that the product of 900×0.496 (about 1/2) is about 450; therefore, (b) 465.496 is a reasonable answer—this strategy indicates that the students have good number sense.

It should be noted that C-1 is the basic knowledge of the other four components. For example, when comparing $\frac{14}{15}$ and $\frac{15}{16}$, students with good number sense should be able to show that $\frac{15}{16} + \frac{1}{16} = \frac{14}{15} + \frac{1}{15}$ and $\frac{1}{16} < \frac{1}{15}$. Therefore, $\frac{15}{16}$ is larger than $\frac{14}{15}$. In this case, students should know the basic definition of $\frac{1}{16}$ and $\frac{1}{15}$ (i.e. understanding the basic meaning of numbers and operations [C-1]) to decide $\frac{1}{16} < \frac{1}{15}$.

Number Sense-Related Studies

Many studies in different countries have reported that elementary students in several countries had low number sense. In Philippines, Facun and Nool (2012) reported that

47 randomly selected sixth grade students had poor number sense and merely worked on the computational aspect when solving a mathematical problem, without paying attention to the context. Kuhn and Holling (2014) reported that German students in grades 3–4 had unsatisfactory performance on number sense-related problems, with average scores of 51.35 and 54.60 (out of 100) on the given pretest and posttest, respectively. Some studies examining approximately 2000 elementary students in Hong Kong and Taiwan also revealed that most of the students performed poorly on number sense-related problems and hold considerable misconceptions about related concepts in numbers and operations domain (e.g. Cheung & Yang, 2020; Yang & Lin, 2015; Yang & Sianturi, 2019a, 2019b). For instance, students often did not realize that there were infinite decimals between two decimals on a number line; they simply added or subtracted the numerators and the denominators when dealing with fractions, and etc. McIntosh et al.'s (1997) study which involved students from the United States, Australia, Taiwan, and Sweden showed that the students also exhibited low performance when solving number sense-related problems. In Turkey, Şengül and Gülbağcı (2012) examined 125 fifth grade students' performance on number sense test and reported that the students had poor performance; the lowest performance was on questions involving multiple representations.

While number sense has attracted the attention of many researchers worldwide, virtually only a few studies have examined Indonesian students' number sense and those studies only involved students in a regional area of Indonesia. For instance, using a paper-based test, Purnomo et al. (2014) examined 80 sixth graders' performance on three number sense components including (1) knowledge of and facility with numbers, (2) knowledge of and facility with operations, and (3) knowledge of and facility with numbers and operations to computational settings. Their results showed that the students performed poorly on the given test items, in which only 26.35%, 49.75%, and 42.19% of the students had the knowledge of and facility with numbers, knowledge of and facility with operations, and knowledge of and facility with numbers and operations to computational settings, respectively. Helmy, Johar, and Abidin (2018) examined 19 seventh graders' strategy to solve numbers and operations-related questions and reported that most of the students could not determine an appropriate strategy to answer the given questions and no one used number sense strategy.

Misconceptions when Solving Number Sense-Related Problems

Students' misconceptions when solving number sense-related problems typically stem from irrelevant conceptions of fractions, decimals, estimations, and four basic operations (Cheung & Yang, 2020; Widjaja, Stacey, & Steinle, 2011; Yang & Lin, 2015; Yang & Sianturi, 2019a, 2019b). Misconceptions refer to the ideas that are incompatible with currently accepted knowledge and conceptualization, and are regarded as symptomatic of a faulty line of thinking that causes systematic errors (Green, Piel, & Flowers, 2008). Thus, misconceptions indicate students' misunderstanding when they form faulty thinking and display shortcuts that remove the development of mathematics concepts. Helme and Stacey (2000) asserted that the misconceptions on numbers and operations are a long-standing problem held by elementary students across the globe. Previous studies have highlighted that students continually hold various misconceptions when solving problems involving numbers and operations (Merenluoto &

Lehtinen, 2002, 2004; Şengül & Gülbağcı, 2012; Yang & Sianturi, 2019a, 2019b). For instance, the majority of students at the upper secondary level (Merenluoto & Lehtinen, 2002) and more than one-tenth of students majoring in mathematics at the university level (Merenluoto & Lehtinen, 2004) exhibited several misconceptions when solving computational problems. This shows that a misconception can persist over a long period and become entrenched (Eryilmaz, 2002; McNeil & Alibali, 2005).

In Indonesia, students often exhibited several misconceptions when solving number sense-related problems because their strategies in solving the problems were dominated by the tendency to use written algorithms and procedures to arrive at a requested answer (Herman, 2001; Purnomo et al., 2014). For example, students often find out the solution of 27 × 182 or 65 ÷ 12 with written algorithms and often displayed some errors when working with the algorithms (Herman, 2001; Purnomo et al., 2014). In addition, most school textbooks in Indonesia contained a collection of rules and algorithms offered to solve a mathematical problem (Sumarto, van Galen, Zulkardi, & Darmawijoyo, 2014). This partially indicated that most students in Indonesia encountered a mechanistic instruction of stiff procedures of algorithms to solve a mathematical problem.

Methods

Study Participants

A total of 308 sixth grade students (approximately 11–13 years old) of varied socio-economic status from different elementary schools in the city, rural, and small town areas of Indonesia participated in accord with the guidelines outlined by an institutional review board. The test was conducted in Indonesian language [Bahasa Indonesia], which all of the students were fluent in. The sixth grade students were chosen because we intended to examine elementary students' number sense and their misconceptions about numbers and operations after acquiring the teaching and learning of numbers and operations in the first to sixth year of elementary school—before they enroll in a lower secondary school. The students took the test at the end of the school year before joining a national examination required for graduation.

Study Instrument

We designed an online TTNST to examine students' performance (indicating number sense), misconceptions, and confidence ratings when solving number sense-related questions. The TTNST comprised a three-tier sequential test: the first-tier examined content knowledge (i.e. questions and answers), the second examined the reasons used to decide the answers, and the third examined confidence ratings about responses to the first two tiers using a five-point Likert scale (i.e. very confident [5], confident [4], neutral [3], unconfident [2], and very unconfident [1]). The TTNST comprised five pivotal components of number sense and each component contains eight questions: A total of 40 questions. We provided an example of the instrument in data repository at http://osf.io/y6shc for a reference.

The questions and reasons provided in the first two tiers were designed and developed on the basis of related studies (e.g. Cheung & Yang, 2020; Markovits & Sowder, 1994; McIntosh et al., 1997; NCTM, 2000; Reys & Yang, 1998; Şengül & Gülbağcı, 2012; Yang, 2019; Yang & Huang, 2004; Yang & Hsu, 2009; Yang & Lin, 2015; Yang & Sianturi 2019a, 2019b; Yang et al., 2008), which are mostly designed in a context-based situation. The reasons were produced based on the data collected from thousands of students in the past two decades through paper-and-pencil tests, open-ended questions, online two-tier and three-tier tests, and interviews (e.g. Cheung & Yang, 2020; Reys & Yang, 1998; Yang, 2019; Yang & Huang, 2004; Yang & Hsu, 2009; Yang & Lin, 2015; Yang & Sianturi, 2019a, 2019b; Yang et al., 2008). We designed the reasons by analyzing students' responses and their most frequent misconceptions in the data collection (Yang & Lin, 2015). Based on the collected data, we identified that some students found it difficult to organize their reasons in words, thereby leaving the reason section blank. However, by conducting the interviews, we observed that some of them actually had good number sense and could provide acceptable reasons for their answer to the given number sense-related questions after we provided some clues relevant to the reasons. Therefore, we initiated to design and provide the reasons in the TTNST based on the collected data; only the reasons relevant and compatible for each answer in the first-tier were included. Such a design would allow students to focus on fewer but more relevant and compatible reasons instead of posting their own reasons, thus resulting in more meaningful results.

Procedure, Data Collection, and Analysis

The test was administered online and the students used computers to respond to the test questions. The students were provided with test instructions. To answer each question, students were required to select (carefully choose as being the best or most suitable) only an answer, a single reason for the selected answer, and a confidence level for the answer and reason, within 40, 60, and 20 s, respectively. To analyze students' responses to the TTNST, we developed a scoring rule (Table 1) based on earlier studies on a three-

Table 1 Scoring rules applied to students' responses to the TTNST

Number sense three-tier test						
1st stage	Answer	Correct				Incorrect
	Score	4				0
2nd stage	Reason	NS-based	Rule-based	Misconception	Guessing	0
	Score	4	2	1	0	
Total score		8	6	5	4	0
3rd stage	Confidence Rate Index (CRI)					
	CRI	Very confident	Confident	Neutral	Unconfident	Very unconfident
	Score given	5	4	3	2	1

Note. NS-based, number sense-based method; Rule-based, rule-based method; if a student answered correctly with a NS-based method, the student would be given 8 points, and so on; conversely, an incorrect answer was given 0 point, regardless of the reason selected for the incorrect answer tier test (Caleon & Subramaniam, 2010; Cetin-Dindar & Geban, 2011; Cheung & Yang, 2020; Pesman & Eryilmaz, 2010; Yang & Sianturi, 2019a, 2019b).

Using SPSS 21.0, we report the results of descriptive statistics of students' responses to the TTNST to demonstrate their performance and confidence level in each component of number sense. We classified students' number sense and their confidence levels based on a classification used in earlier studies on a three-tier test that examined students' performance (including number sense) and confidence level in the field of mathematics (e.g. Cheung & Yang, 2020; Yang & Sianturi, 2019a, 2019b) and science education (e.g. Caleon & Subramaniam, 2010; Cetin-Dindar & Geban, 2011; Pesman & Eryilmaz, 2010). The students' number sense was divided into four groups: (1) high number sense: mean score ≥ 6 (out of 8); (2) high-medium number sense: $4.8 <$ mean score < 6; (3) medium-low number sense: $3 <$ mean score < 4.8; (4) low number sense: mean score ≤ 3. For the third-tier test (confidence level), the students' responses were divided into three groups: (1) high confidence: mean CRI ≥ 3.3 (out of 5); (2) medium confidence: $2.9 <$ mean CRI < 3.3; and (3) low confidence: mean CRI ≤ 2.9.

Previous studies have established a relationship between students' performance and their confidence level, which demonstrated whether students' had conceptual understanding of a topic, lacked related knowledge, hold misconception, or solely guessed an answer (Caleon & Subramaniam, 2010; Cetin-Dindar & Geban, 2011; Cheung & Yang, 2020; Pesman & Eryilmaz, 2010; Stankov & Crawford, 1997; Yang & Sianturi, 2019a, 2019b). Based on the relationship, we provided an implication which helped this study to elucidate students' knowledge acquisition about numbers and operations: (1) high number sense—high confidence level indicates a profound understanding of related concepts (Caleon & Subramaniam, 2010; Cetin-Dindar & Geban, 2011), (2) low number sense—low confidence level indicates lack of knowledge (Caleon & Subramaniam, 2010; Stankov & Crawford, 1997), (3) low number sense—high confidence level indicates the presence of a misconception (Caleon & Subramaniam, 2010), and (4) high number sense—low confidence level is regarded due to students' personality style (Yang & Sianturi, 2019a, 2019b) or lack of knowledge leading to the need for guessing an answer (Pesman & Eryilmaz, 2010).

We identified students' strong misconceptions based on their incorrect responses and confidence levels. Previous studies have established that a significant misconception is indicated by the emergence of an incorrect response to the first two tiers selected by at least 10% of the students (Caleon & Subramaniam, 2010; Cheung & Yang, 2020; Yang & Sianturi, 2019a, 2019b). In this study, as the first two tiers (four options in each tier) contained 16 pairs of answers and reasons for a question, the probability of a paired answer and reason to be chosen (as a particular response) was 1/16. Thus, an incorrect response to a question over 16.25% ($1/16 \times 100\% + 10\%$ [as suggested by previous studies]) of the total responses to the question was regarded as a significant misconception. A significant misconception with a mean confidence level higher than 3.3 (out of 5) was defined as a strong misconception in this study (Caleon & Subramaniam, 2010; Cheung & Yang, 2020; Yang & Sianturi, 2019a, 2019b).

Reliability and Validity

The Cronbach's α coefficient of the three-tier test for each component of number sense and the whole test were .856, .840, .873, .857, .833, and .902, respectively

(Yang, 2009). The results of the structural equation modeling-based construct reliability of the three-tier test on each component and the whole test were .827, .836, .840, .828, .839, and .905, respectively (Yang, 2019). This showed that the three-tier test has good reliability. A semi-structured interview with 20 students was conducted to determine whether the students could fully understand each question and whether each question was appropriate for sixth grade students. The interview lasted around 45 min to 1 h for each student. The students were mainly requested to figure out whether they could understand the questions (contents clarity). Hence, based on the results of the interviews, we revised several questions to rectify ambiguous contents and make all questions sufficiently clear for sixth grade students.

Furthermore, to ensure that the questions and answers are representative and not beyond the curriculum scope for sixth grade students, two mathematics researchers and three elementary school teachers were invited to review the questions. They unanimously agreed that the questions were effectively designed, *reflected number sense*, and were not beyond the curriculum scope for sixth grade students. Moreover, Yang's (2019) study demonstrated that the five components are valid to measure elementary students' number sense by evaluating construct validity of a three-tier test—the five components were established as the five constructs of the test. This showed that the TTNST has satisfactory reliability and validity.

Results

Sixth Grade Students' Performance on TTNST

Table 2 displays the students' performance and their confidence level in each component of number sense. The results revealed that the students performed unsatisfactorily on the number sense test, in which their average scores ranged from 26.99 to 35.01 (out of 64) in the five components of number sense, indicating that most of the students had poor number sense. In particular, the students exhibited the highest performance on understanding the basic meaning of numbers and operations [$M = 35.01$, SD = 12.74].

Table 2 Students' performance and confidence level on the TTNST ($N = 308$)

Number sense components	Number of questions	Performance on the first two tiers		Confidence level
		M (SD)	Correct %	*M* (SD)
C-1	8	35.01 (12.74)	54.70	4.04 (1.25)
C-2	8	29.32 (15.77)	45.81	4.20 (1.19)
C-3	8	29.54 (15.82)	46.16	*4.03* (*1.26*)
C-4	8	27.87 (12.54)	43.55	4.11 (1.23)
C-5	8	26.99 (13.46)	42.17	4.11 (1.21)
Total score		148.72 (59.03)	46.48	4.10 (1.23)

Note. M, mean; SD, standard deviation; C-1, C-2, C-3, C-4, and C-5: the five components of number sense in the TTNST

Conversely, the students had the lowest performance on judging the reasonableness of computational results [$M = 26.99$, SD = 13.46].

Concerning confidence level, most of the students had a relatively high level of confidence about their responses to the number sense-related questions (Table 2), regardless of whether their responses were correct or not. The responses with the highest confidence level were given to questions that examined the ability to recognize the size of numbers ($M = 4.20$, SD = 1.19). In contrast, the students had the lowest confidence level when responding to questions that examined the ability to use multiple representations of numbers and operations ($M = 4.03$, SD = 1.26).

Based on students' responses to the first two tiers, there was a significant difference of students' performance among the five components of number sense in the TTNST as determined by one-way ANOVA ($F_{(4,303)} = 15.063$, $p = .000$). Nevertheless, a Tukey post hoc test revealed that there was no significant difference between students' performance on C-2 and C-3 ($p = .078$). The ANOVA and Tukey post hoc test results are provided in data repository at http://osf.io/y6shc for a reference. Furthermore, it is crucial to examine the correlation among the five components based on the students' responses (Table 3). The results suggested that there was a strong positive correlation among the five components of number sense, $r_{(306)} = .720$, $p < .05$.

Furthermore, as the questions in the TTNST were designed mostly in a context-based situation, the low performance on the given test provided evidence that the sixth graders had certain difficulties in solving context-based problems. This particular evidence strengthens the arguments in Wijaya, van den Heuvel-Panhuizen, and Doorman's (2015) study in which they asserted that Indonesian students encountered some difficulties to solve context-based problems because they had inadequate experience to engage with the context-based problems. Furthermore, based on students' low performance with a high confidence level, they apparently hold various misconceptions (Caleon & Subramaniam, 2010; Yang & Sianturi, 2019a, 2019b).

Distribution of Students Across Their Performance and Confidence Level

Table 4 displays the distribution of the students across their performance (indicating number sense) and confidence level. The results showed that only 25.33% (8.12% + 17.21%) had either high or medium-high number sense with a high confidence level, indicating that these students had either profound or sufficient

Table 3 Pearson correlation among the five components of number sense

Number sense component	C-1	C-2	C-3	C-4	C-5
C-1	1				
C-2	.723**	1			
C-3	.747**	.709**	1		
C-4	.728**	.729**	.724**	1	
C-5	.682**	.723**	.721**	.718**	1

Note. **. Correlation is significant at .05 level (2-tailed)

Table 4 Distribution of students across the category of number sense and confidence level

Category	High confidence	Medium confidence	Low confidence	Total
High number sense	25 (8.12%)	0 (0.00%)	0 (0.00%)	25 (8.12%)
High-medium number sense	53 (17.21%)	1 (0.32%)	3 (0.97%)	57 (18.51%)
Medium-low number sense	95 (30.84%)	17 (5.52%)	9 (2.92%)	121 (39.29%)
Low number sense	78 (25.32%)	14 (4.55%)	13 (4.22%)	105 (34.09%)
Total	251 (81.49%)	32 (10.39%)	25 (8.12%)	308 (100%)

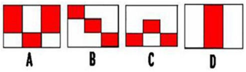
understanding of related concepts in numbers and operations domain. However, 73.38% of the students exhibited either low or medium-low number sense, regardless of their confidence level. This implies that an effective mathematics teaching and learning to develop number sense should be provided for these students. The results also revealed that 56.16% (25.32% + 30.84%) of the students had either low or medium-low number sense with a high confidence level, indicating that these students had strong misconceptions about related concepts in numbers and operations domain. In addition, 14 students (4.55%) had low number sense and medium confidence levels, which indicated that the students had moderate misconceptions. Thus, 60.71% (56.16% + 4.55%) of the sixth grade students had either strong or moderate misconceptions. Moreover, 13 students (4.22%) had low number sense and low confidence, which implies that these students may lack number sense-related knowledge. Furthermore, 3 students (0.97%) had high-medium number sense associated with a low confidence level, which we argued might be because of their personality style or lack of knowledge, leading to the need for guessing a correct answer (Yang & Sianturi, 2019a, 2019b).

Students' Strong Misconceptions

The aforementioned results indicated that the majority of the students (60.71%) hold either strong or moderate misconceptions. Due to the length of the article, we solely present the strong misconception exhibited by the highest number of students in each component (Table 5). In C1–1, the strong misconception indicated that the students were able to perform mathematical computation, but unable to apply relevant concepts to arrive at the requested answer (i.e. the reason for the selected answer was "16 mangoes are divided into 8 sections, so each section has 2 mangoes. Hence, five sections make 10 boxes." However, the correct answer is $\frac{10}{16}$ box). Although they could perform multiplication and division correctly, they apparently lacked conceptual understanding of fraction concepts in using four basic operations. The reason indicated that they failed to notice and understand the properties of the questions, and rushed to determine the answer. In addition, the students lacked adaptive reasoning ability (Alajmi & Reys, 2010) to observe and explain the relationship between the premise and conclusion in the given reason.

In C2–4, the strong misconception displayed that the students could not grasp the number magnitudes when recognizing the size of numbers. The given premise, "The difference between 14 and 15 and the difference between 15 and 16 are both equal to 1"

Table 5 Students' strong misconception in each number sense component

Component	Question	Answer and Reason	N	Mcp	CRI
C1-1	If 16 mangos are packed into one box and each box is equally divided into 8 sections, then 5 sections can be expressed as __ box(es).	Students' answer: 10 boxes Students' reason: 16 mangoes are divided into 8 sections, so each section has 2 mangoes. Hence, five sections make 10 boxes.	78	25.32%	4.16
C2-4	Without using paper and pencil, which fraction is larger: $\frac{14}{15}$ or $\frac{15}{16}$?	Students' answer: $\frac{15}{16}$ Students' reason: The difference between 14 and 15 and the difference between 15 and 16 are both equal to 1. Therefore, the fraction with bigger numbers is larger.	82	26.62%	4.39
C3-8	Which figure does have a shaded area that is not equivalent to $\frac{1}{3}$ of the whole figure? 	Students' answer: A Students' reason: Figure A has 6 pieces and 3 pieces are shaded. Hence, the answer is 1/2.	97	31.49%	4.28
C4-5	Given 103×236 = 24308, which of the following equals to 103×235? (a) 24308-103; (b) 24308-236 (c) 24308÷103; (d) 24308-1	Students' answer: 24308-236 Students' reason: The number 24308 is very large. Thus, it has to be deducted by a large number to be similar to 103×235.	68	22.08%	4.30
C5-8	Fruit A is sold at \$50 per 3 kg, fruit B is sold at \$100 per 7 kg, and fruit C is sold at \$35 per kg. <u>Without using pencil and paper</u>, estimate which fruit is the cheapest?	Students' answer: Fruit C Students' reason: \$35 is the smallest and C is sold in 1 kg. Hence, C is the cheapest fruit.	72	23.38%	4.07

Note. CRI, Confidence Rate Index; N, number of students who held the misconception; Mcp, misconception

showed that the students could perform mental computation, but lacked related conceptions of fractions. The reason, "The fraction with bigger numbers is larger" indicated that the students treated fractions as whole numbers. In addition, they apparently had no capabilities and qualities related to strategic competence that demonstrates flexibility when working with numbers and fractions (Alajmi & Reys, 2010; Yang & Sianturi, 2019b).

In C3–8, the strong misconception showed that the students had inadequate understanding of numbers and operations when working with multiple representations. The reason, "Figure A has 6 pieces and 3 pieces are shaded, so the answer is 1/2" indicated that the students apparently had a conceptual understanding of fraction concepts. However, they hold an irrelevant fraction concept because they ignored that a shaded area of figure A is not similar to the other shaded areas. Accordingly, they apparently had difficulties in making the necessary connection between a visual representation of mathematical concepts and a formal definition of the relevant concepts.

In C4–5, the strong misconception indicated that the students were not able to demonstrate an acceptable estimation using relevant concepts and did not apply any related conceptual knowledge on numbers and operations to arrive at a compatible argument for judging the reasonableness of a computational result. In addition, the students did not exhibit strategic competence (flexibility when working with numbers) to mentally estimate the result of 103×235 based on the given result of $103 \times 236 = 24{,}308$. These students apparently have lacked related knowledge of numbers and operations. For instance, they could not think that $103 \times 235 = 103 \times (236 - 1)$ which would help them to arrive at the requested answer easily. Accordingly, the students could not demonstrate flexible strategies and lacked procedural fluency and multiplication-related knowledge.

In C5–8, the strong misconception indicated that the students were unable to perform productive disposition (i.e. making connections to a context-based situation) (Alajmi & Reys, 2010). In addition, the students failed to estimate the requested answer (without using pencil and paper) because of a hasty decision as they assumed that the lowest collective price and weight of the fruit given in the question would produce the cheapest price. This indicated that most students (23.38%) lacked related knowledge and had insufficient understanding of related concepts on multiplication and division.

Based on the students' strong misconceptions, it is obvious that most of them have lacked conceptual understanding of related concepts in numbers and operations domain. Studies showed that the misconceptions when solving number sense-related problems typically stem from irrelevant conceptions of fractions, decimals, estimations, and four basic operations (Cheung & Yang, 2020; Widjaja et al., 2011; Yang & Lin, 2015; Yang & Sianturi, 2019a, 2019b). If we look at the questions in Table 5, the questions are not related to decimals explicitly because we only present the strong misconceptions exhibited by the highest number of students in each component—due to the length of the article. To clarify this concern, we scrutinized the strong misconceptions exhibited in the 27 out of 40 questions. Findings showed that the students also held strong misconceptions related to decimals. In this study, we solely present the strong misconception held by the highest number of students, which would demand attention in mathematics teaching and learning in the future. The strong misconception is related to misunderstanding the density of decimals (i.e. 3.8 is the only decimal between 3.7 and 3.9—16.88% of total responses and mean confidence level = 4.08). This strong misconception is similar to that in Reys et al.'s (1999) study, in which students often did not realize that there were infinite decimals between two decimals on a number line.

Discussion and Conclusion

This study examined performance, misconception, and confidence level of Indonesian sixth grade students when solving number sense-related questions using a three-tier test. The results showed that most of the students performed unsatisfactorily on the test, indicating that most of them had low number sense. This finding is consistent with earlier studies that examined elementary students' performance in some regional areas of Indonesia (Helmy et al., 2018; Purnomo et al., 2014) and in other countries (Facun & Nool, 2012; Kuhn & Holling, 2014; McIntosh et al., 1997; Şengül & Gülbağcı, 2012; Yang & Sianturi, 2019a, 2019b). The average of the correct percentage of answers and reasons was approximately 46.48% for the 40 questions, which is somewhat better than students' performance on numbers domain reported in PISA and TIMSS studies. It is imperative to highlight that most of the participants (Indonesian sixth grade students) did not have good number sense and exhibited various strong misconceptions. This situation would hinder the learning of more complex mathematical concepts in secondary school. Thus, educational stakeholders in Indonesia should pay more attention to this situation. Teachers' lack of pedagogical knowledge related to the development of students' number sense and an emphasis on algorithmic calculations in mathematics teaching and learning and in mathematics textbook could have generated students' low number sense and the existence of their related misconceptions (Alajmi & Reys, 2010; Herman, 2001; Purnomo et al., 2014).

There was a significant difference of students' performance among the five components of number sense in the TTNST as determined by one-way ANOVA ($F_{(4,303)}$ = 15.063, p = .000). More specifically, a Tukey post hoc test revealed that students' performance was significantly higher when responding to questions in C-1 than in each of the other four components (C2–C5). This is reasonable as the questions in C-1 were relatively easier than those in the other components because most of the questions in C-1 solely examined students' understanding of basic knowledge in numbers and operations domain. In addition, students' performance was significantly higher on C-3 than on C-4 or C-5, and on C-4 than on C-5. This implied that the students had certain difficulties when solving questions in C-5 that examined ability to judge the reasonableness of a computational result. However, there was no significant difference of students' performance when responding to questions that examined their abilities to recognize the size of numbers and to use multiple representations of numbers and operations. We argued that this particular result might occur because the questions in the two components are different in its contexts (C-2 was about recognizing number size, while C-3 was about using multiple representations of numbers and operations). Furthermore, findings demonstrated that the five components had a strong correlation (Table 3), indicating that the five components are highly related to each other reflecting students' number sense. This finding strengthens the result in the study of Yang and Sianturi (2019b) examining number sense of elementary students in Hong Kong, in which the result also showed that there was a positive correlation among the five components. This provides evidence that the five components can be used as a construct for assessing elementary students' number sense.

The students had the lowest performance in judging the reasonableness of a computational result among the five components of number sense. The teaching and learning of mathematics that have disregarded the importance and practicality of judging reasonableness is acknowledged as the most essential factor generating inadequate ability to judge the reasonableness of a computational result (Clarke, Clarke, & Sullivan, 2012; Yang & Sianturi, 2019a). In addition, greatly relying on written algorithms and procedures to solve mathematical problems in mathematics teaching and learning also influences students' ability to judge the reasonableness of a computational result (Alajmi & Reys, 2010). Although earlier studies have shown that mathematics teaching and learning in Indonesia tend to emphasize the use of written algorithms and procedures to solve mathematical problems (e.g. Herman, 2001; Purnomo et al., 2014), the participants of this study exhibited the second lowest performance in recognizing the relative effects of operations on numbers. This partially indicates that the teaching and learning of numbers and operations received by the participants apparently focus on procedural knowledge using written algorithms instead of understanding the meaning of using such procedural knowledge. This situation is not found only in Indonesia but also overseas (e.g. German, Philippines, Turkey, Taiwan, and Hong Kong).

Findings revealed that most of the students had a relatively high level of confidence about the correctness of their answers and reasons ($M = 4.10$ (out of 5), $SD = 1.23$). Apart from using confidence level to predict students' conceptual understanding of numbers and operations (discussed hereafter), it is imperative to notice that most of the participants have a high confidence level about their mathematical ability when responding to the TTNST. Earlier studies revealed that students with high confidence level about their ability when solving a problem tend to demonstrate good performance to handle society's challenging problem in the future (Burton, 2004). The teachers in Indonesia should maintain the confidence level of their students because students who expressed lower confidence (negative appraisals) of their mathematics ability would show lower performance and lack of motivation in mathematics teaching and learning (Burton, 2004).

The results obtained in this study revealed that only 8.12% of the students exhibited good number sense and profound understanding of numbers and operations, while 60.71% of them had either strong or moderate misconceptions. These results are similar to those in previous studies involving students in Hong Kong and Taiwan (Cheung & Yang, 2020; Yang & Sianturi, 2019a, 2019b), in which most of the students exhibited either strong or moderate misconceptions. These students should receive immediate effective instructions to enhance their understanding of related concepts in numbers and operations domain before the misconceptions entranced (Eryilmaz, 2002; McNeil & Alibali, 2005). A teaching instruction that emphasizes the development of students' number sense has been shown effective and efficient to help them recover their misconceptions of numbers and operations, decimals, fractions, and estimations (Helmy et al., 2018; Kuhn & Holling, 2014). Findings showed that the highest number of students who exhibited the strong misconception was found in the component related to the ability to use multiple representations (31.49%). The teachers in Indonesia should be cognizant of this particular result. Earlier studies demonstrated that the use of doing mathematics activities using multiple representations in mathematics teaching and learning could foster students' ability to understand related concepts when solving problems involving multiple presentations and make the connections among the concepts efficiently (Alajmi & Reys, 2010; NCTM, 2000). National Council of Teachers of Mathematics [NCTM] (2000) asserted, "When students gain access to mathematical representations and the ideas they represent, they have a set of tools that significantly expand their capacity to think mathematically" (p. 67).

The existence of strong misconceptions in 27 out of 40 questions indicated that the sixth grade students had held ideas that are incompatible with currently accepted knowledge and conceptualization throughout the elementary school grades. For instance, some of the misconceptions were relevant to the teaching contents that are not provided at the sixth grade level in Indonesia. To exemplify, in comparing $\frac{14}{15}$ or $\frac{15}{16}$, students only considered, "The difference between 14 and 15 and the difference between 15 and 16 are both equal to 1"; then, they decided that $\frac{15}{16}$ is larger than $\frac{14}{15}$. Although the answer is correct, the reason indicated that the students hold a misconception about fractions because they solely considered the difference between the numerator and denominator of each of the two fractions to compare the fractions.

However, students with profound understanding of fractions would be able to think that $\frac{15}{16}+\frac{1}{16}$ is equal to $\frac{14}{15}+\frac{1}{15}$ and $\frac{1}{16}<\frac{1}{15}$. Hence, $\frac{15}{16}$ is larger than $\frac{14}{15}$. The students' misconception indicates lack of conceptual knowledge on fractions, which are usually taught at the third- and fourth-grade level of elementary school in Indonesia. This situation shows the evidence that a misconception can persist over a long period and become entrenched (Eryilmaz, 2002; McNeil & Alibali, 2005).

Findings showed that the students had the lowest mean score in judging the reasonableness of a computational result; the students had difficulties when dealing with estimation-related problems (Table 5). Accordingly, the teachers, not only in Indonesia but also worldwide, may consider designing mathematics teaching and learning that provide more exposure to engage with such problems because the capability to estimate a result allows students to compare their works with prior predictions, and the methods used in solving estimating-related problems could foster a connection between solution and the reasonableness of a result (Bonotto, 2005). Although some students could solve the problems computationally, they encountered difficulty in applying related conceptual knowledge to solve the number sense-related questions, leading them to produce incorrect reasons. Perhaps, this is one of the reasons why NCTM highlighted the necessity of developing number sense in the *Principles and Standards for School Mathematics* (NCTM, 2000).

In conclusion, most of the sixth grade students had low number sense, partly due to the existence of misconceptions and overconfidence, which can be associated with personality style, or lack of knowledge leading to guessing. This situation arose because the students did not understand the concepts that they learned in school or lacked related conceptual knowledge. In addition, they might receive less exposure to mathematical activities that develop their number sense and foster their ability to make connections among related concepts in numbers and operations domain. Moreover, most of the students exhibited difficulties and misconceptions when solving mathematical problems that require an ability to judge the reasonableness of a computational result and to demonstrate an understanding of related concepts in using multiple representations. Furthermore, the students might have had difficulties in solving mathematical problems designed with a context-based situation, indicating that the students had less exposure to such problems.

The major contributions of this study to the field of mathematics education are specifically described in the following directions:

1. The findings of this study supported the results in the earlier studies demonstrating the effective use of TTNST to examine students' number sense, relevant misconceptions, and confidence levels about their answers and reasons.
2. This study provided empirical evidence supporting the earlier studies which revealed that many students in several countries (detailed in the "Theoretical Background" section) did not have good number sense and exhibited several misconceptions. This highly encourages policy makers and teachers to emphasize the development of number sense.
3. Students' confidence about their answers and reasons could be truly considered as a non-cognitive predictor of their conceptual knowledge as well as their understanding of a topic or domain in mathematics.

Implication, Recommendation, and Future Research

The output of this study confirmed that sixth grade Indonesian students have not developed good number sense. The elementary mathematics teachers in Indonesia could reflect on the results of this study, which would be beneficial for improving their teaching strategies and for developing specific instructions that could help students foster their conceptual understanding of numbers and operations and recover their related misconceptions. In particular, the educational stakeholders in Indonesia could use or develop mathematical activities that emphasize more on the development of students' number sense. In addition, it is pivotal that further studies could be conducted to examine and analyze mathematics textbooks that would reveal the development of number sense provided in the textbooks, especially textbooks used in Indonesia.

In the twenty-first century, developing conceptual understanding to enable individuals to solve society's most challenging problems (Karp et al., 2016) is highly emphasized in mathematics education instead of focusing on abstract mathematical concepts (Sarwadi & Shahrill, 2014). Such emphasis should be the priority of mathematics teaching and learning in Indonesia to foster students' number sense and their understanding of numbers and operations. As previous studies have shown that mathematics textbooks used by teachers and students affect students' opportunities to learn and their performance in mathematics (Fan, 2013; Tornroos, 2005; Xin, 2007), educational stakeholders worldwide may consider to select or develop textbooks that emphasize more on judging and recognizing reasonableness and using multiple presentations to solve a particular problem. Correspondingly, further studies may examine and analyze mathematics textbooks concerning judging reasonableness, multiple representation, and context-based problem. Moreover, we highly recommend that conceptual understanding of related topics in numbers and operations domain should be prioritized over procedural knowledge and mathematical rules. Thus, we encourage the deliberate use of context-based problems, estimation questions, multiple presentations, and judging the reasonableness of a computational result to help learners develop their number sense. Ultimately, the results obtained in this study would help educational stakeholders worldwide developing an effective approach to the teaching and learning of numbers and operations and improving the mathematics curriculum and textbooks. Nevertheless, it should be noted that the students were solely requested to select the reasons determined for each answer in the TTNST (detailed in the "Study Instrument" section), leading to unavailability in capturing their other possible reasons. Thus, more studies that could evoke student reasoning other than those we used in the test would be required to reinforce the results of this study. Moreover, there is a need for studies examining students' number sense in Indonesia.

Acknowledgments This work was supported by the Ministry of Science and Technology, Taiwan, with grant no. MOST 105-2511-S-415-003-MY3. Any opinions expressed here are those of the authors and do not necessarily reflect the views of the Ministry of Science and Technology. The authors thank Cheng-Yi Chuang, M.Ed. (National Chiayi University, Taiwan), Muhammad Taqiyuddin, B.Ed. (Universitas Pendidikan Indonesia), and Erna Siregar, B.Ed. (Universitas Pelita Harapan, Indonesia) for their support in handling and collecting the data for the study.

References

Alajmi, A., & Reys, R. (2010). Examining eighth grade Kuwaiti students' recognition and interpretation of reasonable answer. *International Journal of Science and Mathematics Education, 8*(1), 117–139. https://doi.org/10.1007/s10763-009-9165-z.

Berch, D. B. (2005). Making sense of number sense: Implications for children with mathematical disabilities. *Journal of Learning Disabilities, 38*(4), 333–339. https://doi.org/10.1177/00222194050380040901.

Bonotto, C. (2005). How informal out-of-school mathematics can help students make sense of formal in-school mathematics: The case of multiplying by decimal numbers. *Mathematical Thinking & Learning: An International Journal, 7*(4), 313–344. https://doi.org/10.1207/s15327833mtl0704_3.

Burton, L. (2004). Confidence is everything–Perspectives of teachers and students on learning mathematics. *Journal of Mathematics Teacher Education, 7*(4), 357–381. https://doi.org/10.1007/s10857-004-3355-y.

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.

Cheung, K. L., & Yang, D. -C. (2020). Performance of sixth graders in Hong Kong on a number sense three-tier test. *Educational Studies, 46*(1), 39–55. https://doi.org/10.1080/03055698.2018.1516631

Chinnappan, M. (1998). Schemas and mental models in geometry problem solving. *Educational Studies in Mathematics, 36*(3), 201–217. https://doi.org/10.1023/A:1003134323371.

Clarke, D. M., Clarke, D. J., & Sullivan, P. (2012). Reasoning in the Australian curriculum: Understanding its meaning and using the relevant language. *Australian Primary Mathematics Classroom, 17*(3), 28–32.

Devlin, K. (2017). Number sense: The most important mathematical concept in 21st century K-12 education. *HUFFPOST.* From https://www.huffingtonpost.com/entry/number-sense-the-most-important-mathematical-concept_us_58695887e4b068764965c2e0. Accessed 11 Mar 2019.

Eryilmaz, A. (2002). Effects of conceptual assignments and conceptual change discussions on students' misconceptions and achievement regarding force and motion. *Journal of Research in Science Teaching, 39*(10), 1001–1015. https://doi.org/10.1002/tea.10054.

Facun, R. D., & Nool, N. R. (2012, May). *Assessing the number sense of grade 6 pupils*. Paper presented at International Conference on Education and Management Innovation, Singapore. Accessed from http://www.ipedr.com/vol30/58-ICEMI%202012-M10058.pdf

Fan, L. (2013). Textbook research as scientific research: Towards a common ground on issues and methods of research on mathematics textbooks. *ZDM Mathematics Education, 45*(5), 765–777. https://doi.org/10.1007/s11858-013-0530-6.

Gersten, R., Jordan, N. C., & Flojo, J. R. (2005). Early identification and intervention for students with mathematics difficulties. *Journal of Learning Disabilities, 38*(4), 293–304. https://doi.org/10.1177/00222194050380040301.

Gogus, A. (2013). Evaluating mental models in mathematics: A comparison of methods. *Educational Technology Research and Development, 61*(2), 171–195.

Green, M., Piel, J., & Flowers, C. (2008). Reversing education majors' arithmetic misconceptions with short-term instruction using manipulatives. *The Journal of Educational Research, 101*(4), 234–242. https://doi.org/10.3200/JOER.101.4.234-242.

Helme, S., & Stacey, K. (2000). Can minimal support for teachers make a difference to students' understanding of decimals? *Mathematics Teacher Education and Development, 2*, 105–120.

Helmy, N. F., Johar, R., & Abidin, Z. (2018). Student's understanding of numbers through the number sense strategy. *Journal of Physics Conference Series, 1088*, 1–6. https://doi.org/10.1088/1742-6596/1088/1/012098.

Herman. (2001, April). *Strategi mental yang digunakan siswa sekolah dasar dalam berhitung* [Number sense strategies used by elementary students in counting]. Paper presented at National Conference on Mathematics Education, Yogyakarta, Indonesia.

Karp, A., Dubrovsky, V. N., Bulychev, V. A., Garfunkel, S., Ovsyannikova, I., Polikarpov, S. A., . . . Wasserman, N. H. (2016). *Current issues in mathematics education*. Bedford, MA: COMAP.

Kuhn, J.-T., & Holling, H. (2014). Number sense or working memory? The effect of two computer-based trainings on mathematical skills in elementary school. *Advances in Cognitive Psychology, 10*(2), 59–67. https://doi.org/10.5709/acp-0157-2.

Lodge-Scharf, S. E. (2017). *Investigating student mental models at the intersection of mathematics and physical reasoning in physics* (Master's Thesis). University of Maine, MD. From https://digitalcommons.library.umaine.edu/etd/2718.. Accessed 18 Mar 2019

Markovits, Z., & Sowder, J. T. (1994). Developing number sense: An intervention study in grade 7. *Journal for Research in Mathematics Education, 25*(1), 4–29. https://doi.org/10.2307/749290.

McIntosh, A., Reys, B. J., & Reys, R. E. (1992). A proposed framework for examining basic number sense. *For the Learning of Mathematics, 12*, 2–8.

McIntosh, A., Reys, B. J, Reys, R. E, Bana, J. & Farrel, B. (1997). *Number sense in school mathematics: Student performance in four countries*. Perth, Australia: Edith Cowan University.

McNeil, N. M., & Alibali, M. W. (2005). Why won't you change your mind? Knowledge of operational patterns hinders learning and performance on equations. *Child Development, 76*(4), 883–899. https://doi.org/10.1111/j.1467-8624.2005.00884.x.

Merenluoto, K., & Lehtinen, E. (2002). Conceptual change in mathematics: Understanding the real numbers. In M. Limo'n & L. Mason (Eds.), *Reconsidering conceptual change. Issues in theory and practice* (pp. 233–258). Dordrecht, The Netherlands: Kluwer Academic publishers.

Merenluoto, K., & Lehtinen, E. (2004). Number concept and conceptual change: Towards a systemic model of the processes of change. *Learning and Instruction, 14*(5), 519–534. https://doi.org/10.1016/j.learninstruc.2004.06.016.

Ministry of Education and Culture of Indonesia. (2013). *Silabus Sekolah Dasar/Madrasah Ibtidaiyah (SD/MI): Mata pelajaran Matematika* [Elementary school syllabus: Mathematics]. Jakarta, Indonesia: Authors.

Mohamed, M., & Johnny, J. (2010). Investigating number sense among students. *Procedia-Social and Behavioral Sciences, 8*, 317–324. https://doi.org/10.1016/j.sbspro.2010.12.044.

Moseley, C., Desjean-Perrotta, B., & Utley, J. (2010). The draw-an-environment test rubric (DAET-R): Exploring pre-service teachers' mental models of the environment. *Environmental Education Research, 16*(2), 189–208. https://doi.org/10.1080/13504620903548674.

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2016). TIMSS 2015 international results in mathematics. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: http://timssandpirls.bc.edu/timss2015/international-results/. Accessed 18 Mar 2019

National Council of Teachers of Mathematics [NCTM]. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

National Research Council. (2001). Adding it up: Helping children learn mathematics. In J. Kilpatrick, J. Swafford, & B. Findell (Eds.), *Mathematics learning study committee, center for education, division of behavioral and social sciences and education*. Washington, DC: National Academy Press.

Odom, A. L., & Barrow, L. H. (2007). High school biology students' knowledge and certainty about diffusion and osmosis concepts. *School Science and Mathematics, 107*(3), 94–101. https://doi.org/10.1111/j.1949-8594.2007.tb17775.x.

Organisation for Economic Co-operation and Development [OECD]. (2010). *PISA 2009 results: What students know and can do. Student performance in reading, mathematics, and science (Vol. 1)*. Paris, France: Author.

Organisation for Economic Co-operation and Development [OECD]. (2016). PISA 2015 results in focus. From https://www.oecd.org/pisa/pisa-2015-results-in-focus.pdf. Accessed 7 Jan 2019

Pesman, H., & Eryilmaz, A. (2010). Development of a three-tier test to assess misconceptions about simple electric circuits. *The Journal of Educational Research, 103*(3), 208–222. https://doi.org/10.1080/00220670903383002.

Purnomo, Y. W., Kowiyah, K., Alyani, F., & Assiti, S. S. (2014). Assessing number sense performance of Indonesian elementary school students. *International Education Studies, 7*(8), 74–84.

Renner, C. H., & Renner, M. J. (2001). But I thought I knew that: Using confidence estimation as a debasing technique to improve classroom performance. *Applied Cognitive Psychology, 15*, 23–32.

Reys, R. E., & Yang, D. C. (1998). Relationship between computational performance and number sense among sixth- and eighth-grade students in Taiwan. *Journal for Research in Mathematics Education, 29*(2), 225–237.

Reys, R. E., Reys, B. J., McIntosh, A., Emanuelsson, G., Johansson, B., & Yang, D. C. (1999). Assessing number sense of students in Australia, Sweden, Taiwan, and the United States. *School Science and Mathematics, 99*(2), 61–70.

Sahdra, B., & Thagard, P. (2003). Procedural knowledge in molecular biology. *Philosophical Psychology, 16*(4), 477–498.

Sarwadi, H. R. H., & Shahrill, M. (2014). Understanding students' mathematical errors and misconceptions: The case of year 11 repeating students. *Mathematics Education Trends and Research, 2014*, 1–10. https://doi.org/10.5899/2014/metr-00051.

Şengül, S., & Gülbağcı, H. (2012). Evaluation of number sense on the subject of decimal numbers of the secondary stage students in Turkey. *International Online Journal of Educational Sciences, 4*(2), 296–310.

Şengül, S., & Gülbağcı, D. H. (2014). The strategies of mathematics teachers when solving number sense problems. *Turkish Journal of Computer and Mathematics Education (TURCOMAT), 5*(1), 73–88. https://doi.org/10.16949/turcomat.67936.

Stankov, L., & Crawford, J. D. (1997). Self-confidence and performance on test of cognitive abilities. *Intelligence, 25*(2), 93–109.

Sumarto, S. N., van Galen, F., Zulkardi, Z., & Darmawijoyo, D. (2014). Proportional reasoning: How do the 4th graders use their intuitive understanding? *International Education Studies, 7*(1), 69–80. https://doi.org/10.5539/ies.v7n1p69.

ten Berge, T., & van Hezewik, R. (1999). Procedural and declarative knowledge: An evolutionary perspective. *Theory & Psychology, 9*(5), 605–624.

Tornroos, J. (2005). Mathematics textbooks, opportunity to learn and student achievement. *Studies in Educational Evaluation, 31*(4), 315–327. https://doi.org/10.1016/j.stueduc.2005.11.005.

Tsai, C.-C., & Chou, C. (2002). Diagnosing students' alternative conceptions in science. *Journal of Computer Assisted Learning, 18*(2), 157–165. https://doi.org/10.1046/j.0266-4909.2002.00223.x

Widjaja, W., Stacey, K., & Steinle, V. (2011). Locating negative decimals on the number line: Insights into the thinking of pre-service primary teachers. *The Journal of Mathematical Behavior, 30*(1), 80–90.

Wijaya, A., van den Heuvel-Panhuizen, M., & Doorman, M. (2015). Opportunity-to-learn context-based tasks provided by mathematics textbooks. *Educational Studies in Mathematics, 89*(1), 41–65. https://doi.org/10.1007/s10649-015-9595-1.

Xin, Y. P. (2007). Word problem solving tasks in textbooks and their relation to students' performance. *Journal of Educational Research, 100*(6), 347–359. https://doi.org/10.3200/JOER.100.6.347-360.

Yang, D. C. (2019). Development of a three-tier number sense test for fifth-grade students. *Educational Studies in Mathematics, 101*(3), 405–424. https://doi.org/10.1007/s10649-018-9874-8

Yang, D. C., & Huang, F. Y. (2004). Relationships among computational performance, pictorial representation, symbolic representation, and number sense of sixth grade students in Taiwan. *Educational Studies, 30*(4), 373–389.

Yang, D. C., Li, M. N., & Lin, C. I. (2008). A study of the performance of 5th graders in number sense and its relationship to achievement in mathematics. *International Journal of Science and Mathematics Education, 6*(4), 789–807.

Yang, D. C., & Hsu, C. J. (2009). Teaching number sense for 6th graders in Taiwan. *International Electronic Journal of Mathematics Education, 4*(2), 92–109.

Yang, D. C., & Lin, Y. C. (2015). Assessing 10- to 11-year-old children's performance and misconceptions in number sense using a four-tier diagnostic test. *Educational Research, 57*(4), 368–388.

Yang, D. C., & Sianturi, I. A. J. (2019a). Sixth grade students' performance, misconceptions, and confidence when judging the reasonableness of computational results. *International Journal of Science and Mathematics Education, 17*(8), 1519–1540. https://doi.org/10.1007/s10763-018-09941-4

Yang, D. C., & Sianturi, I. A. J. (2019b). Assessing students' conceptual understanding using an online three-tier diagnostic test. *Journal of Computer Assisted Learning, 35*(5), 678–689. https://doi.org/10.1111/jcal.12368

Yilmaz, I., & Yalcin, N. (2012). The relationship of procedural and declarative knowledge of science teacher candidates in Newton's laws of motion to understanding*. *American International Journal of Contemporary Research, 2*(3), 50–56.