

## Validation of Mental Calculation Strategy Questions for Phase A Elementary School Students

Eni Mulyani<sup>1</sup>, Zetra Hainul Putra<sup>2\*</sup>, Jesi Alexander Alim<sup>3</sup>, Yesi Martha Afrillia<sup>4</sup>

Elementary Education Programme, University of Riau, Indonesia<sup>1,2\*,3</sup>

Elementary Education Programme, Yogyakarta State University, Indonesia<sup>4</sup>

eni.mulyani6883@grad.unri.ac.id<sup>1</sup>, zetra.hainul.putra@lecturer.unri.ac.id<sup>2\*</sup>,

jesi.alexander@lecturer.unri.ac.id<sup>3</sup>, yesimartha.2023@student.uny.ac.id<sup>4</sup>

### ABSTRACT

Mental calculation strategy is essential for developing young learners' number sense, flexibility, and efficiency in arithmetic. However, assessment instruments specifically designed to evaluate mental calculation performance in Indonesian primary education remain limited. This study aimed to develop and conduct a preliminary psychometric evaluation of an instrument for assessing second-grade students' performance on mental calculation tasks in addition and subtraction. Using a quantitative approach with a preliminary instrument development and psychometric testing design, the study involved 34 second-grade students from *Witama Nasional Plus* Elementary School, Pekanbaru. The instrument consisted of 30 written items, including context-free and context-based tasks. Data were analyzed using item-total correlation, Cronbach's Alpha, item difficulty, and point-biserial discrimination. The results showed that 20 items were valid and the instrument achieved acceptable internal consistency (Cronbach's Alpha = 0.747). Most items were of moderate difficulty, while the discriminating power ranged from poor to very good. These findings indicate that the instrument has promising initial psychometric quality, although several items require revision before broader validation and application.

**Keyword** : elementary school students, mental calculation strategy, question validation

### ABSTRAK

Strategi berhitung mental penting untuk mengembangkan pemahaman bilangan, keluwesan berpikir, dan kecepatan berhitung siswa sekolah dasar. Namun, instrumen khusus untuk mengukur kemampuan tersebut di Indonesia masih terbatas. Penelitian ini bertujuan mengembangkan dan menguji kualitas psikometrik awal instrumen penilaian berhitung mental siswa kelas II pada materi penjumlahan dan pengurangan. Penelitian ini menggunakan pendekatan kuantitatif dengan desain pengembangan instrumen dan uji psikometrik awal. Subjek penelitian adalah 34 siswa kelas II SD *Witama Nasional Plus*, Pekanbaru. Instrumen terdiri atas 30 soal tertulis, meliputi soal tanpa konteks dan soal berbasis konteks. Data dianalisis menggunakan korelasi item-total, Cronbach's Alpha, tingkat kesukaran, dan daya pembeda point-biserial. Hasil penelitian menunjukkan bahwa 20 soal dinyatakan valid dengan konsistensi internal cukup baik, yaitu Cronbach's Alpha sebesar 0,747. Sebagian besar soal memiliki tingkat kesukaran sedang, sedangkan daya pembeda berada pada kategori kurang hingga sangat baik. Dengan demikian, instrumen ini memiliki kualitas psikometrik awal yang cukup baik, namun beberapa soal masih perlu direvisi.

**Kata Kunci** : siswa sekolah dasar, strategi berhitung mental, validasi soal

## INTRODUCTION

Mental Calculation Strategy is an approach to mathematics learning that emphasises students' ability to solve mathematical problems without using external aids, instead utilising their understanding of the relationships between numbers (Csíkos, 2016; Juric & Pjanic, 2023; Liu et al., 2024; Loja-deleg et al., 2024). This approach is crucial in supporting cognitive flexibility, adaptive thinking skills, and agility in solving more complex mathematical problems. Through this strategy, students are expected to improve their efficiency in mathematical reasoning (Csíkos, 2016; Loja-deleg et al., 2024; Pena-Becerril et al., 2025; Rahayuningsih et al., 2021; Rathgeb-Schnierer & Green, 2019). Previous research also shows that the application of mental calculation strategies can significantly improve number sense, understanding of number structure, and selection of appropriate procedures for solving mathematical problems (Csíkos, 2016; Juric & Pjanic, 2023; Liu et al., 2024). Therefore, this strategy not only strengthens understanding of basic concepts but also improves the speed, accuracy, and stability of students' calculation results (Liu et al., 2024; Loja-deleg et al., 2024).

Despite their proven benefits, mental calculation strategies are not widely implemented in schools. Instruction often prioritizes standard written algorithms, limiting students' flexibility in mental strategies (Csíkos, 2016; Hickendorff & Zanten, 2024; Juric & Pjanic, 2023). Students frequently visualize these algorithms, which can hinder flexible and efficient mental problem-solving. Additionally, many textbooks present only a single universal strategy, further restricting opportunities for flexible thinking (Hickendorff & Zanten, 2024).

On the other hand, international mathematics textbooks often present addition and subtraction material using visual representations, which help students think more flexibly. An example is the use of number lines to support the "jump" and "compensation" strategies in addition and subtraction. This approach allows students to think more flexibly, rather than just imitating the algorithm of writing (Joanna & Laski, 2018; Rogers, 2020). Additionally, the use of spatial linear representations, such as number lines, has been shown to improve accuracy and encourage the use of more advanced calculation strategies, such as counting on and decomposition, particularly in addition and subtraction operations up to 10 (Astodillo, 2024; Joanna & Laski, 2018). One strategy widely applied abroad is making 10 by structuring problems such as  $7 + 4 \rightarrow (7 + 3) + 1$  (Joanna & Laski, 2018; Liu et al., 2024; Mariana & Sasmita, 2024).

In Indonesia, mental calculation strategies are not widely practiced. Both students and teachers are often unfamiliar with these methods, and instruction typically relies on standard algorithms and mechanical procedures. This results in a gap between international theory and local practice, as many teachers and textbooks have yet to recognize the importance of developing students' mental calculation strategies (Mariana & Sasmita, 2024; Prahmana et al., 2020; Pramudiani & Herman, 2021).

Mastery of mental calculation strategies is essential for preparing students for advanced mathematics. However, mathematics instruction in Indonesia remains conventional, focusing on teacher-centered and procedural methods that do not emphasize mental strategies. Studies on Realistic Mathematics Education in Indonesia highlight that lessons often lack exploration of number concepts and operations (Prahmana et al., 2020). Consequently, many students, especially in early grades, are

not accustomed to mental calculation strategies, which may impact their future success in mathematics.

This study addresses this gap by developing and validating an instrument to measure Mental Calculation Strategy in second-grade elementary students. The study focuses on evaluating the instrument's validity, reliability, difficulty, and discriminating power. A valid and reliable instrument will provide a clearer understanding of students' mental calculation abilities and support more effective assessment in Indonesian mathematics education. The findings are expected to contribute to improved teaching quality and classroom practices.

Mental Calculation Strategies are mathematical approaches performed mentally, utilizing an understanding of the relationships between numbers to select efficient procedures. This aims to enhance flexibility and adaptability in mathematical thinking by applying various strategies (DeCaro, 2016; Garcia Coppersmith & Star, 2022). This theory informs the design of measurement instruments in this study, which are tailored to assess the efficiency, accuracy, and flexibility in students' problem-solving. The instruments are developed based on these theoretical principles to evaluate not only the correctness of calculations but also the students' ability to select and apply appropriate strategies in different contexts, fostering adaptive thinking and independent problem-solving in mathematics education (Hopkins et al., 2025).

The characteristics of the Mental Calculation Strategy include various methods used to manipulate numbers. One commonly used technique is number decomposition, both standard (e.g., breaking 28 into  $20 + 8$ ) and non-standard (e.g., 28 into  $3 + 25$  or  $30 - 2$ ) (Hopkins et al., 2025; Rathgeb-Schnierer & Green, 2019). In addition, rounding and compensation strategies are frequently applied, in which numbers are rounded first and then corrected (e.g.,  $39 + 18$  becomes  $40 + 17$ ) (Pjanić et al., 2025). This strategy also involves using known numerical facts and conventional patterns, such as counting from left to right and prioritising larger numbers first. Often, these steps are performed gradually and sometimes involve changing the form of the problem to simplify the calculation, such as using repeated addition as a representation of multiplication (Vásquez et al., 2022). By embedding tasks that require the use of known numerical facts and conventional patterns, such as counting from left to right and prioritizing larger numbers, the instrument measures students' ability to choose the most appropriate strategy in the context of real-time problems, reflecting their flexibility in mental calculations and enhancing their problem-solving skills.

The three main factors influencing the success of the Mental Calculation Strategy are efficiency, accuracy, and flexibility. Efficiency refers to the use of a faster strategy with fewer and simpler steps, although what is considered most efficient may vary between individuals (Garcia Coppersmith & Star, 2022; Hickendorff, 2022). Accuracy depends on the correct execution of strategies and a stable understanding of basic mathematical facts, where flexibility without accuracy is not considered mature procedural flexibility (DeCaro, 2016). Flexibility/adaptivity is measured by two aspects: knowledge of several strategies and the ability to choose the most efficient strategy in various problem contexts (potential flexibility/adaptivity), as well as the spontaneous use of strategies appropriate to the problem context faced (practical flexibility/adaptivity) (Shinno & Mizoguchi, 2021).

The Mental Calculation Strategy not only improves students' ability to perform mental calculations more efficiently but also facilitates the development of adaptive thinking skills and independent problem-solving, and builds a stronger foundation in

mathematics. However, its limited application in learning in Indonesia highlights the need for the development of effective instruments that can both measure and train these mental calculation skills in students. The creation of such instruments is crucial to assessing students' proficiency in applying various mental calculation strategies, promoting adaptive thinking, and enhancing their overall mathematical competence. By integrating these instruments into the educational system, especially in Indonesia, educators will be able to better track students' progress, provide targeted interventions, and foster a deeper understanding of mathematical concepts through practical problem-solving skills. This approach can bridge the gap in the current mathematical education framework and support the holistic development of students' cognitive abilities in mathematics.

Addition and subtraction are essential basic skills for pupils, especially in Phase A of elementary school (Krishnan et al., 2024). Addition is a mathematical operation that combines two or more numbers to form a new number, while subtraction is an operation that separates part of a number to obtain another result. These two operations form the foundation of most other arithmetic and mathematical concepts, which is why they are prioritized in early education (Cowan et al., 2011). In Phase A, the teaching of addition and subtraction is not only aimed at mastering these operations but also at developing students' conceptual understanding of how numbers interact, as well as how the sum relates to its components (Temnikova, 2021). The choice of addition and subtraction as foundational operations stems from their simplicity and ubiquity in daily life, making them ideal for introducing mathematical concepts in an engaging and practical way. This aligns with the thinking of Kullberg et al. (2024), who emphasises the importance of understanding the meanings of 'combining' and 'taking away' before moving on to formal symbols. Without this foundational understanding, students may struggle to grasp more complex mathematical operations in later stages.

In terms of how students solve addition and subtraction problems mentally, Gersten et al. (2009) research shows that children often start by using concrete strategies such as counting objects, grouping items, or visualizing problems in real-world contexts. As they develop, they develop more efficient mental strategies, such as decomposition (for example, breaking 15 into  $10 + 5$  for addition) and compensation (for example, adjusting one number and correcting the result). Mental calculation strategies like these are crucial in helping students solve problems quickly and flexibly, without relying on physical tools like counters or fingers (Chai et al., 2019).

Addition and subtraction were selected in this study because they are the most instructionally appropriate and theoretically informative operations for students in Phase A. First, these operations constitute the central numerical content taught in the early years of primary school and therefore align directly with students' developmental and curricular experiences. Second, compared with more advanced operations, addition and subtraction offer richer opportunities to observe variation in mental strategies, because the same problem can often be solved through multiple approaches. Third, these operations are closely associated with the development of number sense, part-whole understanding, and adaptive expertise, all of which are core elements of mental calculation. Therefore, focusing on addition and subtraction allows the instrument to capture strategic thinking at a stage when students are beginning to internalize number relationships and shift from concrete procedures to more abstract reasoning.

This theoretical position directly informed the construction of the research instrument. The instrument was designed not only to determine whether students could obtain correct answers, but also to represent item situations in which different mental strategies could plausibly be used. In this way, addition and subtraction served as the content domain through which the dimensions of mental calculation strategy particularly accuracy, efficiency, and flexibility could be observed and evaluated. Items were therefore developed to reflect numerical structures that may invite decomposition, compensation, counting on, or bridging through ten, so that students' responses could provide evidence of the strategic quality of their thinking rather than simple answer production alone.

Accordingly, the use of addition and subtraction in this study is closely tied to the formulation of the research problem. Since early mathematics instruction in Indonesia still tends to emphasize standard written procedures, there remains limited attention to how young learners mentally solve basic numerical problems and how such ability can be assessed with valid instruments. This creates a methodological and pedagogical gap: although mental calculation is important for later mathematical learning, there is still insufficient validated measurement for identifying students' strategic competence in basic operations. Therefore, this study focuses on developing and validating an instrument based on addition and subtraction tasks in order to measure mental calculation strategies among second-grade elementary students in a manner that is developmentally appropriate, psychometrically sound, and educationally meaningful.

## **METHOD**

This study employed a quantitative approach using a preliminary instrument development and psychometric testing design. The study was not intended to provide final large-scale validation, but rather to develop an initial set of test items and examine their early psychometric quality for assessing second-grade students' performance on mental calculation tasks. In line with the theoretical framework of this study, the instrument was designed to elicit students' use of mental calculation in addition and subtraction through items that reflect numerical relations, efficiency, accuracy, and flexibility in solving basic arithmetic problems.

The study was conducted at *Witama Nasional Plus* Elementary School in Pekanbaru during the 2024/2025 academic year. The participants were 34 second-grade students who had completed basic instruction in addition and subtraction relevant to the content domain of this study. These students were selected because Phase A learners are at a developmental stage in which number relations, part-whole understanding, and early mental calculation strategies begin to emerge and can therefore be meaningfully examined through structured test items. To maintain confidentiality, only students' answer sheets were recorded and analyzed, and no identifying personal information was included in the dataset.

The study used total sampling within the selected class context, meaning that all students who met the inclusion criteria were involved in the initial field test. This sampling decision was taken because the study focused on a limited school-based trial of an instrument rather than population-level generalization. The sample size of 34 students was considered adequate for a preliminary psychometric trial aimed at identifying the initial functioning of the items, including item validity, internal consistency, difficulty level, and discriminating power (Raykov & Marcoulides,

2011). Accordingly, the findings of this study should be interpreted as evidence from an initial validation stage and not as definitive evidence of final instrument validity across broader populations.

The instrument was developed based on the theoretical framework of Mental Calculation Strategy presented earlier in this study. The development process began by identifying the main dimensions emphasized in the literature, particularly the use of number relationships, decomposition, compensation, counting on, and other flexible approaches that allow students to solve addition and subtraction problems mentally. These conceptual foundations informed the construction of items intended to capture students' performance on tasks that could plausibly invite mental calculation strategies. The content domain was limited to addition and subtraction, because these operations are the most instructionally relevant for Phase A elementary school students and provide rich opportunities for observing variation in mental calculation performance. Based on this rationale, 30 items were constructed to represent numerical situations that vary in structure and level of challenge. The items were then organized into two forms: context-free items, which directly assess basic numerical processing, and context-based items, which embed addition and subtraction in simple real-life situations (see figure 1 and 2).



Figure 1. Example of a Context-Free Question

2. Perhatikan gambar dibawah ini!



Siswa kelas 2 SD Witama akan melakukan perjalanan liburan ke pantai Koneng di Dumai dengan mengendarai bus sekolah. Pada bus pertama ada 34 siswa, pada bus kedua ada 30 siswa. Berapa jumlah seluruh siswa SD Witama yang melakukan perjalanan ke pantai Koneng?

Figure 2. Example of Context-Based Questions

This distinction was intended to examine students' performance both in direct arithmetic tasks and in applied situations that still rely on mental calculation. Because this study was positioned as a preliminary instrument development study, the emphasis of the procedure was on producing an initial item set and examining how well the items functioned empirically in a small-scale school trial.

The instrument consisted of a 30-item written test designed to assess students' performance on mental calculation tasks in addition and subtraction. Each item required students to solve the problem mentally without relying on external aids. In accordance with the theoretical orientation of the study, the items were designed not merely to assess answer correctness, but to represent numerical structures that may encourage the use of mental strategies such as decomposition, compensation, counting on, and bridging through ten.

The instrument included two categories of items. Context-free items focused on basic addition and subtraction computations presented without a story context, whereas context-based items presented simple everyday situations requiring students

to apply arithmetic mentally. This structure was intended to reflect both direct numerical competence and the application of mental calculation in meaningful situations. Each item was scored dichotomously: 1 for a correct answer and 0 for an incorrect answer. Thus, each student could obtain a total score ranging from 0 to 30. In this study, the total score represented the student’s overall performance on the set of mental calculation tasks.

Data were collected through a written test administered in the school setting. Before the test began, students were given brief instructions to ensure that they understood how to complete the questions. The administration was carried out without strict time pressure so that students could focus on solving the items independently. During the test, supervision was maintained to ensure that students worked on the questions by themselves and did not receive assistance from peers, teachers, or external tools. All 30 items were administered in a single session. Students’ written responses were collected immediately after the test for scoring and analysis. The data collection procedure was intended to obtain students’ individual responses to the developed items under regular classroom conditions while minimizing external interference.

Data were analyzed using IBM SPSS Statistics 25, with item analysis focusing on four main aspects of the instrument’s preliminary psychometric quality, as summarized in Table 1.

**Table 1.** The instrument’s preliminary psychometric quality

Analysis Aspect	Statistical Technique	Decision Rule / Classification	Purpose
Item validity	Item-total correlation	Items with a correlation coefficient higher than the r-table value of 0.279 were classified as valid.	To examine whether each item functioned consistently with the overall test.
Internal consistency reliability	Cronbach’s Alpha	A coefficient above 0.70 was interpreted as indicating acceptable reliability for this preliminary stage of instrument testing.	To assess the internal consistency of the instrument.
Item difficulty	Proportion of correct responses	Easy = 0.61–1.00; Moderate = 0.31–0.60; Difficult = 0.00–0.30	To determine the level of difficulty of each item.
Item discrimination	Point-biserial correlation	Very good = 0.71–1.00 Good = 0.41–0.70; Fair = 0.21–0.40; Poor = 0.00–0.20	To determine how well each item distinguished between students with higher and lower total scores.

## RESULTS AND DISCUSSION

The study produced four main findings regarding the instrument's quality: validity, reliability, question difficulty, and discrimination, as summarized in Table 2.

**Table 2.** Results of Mental Calculation Strategy Instrument Testing

1	0.549	0.279	Valid	0.747	0.558	Moderate	0.643	Good
2	0.384	0.279	Valid	0.747	0.500	Moderate	0.357	Fair

3	0.555	0.279	Valid	0.747	0.500	Moderate	0.643	Good
4	0.560	0.279	Valid	0.747	0.559	Moderate	0.500	Good
5	0.651	0.279	Valid	0.747	0.471	Moderate	0.714	Very Good
6	0.474	0.279	Valid	0.747	0.559	Moderate	0.500	Good
7	0.014	0.279	Not Valid	0.747	0.647	Easy	0.143	Poor
8	0.244	0.279	Invalid	0.747	0.706	Easy	0.286	Fair
9	0.598	0.279	Valid	0.747	0.588	Moderate	0.714	Very Good
10	0.420	0.279	Valid	0.747	0.559	Moderate	0.500	Good
11	0.538	0.279	Valid	0.747	0.559	Moderate	0.500	Good
12	0.414	0.279	Valid	0.747	0.588	Moderate	0.429	Good
13	0.399	0.279	Valid	0.747	0.559	Moderate	0.357	Fair
14	0.134	0.279	Not Valid	0.747	0.618	Easy	0.071	Poor
15	0.560	0.279	Valid	0.747	0.559	Moderate	0.786	Very Good
16	0.334	0.279	Valid	0.747	0.559	Moderate	0.357	Fair
17	0.244	0.279	Not Valid	0.747	0.618	Easy	0.357	Fair
18	0.138	0.279	Not Valid	0.747	0.529	Moderate	0.286	Fair
19	0.097	0.279	Not Valid	0.747	0.529	Moderate	0.286	Fair
20	0.437	0.279	Valid	0.747	0.500	Moderate	0.357	Fair
21	0.555	0.279	Valid	0.747	0.500	Moderate	0.643	Good
22	0.560	0.279	Valid	0.747	0.559	Moderate	0.500	Good
23	0.651	0.279	Valid	0.747	0.471	Moderate	0.714	Very Good
24	0.474	0.279	Valid	0.747	0.559	Moderate	0.500	Good
25	0.014	0.279	Not Valid	0.747	0.647	Easy	0.143	Poor
26	0.375	0.279	Valid	0.747	0.529	Moderate	0.429	Good
27	0.064	0.279	Not Valid	0.747	0.676	Easy	0.071	Poor
28	0.449	0.279	Valid	0.747	0.676	Easy	0.357	Fair
29	0.234	0.279	Not Valid	0.747	0.471	Moderate	0	Poor
30	0.240	0.279	Invalid	0.747	0.588	Moderate	0.286	Fair

Validity testing showed that 20 out of 30 items were valid, based on the correlation coefficient (R-count) exceeding the R-table value of 0.279. Items 7, 8, 14, 17, 18, 19, 25, 27, 29, and 30 were invalid and require revision. Examples of valid and invalid questions are shown in Figures 3 and 4.

6.  $70 - 34 =$        7.  $65 - 43 =$

**Figure 3.** Example of a Valid Question

**Figure 4.** Example of an Invalid Item

The question in Figure 5 involves calculating the distance between cities and requires students to understand contextual information and relationships. Although not mathematically complex, it demands attention to detail and concentration, making it moderately difficult. In contrast, the question in Figure 6 is easy, involving a simple purchasing scenario that requires only basic calculations in a familiar context.

Discriminating power, measured by the point-biserial index, showed that 4 items were excellent, 11 were good, and 10 were fairly good. Items 7, 14, 25, and 29 had poor discriminating power and are less effective at distinguishing student abilities. Examples are shown in Figures 5 and 6.



**Figure 5.** Example of an Excellent Question

**Figure 6.** Example of Poor Questions

The question in Figure 5 demonstrates excellent discriminating power. It uses simple numbers to assess students' ability to compare values, effectively distinguishing those with basic number comparison skills from those who struggle with fundamental concepts. In contrast, the question in Figure 8 has poor discriminating power. The use of a pencil image may distract students from the mathematical task, reducing the question's effectiveness in distinguishing mental calculation skills.

This study provides preliminary evidence that the developed instrument has adequate psychometric potential for assessing second-grade students' performance on mental calculation tasks in addition and subtraction, although several items require substantial revision before the instrument can be used more broadly. The finding that 20 out of 30 items met the validity criterion, combined with a Cronbach's Alpha of 0.747, indicates that the instrument already functions with an acceptable level of internal consistency for an early-stage school-based trial. However, these results should not be interpreted as evidence of final instrument quality. Rather, they suggest that the instrument has a sufficiently stable initial structure to support further refinement and larger-scale validation. This point is important because instrument development in mathematics education requires not only statistical acceptability, but also strong alignment between item structure and the targeted cognitive processes (Prasad, 2017)

The overall pattern of results is also consistent with the theoretical framework underlying this study. Mental calculation strategy is not limited to answer production it involves the use of number relations, efficiency, flexibility, and adaptive selection of procedures in response to problem structure (Baranyai et al., 2019; Thurm & Barzel, 2022). From this perspective, the fact that most valid items were in the moderate difficulty range is meaningful. Items of moderate difficulty are more likely to activate strategic thinking because they cannot be solved solely through immediate fact retrieval, yet they are still accessible enough for young learners to attempt mentally. This interpretation supports the argument that an instrument designed around numerical structure, rather than around answer correctness alone, is more likely to capture meaningful variation in mental calculation performance (Månsson, 2022).

The reliability coefficient obtained in this study also deserves careful interpretation. A Cronbach's Alpha of 0.747 indicates acceptable internal consistency for a preliminary trial, suggesting that the retained items are measuring a relatively coherent domain. This finding is broadly in line with previous psychometric studies in educational assessment showing that coefficients above 0.70 may be considered adequate at early stages of instrument testing, particularly when the goal is initial screening and refinement rather than high-stakes decision making (Raykov & Marcoulides, 2011). At the same time, the coefficient is not yet high enough to justify strong claims that the instrument is fully stable across contexts or populations.

The dominance of moderately difficult items is one of the strongest aspects of the instrument. Twenty-three items fell into the moderate category, while only seven were classified as easy and none were difficult. For second-grade students, this

distribution is desirable because an overly easy test tends to compress performance at the upper end, while an overly difficult test may underestimate developing competence. In this study, the predominance of moderate items indicates that the test was generally targeted at an appropriate developmental level for Phase A learners. This finding is consistent with work showing that well-functioning mathematics items for young children should provide enough challenge to differentiate levels of understanding without exceeding the learners' conceptual reach (Marconi et al., 2025). At the same time, the absence of difficult items may indicate that the item pool has not yet fully covered the upper range of emerging mental calculation ability

A more critical issue concerns the invalid items, because their pattern reveals where the instrument still fails to represent the construct effectively. Several invalid items, such as Items 7, 14, 25, and 27, were not only invalid but also easy and had poor discrimination. This combination suggests that these items were too simple to generate meaningful variation among students. When most students can answer an item correctly with minimal cognitive effort, the item no longer functions as a useful indicator of mental calculation performance. Instead, it merely confirms that the task is within basic reach. In theoretical terms, such items are weak because they do not sufficiently activate the strategic dimensions emphasized in the framework, especially flexibility and efficiency. They are therefore more likely to measure basic recall or routine computation than genuine mental calculation quality.

However, not all invalid items failed for the same reason, and this distinction should be made explicit. Items 18, 19, 29, and 30 were not classified as easy, yet they still failed the validity criterion and showed only fair or poor discrimination. This indicates that invalidity in this instrument was not caused solely by low difficulty. In these cases, other factors may have reduced item functioning, such as ambiguous wording, mismatch between the intended numerical structure and students' interpretation, or contextual features that introduced irrelevant processing demands. Item 29 is especially notable because it showed moderate difficulty but zero discrimination, meaning that it did not differentiate stronger and weaker students at all. An item with this profile may contain construct-irrelevant variance or may be interpreted inconsistently across students. This finding is methodologically important because it shows that item revision should not focus only on increasing difficulty, but also on improving clarity, representational relevance, and alignment between item form and the cognitive operation intended to be elicited.

The contrast between stronger and weaker items also supports the study's theoretical rationale for choosing addition and subtraction as the content domain. These operations are not only foundational in early mathematics, but also especially suitable for observing emerging strategic behavior because the same problem can often be solved through multiple mental routes. Prior studies have shown that children's performance in early arithmetic depends strongly on how well they understand number structure and can adapt their procedures to the task at hand (Kullberg et al., 2024; Liu et al., 2024) The present findings extend this argument by showing that not every addition or subtraction item is equally useful for assessment. Items function better when they are structured in ways that can plausibly evoke strategic processing, rather than merely presenting routine computation. Thus, the results support the broader claim that content selection alone is insufficient item design must also reflect the cognitive architecture of mental calculation (Mariana & Sasmita, 2024). In the Indonesian where instruction still tends to emphasize written algorithms and

procedural repetition (Mariana & Sasmita, 2024; Putra et al., 2023), such instruments may help teachers identify whether students are beginning to reason flexibly with numbers rather than merely reproducing taught procedures. Therefore, the contribution of this study is not limited to item analysis it also lies in showing how assessment can support a gradual shift toward more strategy-sensitive mathematics learning in the early grades.

This study also has several limitations that should be acknowledged explicitly. First, the sample was small and drawn from a single school, so the findings cannot yet be generalized to broader populations of elementary students. Second, the study relied on dichotomous written responses, which means that students' actual strategic processes were inferred indirectly rather than observed directly. This limits the extent to which the instrument can claim to measure flexibility or adaptivity in a full sense. Third, the study represents only a preliminary stage of instrument development, and therefore stronger validation procedures, such as expert review, larger field trials, factor-analytic evidence, or think-aloud protocols, remain necessary. Future studies should expand the sample, refine weak items, and incorporate process-oriented evidence so that the instrument can more accurately distinguish between answer accuracy and genuine strategy use.

Overall, the findings indicate that the instrument has promising initial quality, but its value lies as much in the weaknesses it reveals as in the strengths it demonstrates. The valid and moderately difficult items show that young learners' mental calculation performance can be assessed in a structured way through addition and subtraction tasks. At the same time, the invalid and weakly discriminating items show that task simplicity, distracting representations, and insufficient alignment with strategic demands can undermine measurement quality. Therefore, the main contribution of this study is not to claim that a final instrument has been achieved, but to provide an empirically grounded starting point for building a more robust assessment of mental calculation strategy in early elementary mathematics.

## CONCLUSION

This study indicates that the developed instrument shows promising initial psychometric quality for assessing second-grade students' performance on mental calculation tasks in addition and subtraction. The valid items with moderate difficulty suggest that young learners' mental calculation performance can be assessed in a structured manner through carefully designed arithmetic tasks. At the same time, the invalid and weakly discriminating items reveal that overly simple tasks, distracting representations, and insufficient alignment with strategic demands may reduce measurement quality. Therefore, the main contribution of this study is not to claim that a final instrument has been established, but to provide an empirically grounded starting point for developing a more robust assessment of mental calculation strategy in early elementary mathematics.

This study also has several limitations. First, the sample was small and drawn from a single school, so the findings cannot yet be generalized to a broader population of elementary students. Second, the instrument relied on dichotomous written responses, meaning that students' actual strategic processes were inferred indirectly rather than observed directly; this limits the extent to which the instrument can claim to capture flexibility and adaptivity in a comprehensive sense. Third, the present study represents only an initial stage of instrument development. Stronger validation

procedures, such as expert review, larger-scale field testing, factor-analytic evidence, and think-aloud protocols, are still needed. Future studies should therefore expand the sample, revise weak items, and incorporate process-oriented evidence so that the instrument can more accurately distinguish between answer accuracy and genuine strategy use.

## ACKNOWLEDGEMENT

The authors would like to thank KEMDIKTIRISTEK and LPPM Universitas Riau for funding this research under DPPM grant number 102/C3/DT.05.00/PL/2025 (Main Contract) and 19614/UN.19.5.1.3/AL.04/2025 (Derivative Contracts).

## DAFTAR PUSTAKA

- Baranyai, T., Egri, E., Molnár, A., & Zsoldos-Marchiş, I. (2019). Mental Calculation Strategies Used by Pre-Service Primary School Teachers. *Edulearn19 Proceedings*. <https://doi.org/10.21125/edulearn.2019.2167>
- Chai, C. S., Jong, M., Yin, H., Chen, M., & Zhou, W. (2019). Validating and Modelling Teachers' Technological Pedagogical Content Knowledge for Integrative Science, Technology, Engineering and Mathematics Education. *Journal of Educational Technology & Society*, 22(3), 61–73. JSTOR. <https://www.jstor.org/stable/26896710>
- Cowan, R., Donlan, C., Shepherd, D.-L., Cole-Fletcher, R., Saxton, M., & Hurry, J. (2011). Basic calculation proficiency and mathematics achievement in elementary school children. *Journal of Educational Psychology*, 103, 786–803. <https://doi.org/10.1037/a0024556>
- DeCaro, M. S. (2016). Inducing mental set constrains procedural flexibility and conceptual understanding in mathematics. *Memory & Cognition*, 44(7), 1138–1148. <https://doi.org/10.3758/s13421-016-0614-y>
- Garcia Coppersmith, J., & Star, J. R. (2022). A Complicated Relationship: Examining the Relationship Between Flexible Strategy Use and Accuracy. *Journal of Numerical Cognition*, 8(3). <https://doi.org/10.5964/jnc.7601>
- Gersten, R., Beckmann, S., Clarke, B., Foegen, A., Marsh, L., Star, J., & Witzel, B. (2009). *Assisting Students Struggling with Mathematics: Response to Intervention (RTI) for Elementary and Middle Schools*. NCEE 2009-4060. <https://consensus.app/papers/assisting-students-struggling-with-mathematics-response-gersten-beckmann/39526e636bb25999a60ed99c3c89660c/>
- Hickendorff, M. (2022). Flexibility and adaptivity in arithmetic strategy use: What children know and what they show. *Journal of Numerical Cognition*, 8(3), 367–381. <https://doi.org/10.5964/jnc.7277>
- Hopkins, S., Russo, J., & Downton, A. (2025). Assessing children's capacity for procedural flexibility with mental addition. *Mathematics Education Research Journal*, 37(4), 773–795. <https://doi.org/10.1007/s13394-024-00513-0>
- Krishnan, P., Somagutta, V., Simha, K., & Herur, V. (2024). Addition and Subtraction of Numbers Through Visualisation Using the Number Grid Technique Among Primary School Students. *Qeios*. <https://doi.org/10.32388/xf403>
- Kullberg, A., Björklund, C., Kempe, U. R., Brkovic, I., Nord, M., & Maunula, T. (2024). Improvements in learning addition and subtraction when using a structural approach in first grade. *Educational Studies in Mathematics*, 117,

- 399–417. <https://doi.org/10.1007/s10649-024-10339-z>
- Liu, D., Tan, X., Yan, H., & Li, W. (2024). Improving mental arithmetic ability of primary school students with schema teaching method: An experimental study. *Plos One*, *19*. <https://doi.org/10.1371/journal.pone.0297013>
- Månsson, A. (2022). Preservice elementary teachers' mental computation strategy use in subtraction on two[1]digit natural numbers. *Acta Didactica Napocensia*. <https://doi.org/10.24193/adn.15.2.7>
- Marconi, C., De León, D., Luzardo, M., & Maiche, A. (2025). An Early Numeracy Digital Brief Assessment: Parametric and Non-parametric Item Response Theory Models. *International Journal of Educational Methodology*. <https://doi.org/10.12973/ijem.11.2.245>
- Mariana, N., & Sasmita, F. E. (2024). Designed activities on developing mental strategies using alms value-based learning trajectory in elementary school. *Journal on Mathematics Education*. <https://doi.org/10.22342/jme.v15i3.pp859-882>
- Pjanić, K., Jurić, J., & Mišurac, I. (2025). The Use of Different Strategies and Their Impact on Success in Mental Calculation. *Education Sciences*. <https://doi.org/10.3390/educsci15091098>
- Prasad, R. (2017). Trainee Teachers' Feedback And Reflection On Math Lesson Study. *International Journal of Advanced Research*, *5*(12), 237–242. <https://doi.org/10.21474/IJAR01/5959>
- Putra, Z. H., Sari, I. K., & Dahnilyah, D. (2023). Online Formative Assessment in Mathematics Education: Prospective Primary Teachers' Understanding of Rational Numbers. *Revija Za Elementarno Izobraževanje*, *16*(2), 169–188. <https://doi.org/10.18690/rei.16.2.1232>
- Rathgeb-Schnierer, E., & Green, M. (2019). Developing Flexibility in Mental Calculation. *Educação & Realidade*. <https://doi.org/10.1590/2175-623687078>
- Raykov, T., & Marcoulides, G. (2011). Introduction to Psychometric Theory. In *Introduction to Psychometric Theory* (p. 335). <https://doi.org/10.4324/9780203841624>
- Shinno, Y., & Mizoguchi, T. (2021). Theoretical approaches to teachers' lesson designs involving the adaptation of mathematics textbooks: Two cases from kyozai kenkyuu in Japan. *ZDM – Mathematics Education*, *53*(6), 1387–1402. <https://doi.org/10.1007/s11858-021-01269-8>
- Temnikova, M. (2021). Specifics of Teaching The Inverse Relations Between the Arithmetic Operations Addition and Subtraction During the Education in Mathematics in Grade 1. *Proceedings of CBU in Social Sciences*. <https://doi.org/10.12955/pss.v2.242>
- Thurm, D., & Barzel, B. (2022). Teaching mathematics with technology: A multidimensional analysis of teacher beliefs. *Educational Studies in Mathematics*, *109*(1), 41–63. <https://doi.org/10.1007/s10649-021-10072-x>
- Vásquez, C., Piñeiro, J. L., & García-Alonso, I. (2022). What Challenges Does the 21st Century Impose on the Knowledge of Primary School Teachers Who Teach Mathematics? An Analysis from a Latin American Perspective. *Mathematics*, *10*(3), 391. <https://doi.org/10.3390/math10030391>