Jurnal Elemen, 12(3), 962-979, July 2026
https:/doi.org/10.29408/jel.v12i3.34496

e- ISSN: 2442 - 4226

Students' mathematical thinking in solving KenKen puzzles based on Stacey's theory

Raudhatul Khaira *, **Aiyub**

Department of Mathematics Education, Universitas Islam Negeri Ar-Raniry, Aceh, Indonesia

* Correspondence: khairaraudhatul@gmail.com
© The Author(s) 2026

Abstract

Mathematical thinking in non-routine problem solving involves not only obtaining correct answers but also exploring possibilities, connecting information, and justifying decisions. This study examined Grade 8 students' mathematical thinking processes in solving KenKen puzzles using Stacey's framework of specialization, generalization, conjecturing, and convincing. Employing a qualitative descriptive design, eight students were selected based on contrasting performance patterns in an introductory KenKen task. Data were collected through written solutions to four puzzles and retrospective interviews, and then analyzed deductively using the four processes. Findings revealed that these processes were interrelated and varied across students' performance profiles. Higher-performing students consistently coordinated puzzle constraints, evaluated possible combinations, and justified their solutions. Intermediate-performing students demonstrated similar processes but required more revisions, particularly in 4×4 puzzles. Lower-performing students struggled to coordinate permissible numbers, cage conditions, and row–column restrictions. This study provides qualitative evidence of how Stacey's mathematical thinking processes emerge during problem-solving involving multiple simultaneous constraints. The findings highlight KenKen's potential as a structured instructional context for observing students' mathematical reasoning, identifying difficulties, and informing targeted instructional support.

Keywords: game-based learning; KenKen puzzle; mathematical thinking; non-routine problem-solving; Stacey's framework

How to cite: Khaira, R., & Aiyub, A. (2026). Students' mathematical thinking in solving kenken puzzles based on Stacey's theory. *Jurnal Elemen, 12*(3), 962-979. https://doi.org/10.29408/jel.v12i3.34496

Received: 6 April 2026 | Revised: 4 July 2026
Accepted: 31 July 2026 | Published: 8 August 2026

Jurnal Elemen is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

Introduction

Mathematics education is expected to develop students' ability to reason, make connections, and solve problems that cannot be addressed through routine procedures alone. These abilities are particularly important in the context of 21st-century learning, where students are expected to apply mathematical knowledge flexibly rather than only perform calculations (Vural & Vural, 2021). At the junior high school level, students encounter increasingly complex mathematical situations that require them to coordinate numerical information, recognize relationships, and justify their decisions. However, Indonesian students' performance in international assessments continues to indicate difficulties in applying mathematics to problems that require reasoning and problem solving beyond familiar procedures (Nusantara et al., 2025). This suggests the importance of examining not only whether students obtain a correct answer, but also how they reason while working through a mathematical problem.

Mathematical thinking involves more than computational fluency. It includes the ways students explore particular cases, identify relationships, consider possible solutions, and provide reasons for their conclusions. Schoenfeld (1992) described mathematical thinking as an intellectual activity in which students make sense of mathematical situations and regulate their problem-solving activity. Such processes are particularly relevant when students encounter non-routine problems, because a known procedure may not directly determine the next step. Research has shown that students who are accustomed to procedural learning may continue to apply familiar procedures when faced with unfamiliar problems, even when those procedures are not appropriate (Nguyen et al., 2020). Examining students' mathematical thinking during non-routine tasks can therefore provide information that is not visible from final answers alone.

Game-based learning provides one context in which such problem-solving activity can be examined. Mathematical games can engage students in tasks that require exploration, decision making, and repeated checking while also providing a structured set of rules and goals (Čujdíková & Vankúš, 2021). Recent research has reported positive associations between game-based learning and outcomes such as engagement, motivation, creativity, and reflective thinking (Khoo et al., 2025). However, much of this literature has focused on learning outcomes or affective responses. Less attention has been given to the reasoning processes that occur while students are actually solving mathematical game tasks, particularly how students move between considering particular cases, identifying relationships, proposing possibilities, and explaining why a solution is acceptable.

KenKen provides a useful setting for examining these processes because solving a puzzle requires students to coordinate several constraints at the same time. Each cage contains an arithmetic operation and a target value, while the grid imposes restrictions on the numbers that may be placed and prohibits repetition within rows and columns. A student may therefore need to begin with a highly constrained part of the puzzle, use information from already placed numbers to reduce subsequent possibilities, consider alternative number combinations, and check whether a proposed placement satisfies the relevant constraints. When a placement creates a conflict, the student may need to revise the proposed solution and reconsider the

available possibilities. These features make the reasoning process observable through both students' written work and their explanations of how they solved the puzzle.

This characteristic of KenKen is particularly relevant to Stacey's framework of mathematical thinking, which identifies four processes: specialization, generalization, conjecturing, and convincing (Mason et al., 2010). In solving a KenKen puzzle, specialization can be observed when students select a particular or highly constrained case as a starting point; generalization when they use relationships among previous placements and puzzle constraints to determine further possibilities; conjecturing when they propose and test a possible value or combination; and convincing when they provide a mathematical reason for accepting or rejecting a placement. The four processes are not necessarily separate stages. A student may move between proposing a possibility, checking it against constraints, revising it, and explaining the decision during the same problem-solving episode.

Previous research has recognized KenKen as a mathematical puzzle that can support mathematical reasoning and provide opportunities for students to engage with arithmetic and logical relationships (Lipscomb, 2015; Reiter et al., 2011). However, the existing literature does not sufficiently describe how students' mathematical thinking processes are manifested while they solve KenKen puzzles, particularly when the analysis is organized around Stacey's four processes. There is therefore a need for qualitative evidence showing what students actually do and explain when these processes occur, and how the evidence differs among students with different initial performance patterns. In particular, less is known about how these four processes interact within the same KenKen problem-solving episode as students coordinate multiple puzzle constraints. This is especially relevant for junior high school students, for whom the transition from procedural calculation to more flexible mathematical reasoning is an important part of mathematical learning.

The present study addresses this gap by examining Grade 8 students' mathematical thinking processes while solving KenKen puzzles through Stacey's framework of specialization, generalization, conjecturing, and convincing. Rather than evaluating KenKen primarily in terms of achievement or motivation, the study focuses on the reasoning visible in students' written solutions and task-based retrospective interview explanations. The study therefore asks how these four mathematical thinking processes appear in students' problem solving and how their manifestations differ across the initial KenKen performance profiles identified from the introductory task. By focusing on the processes underlying students' solutions, the study aims to provide a more detailed account of mathematical thinking during KenKen problem solving and to clarify the pedagogical potential of KenKen as a context for observing and supporting such thinking.

Methods

This study employed a qualitative descriptive design to examine Grade 8 students' mathematical thinking while solving KenKen puzzles. This design was selected because the study focused on describing participants' written problem-solving actions and retrospective

explanations rather than testing an intervention or establishing causal relationships (Doyle et al., 2009).

The analysis drew on Stacey's four mathematical thinking processes: specialization, generalization, conjecturing, and convincing (Aiyub et al., 2024). These processes were treated as interrelated analytical categories rather than as stages that necessarily occurred in a fixed sequence. The unit of analysis was each participant's problem-solving process as represented in the written task sheets and audio-recorded task-based retrospective interviews. Table 1 presents the operational definitions used in this study.

Table 1. Operational definitions of Stacey's mathematical thinking processes

Indicator	Operational Definition	Evidence examined in KenKen tasks
Specialization	Trying special cases or working from simpler, highly constrained elements of a problem to gather information and establish a starting point.	Selection of a starting cage, prioritized use of single-cell or heavily constrained cages, and explanations of why the solution process began at a specific point.
Generalization	Identifying patterns, underlying relationships, or general rules from observed instances and applying them to deduce further constraints.	Elimination of invalid numbers across rows/columns, reuse of remaining permissible values, and coordination of local cage rules with broader grid constraints.
Conjecture	Formulating a plausible statement, hypothesis, or tentative solution based on available mathematical evidence and testing it against constraints	Proposal of feasible number combinations, trial placements, notation of alternative possibilities, and revisions made after encountering a logical conflict.
Convincing	Providing mathematical reasons to justify why a placement is accepted or rejected	Verbal or written reasoning grounded in cage arithmetic operations, target values, domain of permissible numbers, and row–column non-repetition rules.

The study was conducted during the odd semester of the 2025/2026 academic year in one Grade 8 class at SMP Negeri 8 Banda Aceh, Indonesia. The university issued a research-permission letter, and the school approved the research activities. Students were informed about the follow-up tasks, interviews, and audio recording, and participation was voluntary. The participating students agreed to take part in the interviews and audio recording. Students' identities were protected through participant codes (S-1 to S-8), and the recorded interview data were used only for research purposes.

The class consisted of 30 students. Before participant selection, the first author reviewed basic arithmetic operations and introduced the KenKen rules to the class. These included the permissible number range for each grid, the arithmetic operation and target within each cage, and the prohibition against repeating a number in the same row or column. Students were also advised to begin with an easy or highly constrained cage, such as a single-cell cage.

All 30 students then attempted an introductory 3×3 addition–subtraction KenKen puzzle. The researcher noted differences in correctness, relative independence, errors, revisions, and need for clarification. Students showing contrasting response patterns were invited to take part in the main tasks and interviews, and eight agreed to participate.

For analysis and reporting, the eight participants were retrospectively organized into three initial KenKen performance profiles. S-1, S-2, and S-3 formed the higher initial performance profile because they completed the introductory task correctly and relatively independently. S-4, S-5, and S-6 formed the intermediate profile because they generally understood the rules but made errors or revisions. S-7 and S-8 formed the lower profile because they had difficulty coordinating several KenKen rules and required repeated clarification or correction.

The profiles were based only on response patterns observed during the introductory KenKen activity. They were not derived from standardized mathematics tests, numerical cut-offs, report-card grades, or teacher recommendations and did not represent the participants' general mathematical ability. The teacher was not involved in participant selection or categorization. Codes S-1 to S-8 were used to protect the participants' identities.

The main instrument consisted of four KenKen puzzles obtained from the collection published by the Michigan State University Department of Mathematics (2023). The numerical values and cage structures were used without modification. The task set comprised a 3×3 addition–subtraction puzzle, a 3×3 multiplication–division puzzle, a 4×4 addition–subtraction puzzle, and a 4×4 multiplication–division puzzle. The four tasks were selected because their cage structures and grid constraints required students to coordinate arithmetic conditions with row–column restrictions, thereby providing opportunities for observing the four Stacey processes.

The tasks were administered in the same order to all participants, beginning with the 3×3 puzzles and continuing with the 4×4 puzzles. Each puzzle required participants to coordinate the permissible number range, cage operation, cage target, and row–column non-repetition rule.

No pilot study was conducted before data collection. During manuscript preparation, the second author checked the alignment among the task characteristics, the available student evidence, Stacey's analytical processes, and the reported interpretations.

S-1 to S-6 completed all four tasks. S-7 and S-8 attempted both 3×3 tasks but did not complete the two 4×4 tasks before the allocated classroom time ended. Their incomplete work was retained as evidence of their problem-solving activity.

The primary data sources were the participants' written KenKen task sheets and documented excerpts from individual task-based retrospective interviews. Task-based interviews are used in mathematics education to investigate students' thinking through interaction around one or more mathematical tasks (Johnson et al., 2025; Maher & Sigley, 2020).

Data collection took place in two stages. First, participants worked individually on the four KenKen tasks. The researcher observed their starting points, number placements, questions, erasures, corrections, and revisions. The written sheets preserved visible traces such as crossed-out numbers, rewritten values, alternative combinations, and incomplete placements.

Clarification was provided when participants had difficulty recalling or applying the KenKen rules. For participants in the higher and intermediate profiles, incorrect or nearly completed solutions were usually followed by a general prompt: “There is still an error; please check again.” The researcher did not identify the incorrect cell or provide the correct answer. Most participants located and revised the error after this prompt, although one intermediate-profile participant required more direct guidance. Participants in the lower profile received additional clarification when they had difficulty applying the permissible-number, cage-operation, target, or row–column rules.

After completing or attempting the tasks, each participant took part in an individual retrospective interview based on their written work. Participants were asked to explain how they began solving a puzzle, why they chose particular cages or values, which combinations they considered, how they recognized an invalid placement, and why they revised an answer.

The interviews were conducted after task completion and were therefore treated as retrospective interviews rather than concurrent think-aloud sessions. Observations made during the tasks were used to identify particular actions or revisions for further discussion, but they were not analyzed as a separate source of data. The interview excerpts presented in the Results were drawn from the interview records and corresponding written task sheets. They were translated into English and lightly edited for clarity while preserving the intended meaning.

The written task sheets and interview excerpts were analyzed deductively using the four categories presented in Table 1. The analysis followed the Miles and Huberman model, involving data condensation, data display, and conclusion (Mezmir, 2020).

The data were organized by participant and task. Each task sheet was examined for the participant’s starting point, number placements, possible combinations, erasures, corrections, conflicts, and incomplete work. These written traces were compared with the corresponding interview explanations and coded as specialization, generalization, conjecturing, or convincing.

The coding focused on the reasoning visible in the available evidence rather than on the correctness of the final answer alone. A correct solution was not considered sufficient evidence of convincing unless the participant provided a reason for the placement. A placement was coded as conjecturing only when the written work or interview explanation showed that the participant had proposed and checked a possible value or combination.

The analysis first examined how the four processes appeared across each participant’s available tasks, including differences between the 3×3 and 4×4 puzzles. Participants were then compared across the three initial KenKen performance profiles, with attention to how they coordinated permissible numbers, cage operations and targets, and row–column constraints. Differences among participants within the same profile were retained.

The profiles were used to organize the comparison and were not treated as explanations for participants’ performance. Stacey’s four processes were also not assumed to follow a fixed sequence, as participants could move between proposing, checking, revising, and explaining a placement.

To check the interpretations, the first author compared the interview explanations with the corresponding written work. The second author reviewed whether the available evidence supported the assigned categories and the interpretations reported in the manuscript.

Differences were discussed and the analysis was revised where necessary. Cases that did not follow the broader pattern of a profile were retained.

The university issued a research-permission letter, and the school approved the research activities. Students were informed about the follow-up tasks, interviews, and audio recording, and participation was voluntary. The participating students agreed to take part in the interviews and audio recording. Participant identities were protected using codes (S-1 to S-8), and the recordings were used solely for research purposes. Eight students agreed to participate. Codes S-1 to S-8 were used in the research records and manuscript to protect their identities.

Results

The analysis identified specialization, generalization, conjecturing, and convincing in the participants' solutions to the KenKen tasks. The four processes were not equally evident across participants. Differences were observed in how participants selected starting points, used relationships among constraints, considered possible values, and explained their decisions. The three initial KenKen performance profiles are used below to organize these differences rather than to represent general levels of mathematical ability.

Table 2. Initial KenKen performance profiles

Participant Code	Initial KenKen Performance Profile	Characteristics
S-1	Higher	Completed the task correctly and relatively independently
S-2	Higher	Completed the task correctly and relatively independently
S-3	Higher	Completed the task correctly and relatively independently; recorded possible combinations before placement
S-4	Intermediate	Understood the basic rules but made errors and revisions
S-5	Intermediate	Understood the basic rules but made errors and revisions
S-6	Intermediate	Understood the basic rules but made errors and revisions
S-7	Lower	Had difficulty coordinating several KenKen constraints and required repeated clarification
S-8	Lower	Had difficulty coordinating several KenKen constraints and required repeated clarification

Specialization

In this study, specialization was observed in how participants selected a particular part of the puzzle as a starting point. Because all participants had previously been advised to begin with

an easy or highly constrained cage during the introduction to KenKen, the analysis focused on how they applied this strategy when solving the research tasks.

The higher-profile participants generally began with highly constrained cages. S-1 explained, “I started with the one that only has one box.” S-2 similarly stated, “I chose the easiest one first.” Their written work showed that these initial choices were followed by further placements with fewer visible revisions on the 3×3 tasks. On the 4×4 tasks, they made occasional errors but were generally able to locate and revise them after receiving the general prompt to check their work. The researcher did not identify the incorrect cell or provide the answer. As shown in Figure 1, Sample of KenKen Solutions Produced by Higher-Profile Participants on the 3×3 Grid.

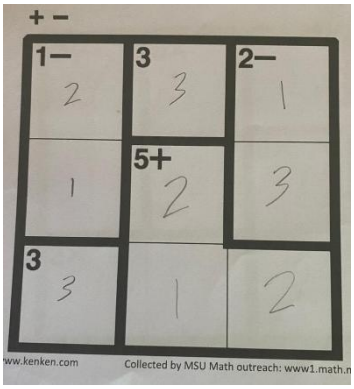


Figure 1. Sample of a KenKen solution produced by a higher-profile participant on the 3×3 grid

The intermediate participants were also able to begin solving the puzzles, although their explanations for the initial choice were less explicit. S-4 stated, ‘I just tried this part first.’ The corresponding written work showed that the participant began with the selected cage before moving to other parts of the grid. This was treated as evidence of specialization because the participant used a particular part of the puzzle as the starting point, although the interview did not provide a clear explanation of why that cage was selected. The written solutions of S-4, S-5, and S-6 contained more revisions than those of the higher-profile participants, particularly on the 4×4 tasks. As shown in Figure 2, the students’ written work contained incorrect number placements and subsequent revisions, alongside less explicit evidence of a clear starting strategy.

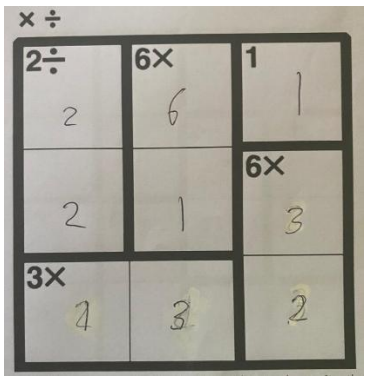


Figure 2. Sample of a KenKen solution produced by an intermediate-profile participant on the 3×3 grid

The lower-profile participants showed greater difficulty deciding where to begin. S-7 stated, "While working on it, I did try many times. I was confused about where to start." S-8 also showed difficulty applying the permissible-number rule together with the row–column restriction, stating, "I thought any number could be used, as long as it wasn't the same in the same row or column." Their written work contained repeated changes and conflicts. One participant also considered a number outside the permissible range of the 3×3 grid. Thus, although the starting strategy had been introduced to all participants, its application was less consistent in the lower profile. As shown in Figure 3, students' work contains frequent errors and repeated revisions, showing repeated trial-and-error attempts and revisions alongside less consistent use of a clear starting strategy.



Figure 3. Sample of a KenKen solution produced by a lower-profile participant on the 3×3 grid

Generalization

Generalization was examined through participants' use of relationships among placed numbers, cage conditions, and row–column restrictions to determine subsequent possibilities. The analysis therefore focused on whether information from one part of the puzzle was used to constrain decisions in another part.

The higher-profile participants frequently used previously established values to restrict what could be placed next. S-1 explained, "I noticed that when I chose the number 2 and then 1, then automatically the last remaining cell in the grid was right for the number 3." Here, the participant used values that had already been determined to identify the remaining permissible value. S-2 similarly considered the row and column before deciding on a placement: "There is no 2 in this row or column yet, and the other numbers are already there, so this cage must be 2, so it doesn't clash." The written work also showed that these participants used existing placements to reduce the possible values for subsequent cells. Similar use of existing placements was also observed in the 4×4 tasks, although occasional incorrect placements were subsequently revised.

The intermediate participants were able to use individual constraints, but their coordination of several constraints was less consistent. S-5, for example, recognized a problem after satisfying the arithmetic condition of a cage: "Oh yes, there is a clash again even though the operation and rules are already correct." This response shows that the participant recognized that satisfying the cage operation alone was not sufficient when another KenKen constraint was

violated. In several instances, however, the conflict was noticed only after a number had already been placed. The participants then erased or revised the placement.

For S-7 and S-8, conflicts were also commonly addressed after they appeared. S-7 stated, "Because there is a clash, I'll change it again later." S-8 similarly said, "When there is a clash, I erase it first, then think again." These responses show awareness that a conflict required revision, but the available evidence showed less consistent coordination of cage conditions, permissible numbers, and row–column restrictions before a placement was made.

Conjecturing

Conjecturing was identified when a participant proposed a possible value or combination and considered it against the available constraints. Tentative placements, alternative combinations, and revisions following a conflict were examined as evidence of this process.

The higher-profile participants generally proposed feasible combinations before making or confirming a placement. S-2 stated, "This must be 3 and 1." The proposed combination was considered in relation to the cage requirement and the surrounding grid. When a placement led to a conflict, participants in this profile were generally able to revise it without the researcher identifying the source of the error.

S-3 showed a somewhat different approach. The participant recorded several possible number combinations before placing values in the grid. This provided more direct written evidence of considering alternatives. However, this strategy was not observed in exactly the same form in S-1 and S-2. It therefore represents a within-profile difference rather than a characteristic shared by all higher-profile participants.

The intermediate participants also generated possible combinations, although their consideration of alternatives was less consistent. S-5 explained, "I tried 3 and 1 first." S-6 gave a more explicit account of another possibility: "I started with that one first (3 + 1), I didn't think about the 2 + 2 also as another option." S-6's explanation shows that one feasible combination was considered first, while another mathematically possible combination was recognized only afterwards. The written work of the intermediate participants similarly contained placements that were later changed when conflicts became apparent.

The lower-profile participants provided less explicit evidence of proposing and checking feasible alternatives. S-7 stated, "I just put a number." S-8 referred to the conflict that followed a placement: "Because there is a clash." In these cases, the available evidence was more closely associated with placing a value and responding to a conflict than with explicitly considering several constraint-based possibilities beforehand. Nevertheless, both participants continued to revise their work and attempt alternative placements when an initial choice did not work.

Convincing

Convincing was identified when participants provided a mathematical reason for accepting a placement or rejecting an alternative. The interview explanations were considered together with the corresponding written work.

The higher-profile participants generally provided clearer reasons based on the arithmetic operation of a cage and the row–column restrictions. S-2 explained: "You mainly look for this first how much is 2 plus 1. For example, say 3 and 1, but 3 is already here, so I look for something else."

The explanation connects the arithmetic requirement with the non-repetition rule. The participant first considered a combination that could satisfy the cage and then rejected it because one of the values was already present in the relevant row or column.

The intermediate participants were also able to explain some of their decisions, although the explanations varied in detail. S-5 said, "You just erase it when there is a clash and then think again." This response shows an understanding that a placement must be reconsidered when it violates a constraint. However, the response explains the corrective action more clearly than the mathematical basis for selecting the next value. The evidence of convincing was therefore present but less explicit in some of the intermediate participants' responses.

The lower-profile participants had greater difficulty articulating the reasons behind their placements. In some interview exchanges, they could describe what they did but were unable to provide a complete mathematical explanation for why a particular value was acceptable. For example, S-7 stated, "If my friend asks why my answer is correct, I tell them to calculate it themselves." This response does not provide a mathematical reason for the participant's answer. Similar instances occurred in which the written work showed attempts to solve the puzzle, while the subsequent interview did not provide sufficient evidence to classify the explanation as a complete justification. The absence of an explicit explanation was therefore treated as limited evidence of convincing, rather than as evidence that the participant had no reasoning.

Differences across tasks and profiles

The 3×3 and 4×4 tasks differed in the demands they placed on participants' problem solving. The higher-profile participants generally completed the 3×3 tasks accurately and relatively independently. On the 4×4 tasks, they occasionally made incorrect placements but were generally able to identify and revise them after receiving the general prompt to check their solutions.

The intermediate participants showed a similar understanding of the basic rules but required more revisions as the tasks became more demanding. These revisions were generally related to conflicts among the cage operation, target value, permissible numbers, and row–column restrictions. In one instance, a participant required more direct clarification to locate the source of an error.

S-7 and S-8 showed the greatest difficulty when several constraints had to be considered together. They required repeated clarification during the task session and continued to revise their placements. Both attempted the 4×4 tasks, but neither completed them before the allocated classroom time ended. Their incomplete work was retained as evidence of their problem-solving activity.

Overall, the four mathematical thinking processes were evident across the participants, but the evidence differed in clarity and consistency. The higher-profile participants more often

used available constraints to guide their next step before making a placement and provided clearer explanations of their decisions. The intermediate participants demonstrated similar processes with more revisions and greater variation across tasks. The lower-profile participants showed greater difficulty coordinating several constraints at the same time and more often identified problems after a conflict had already occurred.

Importantly, these differences were observed within the KenKen problem-solving activities. The initial profiles were based on response patterns in the introductory KenKen task and should therefore not be interpreted as measures of participants' general mathematical ability.

Discussion

This study examined how specialization, generalization, conjecturing, and convincing appeared in Grade 8 students' mathematical thinking while solving KenKen puzzles. The findings do not suggest a simple progression from one process to the next. Rather, the four processes appeared in interaction as students selected starting points, used available constraints, proposed or revised values, and explained their decisions. The main contribution of the study is therefore not simply that different students showed different levels of mathematical thinking, but that KenKen made visible differences in how students coordinated several forms of reasoning within the same problem-solving activity.

The findings concerning specialization show that the starting point selected by a student can shape the subsequent problem-solving process. Participants in the higher profile generally used highly constrained cages as entry points, whereas participants in the lower profile had greater difficulty deciding where to begin. This distinction should be interpreted in light of the instructional context: all participants had previously been introduced to the strategy of beginning with an easy or highly constrained cage. Thus, the finding does not show that higher-profile students independently discovered specialization. Instead, it shows differences in how readily students applied an available strategy when faced with a new puzzle. This extends the discussion of specialization beyond the identification of a "good" starting point. In the KenKen context, specialization functioned as an early decision that reduced the number of possibilities available for later reasoning. This interpretation is consistent with research emphasizing the importance of context-specific strategies in non-routine problem solving (Foster, 2023).

The differences in specialization also help explain why the participants did not simply differ in the amount of trial-and-error they used. The lower-profile participants did engage in repeated attempts, but their difficulty was often visible earlier, when deciding where to start and which constraints to use. This is relevant to Stacey's account of specialization because it suggests that the usefulness of a special case depends not merely on selecting a simpler part of a problem, but on recognizing why that part provides useful information for subsequent steps (Stacey, 2006). Research on weaker problem-solving performance similarly points to difficulties in representing a problem and establishing an effective starting point (Schoenfeld, 1992). In the present study, however, the evidence suggests that students' difficulties were particularly apparent when several KenKen constraints had to be considered together.

Generalization provided a more distinctive view of the differences among participants. The higher-profile participants did not simply identify numerical patterns. They used previously established values to restrict later possibilities and connected information from cages with row and column constraints. In other words, the relevant form of generalization in this setting was relational: information obtained in one part of the puzzle was used to determine what remained possible elsewhere. This finding is important because it places generalization in a concrete problem-solving setting rather than treating it only as the recognition of a broad mathematical rule. It is consistent with accounts of generalization as the identification and extension of relationships from particular cases (Ellis et al., 2022; Hashemi et al., 2013), but the KenKen tasks show how such relationships can operate simultaneously at local and grid levels.

The intermediate participants illustrate why this coordination matters. They could often satisfy the arithmetic requirement of a cage but sometimes recognized a row or column conflict only after placing a number. Their revisions therefore did not necessarily reflect a lack of understanding of the individual rules. Rather, the difficulty was in coordinating rules that had to be considered at the same time. This distinction is important because describing these students simply as "trial-and-error" would overlook the reasoning that was already present in their work. Their responses are more consistent with developing planning and monitoring, as described in research on mathematical problem solving (Schoenfeld, 1992). The lower-profile participants showed this difficulty more strongly: conflicts were often noticed after they occurred, and several constraints had to be clarified repeatedly. The findings therefore suggest that, in KenKen, the development of generalization is closely connected to the ability to coordinate local and global constraints rather than merely to recognize a pattern.

The conjecturing findings further complicate a simple high–medium–low interpretation. Participants in the higher profile generally proposed feasible combinations and checked them against the puzzle constraints. However, S-3 differed from S-1 and S-2 by explicitly recording several possible combinations before placing values in the grid. This within-profile variation is theoretically relevant. It shows that participants who were initially grouped within the same performance profile did not necessarily use the same conjecturing strategy. Thus, the profiles provide a useful way of organizing the data, but they should not be treated as fixed cognitive types.

The intermediate participants also provided evidence of conjecturing, although their consideration of alternatives was less consistent. S-6, for example, initially considered one feasible combination and only later recognized another possible combination. This suggests that conjecturing was present even when the process was not fully developed or explicit. The lower-profile participants more often placed a value and responded to a conflict afterwards, providing less evidence of an explicitly constraint-based conjecture before placement. The distinction is therefore not between students who conjectured and students who did not. Rather, the evidence suggests differences in how explicitly and effectively possible values were generated, tested, and revised. This interpretation is compatible with research describing conjecturing as an iterative process of proposing, testing, and revising mathematical possibilities (Anwar et al., 2024; Komatsu & Jones, 2022). The present study adds that, in

KenKen, this process can be observed directly through the combination of written alternatives, placements, conflicts, and retrospective explanations.

The findings on convincing also require a cautious interpretation. Higher-profile participants more often provided reasons that connected cage arithmetic with row–column restrictions. Their explanations showed that a placement could be accepted not simply because it produced the required target, but because competing possibilities had been ruled out by another constraint. This represents a form of mathematical justification grounded in the structure of the puzzle. It is more appropriate to describe this as constraint-based or informal justification than as proof, since the participants were not constructing formal mathematical arguments.

The intermediate participants also provided reasons for some decisions, but their explanations varied in completeness. In several cases, they could explain that a number had to be changed because it produced a conflict, while providing less detail about why the replacement was valid. This difference suggests that convincing was not independent of the earlier processes. A student may be able to identify a possible value and detect a conflict without being able to articulate a complete reason for the final placement. The lower-profile participants showed the clearest difficulty in articulating such reasons. Importantly, the absence of a complete verbal explanation should not be interpreted as proof that no reasoning occurred. The available evidence only supports the more limited conclusion that the written work and retrospective interviews did not consistently provide sufficient evidence of a complete justification.

Taken together, these findings suggest a relationship among the four processes that is more useful than viewing Stacey's indicators as a sequence of four stages. In the KenKen tasks, a starting point selected through specialization provided information for subsequent deductions; those deductions restricted possible values and supported conjectures; conjectures were then tested against the same constraints; and convincing involved explaining why a placement could be accepted or rejected. A difficulty at one point could therefore affect later reasoning. For example, when a student selected an uninformative starting point, subsequent possibilities could remain broad. When a student considered only one combination, a later conflict could require revision. When a student detected a conflict but could not explain why the replacement was valid, convincing remained limited. This interaction among processes is one of the more important findings of the study because it illustrates how mathematical thinking processes can operate together within a single non-routine task.

The role of task complexity also provides an important qualification to the profile differences. The 3×3 tasks were generally completed more easily, whereas the 4×4 tasks required participants to coordinate more information and resulted in more revisions, particularly among the intermediate and lower profiles. The higher-profile participants were also not error-free on the more demanding tasks; however, they generally identified and corrected errors after a general prompt. The lower-profile participants attempted the 4×4 tasks but did not complete them before the available time ended. These observations suggest that the difficulty of the task influenced how clearly the four processes could be observed. They also caution against

interpreting performance on the introductory 3×3 task as a complete representation of a student's mathematical thinking.

The findings have implications for the use of KenKen in mathematics education. The value of KenKen is not only that it provides an engaging game-based activity. Its structure requires students to coordinate arithmetic operations, permissible values, cage targets, and row–column restrictions within the same problem. This makes students' reasoning processes more visible than a final correct or incorrect answer alone. Research on game-based learning has reported positive effects on cognitive outcomes in mathematics (Anggoro et al., 2025; Yao, 2026), while research on non-routine problem solving emphasizes the importance of exploration, reasoning, and appropriate support (Öztürk et al., 2020; Santos-Trigo, 2024). The present findings add a more specific point: KenKen can provide a context in which teachers can observe not only whether students solve a puzzle, but also where difficulties arise in the coordination of mathematical thinking processes.

The need for support was particularly visible among the lower-profile participants. However, the findings do not suggest that these students simply lacked mathematical reasoning. They continued to attempt placements, respond to conflicts, ask questions, and revise their work after clarification. The more defensible interpretation is that they required more external support to coordinate the constraints independently. This is consistent with the principle underlying the Zone of Proximal Development, in which appropriate guidance can support students in performing tasks that they cannot yet manage independently (Utomo & Santoso, 2021). Guided questioning and gradual transfer of responsibility may therefore be useful when students have difficulty coordinating multiple constraints, rather than simply providing the correct placement.

Overall, the study contributes to the understanding of mathematical thinking in KenKen by showing that Stacey's four processes are observable not as isolated stages, but through the ways students coordinate information throughout problem solving. The findings also show that students within the same initial performance profile may use different strategies, as illustrated particularly by S-3's explicit recording of alternative combinations. The three profiles are therefore useful for describing differences in the available evidence, but they should not be interpreted as fixed categories of mathematical thinking. More broadly, the study suggests that examining students' written traces together with their retrospective explanations can reveal differences in how they begin a problem, connect constraints, generate possibilities, revise decisions, and justify placements. This provides a more detailed account of mathematical thinking than final task accuracy alone.

However, this study has several limitations. First, the study involved only eight students from one Grade 8 class in a single school, which limits the transferability of the findings to other contexts. Second, the initial performance profiles were based on students' response patterns in an introductory KenKen task rather than on standardized measures of mathematical ability. Third, the analysis relied on students' explanations obtained through retrospective interviews rather than concurrent think-aloud data. These limitations should be considered when interpreting the findings and suggest the need for further studies involving more diverse participants and research settings.

Conclusion

This study examined Grade 8 students' mathematical thinking processes while solving KenKen puzzles using Stacey's framework of specialization, generalization, conjecturing, and convincing. The findings show that these processes were interrelated and became visible as students selected starting points, coordinated information across cages and rows or columns, considered possible values, revised placements, and explained their decisions.

The contribution of this study is to provide qualitative evidence of how Stacey's four mathematical thinking processes are manifested in KenKen problem solving when students must coordinate multiple mathematical constraints. The findings show that differences among students were evident not only in whether they reached a correct solution, but also in how they used and coordinated the available constraints, responded to conflicts, considered alternatives, and justified their decisions. Written work together with task-based retrospective interviews therefore provided complementary evidence for examining students' mathematical thinking processes.

For classroom practice, KenKen can be used not only as a game-based activity but also as a structured mathematical task for observing students' reasoning and identifying areas where instructional support may be needed. In particular, students who experience difficulty coordinating several constraints may benefit from prompts that direct attention to relationships among permissible numbers, cage conditions, and row–column restrictions without directly providing the solution.

This study does not establish that KenKen improves students' mathematical thinking, and the initial performance profiles should not be interpreted as measures of students' general mathematical ability. Rather, the findings describe how mathematical thinking processes appeared in the observed KenKen problem-solving activity. Future research could examine these processes with larger and more diverse groups of students and investigate how they develop over a longer period of KenKen-based instruction.

Declarations

Conflicts of Interest	:	The authors declare no conflict of interest.
Generative AI Statement	:	AI Used for Limited, Non-Substantive Support: Generative AI tools, such as Grammarly and Microsoft Copilot, were employed solely for language editing and minor phrasing enhancements. All conceptualization, analysis, and scholarly content were independently developed and verified by the authors.
Funding Statement	:	This work received no specific grant from any public, commercial, or not-for-profit funding agency.
Author Contributions	:	**Raudhatul Khaira:** Conceptualization, methodology, writing - original draft, editing, and visualization; **Aiyub:** Writing - review & editing, formal analysis, validation and supervision.

References

Aiyub, A., Suryadi, D., Fatimah, S., Kusnandi, K., & Abidin, Z. (2024). Investigation of Students' Mathematical Thinking Processes in Solving Non-routine Number Pattern Problems: A Hermeneutics Phenomenological Study. *Mathematics Teaching Research Journal 54 Early Spring*, *16*(1), 54–78. https://eric.ed.gov/?id=EJ1427415

Anggoro, B. S., Dewantara, A. H., Suherman, S., Muhammad, R. R., & Saraswati, S. (2025). Effect of game-based learning on students' mathematics high order thinking skills: A meta-analysis. *Revista de Psicodidáctica*, *30*(1), 500158. https://doi.org/10.1016/j.psicod.2024.500158

Anwar, L., Mali, A., & Goedhart, M. (2024). Formulating a conjecture through an identification of robust invariants with a dynamic geometry system. *International Journal of Mathematical Education in Science and Technology*, (7), 1681–1703. https://doi.org/10.1080/0020739X.2022.2144517

Čujdíková, M., & Vankúš, P. (2021). Development of mathematical thinking through playing video games. *Proceedings of the European Conference on Games-Based Learning*, *2021-September*, 160–169. https://doi.org/10.34190/GBL.21.039

Doyle, L., Brady, A. M., & Byrne, G. (2009). An overview of mixed methods research. *Journal of Research in Nursing*, *14*(2), 175–185. https://doi.org/10.1177/1744987108093962

Ellis, A. B., Lockwood, E., Tillema, E., & Moore, K. (2022). Generalization across multiple mathematical domains: Relating, forming, and extending. *Cognition and Instruction*, *40*(3), 351–384. https://doi.org/10.1080/07370008.2021.2000989

Foster, C. (2023). Problem solving in the mathematics curriculum: From domain-general strategies to domain-specific tactics. *Curriculum Journal*, *34*(4), 594–612. https://doi.org/10.1002/curj.213

Hashemi, N., Abu, M. S., Kashefi, H., & Rahimi, K. (2013). Generalization in the Learning of Mathematics. In *2 nd International Seminar on Quality and Affordable Education*.

Johnson, H. L., Mcclintock, E., & Knurek, R. (2025). Students are experts in their mathematical thinking: Positioning principles for task-based interviews. *Hiroshima Journal Of Mathematics Education*, *18*, 15–29. https://doi.org/10.24529/hjme.1802

Khoo, Y. Y., Ramdan, M. R., Abdullah, N. L., Abd Aziz, N. A., & Mahjom, N. (2025). The impacts of game-based learning on thinking and learning in higher education context: A scoping review. *International Journal of Education in Mathematics, Science and Technology*, *13*(3), 623–637. https://doi.org/10.46328/ijemst.4776

Komatsu, K., & Jones, K. (2022). Generating mathematical knowledge in the classroom through proof, refutation, and abductive reasoning. *Educational Studies in Mathematics*, *109*(3), 567–591. https://doi.org/10.1007/s10649-021-10086-5

Lipscomb, K. (2015). *Math reasoning and vocabulary integration with KenKen®*. Charlotte Teachers Institute.

Maher, C. A., & Sigley, R. (2020). Task-based interviews in mathematics education. *Encyclopedia of Mathematics Education*, 821–824. https://doi.org/10.1007/978-3-030-15789-0_147

Mason, J., Burton, L., & Stacey, K. (2010). *Thinking mathematically (2nd Ed.)*. Pearson Education.

Mezmir, E. A. (2020). Qualitative data analysis: An overview of data reduction, data display and interpretation. *Research on Humanities and Social Sciences*, *10*, 15–27. https://doi.org/10.7176/rhss/10-21-02

Michigan State University. (2023). *KenKen Puzzles*. https://math.msu.edu/Outreach/resources/kenken-puzzles.aspx

Nguyen, H. A., Guo, Y., Stamper, J., & McLaren, B. M. (2020). Improving students' problem-solving flexibility in non-routine mathematics. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *12164 LNAI*, 409–413. https://doi.org/10.1007/978-3-030-52240-7_74

Nusantara, D. S., Zulkardi, Z., & Putri, R. I. I. (2025). Investigating students' mathematical argumentation ability in solving PISA-like mathematics tasks on change and relationships. *Jurnal Pendidikan Matematika*, *19*(4), 807–826. https://doi.org/10.22342/mej.v19i4.pp807-826

Öztürk, M., Akkan, Y., & Kaplan, A. (2020). Reading comprehension, mathematics self-efficacy perception, and mathematics attitude as correlates of students' non-routine mathematics problem-solving skills in Turkey. *International Journal of Mathematical Education in Science and Technology*, *51*, 1042–1058. https://doi.org/10.1080/0020739X.2019.1648893

Reiter, H., Thornton, J., & Vennebush, P. (2011). Using KenKen to build reasoning skills. *Mathematics Teacher Learning and Teaching PK-12, 107*(5),341-347. https://doi.org/10.5951/mathteacher.107.5.0341

Santos-Trigo, M. (2024). Problem solving in mathematics education: tracing its foundations and current research-practice trends. *ZDM - Mathematics Education*, *56*(2), 211–222. https://doi.org/10.1007/s11858-024-01578-8

Schoenfeld, A. H. (1992). *Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics (Reprint)*.

Stacey, K. (2006). What is mathematical thinking and why is it important? *Proceedings of the APEC- Tsukuba International Conference 2007: Innovative Teaching Mathematics through Lesson Study*.

Utomo, D. P., & Santoso, T. (2021). Zone of proximal development and scaffolding required by junior high school students in solving mathematical problems. In *Obrazovanie i Nauka* (Vol. 23, Number 9, pp. 186–202). Russian State Vocational Pedagogical University. https://doi.org/10.17853/1994-5639-2021-9-186-202

Vural, Ö. F., & Vural, S. (2021). An Examination of 5th Grade Mathematics Curriculum in Terms of 21st Century Skills. *International Journal of Educational Research Review*, (2), 82–92.

Yao, D. (2026). The impact of gamified learning in mathematics education on students' cognitive and affective outcomes: Insights from a second-order meta-analysis. *British Education Research Journal*, 1–30. https://doi.org/10.1002/berj.70144