



A rubric for assessing teachers' game pedagogical content knowledge in teaching mathematics

Ngoc Dan Nguyen ¹

 0009-0002-6489-5109

Nguyen-Ai-Thu Duong ¹

 0009-0008-2070-4001

Minh Dung Tang ^{1*}

 0000-0001-5401-1395

¹ Ho Chi Minh City University of Education, Ho Chi Minh City, VIETNAM

* Corresponding author: dungtm@hcmue.edu.vn

Citation: Nguyen, N. D., Duong, N.-A.-D., & Tang, M. D. (2026). A rubric for assessing teachers' game pedagogical content knowledge in teaching mathematics. *Contemporary Educational Technology*, 18(4), Article ep692. <https://doi.org/10.30935/cedtech/19470>

ARTICLE INFO

Received: 30 Apr 2026

Accepted: 2 Sep 2026

ABSTRACT

Digital games can support mathematics learning when their features are aligned with mathematical content and instructional goals. This methodological study adapted the technological pedagogical content knowledge-practical rubric to examine game pedagogical content knowledge in digital game-based mathematics lesson plans. The seven dimensions and four performance levels of the original rubric were retained, while the descriptors were revised for game mechanics, mathematical representation, instructional guidance, and game-supported assessment. The adapted descriptors underwent qualitative review by three experts. Two trained raters then used the rubric to score 50 lesson plans, and agreement for each dimension was estimated using weighted Cohen's kappa. The coefficients ranged from .710 to .885, indicating substantial to almost perfect agreement. These coefficients describe agreement between the two raters for this sample; they do not establish construct or criterion-related validity. The rubric may provide teacher educators and professional development facilitators with a structured basis for reviewing and revising digital game-based mathematics lesson plans. Further studies should test the rubric with an independent expert panel, additional raters, new samples from multiple institutions, and evidence from classroom practice.

Keywords: digital game-based learning, rubric, assessment, teacher, TPACK-G, mathematics education

INTRODUCTION

Digital game-based learning (DGBL) has attracted growing interest in mathematics education (Chen et al., 2022; Dan et al., 2024; Hussein et al., 2022). Digital games can provide settings for exploration and problem solving, while also supporting motivation and engagement. Previous studies have reported greater motivation and cognitive engagement when games are integrated into instruction (Bang et al., 2023; Hussein et al., 2022; Jaiswal, 2021).

However, using games does not necessarily lead to better learning outcomes. Learning outcomes in DGBL depend partly on how gameplay is linked to learning objectives, mathematical content, and instructional activities. Without these links, games may function mainly as sources of motivation or drill-and-practice, offering little support for conceptual understanding (Dan et al., 2024; Qian & Clark, 2016). This concern is particularly relevant to mathematics, where students often need help connecting representations, procedures, and reasoning. Teachers therefore require subject-specific knowledge in addition to general knowledge of technology integration (Hayak & Avidov-Ungar, 2023).

To address this issue, recent scholarship has paid attention to constructs that describe how teachers integrate digital games into instruction in meaningful ways. In this study, we focus on game pedagogical content knowledge (GPCK), a construct derived from the TPACK-G framework proposed by Hsu et al. (2013). GPCK refers to teachers' integrated understanding of pedagogy, content, and digital games in specific instructional contexts. In mathematics, GPCK involves selecting, adapting, and using games in ways that support conceptual understanding and mathematical reasoning. In this sense, GPCK is not simply technological pedagogical content knowledge (TPACK) applied to a game setting. It refers more specifically to teachers' ability to align game mechanics, mathematical meaning, and pedagogical support so that gameplay contributes to learning rather than serving only as motivation or repetitive practice.

Despite its importance, GPCK remains difficult to assess. Most studies have relied on self-report measures, such as questionnaires and surveys (e.g., Hsu et al., 2013; Hsu et al., 2015; Kuo & Kuo, 2024). Although these instruments are useful for examining teachers' perceptions, they provide limited evidence of how GPCK is represented in instructional planning. A discrepancy may therefore exist between teachers' reported knowledge and the evidence presented in their lesson designs.

This limitation points to the need for performance-based approaches that assess GPCK through concrete instructional artifacts, such as lesson plans. Within the broader TPACK literature, the TPACK-practical (TPACK-P) rubric developed by Yeh et al. (2015) provides a useful model. That rubric was designed to assess preservice teachers' practical application of TPACK in science teaching with information and communications technologies (ICTs) through lesson plans and microteaching. However, to our knowledge, no comparable rubric has been developed specifically to assess GPCK in digital game-based mathematics instruction.

Against this background, the study had two objectives. The first was to adapt the TPACK-P rubric to examine GPCK-related evidence in digital game-based mathematics lesson plans. The second was to examine agreement between two raters when applying the adapted rubric to a sample of lesson plans. The study addressed the following research questions:

1. *How was the TPACK-P rubric adapted to examine GPCK-related evidence in digital game-based mathematics lesson plans?*
2. *To what extent did two trained raters agree when applying the adapted rubric to the lesson plans?*

THEORETICAL FRAMEWORK: TPACK-G

To understand the knowledge required to integrate digital games into subject-specific teaching, GPCK should be situated within the broader context of technology integration in education. The TPACK framework (Mishra & Koehler, 2006) conceptualizes technology integration as the interaction among three core knowledge domains: content knowledge, pedagogical knowledge, and technological knowledge. Their intersections give rise to four related constructs: pedagogical content knowledge, technological content knowledge, technological pedagogical knowledge, and TPACK. Together, these domains describe the knowledge teachers need to integrate technology into subject teaching.

Building on TPACK, Hsu et al. (2013) proposed the TPACK-G framework to describe the knowledge needed for integrating digital games into teaching. This framework includes game knowledge (GK), game pedagogical knowledge (GPK), and game content knowledge (GCK). GK refers to knowledge of game systems, mechanics, and operation. GPK refers to knowledge of how to organize gameplay in the classroom, support student engagement, and incorporate games into teaching strategies. GCK refers to knowledge of how games represent and support domain-specific content, such as mathematical ideas or problem-solving processes. Within this framework, GPCK refers to teachers' knowledge of how to connect digital games with subject content and instructional approaches.

Although GPCK is grounded in the broader TPACK tradition, it is more specific to the instructional demands of digital game-based teaching. In TPACK, technology is often treated as a broad instructional resource. In contrast, digital games are structured environments with rules, goals, feedback systems, and motivational features that shape how students participate and learn. As a result, teachers need not only technical knowledge about using games, but also knowledge of how to align game mechanics with instructional goals

and content learning (Garris et al., 2002). In mathematics, this includes using games in ways that support conceptual understanding rather than relying on them only for engagement or repetitive practice.

GPCK also includes the pedagogical decisions required during game-based instruction. Teachers need to organize gameplay, guide students, anticipate responses to game challenges, and connect those experiences to disciplinary ideas. They must also consider affective factors such as engagement, challenge, and flow (Hamari et al., 2016). These demands make GPCK a context-specific extension of TPACK (Kuo & Kuo, 2024), and its assessment should examine evidence in instructional design and practice rather than rely only on reported beliefs.

Assessing Teachers' GPCK

Despite growing interest in GPCK, most studies have relied on self-report questionnaires. These questionnaires provide a practical means of collecting data in research and professional development settings. For example, Hsu et al. (2015) and Hsu et al. (2020) used self-report scales to examine changes in teachers' knowledge following DGBL-related training. Similarly, Villa et al. (2023) used pre- and post-intervention surveys to assess teachers' perceived growth in TPACK-G after participating in structured game-based learning modules. These instruments typically ask teachers to rate their confidence or perceived competence across domains related to pedagogy, content, technology, and game integration.

However, self-report measures have an important limitation: they do not necessarily reflect how teachers apply knowledge in practice. They are useful for capturing perceptions, but they provide limited evidence of how teachers design or perform game-based instruction. Previous studies have reported possible differences between self-reported knowledge and evidence obtained from lesson plans or observations (e.g., Canbazoglu Bilici et al., 2016; Hsu et al., 2015; Yeh et al., 2015). For example, a participant may report confidence in using digital games for mathematics instruction, while the corresponding lesson plan may use the game only for motivation rather than supporting students' learning. Self-report data may not fully capture GPCK. Lesson plans, teaching videos, and classroom observations can provide additional evidence of how this knowledge is reflected in instructional planning and practice.

In response to this need, researchers have increasingly called for performance-based assessments of TPACK and related constructs. One relevant example is the TPACK-P rubric developed by Yeh et al. (2015), which was designed to assess preservice science teachers' practical application of TPACK through lesson-plan analysis and microteaching. The rubric was developed through expert review and organized into seven dimensions:

- (a) using ICTs to understand students and support their learning,
- (b) constructing ICT-infused curricula,
- (c) using ICT-supported representations to present content,
- (d) employing ICT-integrated teaching strategies,
- (e) providing explicit guidance when using ICTs,
- (f) preparing for or troubleshooting technical problems in instructional contexts, and
- (g) using ICTs to assess students.

Together, these dimensions provide a useful model for assessing practical technology-related teaching knowledge.

To our knowledge, no rubric has been developed specifically for examining GPCK in digital game-based mathematics lesson plans. We therefore adapted the TPACK-P rubric for this purpose.

METHOD

Research Design

This methodological study adapted an existing rubric and examined agreement between two raters who applied it to lesson plans. The study comprised descriptor adaptation, qualitative expert review, rater training,

Table 1. Summary of rubric adaptations and their rationale

Dimension	Retained element	Main adaptation	Rationale
A	Understanding students through technology	Gameplay data, scores, feedback and recordings	Digital games generate specific evidence of student strategies
B	Curriculum integration	Purpose and depth of game integration	Distinguishes motivation/drill from conceptual and inquiry-oriented use
C	Technology-supported representation	Representation of mathematical concepts through game mechanics	Mathematics requires attention to conceptual representation
D	Technology-integrated pedagogy	Game-integrated inquiry, collaboration and exploration	Captures pedagogical use of gameplay
E	Explicit guidance	Questions and prompts connecting gameplay with mathematics	Focuses on mathematical reflection and sense-making
F	Technical preparation	Technical and pedagogical contingencies	DGBL requires both technical and instructional preparation
G	Technology-supported assessment	In-game feedback, progress data and self-assessment	Games can produce immediate and process-based assessment evidence

initial scoring, rubric refinement, and independent rescoring. It did not involve an instructional intervention or evaluate intervention outcomes.

Study Context and Data Source

The data comprised 50 lesson plans prepared by graduate students enrolled in a Master of Education program in primary education at a university of education in Vietnam. Each lesson plan described a digital game-based mathematics activity for elementary-level instruction. The plans were treated as evidence of instructional planning, not as evidence of classroom practice. The findings therefore concern GPCK-related features documented in lesson plans and do not show how participants implemented the activities in classrooms.

Rubric Adaptation

The study adapted the TPACK-P rubric developed by Yeh et al. (2015). The seven original dimensions and four performance levels were retained. However, the descriptors were revised to address the specific characteristics of digital game-based mathematics instruction.

Three principles guided the adaptation. First, the conceptual purpose of each original dimension was retained. Second, references to general ICT use were replaced with language specific to digital games. Third, the descriptors were revised for mathematics instruction, with attention to mathematical representation, conceptual understanding, guidance during gameplay, and game-generated assessment evidence.

Table 1 summarizes the relationship between the original dimensions and their adaptations for digital game-based mathematics lesson plans.

In dimension A, the original rubric focused on using ICTs to understand students' learning through interaction or feedback. In the adapted version, this dimension was revised to reflect the specific ways digital games can provide information about students' learning, such as through gameplay interaction, in-game feedback, game logs, scores, outcomes, or screen recordings. The focus of this dimension is therefore on whether teachers can interpret these sources of information to identify students' strategies, difficulties, and learning progress, and use that information to support instruction.

Dimension C required more specific revision. In the original rubric, this dimension focused on using ICTs to represent instructional content. In the context of digital game-based mathematics instruction, this was revised to capture whether teachers select or design games that meaningfully represent mathematical ideas. In other words, the focus is not simply on using games to present tasks, but on whether the game provides a model or representation that helps students make sense of the target concept. For example, a game in which students divide a pizza into equal parts can support the representation of fractions through a part-whole model. By contrast, a game that only presents fraction calculations without any visual or conceptual model would show weaker alignment with this dimension. This revision was intended to reflect the importance of representation in mathematics instruction.

Dimension E and dimension G were also revised. In the original rubric, dimension E focused on teachers' provision of explicit guidance when using ICTs. In the present study, this dimension was revised to capture how teachers guide students in connecting gameplay with mathematical ideas, for example through questions, prompts, or reflective activities. Dimension G was also revised to reflect the assessment opportunities provided by digital games, including the use of in-game feedback, performance data, or student self-reflection to support learning.

Dimensions B, D, and F required only minor revision. In these dimensions, the main changes involved replacing ICT-related wording with references to digital games and revising examples to fit mathematics instruction.

Qualitative Expert Review and Rubric Refinement

Following the initial adaptation, the rubric was independently reviewed by three experts: one specializing in mathematics education, one in DGBL, and one in teacher education and instructional design. The experts were asked to comment on the relevance of each dimension to GPCK, the clarity of the performance descriptors, and the suitability of the rubric for DGBL in mathematics.

The experts' comments led to revisions in wording, clearer distinctions between performance levels, and closer alignment with digital game-based mathematics instruction. Because the review did not collect item-level numerical ratings, neither a content validity index nor an expert-agreement coefficient was calculated. We therefore treat the review as qualitative input to rubric refinement, not as evidence of formal content validity.

Coders and Coder Training

Two coders were selected based on their academic backgrounds and professional experience. Coder 1 held a master's degree and had eight years of experience in primary mathematics education. Coder 2 also had a master's degree and had three years of experience in primary education. At the time of the study, both coders worked at the university from which the lesson plans were obtained.

Before scoring, the coders reviewed the seven dimensions and four performance levels, discussed the descriptors, and practiced applying the rubric to sample lesson plans. They also established decision rules for distinguishing adjacent levels. After training, they independently scored all 50 lesson plans without discussing individual cases or scores.

Scores were based only on evidence explicitly documented in the lesson plans. The coders did not infer intentions or classroom practices that were not described. For each dimension, they assigned the highest level whose defining features were clearly met. When the evidence fell between two levels, they assigned the lower level.

Ethical Considerations

Ethical approval for the study was granted by the university where authors collected lesson plans. The university authorized using the lesson plans for research purposes.

Participants were informed of the study purpose, the proposed use of their lesson plans, the voluntary nature of participation, and the confidentiality procedures. Written informed consent was obtained before the lesson plans were included in the dataset. Participation did not affect course grades or academic standing. Names, student numbers, course records, and other direct identifiers were removed before analysis. The coders received only de-identified lesson plans.

Inter-Rater Agreement Analysis

Because the rubric uses ordinal performance levels, weighted Cohen's kappa was calculated separately for each dimension. Quadratic weights assigned larger penalties to disagreements spanning more performance levels. The analysis was conducted in RStudio (version 2025.05.1+513) with R version 4.5.1 and the irr package. Kappa coefficients were interpreted using the categories proposed by Landis and Koch (1977). In this study, the coefficients describe agreement between the two raters for the present sample and scoring procedure; they are not estimates of validity or reliability across other raters and contexts.

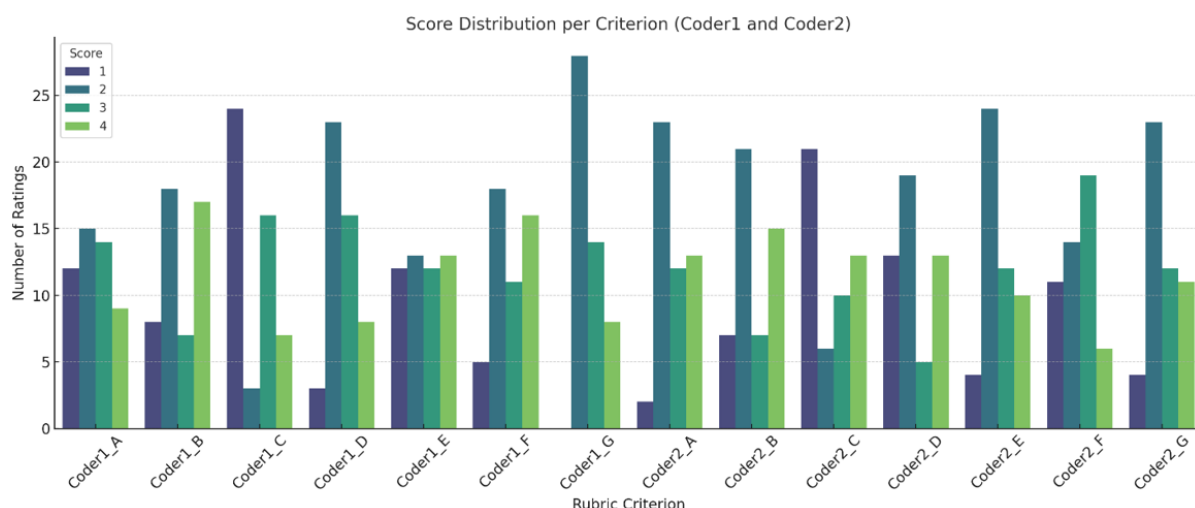


Figure 1. Score distributions assigned by coder 1 and coder 2 across the seven rubric dimensions (the authors' own elaboration)

Table 2. Weighted Cohen's kappa coefficients for agreement between the two raters

Dimension	Weighted Kappa	Interpretation
A	0.718	Substantial
B	0.855	Almost perfect
C	0.885	Almost perfect
D	0.710	Substantial
E	0.825	Almost perfect
F	0.730	Substantial
G	0.714	Substantial

Rubric Refinement and Rescoring

After the initial scoring round, the research team reviewed descriptors that the coders had found difficult to distinguish. Minor wording changes clarified these descriptors without altering the seven dimensions or the meaning of the four performance levels. The coders then independently rescored the same 50 lesson plans.

The reported kappa coefficients refer to the refined rubric. Because the same lesson plans were used in both scoring rounds, familiarity with the cases may have increased agreement. This limitation is discussed below.

RESULTS

Score Distribution per Dimension

The histograms in [Figure 1](#) show that all four rating categories were used for most dimensions, indicating variability in the scores assigned within this sample. The distributions are presented descriptively and are not interpreted as evidence of construct validity, discriminant validity, or the absence of ceiling or floor effects.

Inter-Rater Agreement

Weighted kappa coefficients ranged from .710 to .885 across the seven dimensions ([Table 2](#)). Agreement was almost perfect for dimensions B ($\kappa = .855$), C ($\kappa = .885$), and E ($\kappa = .825$), and substantial for dimensions A ($\kappa = .718$), D ($\kappa = .710$), F ($\kappa = .730$), and G ($\kappa = .714$), based on the interpretive categories proposed by Landis and Koch (1977). These results indicate that the two raters generally assigned similar scores when applying the refined rubric to the 50 lesson plans.

The complete adapted rubric is presented in [Table 3](#).

Table 3. Adapted rubric for examining GPCK-related evidence in digital game-based mathematics lesson plans

Dimension	Level 1	Level 2	Level 3	Level 4
A. Using games to understand and support students' learning	Not using games to understand or support students	Using games only as supplementary activities without observing students' learning strategies	Observing students' gameplay to understand their strategies and learning difficulties	Using in-game feedback to understand students and guide them toward self-directed learning
B. Integrating games into the curriculum	Using games solely to boost student motivation or engagement	Using games to support instruction and partially help students meet learning objectives	Using games to reinforce conceptual understanding and clarify students' misconceptions	Using games to guide students in constructing knowledge through exploration and inquiry-based learning
C. Using games to represent mathematical concepts	Presenting mathematical content in traditional ways, without using games as representational tools	Using games to represent mathematical concepts visually	Guiding students to interact with games in ways that help generalize mathematical knowledge	Designing game-based activities that allow students to experiment and explore in order to discover mathematical ideas
D. Employing game-integrated teaching strategies	Using traditional lecture-based instruction, with games serving only as supplementary elements	Using basic pedagogical methods (e.g., teacher-student Q&A) to support understanding of mathematical content	Using appropriate pedagogical strategies (e.g., inquiry-based learning, collaborative problem-solving) to help students explore mathematical ideas through games	Designing open learning environments that encourage student exploration, creativity, and collaboration through DGBL to foster mathematical thinking
E. Providing explicit guidance when using games	Providing no clear instructional guidance	Providing general guidance, but without helping students connect gameplay with mathematical concepts	Listing clear guiding questions that connect gameplay actions with mathematical concepts and rules	Listing clear guiding questions and anticipating student responses to promote reflection and strategic adjustment in learning mathematics through DGBL
F. Preparing for and responding to technical problems in instructional contexts	Having no backup plan or being unable to resolve technical issues	Preparing basic contingency plans for simple technical problems, but without anticipating pedagogical challenges	Preparing thoroughly for both technical and pedagogical contingencies in order to manage DGBL-related situations effectively	Designing multiple backup plans and adaptive strategies to personalize the DGBL experience for individual learners
G. Using games to assess students	Using assessment activities unrelated to games	Using games for basic assessment purposes (e.g., checking correct/incorrect answers)	Using games to provide immediate feedback or to record students' learning progress	Using games for assessment while supporting student self-regulation and self-assessment

DISCUSSION

The two raters reached substantial or almost perfect agreement on all seven dimensions. The higher performance levels retain the original TPACK-P emphasis on active, student-centered uses of technology (Yeh et al., 2015). The adapted descriptors specify how games may represent mathematical ideas, structure exploration, provide feedback, and support assessment. The examples below show how these distinctions were applied when scoring the lesson plans.

One clear example can be seen in dimension G, which focuses on using games to assess students. In the original TPACK-P rubric, assessment with technology is described in broader terms. In the present rubric, this dimension was specified in relation to the feedback and performance data available in digital games. Several lesson plans, such as lesson plan #04 (see [Figure 2](#)), used games mainly to collect correct or incorrect

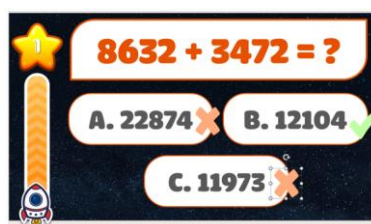


Figure 2. Game used in lesson plan #04 (reproduced with participant consent)

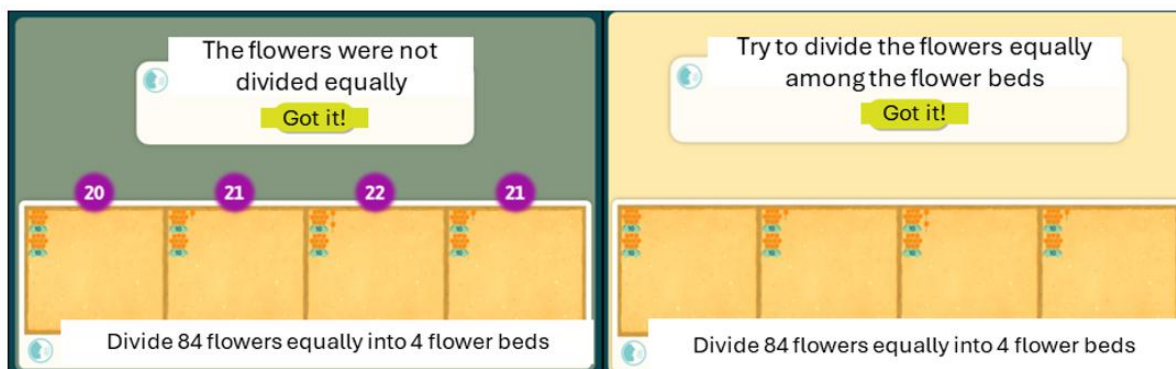


Figure 3. Game used in lesson plan #47 (reproduced with participant consent)

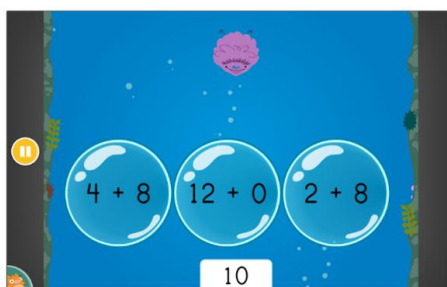


Figure 4. Game used in lesson plan #14 (reproduced with participant consent)

responses in quiz-like formats. In such cases, the game functioned as a simple checking tool and reflected level 2 of dimension G. By contrast, lesson plan #47 was rated at level 3 for dimension G (Figure 3).

The game provided immediate feedback on incorrect moves, enabling students to reconsider and revise their responses. This evidence exceeded level 2, which is limited to checking correct or incorrect answers. However, the lesson plan did not explicitly incorporate student self-assessment or systematic self-regulation; therefore, it did not meet the level 4 descriptor.

A similar pattern can be seen in dimension E, which focuses on guidance during game-based instruction. In the original TPACK-P rubric, this dimension mainly concerned guidance in the use of ICTs. In the present rubric, the focus was revised to capture how teachers guide students in connecting gameplay with mathematical ideas. Lesson plan #47 was rated at level 4. The lesson plan included questions before and after gameplay, anticipated students' responses and misconceptions, and provided prompts for reflection and strategy adjustment. These features distinguish it from level 3, which requires guiding questions but does not include the anticipation of students' responses.

Dimension C also illustrates how the rubric distinguishes between different levels of content representation in game-based mathematics instruction (Figure 4). For example, lesson plan #14 used an undersea race game in which students solved addition problems to make a submarine move faster. Although the game was engaging, it did not represent addition visually or conceptually. Its main role was to motivate repeated practice. This type of use, often described in the literature as drill-and-practice (e.g., Dan et al., 2024) forms that are not related to mathematics representation, hence being rated at level 1 of dimension C.

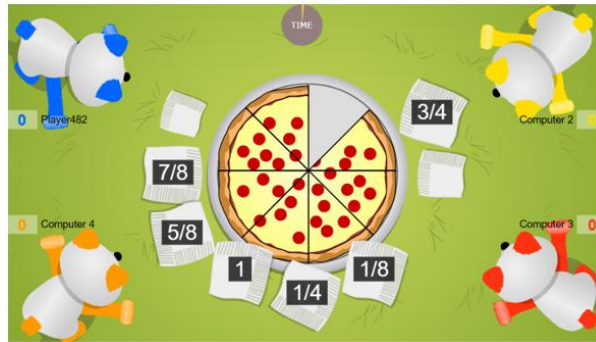


Figure 5. Game used in lesson plan #38 (reproduced with participant consent)

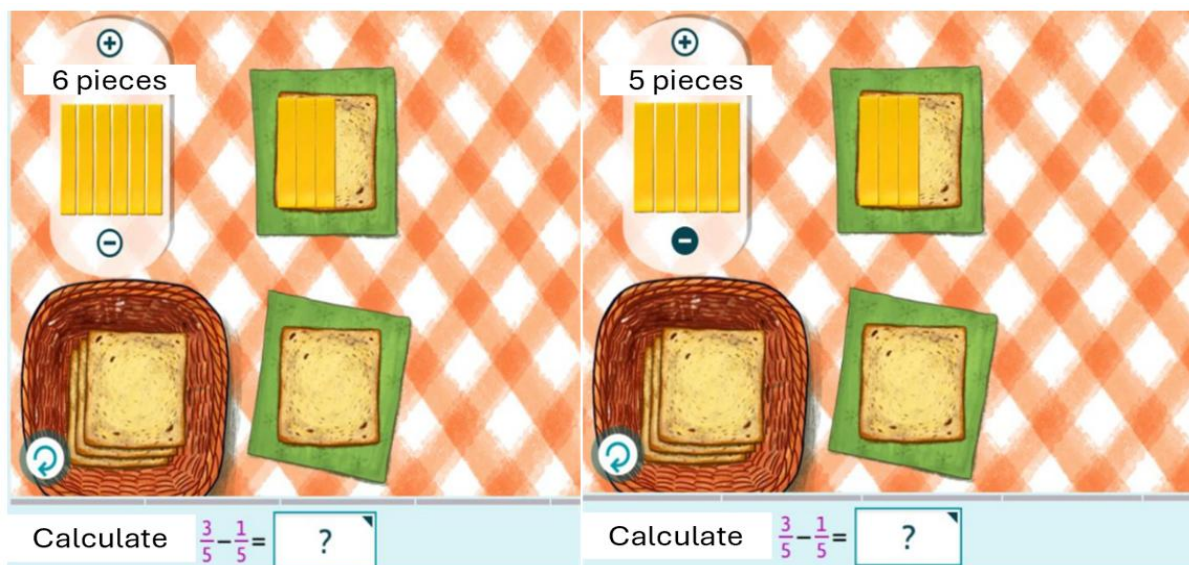


Figure 6. Game used in lesson plan #39 (reproduced with participant consent)

By contrast, lesson plan #38 used a pizza-slicing game to introduce fractions (Figure 5). This lesson plan was rated at level 2 because the game provided a visual representation linked to part-whole relationships, but it offered limited support for generalization beyond the immediate context. Lesson plan #39 was rated at level 4 for dimension C (Figure 6). In the game, students represented fraction subtraction by removing slices of cheese. They could vary the number of slices and compare the resulting fractions, rather than follow a single predetermined procedure. The game therefore functioned as a setting for exploring fraction subtraction, which distinguished this lesson plan from level 3.

Implications

The rubric may have practical uses in professional development. Instructors could ask teachers to score the same lesson plan, compare their ratings, and discuss how the game mechanics relate to the mathematical objectives. Teachers could also use the descriptors to revise a plan that relies on games mainly for motivation or repeated practice. These proposed uses are consistent with research on collaborative design and rubric-guided reflection (Belda-Medina & Calvo-Ferrer, 2022; Fu et al., 2022), but they were not tested in the present study.

In teacher education, the rubric could be used to assess preservice teachers' lesson plans or to structure feedback during lesson revision and microteaching. Because the dimensions refer to observable features of a plan, teacher educators can identify whether and how a game is connected to mathematical representation, instructional guidance, and assessment. This proposed use is consistent with performance-based assessment in higher education (Vlachopoulos & Makri, 2024), although the present study did not examine its effects on preservice teachers' learning.

For research, the rubric offers a way to code GPCK-related evidence in lesson plans. It may complement self-report measures, which do not directly capture instructional design (Canbazoglu Bilici et al., 2016; Hsu et al., 2015). Studies combining lesson plans with observations or teaching videos could examine whether planned uses of games correspond to classroom enactment.

The main contribution of the study is a lesson-plan rubric that translates the seven TPACK-P dimensions into descriptors for digital game-based mathematics instruction. The adaptation specifies how mathematical representation, guidance during gameplay, technical preparation, and game-supported assessment can be identified in lesson plans.

Limitations

Several limitations should be considered. The expert review relied on qualitative comments rather than item-level numerical ratings. Therefore, a content validity index or expert-agreement coefficient was not calculated, and the study does not provide formal quantitative evidence of content validity.

The data were limited to 50 lesson plans prepared by graduate students in one program at one institution. The findings describe this institutional sample and should not be generalized to practicing teachers in other settings. Lesson plans also document intended instruction rather than classroom practice. The rubric scores cannot show how participants would implement the planned activities, respond to students, or manage gameplay in actual classrooms. Moreover, the rubric was examined only with elementary mathematics lesson plans involving digital games, so its applicability to other subjects, educational levels, institutions, and cultural settings remains unknown.

The agreement analysis also has limitations. Only two trained raters scored the lesson plans, and raters with different professional backgrounds may interpret some descriptors differently. In addition, the same lesson plans were used during rubric refinement and final rescoring. Familiarity with these cases may have contributed to the observed agreement. Although weighted kappa provided evidence of agreement between the two raters, the study did not examine construct validity, criterion-related validity, predictive validity, or generalizability across raters and contexts.

Future Directions

Content validity should next be assessed with an independent panel that provides item-level ratings. The rubric also needs testing with new lesson plans, raters who were not involved in its development, and samples from more than one institution. A generalizability study could estimate score variation associated with lesson plans, raters, dimensions, and settings.

Future work should combine lesson plans with classroom observations, teaching videos, interviews, or evidence of student learning. These data would allow researchers to compare planned GPCK with classroom enactment. Construct and criterion-related validity could also be examined by relating rubric scores to theoretically relevant measures of GPCK.

CONCLUSION

We adapted the TPACK-P rubric for use with digital game-based mathematics lesson plans. The revised descriptors cover seven areas of planning, including mathematical representation, guidance during gameplay, and game-supported assessment. When two trained raters applied the refined rubric to 50 lesson plans, weighted kappa values indicated substantial or almost perfect agreement across the seven dimensions.

These coefficients describe agreement for two raters, one sample, and one scoring procedure. They do not show that the rubric fully measures GPCK, predicts classroom practice, or transfers to other groups and settings. The rubric should therefore be tested with independent experts and raters, samples from other institutions, and classroom-based evidence.

Author contributions: NND & N-A-TD: writing – original draft & TMD: conceptualization. All authors approved the final version of the article.

Funding: This research is funded by Ho Chi Minh City University of Education Foundation for Science and Technology under grant number CS.2025.19.43.

Acknowledgments: The authors would like to thank Ho Chi Minh City University of Education for supporting this study.

Ethics declaration: Ethical approval was granted by the Research Ethics Committee of Ho Chi Minh City University of Education. Institutional authorization was obtained for the research use of the lesson plans. All participants provided written informed consent, and the lesson plans were de-identified before analysis.

AI statement: During revision, the authors used Grammarly and ChatGPT to edit English grammar and improve readability. These tools were not used to generate or analyze data, calculate statistical results, or make methodological decisions. The authors reviewed all tool-assisted revisions and took responsibility for the final text.

Declaration of interest: The authors declared no competing interest.

Data availability: Data generated or analyzed during this study are available from the authors on request.

REFERENCES

- Bang, H. J., Li, L., & Flynn, K. (2023). Efficacy of an adaptive game-based math learning app to support personalized learning and improve early elementary school students' learning. *Early Childhood Education Journal*, 51(4), 717-732. <https://doi.org/10.1007/s10643-022-01332-3>
- Belda-Medina, J., & Calvo-Ferrer, J. R. (2022). Preservice teachers' knowledge and attitudes toward digital-game-based language learning. *Education Sciences*, 12(3), Article 182. <https://doi.org/10.3390/educsci12030182>
- Canbazoglu Bilici, S. C., Guzey, S. S., & Yamak, H. (2016). Assessing pre-service science teachers' technological pedagogical content knowledge (TPACK) through observations and lesson plans. *Research in Science & Technological Education*, 34(2), 237-251. <https://doi.org/10.1080/02635143.2016.1144050>
- Chen, P. Y., Hwang, G. J., Yeh, S. Y., Chen, Y. T., Chen, T. W., & Chien, C. H. (2022). Three decades of game-based learning in science and mathematics education: An integrated bibliometric analysis and systematic review. *Journal of Computers in Education*, 9(3), 455-476. <https://doi.org/10.1007/s40692-021-00210-y>
- Dan, N. N., Trung, L. T. B. T., Nga, N. T., & Dung, T. M. (2024). Digital game-based learning in mathematics education at primary school level: A systematic literature review. *Eurasia Journal of Mathematics, Science and Technology Education*, 20(4), Article em2423. <https://doi.org/10.29333/ejmste/14377>
- Fu, Q. K., Zou, D., Xie, H., Cheng, G., & Hwang, G. J. (2022). Effects of a collaborative design approach on pre-service teachers' ability of designing for learning with a digital game. *Education and Information Technologies*, 27(4), 5641-5664. <https://doi.org/10.1007/s10639-021-10818-3>
- Garris, R., Ahlers, R., & Driskell, J. E. (2002). Games, motivation, and learning: A research and practice model. *Simulation & Gaming*, 33(4), 441-467. <https://doi.org/10.1177/1046878102238607>
- Hamari, J., Shernoff, D. J., Rowe, E., Coller, B., Asbell-Clarke, J., & Edwards, T. (2016). Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. *Computers in Human Behavior*, 54, 170-179. <https://doi.org/10.1016/j.chb.2015.07.045>
- Hayak, M., & Avidov-Ungar, O. (2023). Knowledge and planning among teachers integrating digital game-based learning into elementary school classrooms. *Technology, Pedagogy and Education*, 32(2), 239-255. <https://doi.org/10.1080/1475939X.2023.2175719>
- Hsu, C.-Y., Liang, J.-C., & Su, Y.-C. (2015). The role of the TPACK in game-based teaching: Does instructional sequence matter? *The Asia-Pacific Education Researcher*, 24(3), 463-470. <https://doi.org/10.1007/s40299-014-0221-2>
- Hsu, C.-Y., Liang, J.-C., & Tsai, M.-J. (2020). Probing the structural relationships between teachers' beliefs about game-based teaching and their perceptions of technological pedagogical and content knowledge of games. *Technology, Pedagogy and Education*, 29(3), 297-309. <https://doi.org/10.1080/1475939X.2020.1752296>
- Hsu, C.-Y., Liang, J.-C., Chai, C. S., & Tsai, C.-C. (2013). Exploring preschool teachers' technological pedagogical content knowledge of educational games. *Journal of Educational Computing Research*, 49(4), 461-479. <https://doi.org/10.2190/EC.49.4.c>
- Hussein, M. H., Ow, S. H., Elash, M. M., & Jensen, E. O. (2022). Digital game-based learning in K-12 mathematics education: A systematic literature review. *Education and Information Technologies*, 27(2), 2859-2891. <https://doi.org/10.1007/s10639-021-10721-x>
- Jaiswal, A. (2021). Revisiting the historical roots of game-based learning. *TechTrends*, 65(3), 243-245. <https://doi.org/10.1007/s11528-021-00603-x>

- Kuo, Y.-C., & Kuo, Y.-T. (2024). An exploratory study of pre-service teachers' perceptions of technological pedagogical content knowledge of digital games. *Research and Practice in Technology Enhanced Learning*, 19, Article 008. <https://doi.org/10.58459/rptel.2024.19008>
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159-174. <https://doi.org/10.2307/2529310>
- Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, 108(6), 1017-1054. <https://doi.org/10.1111/j.1467-9620.2006.00684.x>
- Qian, M., & Clark, K. R. (2016). Game-based learning and 21st century skills: A review of recent research. *Computers in Human Behavior*, 63, 50-58. <https://doi.org/10.1016/j.chb.2016.05.023>
- Villa, A. M., Sedlacek, Q. C., & Pope, H. Y. (2023). I DiG STEM: A teacher professional development on equitable digital game-based learning. *Education Sciences*, 13(9), Article 964. <https://doi.org/10.3390/educsci13090964>
- Vlachopoulos, D., & Makri, A. (2024). A systematic literature review on authentic assessment in higher education: Best practices for the development of 21st century skills, and policy considerations. *Studies in Educational Evaluation*, 83, Article 101425. <https://doi.org/10.1016/j.stueduc.2024.101425>
- Yeh, Y.-F., Chien, S.-P., Wu, H.-K., & Hsu, Y.-S. (2015). Rubrics of TPACK-P for teaching science with ICTs. In Y.-S. Hsu (Ed.), *Development of science teachers' TPACK* (pp. 53-70). Springer. https://doi.org/10.1007/978-981-287-441-2_4

