MethodsX 14 (2025) 103299

Contents lists available at ScienceDirect

MethodsX

journal homepage: www.elsevier.com/locate/methodsx

A design science approach to mixed-methods evaluation in serious game research ☆

Luis Demetrio Gómez García*, Gloria María Zambrano Aranda, Emerson Jesus Toledo Concha

Academic Department of Accounting Sciences, Pontificia Universidad Católica del Perú, Av. *Universitaria 1801, San Miguel, 15088, Lima, Perú*

A R T I C L E I N F O

Method name:
Design Science Approach to Mixed-Methods Evaluation in Serious Game Research

Keywords:
Business ethics
Board game
Informal business practices
Educational technology
Behavioural studies

A B S T R A C T

For a serious game to be effective, it must undergo rigorous validation process. A Design Science approach advocates the use of quantitative or qualitative research methodologies within the creation and validation of artifacts, an approach suitable for evaluating serious games as educational tools.

This study presents a methodological framework that integrates quantitative measurement and qualitative inquiry to assess the effectiveness of a serious game designed for ethics education. We provide access to the quantitative questionnaire, its codebook, and the dataset generated during the validation of the authors' approach using a serious game for teaching business ethics.

The integration of both methods allowed us to validate the game as a relevant and effective strategy for promoting ethical reflection among university students. These findings support the consistency and reliability of the method used for validating serious games.

Methodological Highlights

- The quantitative assessment is based on the Technology Acceptance Model III (TAM III) and the Theory of Planned Behavior (TPB).
- Qualitative inquiry analyzes students' group work to understand their perceptions of ethical phenomena after gameplay.
- Professors can use insights from students' perceptions as a starting point or framework for takeaways in future game applications.

Specifications table

Subject area:	Economics and Finance
More specific subject area:	Business Ethics and Informal Business Practices
Name of your method:	Design Science Approach to Mixed-Methods Evaluation in Serious Game Research
Name and reference of original method:	NA
Resource availability	https://doi.org/10.5281/zenodo.15103804

Background

This methodology was developed to rigorously validate innovative serious game implementations in university classrooms. When educators introduce new teaching methods and technologies, it is often assumed that these innovations are effective in enhancing

☆ **Related research article:** The original research article associated to this methods paper is under journal evaluation.

* Corresponding author.
E-mail address: l.gomez@pucp.edu.pe (L.D.G. García).
Social media: @demi_luigi (L.D.G. García)

https://doi.org/10.1016/j.mex.2025.103299
Received 13 February 2025; Accepted 1 April 2025
Available online 2 April 2025

students' understanding of key concepts. However, this assumption may not always align with the students' actual learning experiences.

Educators often rely on basic quantitative satisfaction surveys or simplistic qualitative assessment techniques, such as categorizing feedback into "positive, negative, and interesting" aspects. While these methods provide some insights, they lack the depth necessary to determine whether innovation truly enhances learning. Just as any market innovation must gain acceptance to be considered successful, educational innovations must also be validated by their target audience, students, to ensure their effectiveness.

Previous studies on serious game evaluation have often favored either quantitative [1] or qualitative [2] assessment methods, treating them as separate approaches. However, from a Design Science perspective, these methodologies can be integrated to provide both empirical validation and a deeper understanding of students' experiences with serious games.

Method details

In presenting the details of the proposed method, it is important to clarify that the evaluation of a serious game as an educational innovation follows the prior stages of design, development, and classroom implementation. However, this study does not focus on the innovation and implementation process but rather on the evaluation of the game as an educational tool.

For quantitative evaluation, we recommend using an intervention-based experimental design with a single post-test measurement based on the Technology Acceptance Model (TAM) III and the Theory of Planned Behavior.

The Technology Acceptance Model (TAM) III, developed by Venkatesh and Bala [3], extends the original model proposed by Davis [4] to explain technology-related behaviors through the variables Perceived Usefulness, Perceived Ease of Use, and User Satisfaction. TAM III can be considered an extension of Ajzen's [5] Theory of Planned Behavior (TPB) tailored to the context of technology use. While TPB conceptualizes behavior as the dependent variable influenced by Behavioral Intention and Attitude toward the Behavior, TAM III builds on this foundation by incorporating technology-specific determinants that influence individuals' acceptance and usage of technology.

Building on these theoretical foundations, our method for the quantitative assessment of serious games adopts a conceptual framework that combines elements from TAM III and the TPB, as illustrated in Fig. 1. This framework emphasizes the relationship between ease of use, perceived utility, user satisfaction, attitude toward use, and intention to use. Prior research further supports the relevance of applying TAM-based approaches in the context of serious games. Several studies, including those by Malaquias et al. [6], Ting and Ming [7], Kuang et al. [8], and Silva et al. [9], have demonstrated the adaptability and effectiveness of Technology Acceptance Model in educational environments.

On the other hand, the research design involving a single post-test measurement, proposed for quantitative evaluation, can be considered appropriate owing to the nature of the variables employed, which are grounded in our theoretical framework. These variables, applied through a design science approach for artifact validation, do not permit pretest/post-test comparisons, as the relevance and effectiveness of the artifact can only be observed after the game has been played.

Similarly, the use of pure control groups is not deemed pertinent for directly assessing the game itself but rather for evaluating learning outcomes or for comparison with other instructional strategies. Nevertheless, recognizing the limited acceptance of experimental designs based solely on a single post-test measurement, our proposal incorporates a qualitative assessment of the moral sensitivity elicited by the game in students as a form of triangulation. However, according to Holmström et al. [10], design science should be approached as an exploratory form of research in which experimental designs based on a single post-test measurement may be suitable for artifact incubation and refinement.

Quantitative evaluation

Administration of the quantitative questionnaire

Immediately after gameplay, instructors will administer a quantitative questionnaire provided as supplementary material in this study. The questionnaire should be adapted to a specific serious game and adjusted to align with the learning objectives.

The questionnaire is structured around five key variables and includes three control variables:

- Gender and age: These demographic factors can serve as control variables to determine whether they influence dependent variables.

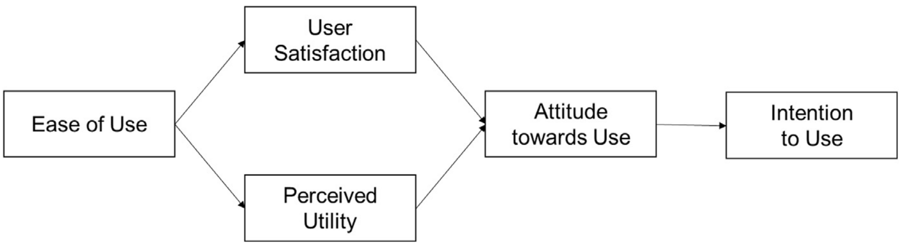

Fig. 1. Conceptual framework.
Note: Source – The authors

- Player role in the game: Depending on the game design, players may assume different roles, which could affect both the evaluation outcomes and learning experience. Controlling for this variable is important as it may indicate the need for modifications in specific roles within the game.

The dependent and independent variables were measured using five Likert-scale questions (Q1–Q5), each containing five statements rated on a five-point scale (1 = strongly disagree to 5 = strongly agree).

- Q1: Intention to use. This question assessed students' willingness to play games again. Based on Ajzen's [5] Theory of Planned Behavior (TPB), intention is the strongest predictor of behavior. In this study, it served as the main dependent variable, reflecting the effectiveness of the game in engaging students.
- Q2: Attitudes toward use. This question measured the students' perceptions of the game after playing. According to TPB, attitude is shaped by an individual's beliefs about a behavior and influences intention. A positive attitude suggests a higher probability of continued use, whereas a negative attitude indicates the need for improvements in the game's design or instructional approach.
- Q3: User satisfaction. This question evaluates students' overall satisfaction with the gaming experience. It is derived from the Technology Acceptance Model III (TAM III) [4], where satisfaction is considered a key antecedent of attitude: if students find the game satisfying, they are more likely to develop a positive perception, and vice versa.
- Q4: Perceived utility. This question measures how useful students perceive the game in supporting their learning process. Unlike previous variables, which focus on engagement and enjoyment, this study assesses the educational value of the game. Based on TAM III, perceived utility is an important predictor of attitude; if students believe the game effectively reinforces learning, they are more likely to have a positive perception.
- Q5: Ease of use. This question evaluates how easily or complexly students find a game. Also derived from TAM III, this variable played a critical role in engagement. If a serious game is perceived as too complex, it may hinder both students' motivation and learning outcomes. This is especially relevant for subjects with high theoretical-practical complexity, where serious games should act as facilitators rather than obstacles.

Data processing and analysis

To analyze quantitative data, we recommend using Partial Least Squares Structural Equation Modeling (PLS-SEM) [11]. This method is particularly suitable because it handles non-normally distributed data and works effectively with a small sample size [12].

PLS-SEM is especially useful for mediation analysis, such as in the relationships proposed in Fig. 1, where User Satisfaction, Perceived Utility, and Attitude Toward Use act as mediators between Ease of Use and Intention to Use.

To ensure statistical validity, the sample size must meet a minimum threshold. According to Hair et al. [13], a model with seven independent variables and an expected R^2 between 0.10 (low) and 0.25 (moderate) requires a sample between 51 and 137 participants. We recommend using a higher threshold (∼ 140 participants) to ensure robust results.

Given the practical challenges of conducting an experiment with 140 students in a single session, we suggest implementing the study across multiple classroom settings, ensuring that the experimental conditions remain constant for all participants. Additionally, "classroom" can be included as a control variable to account for possible contextual effects.

Qualitative inquiry

Administration of reflective questions

Instructors can pose reflection-based questions either individually or in groups to assess the qualitative impact of the game on student learning. These questions should focus on students' learning experiences rather than the direct evaluation of the game. In the following section, we provide our recommendations for structuring the qualitative questions.

- Use open-ended questions to encourage students to elaborate on their thoughts.
- Avoid using questions for grading purposes as they may encourage students to give socially desirable responses rather than authentic reflections.
- Limit the number of questions to three to maintain clarity and engagement.

Data processing and analysis

For qualitative data processing, we suggest applying qualitative content analysis using data-coding techniques [14]. The instructor followed the coding sequence outlined below:

- Start with open coding, identifying key concepts and significant data fragments in student responses.
- Proceeding with axial coding, the initial codes are organized into broader categories, and the relationships between them are explored.
- Concluding selective coding, grouping categories into logical frameworks allows for the integration of results with theoretical or empirical foundations related to the learning objectives of the serious game. This stage may be concluded with the generation of visual representations of insights derived from students' reflections.

Although specialized software exists for qualitative analysis, instructors can conduct coding manually using tools such as word processors, spreadsheets, or slide presentations, which allows for efficient classification and organization of data.

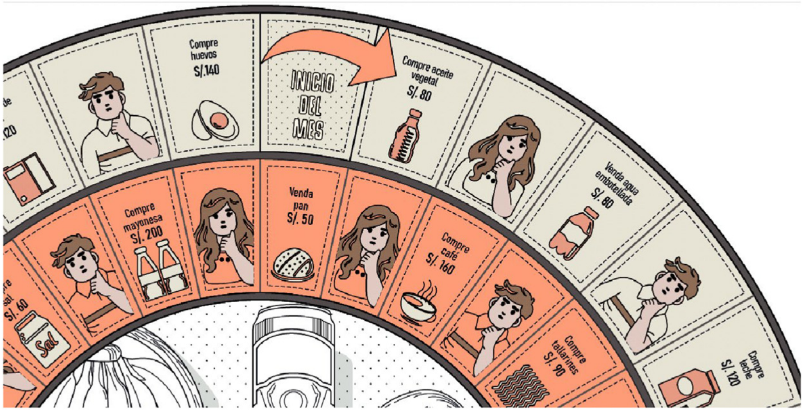

Fig. 2. Fragment of the Bodegus Game Board ©.
Note. Bodegus is the intellectual property of (the University), registered with the National Institute for the Defense of Competition and Protection of Intellectual Property of Peru (INDECOPI) under number (registration number), and authored by two of the authors of this study.

Comprehensive analysis and game improvement

Finally, instructors and game designers should integrate quantitative validation and qualitative insights, as follows:

- Identify behavioral aspects that may be negatively impacting the game's effectiveness.
- Determine whether the game effectively supports student reflection on the intended learning objectives.
- Refine game mechanics, instructional strategies, and role dynamics to enhance learning outcomes.

This methodology aims for serious games to be not only validated for effectiveness, but also analyzed through reflective evaluation aligned with educational goals.

Method validation

The present methodology was applied to an educational innovation involving a serious game called bodegus. Bodegus is designed to encourage ethical reflection among students by exposing them to ethical dilemmas related to informal business practices encountered by small shop owners, commonly known as Bodegas in some Latin American countries (see Fig. 2).

The game was implemented with a sample of 118 accounting students from Peruvian University in the second semester of 2022. Following the guidelines recommended by Hair et al. [13] for PLS-SEM based on Cohen [15], an a priori sample size calculation suggested a minimum sample size of 137 cases. This estimate assumed a conservative scenario with a small anticipated effect size (minimum $R^2 = 0.10$), statistical power of 0.80, significance level of 0.05, and maximum of seven predictors pointing to a dependent construct. However, the practical conditions allowed us to collect data from 118 students who played Bodegus.

Nevertheless, our SEM analysis showed substantial explained variance in the key endogenous constructs (R^2 values ranged from 0.212 to 0.531). Additionally, the predictive relevance of the model (Q^2 values ranging from 0.053 to 0.163) supported its predictive adequacy. Furthermore, the observed path relationships indicated medium to large effect sizes (f^2 ranged from 0.138 to 0.270 for the main paths), suggesting practically meaningful relationships among the constructs.

To further assess the sample adequacy, we conducted a post-hoc statistical power analysis for each individual SEM path using G*Power [16]. The results showed statistical power values above the recommended threshold of 0.80 (ranging from 0.913 to 0.999) for the principal paths of the model. Specifically, key relationships, such as Utility_Perc → attitude (power = 0.999), satisfaction → attitude (power = 0.999), and attitude → intention (power = 0.979), demonstrated sufficient statistical power. Only paths with very small observed effect sizes ($f^2 \leq 0.024$) exhibited a power below 0.80. Given these empirical results, we confirm that a sample size of 118 cases provided adequate statistical power to identify relevant relationships within our exploratory SEM model. Detailed results of the statistical power analysis can be found in the supplementary material provided in this paper.

The students were divided into 25 teams, each playing bodegus. Before participating in the study, 100 % of the participants were asked to provide informed consent. The quantitative questionnaire was administered as outlined in this methodology, and the resulting dataset is included as supplementary material for this article. The dataset, available in an Excel file, also includes the questionnaire validation process conducted using SmartPLS 4, following a reflective measurement model approach.

The first validation criterion for its use in Structural Equation Modeling (SEM) was the analysis of outer loadings. Within the dataset, five observable variables did not meet the 0.70 threshold for outer loadings. Since the latent variables in this study followed a reflective measurement model, the removal of these indicators did not compromise construct validity as long as each latent variable retained at least three indicators. The removal process was based on both statistical criteria (outer loadings < 0.70) and theoretical considerations to ensure construct alignment and sample relevance.

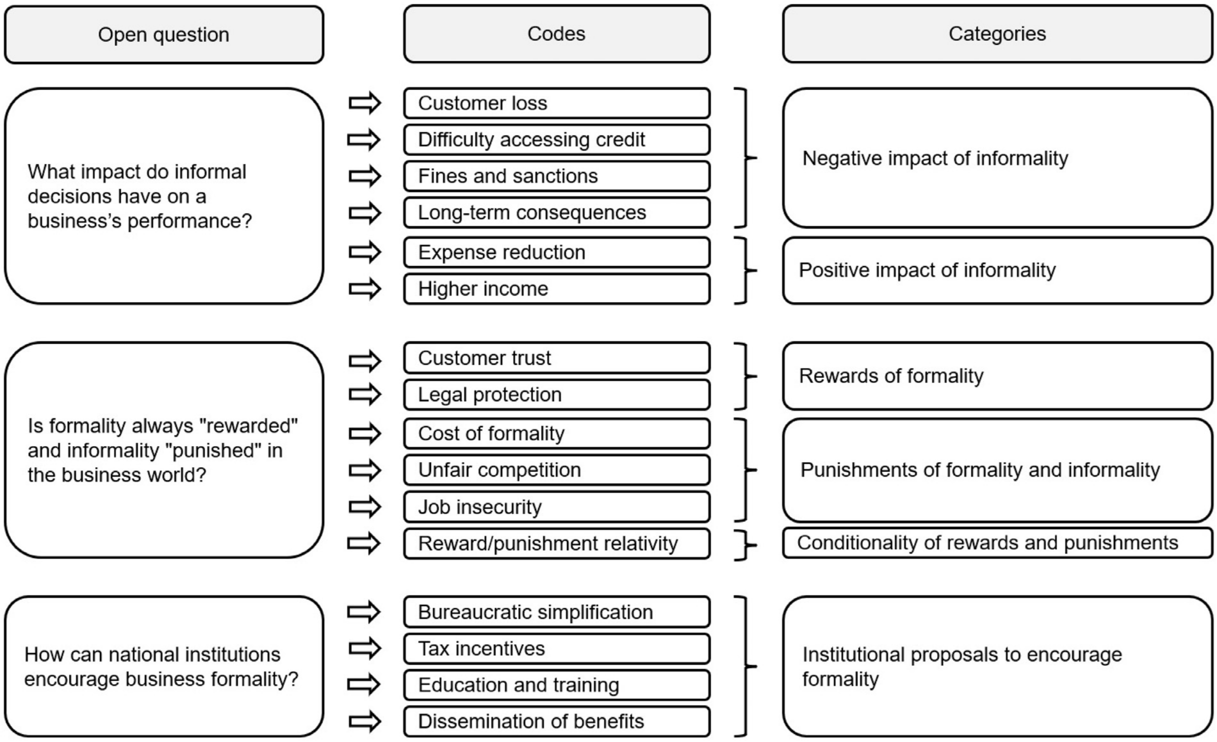

Fig. 3. From open-ended questions to category generation.
Note. Source – The authors

The dataset file shows that after the removal of these items, the outer loadings of all the factors exceeded the 0.70 threshold. Additionally, the dataset includes construct reliability validation based on Cronbach's alpha, rho_A, rho_C, and Average Variance Extracted (AVE) criteria. Discriminant validity was assessed using the heterotrait-monotrait ratio (HTMT) and Fornell-Larcker criteria, and the absence of collinearity was confirmed using the Variance Inflation Factor (VIF) criterion. Construct validity was established for all indicators.

Although not part of this study but rather of the related research article, it is worth mentioning that the structural model associated with the validation of Bodegus as a serious game included two mediation relationships, which were significantly supported, confirming the quantitative validity of our methodology and reinforcing the value of integrating TAM III and TPB within our conceptual evaluation framework.

For the qualitative validation, the three specific questions used were:

1. What impact do informal decisions have on a business's performance?
2. Is formality always "rewarded" and informality "punished" in the business world?
3. How can national institutions encourage business formality?

Qualitative analysis followed the proposed methodology. For the coding process, ATLAS.ti was used for code labeling and category generation, allowing the identification of 16 codes in the first phase and the generation of six categories in the second phase, as shown in Fig. 3. The third phase allowed the authors to synthesize the findings into an Integrative View of Formality and Informality, a graphical representation of students' reflections throughout the study. An Integrative View of Formality and Informality is presented in our original research paper. Microsoft PowerPoint was used to develop this representation.

The following excerpts illustrate how students' reflections align with the codes and categories in Fig. 3. These selected responses highlight the variety and depth of the perspectives shared in relation to each open-ended question.

The open coding analysis of the first question showed how students identified a range of consequences, both negative and positive, stemming from informality.

One group reflected on the cost dynamics associated with informality:

> *"Formality involves costs both to enter this sector—such as lengthy, complex, and expensive registration processes—and to remain in it, including tax payments and compliance with labor and salary regulations, among others. Informal decisions in some cases can mean a reduction in expenses and costs for a business." (Group 11).*

Another group pointed to the financial limitations faced by informal businesses:

> *"Difficulty accessing loans from banks due to the lack of trust an informal business generates." (Group 12).*

The potential for legal repercussions was also acknowledged:

> *"In the same line, if we are informal, there is always the risk of being sanctioned or fined for not complying with tax obligations." (Group 6).*

In response to the second question, the students offered insights that highlight how formality can provide structural advantages in the business environment.

One group emphasized the institutional protections associated with formality:

> *"However, it should be noted that formality also involves positive aspects: police protection against abuses, judicial support in conflicts and contracts." (Group 1).*

Another group highlighted access to government support:

> *"Formal businesses have access to tax incentives, subsidies, and government assistance that are not available to informal businesses, which gives them a competitive disadvantage." (Group 22).*

The third question prompted students to suggest various measures to support formalization.

One group proposed changes in labor policy and social protection frameworks:

> *"It is necessary to implement measures to encourage formalization through the flexibilization of labor regimes and the creation of incentives to expand social protection coverage." (Group 2).*

Another group focused on simplifying bureaucratic procedures:

> *"Authorities should consider the need to simplify and streamline procedures related to business registration and operating licenses." (Group 14).*

Financial support mechanisms were also discussed:

> *"Grants or loans can be provided to offer financial support to small businesses that wish to become formal but also continue to grow, since that is their main goal." (Group 20).*

Finally, one group underscored the importance of promoting legal awareness:

> *"In the same way, it is essential to promote legal/formal awareness through awareness campaigns, highlighting the long-term benefits of operating in accordance with the law." (Group 28).*

Insights from both the quantitative and qualitative results indicated improvements to the game, enabling the authors to refine its rules, expand the set of bodega-related dilemmas, propose new variations, and reconsider its design as a virtual game.

Limitations

Although we previously established the validity of our methodology, it is worth noting that most studies on educational methods typically employ quasi-experimental designs involving pre-/post-test procedures and/or control groups. In this regard, to further strengthen the application of our method, we suggest that future research adopt our methodology using comparison groups rather than strict control groups. We propose two alternatives: a phased implementation of a serious game, or the use of an alternative serious game with a similar educational purpose.

In the first option, instructors could vary the duration of gameplay, adjust the rules, or modify the intensity of the ethical dilemmas presented in the game, making it more or less challenging than in our version. This would allow researchers to compare the game scenarios that lead to better outcomes in the quantitative assessment. Moreover, this approach supports the second stage of design science, namely, artifact refinement, as suggested by Holmström et al. [10].

The second option, although more complex owing to the difficulty of identifying a comparable serious game, would involve using the same quantitative instrument to evaluate two different games across separate groups. This would allow instructors to determine which artifacts better meet their courses' learning objectives. However, a limitation of this variant is the risk of favoring one game over another, potentially discarding a game that might still be pedagogically valuable. Nonetheless, despite this limitation, this approach could also meaningfully contribute to the refinement of serious games.

Another limitation is the potential influence of the instructor on the qualitative analysis. In our methodology, instructors administer qualitative reflection questions, which might introduce unintentional bias into students' responses, particularly if they perceive certain answers as more desirable or aligned with the instructor's perspective. To minimize bias, reflection questions were administered anonymously or moderated by an independent facilitator. Additionally, the use of multiple coders for qualitative content analysis (inter-coder reliability) can enhance objectivity.

Ethics statements

All participants provided informed consent before participating in the study. Prior to data collection, the students were informed about the purpose of the study, the voluntary nature of their participation, the confidentiality of their responses, and their right to withdraw at any time without consequences. Consent was obtained in written form, and 100 % of the students agreed to participate.

Supplementary material *and/or* additional information [OPTIONAL]

Supplementary data for this paper can be found in the following link: https://doi.org/10.5281/zenodo.14866684

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Luis Demetrio Gómez García: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. **Gloria María Zambrano Aranda:** Formal analysis, Investigation, Resources, Validation. **Emerson Jesus Toledo Concha:** Data curation, Resources, Validation.

Data availability

I have shared the link of my data, the questionnaire, and the codebook.

Acknowledgments

Funding: This work was supported by "Concurso Anual de Proyectos de Investigación PUCP 2023″ (Annual Competition of Research Projects PUCP 2023) CAP 2023 of the Pontificia Universidad Católica del Perú. The project is number PI1018 and is entitled "Improving the perception of entrepreneurs towards the formalization of businesses in the convenience store sector of Lima, Peru, year 2023: a research proposal with social impact."

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.mex.2025.103299.

References

[1] S. Arnab, T. Lim, M.B. Carvalho, F. Bellotti, S. De Freitas, S. Louchart, et al., Mapping learning and game mechanics for serious games analysis, Br. J. Educ. Technol. 46 (2) (2015) 391–411.
[2] A.M.A. de Araujo Pistono, A.M.P. Santos, R.J.V. Baptista, An initial framework for adaptive serious games based on a systematic literature review, in: Int. Conf. Technol. Innov. Learn. Teach. Educ, Springer Nature Switzerland, Cham, 2022, pp. 43–51.
[3] V. Venkatesh, H. Bala, Technology acceptance model 3 and a research agenda on interventions, Decis. Sci. 39 (2) (2008) 273–315.
[4] F.D. Davis, Perceived usefulness, perceived ease of use, and user acceptance of information technology, MIS Q 13 (3) (1989) 319–340.
[5] I. Ajzen, The theory of planned behavior, Organ. Behav. Hum. Decis. Process. 50 (2) (1991) 179–211.
[6] R.F. Malaquias, F.F. Malaquias, Y. Hwang, Understanding technology acceptance features in learning through a serious game, Comput. Hum. Behav. 87 (2018) 395–402.
[7] M.P. Ting, C.D. Min, What drives user churn in serious games? An empirical examination of the TAM, SOR theory, and game quality in Chinese cultural heritage games, Entertain. Comput. 52 (2025) 100758.
[8] T.M. Kuang, L. Agustina, Y. Monalisa, Acceptance of digital game-based learning by accounting and business lecturers: empirical evidence from Indonesia based on the extended Technology Acceptance Model, Account. Educ. 33 (4) (2024) 391–413.
[9] R. Silva, R. Rodrigues, C. Leal, Games-based learning in accounting education – which dimensions are the most relevant? Account. Educ. 30 (2) (2021) 159–187.
[10] J. Holmström, M. Ketokivi, A.P. Hameri, Bridging practice and theory: a design science approach, Decis. Sci. 40 (1) (2009) 65–87.
[11] C.M. Ringle, S. Wende, J.M. Becker, SmartPLS 4, SmartPLS, Oststeinbek, 2022 Retrieved March 13, 2023.
[12] J.F. Hair, G.T.M. Hult, C.M. Ringle, M. Sarstedt, A Primer On Partial Least Squares Structural Equation Modeling (PLS-SEM), SAGE, 2021.
[13] J.F. Hair, G.T.M. Hult, C.M. Ringle, M. Sarstedt, A Primer On Partial Least Squares Structural Equation Modeling (PLS-SEM), 2nd ed., SAGE, Thousand Oaks, 2017.
[14] J. Saldaña, The Coding Manual For Qualitative Researchers, SAGE, London, 2021.
[15] J. Cohen, A power primer, Psychol. Bull. 112 (1992) 155–159.
[16] F. Faul, E. Erdfelder, A. Buchner, A.G. Lang, Statistical power analyses using G* Power 3.1: tests for correlation and regression analyses, Behav. Res. Methods. 41 (4) (2009) 1149–1160.