External and Model Validity in A Systematic Review of Game-Based Interventions

Franz Coelho[1,2], Daniel Gonçalves[3], Ana Maria Abreu[1,4]

Universidade Católica Portuguesa, Centro de Investigação Interdisciplinar em Saúde (CIIS), Portugal[1]
Universidade Católica Portuguesa, Faculdade de Ciências da Saúde e Enfermagem (FCSE), Portugal[2]
INESC-ID and Instituto Superior Técnico, University of Lisbon, Portugal[3]
Forward College, Portugal[4]

Abstract

Game-based learning, training, exercises, serious games, and gamification represent distinct approaches, integrating games into diverse contexts. Lately, interventions based on these approaches have gained popularity due to their potential to enhance cognitive outcomes. The term game-based intervention (GBI) was adopted to describe the use of all these playful processes with the goals of cognition and behavior promotion. Here, we present complementary research originating from a comprehensive systematic review examining the influence of GBI on adult cognition. This additional research evaluates the external and model validity of the original studies included in a systematic review, adhering to a registered PROSPERO protocol and PRISMA guidelines. Our systematic methodology covered various databases, resulting in 1398 articles. Following examination, 42 studies (26 randomized control trials and 16 non-randomized control trials) were selected. External and model validity were assessed using the External Validity Assessment Tool (EVAT©). Half of the studies inadequately described recruitment (48%), and most poorly outlined participation (71%) in terms of external validity, obscuring the results' generalizability. However, most studies adequately described model validity (88%), clarifying the comprehension of staff, places, and facilities used. While the systematic review showcased encouraging results regarding the impact of GBI on adult cognition, the evaluation of external and model validity conducted here revealed challenges in generalizing these findings to real-world settings and other populations beyond the laboratory context. However, it underscored that the contextual environment and operational procedures are conducive to replicability.

Keywords: *Games, Technology, Digital, Education, Learning, Cognition*

1. Introduction

Internal validity indicates the validity of the study results and conclusions within the study population [1]. In contrast, external validity pertains to the degree to which the findings and interpretations derived from a study can be applied to a broader population or other similar groups [2]. Factors influencing external validity encompass the representativeness of the study sample, the similarity of study conditions and interventions to real-life settings, and the generalizability of outcomes across various contexts [3]. Model validity, also known as ecological validity, considers elements like etiology, contextual environment, and operational procedures [1]. Together with external validity, it reflects how a study can be replicated for other populations. While empirical research commonly focuses on threats to internal validity, external validity receives less systematic attention [4]. Meta-studies could be used to address research challenges by identifying key issues about external validity and offering recommendations for future studies [5]. Therefore, following a systematic review of the impact of game-based interventions (GBI) on adult cognition [6], we recognized the necessity to assess the external validity of the included studies, prompted by the methodological shortcomings identified in a substantial portion of the considered literature. Our objective is to examine how the findings originating from the literature on GBI can be generalized and replicated across different populations, improving comprehension of how GBI can serve as effective tools for improving adult cognition in real world contexts.

Through our systematic review [6], we aimed to investigate how game-based interventions (GBI) might impact healthy adult cognition. The term GBI was used to cover all different approaches in which skills and behaviors can be promoted through playful processes influenced by games [7], i.e., game-based learning, game-based training, game-based exercises, serious games, and gamification. These approaches can be digital, non-digital, or a combination of both [8], [9]. Interventions based on these approaches have gained popularity in recent years due to their potential to improve cognitive outcomes [10], [11]. Cognition involves vital mental processes like selecting, organizing, and integrating information [12]. These processes include categories such as attention, [13], working memory [14], executive functions (idea generation, resisting temptations, and maintaining focus) [15], long-term memory [14], and learning [16].

Our systematic review was divided into six key analyses. The first analysis covered the bias assessment using the Risk of Bias tool for randomized trials (RoB 2.0), and the Risk of Bias Tool In Non-randomized Studies of Interventions (ROBINS-I) [17]. The second outlined general study characteristics like participants' nationality, gender, and age, GBI types, and cognitive domains. The third examined referenced literature and theories adopted to justify GBI use. The fourth compared digital and analog applications. The fifth analyzed study outcomes, comparing IGs and CG across various categories. The sixth investigated game elements in GBI design and their impact on adult cognition. Each of these analyses revealed distinct findings, yet overall, the reviewed literature provided promising results regarding the influence of GBI on adult cognition.

On the other hand, regarding bias assessment, we identified that the literature lacked stringent methodological standards and scientific rigor, with a mere minority of studies displaying a low risk of bias across all evaluated domains. However, the majority of bias assessments primarily focused on internal validity, overlooking a thorough examination of external validity concerns., i.e., the inclusiveness of the study participants, the alignment of study circumstances and interventions with real-world scenarios, and the applicability of outcomes across diverse environments [3]. Following the same rigorous methodology of the previous systematic review, our objective with this complementary review was to evaluate the external and model validity of the studies included in the systematic review, examining the generalizability of results and the replicability of the model and experimental design.

2. Materials and Methods

We considered the PRISMA guidelines [18] in our systematic review [6] to include the studies for the external and model validity analysis. Our systematic review was registered in PROSPERO (# CRD42021282683). The systematic search strategy, implemented in June 2022, was used across multiple databases: Web of Science, Scopus, ERIC, PubMed, APA, and ACM Digital Library. Our query adhered to the PICO framework [19], except for a specific reference to the population (adults), to avoid significantly limiting the results. Therefore, we manually identified the study sample. The final query was: (gamification OR gameful OR gamified OR "game-based" OR "serious game" OR "game-like") AND (cognition OR cognitive) AND (experiment* OR quasi-experimental OR rct OR randomized OR non-randomized OR trial OR "control group"). The query was customized for each database, considering the unique characteristics of each search engine, to ensure consistent semantics. All the primary searches were conducted by attending to the query in the title and abstract.

The inclusion criteria were determined based on the following categories. Population: Only neurotypical healthy adults over 18 were considered. Intervention: Included studies involved gamification, game-based training, game-based learning, serious games, and exergames as intervention groups (IG). Control: Original non-randomized (NRCT) and randomized control trials (RCT) with active control groups (CG) were included. Outcomes: Studies related to cognition were included.

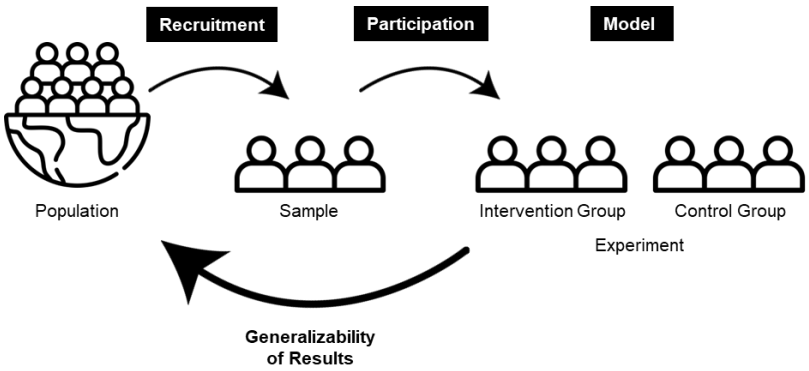
The exclusion criteria were determined based on the following categories. Publication date: Studies before 2011 were excluded. Type of study: Non-peer-reviewed sources such as dissertations were excluded. Language: Non-English studies were excluded.

2.1 External and Model Validity Instrument

Here we assessed the studies included in a previous comprehensive systematic review and focused on the evaluation of external and model validity using the EVAT©. This tool analyses the generalizability of research findings to other individuals (external validity) and different settings (model validity) of both RCT and NRCT [1]. The EVAT instrument comprises three core domains assessing the external validity (Domains 1 and 2) and model validity (Domain 3) of each study [1]. Domain 1

(Recruitment) evaluates the identification and selection of participants from the source population. Domain 2 (Participation) focuses on the description of the study sample representative of the entire source population. Domain 3 (Model Validity) assesses the clear description of staff, setting, and intervention characteristics. Figure 1 demonstrates the relationship between these Domains and the generalizability of results. Each domain is rated as well covered (++), adequately addressed (+), poorly addressed (-), or not applicable (0), based on the study's reporting quality.

Fig.1. EVAT Domains



Note: Adapted from “External Validity and Model Validity: A Conceptual Approach for Systematic Review Methodology,” by R. Khorsan and C. Crawford, 2014, *Evidence-Based Complementary and Alternative Medicine*, vol. 2014, p. 694804

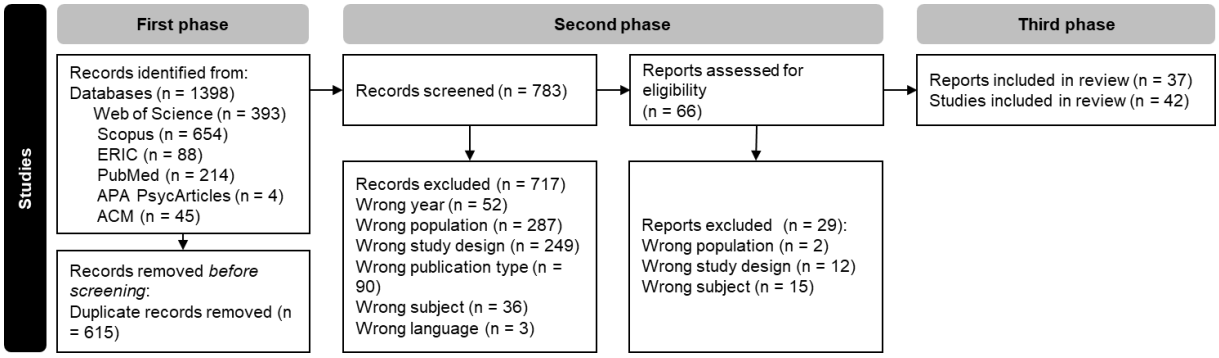
3. Results

3.1 Selection Data

We gathered a total of 1398 articles from various databases. After removing 615 duplicates, we assessed 783 titles and abstracts based on our inclusion and exclusion criteria. Subsequently, we identified 66 reports for further examination, ultimately including 37 in our review after full-text analysis. Among these, one report comprised three studies, and three reports contained two studies each, resulting in a total of 42 studies. The breakdown of study methods revealed that 26 employed RCT (62%) and 16 utilized NRCT (38%). See Figure 2 for a visual representation of this process in the PRISMA flow diagram. The list of all studies included is hosted in the Open Science Framework (OSF) [20].

As depicted in the PRISMA flow diagram, the criterion leading to the most reports excluded during screening and eligibility assessment was the wrong population selection, accounting for 289 records (37% of those 783 screened). The main reasons for these exclusions were: the inclusion of a healthy population under 18 years old, an unhealthy population (with any disease or disorder) over 18 years old, or an unhealthy population under 18 years old.

Fig.2. PRISMA Flow Diagram



3.2 External and Model Validity Assessment

Half of the studies (48%) provided inadequate descriptions of recruitment, while the majority (71%) poorly outlined participation, impacting external validity and obscuring generalizability. However, model validity was effectively described in most studies (88%), enhancing understanding of staff, locations, and facilities involved. EVAT data are shown in Table 1.

Table 1. EVAT

Reference	DOI/URL	Study Design	Recruitment	Participation	Model Validity
Boeker et al. (2013)	https://doi.org/10.1371/journal.pone.0082328	RCT	+	-	++
Baniqued et al. (2015)	https://doi.org/10.1371/journal.pone.0142169	RCT	++	+	++
Wang et al. (2015)	https://doi.org/10.1108/ITP-03-2013-0053	RCT	-	-	++
Schättin et al. (2016)	https://doi.org/10.3389/fnagi.2016.00278	RCT	++	++	-
Smith et al. (2017)	https://doi.org/10.1177/1046878117731888	NRCT	-	-	-
Chang et al. (2017)	https://doi.org/10.1016/j.chb.2017.01.031	NRCT	-	-	++
Kühn et al. (2017)	https://doi.org/10.1016/j.neuroimage.2017.05.026	RCT	++	+	-
Dankbaar et al. (2017a)	https://doi.org/10.1097/SIH.0000000000000194	NRCT	+	+	++
Dankbaar et al. (2017b)	https://doi.org/10.1186/s12909-016-0836-5	RCT	-	-	++
Dunbar et al. (2017) (1)	https://doi.org/10.4018/IJGBL.2017100105	RCT	+	-	+
Dunbar et al. (2017) (2)	https://doi.org/10.4018/IJGBL.2017100105	RCT	+	-	+
Wardaszko & Podgórski (2017)	https://doi.org/10.1177/1046878117704350	RCT	+	++	++
Rhodes et al. (2017) (1)	https://doi.org/10.1177/1555412016686642	RCT	-	-	+
Rhodes et al. (2017) (2)	https://doi.org/10.1177/1555412016686642	RCT	-	-	+
Lumsden et al. (2017)	https://doi.org/10.2196/jmir.8473	RCT	+	-	+
Thomas (2017)	https://files.eric.ed.gov/fulltext/EJ1160632.pdf	RCT	-	-	+
Brom et al. (2018)	https://doi.org/10.1177/0735633118797330	RCT	-	-	+
Kelders et al. (2018)	https://doi.org/10.2196/jmir.9923	RCT	+	-	+
Wu (2018)	https://doi.org/10.1080/14703297.2016.1250662	NRCT	-	-	+
Chang et al. (2018)	https://pdfs.semanticscholar.org/2831/44dda43467a291f7cf8f0163b29a00a993b2.pdf	NRCT	-	-	+
Lee et al. (2018a)	https://doi.org/10.2196/jmir.8987	RCT	++	+	+
Huang & Ho (2018)	https://doi.org/10.1080/10494820.2017.1374979	RCT	-	-	+
Shaw et al. (2018)	https://doi.org/10.1027/1864-1105/a000174	RCT	-	-	+
Lee et al. (2018b)	https://doi.org/10.2196/mededu.9237	RCT	++	-	+
Friehs et al. (2020)	https://doi.org/10.2196/17810	RCT	-	-	-
Legaki et al. (2021)	https://doi.org/10.1016/j.techfore.2021.120725	RCT	+	+	+
Ma et al. (2021)	https://doi.org/10.1016/j.nedt.2021.104923	RCT	+	+	+
Bernecker & Ninaus (2021)	https://doi.org/10.1016/j.chb.2020.106542	NRCT	+	+	+
Bernecker & Ninaus (2021)	https://doi.org/10.1016/j.chb.2020.106542	NRCT	+	+	+
Groening & Binnewies (2021)	https://doi.org/10.1080/10447318.2020.1870828	NRCT	-	-	+
Groening & Binnewies (2021)	https://doi.org/10.1080/10447318.2020.1870828	NRCT	-	-	+
Groening & Binnewies (2021)	https://doi.org/10.1080/10447318.2020.1870828	NRCT	-	-	+
Yang et al. (2021)	https://doi.org/10.1016/j.compedu.2020.104057	RCT	-	-	+
Chan et al. (2021)	https://doi.org/10.1108/IJAIM-07-2021-0136	NRCT	-	-	+
Adams & Toh (2021)	https://doi.org/10.34190/EJEL.19.6.2546	RCT	+	-	+
Luengvilai et al. (2021)	https://doi.org/10.47836/PJSSH.29.1.39	NRCT	+	+	+
Cechella et al. (2021)	https://doi.org/10.1016/j.chbr.2020.100044	NRCT	+	+	++
Bakri et al. (2021)	https://files.eric.ed.gov/fulltext/EJ1288043.pdf	NRCT	-	-	+
Moradi & Noor (2022)	https://doi.org/10.1109/ACCESS.2022.3140434	NRCT	+	-	++
Kyung-Mi (2022)	https://doi.org/10.3991/ijet.v17i03.26349	NRCT	+	-	++
Ye et al. (2022)	https://doi.org/10.1080/10494820.2022.2042032	RCT	-	-	+
Redlinger et al. (2022)	https://doi.org/10.2196/35295	RCT	+	-	-
	% of Poorly Addressed (-)		48%	71%	12%

Also, Table 2 illustrates the primary reasons for classifying some studies as poorly addressed in certain domains, along with a concise description used for each domain classification.

Table 2. EVAT Details

Domain	Description	Main reasons for poor addressing
Recruitment	Identification of the source population for participants and description of how the participants were recruited from that source population	Inadequate description to specify the recruitment source, providing only vague information, e.g., "from a university".
Participation	Representativeness of the participants with the entire source population from which they were recruited	The absence of participant profiles, coupled with a lack of information regarding the inclusion and exclusion criteria of participants, undermines the description of the sample characteristics.
Model validity	Representativeness of the staff, places, and facilities where the patients were treated with the treatment that most patients would typically receive	The utilization of laboratory equipment, i.e., fMRI, eye-tracking, and EEG, during the experiment, challenges the classification of the intervention as representative of real-world treatment experiences for most participants.

4. Discussion

In this study, we conducted an additional analysis of a systematic review [6] focusing on original research investigating how GBI affects adult cognition. In the systematic review, we cast a wide net across multiple databases, retrieving 1398 articles. Employing a meticulous approach, we sifted through the data and selected 42 studies with comprehensive descriptions. Among these, 26 utilized the RCT method, while 16 employed a NRCT approach. Even though the systematic review highlighted promising findings regarding the influence of GBI on adult cognition [6], we recognized the pervasive lack of methodological rigor in the included studies during the risk of bias analysis. Thus, we identified the need to enhance the validity analysis and undertook this focused review to evaluate the external validity of the same studies. Our findings uncovered challenges in external validity within the studies, although model validity is adequately addressed.

Here we revealed deficiencies in the studies' descriptions regarding the domains of recruitment and participation. Inadequate recruitment descriptions lacked specificity about participant origin, i.e., 65% of the studies indicated recruitment from a university but lacked further details, while 35% provided no information about the recruitment source, hindering understanding of the population characteristics. Participation issues predominantly stemmed from vague inclusion and exclusion criteria (70%), resulting in generic or heterogeneous participant profiles, and the population description was overly comprehensive (23%), potentially impacting the representativeness of the participants. These two domains represent the studies' external validity [1], which is crucial for determining the practical implications and broader significance of research findings, allowing researchers to assess their work's potential impact and relevance beyond the immediate study population and setting. The domains assess the applicability and relevance of research findings to real-world situations. Thus, our review highlights challenges in generalizing the studies' findings to broader populations and real-world contexts due to superficial participant descriptions.

Model validity, alternatively termed ecological validity, extends beyond participant eligibility criteria and delves into the conceptual framework, encompassing factors such as etiology, contextual setting, and operational practices [1]. Together with the domains of recruitment and participation, it reflects how a study can be replicated in other populations. We reveal that most studies effectively described model validity, enhancing understanding of staff, locations, and facilities involved. The primary factor contributing to a few studies receiving low scores in model validity was the utilization of laboratory equipment to evaluate specific outcomes during an experiment (80%) and employing IG and CG from different studies with distinct models (20%). This aspect undermines the classification of the intervention as representative of a treatment that most participants would typically encounter in real-world settings. Henceforth, apart from that specific scenario with the minority of studies, the majority experimented within contexts that participants would ordinarily encounter in real life.

Gamification and game science have emerged as an interdisciplinary field spanning education, psychology, human-computer interaction, computer, social, and health sciences, among others [21], [22], [23]. A bibliometric analysis focusing on game concepts in digital learning environments has revealed a substantial expansion in the GBI [24]. The study found that in 2000 only 44 articles were published, whereas by 2017, this number had surged to 1396, indicating significant

growth in GBI research over this period. Consequently, research on games and their diverse applications is relatively new, encountering challenges inherent to an emerging scientific field that requires further development and consolidation to reach high research standards that sustain and explain our findings about external validity. A recent meta-analysis on gamification and education [25] highlights the growing interest in using gamification for learning due to its potential positive impacts on cognitive, motivational, and behavioral outcomes. However, the study also indicates that diverse research findings suggest unresolved factors affecting its effectiveness, emphasizing the requirement of high-quality studies with robust designs and controls to clarify the relationship between gamification and learning.

Conversely, the contextual environment and operational procedures in GBI studies have been effectively implemented, demonstrating high standards in facility quality and design for model validity assessment. Despite the methodological hurdles faced in external validity, GBI research appears to be in conditions conducive to easy replication in other interventions. This could motivate future research in the field to embrace similar experimental frameworks, enhancing data quality through improved recruitment and participation protocols.

In our initial systematic review, we encountered promising findings regarding the influence of GBI on adult cognition [6]. GBI encompass different approaches aimed at promoting skills and behaviors through games [7]. Within this framework, learning processes, including game-based learning interventions, play a pivotal role. However, as evidenced in this analysis, the recruitment and participant selection processes were inadequately addressed in the studies examined, thus compromising the potential efficacy of GBIs on cognition and learning outcomes. This lack of clarity hampers understanding regarding how the characteristics of the study population may influence the observed effects. Nevertheless, GBI appear to exhibit high replicability and feasibility in terms of experimental design quality (model validity), rendering them a viable option for educators and researchers seeking to implement innovative pedagogical approaches in educational settings. To enhance the external validity of future research endeavors, we strongly advocate for the comprehensive documentation of participant profiles and recommend validating previous research designs with diverse student populations. Therefore, in addition to benefiting the scientific community by improving data quality, these game-based approaches may contribute to an enhanced educational environment.

It is important to adhere to PRISMA guidelines [18] and the Cochrane Collaboration's recommendations [17], when conducting systematic reviews, which require evaluation of the internal validity of included studies to assess their risk of bias. The Cochrane Collaboration's recommendations [17] underscore that this risk pertains to the potential of either overestimating or underestimating the true effect of the intervention and it should not be conflated with external validity, which concerns the extent to which study results can be applied to other populations and settings. Thus, there is a notable gap in the analysis of external validity within many systematic reviews, as it's not explicitly addressed in established guidelines. This oversight can hinder the generalizability of findings not only in research concerning GBI and cognition but also across systematic reviews that strictly adhere to these recommendations. Although some systematic reviews do prioritize exploring the external validity of studies [26], [27], [28], there remains a gap in the broader literature following systematic methodologies.

In the end, both the original systematic review [6] and this subsequent review concentrated solely on healthy adults. However, as highlighted in the Results section, a notable portion (37%) of the screened records pertained to unhealthy adults, teenagers, children, or healthy teenagers and children. This underscores the growing interest in applying GBI within these other populations and illustrates how this scientific field extends to numerous other application areas. It presents intriguing paths for further exploration in this promising, interdisciplinary scientific field.

5. Conclusion

Our analysis revealed a notable need for more recruitment and participation domains across the literature related to GBI and cognition identified in a previous comprehensive systematic review, limiting the generalizability of results and external validity. Conversely, the well-documented application of staff, places, and facilities demonstrates adequate model validity, enhancing the applicability of experimental designs. Therefore, it is essential to enhance the caliber of research in the field of GBI and cognition, given that many findings cannot be extrapolated to other populations, thus constraining the practical implications of interventions in real-world contexts. Despite these challenges, the experimental designs appear to incorporate protocols encompassing staff, locations,

and facilities that can be readily replicated in other research endeavors, thereby facilitating the implementation of similar interventions.

Our initial systematic review [6] unveiled encouraging outcomes concerning the impact of GBI on adult cognitive processes. Nevertheless, this scrutiny of external validity exposes certain reservations regarding the generalizability of the identified results. While GBI exhibits promise and potential for replication across diverse contexts (i.e., model validity), uncertainties persist regarding its efficacy within specific populations (i.e., external validity). External and model validity assessment should be more comprehensively examined in original research and meta-studies, as the focus often centers solely on internal validity or risk of bias [4]. As widely recognized guidelines such as PRISMA and the Cochrane Collaboration's recommendations do not mandate specific external validity analysis, questions arise regarding the generalizability of results presented in systematic reviews. This compromises the applicability of interventions analyzed through systematic reviews in real-world settings. Therefore, by integrating this external validity concern into original research about GBI, the generalizability of results may improve, rendering outcomes more representative of real-life educational scenarios and ultimately enhancing their impact on learning processes beyond laboratory confines.

Finally, while our focus remains on the impact of GBI on cognition in healthy adults, a substantial body of literature exists addressing GBI in other populations. This highlights the promising nature of game research as a scientific field, encompassing diverse contexts and evolving treatment and intervention approaches for youth and individuals with disorders or diseases.

6. Limitations and Recommendations for Future Research

This study is an additional research of a comprehensive systematic review adhering to PRISMA guidelines [6], aimed to explore GBI's impact on adult cognition. Thus, the studies included in our analysis were delimited by the search criteria utilized, the scholarly databases referenced, the criteria for inclusion and exclusion, and the time frame applied in selecting articles.

Furthermore, it is acknowledged that certain details may be obscured in some articles to safeguard data privacy, such as the identity of the educational institution from which participants were recruited. This absence of information regarding the participants' institutional affiliations may impact the external validity of the data, as it precludes a comprehensive understanding of the population's origins. Consequently, in such instances, we advocate for further research endeavors aimed at elucidating these privacy considerations and providing alternative contextual details that could offer insights into the participants' backgrounds without compromising confidentiality.

REFERENCES

[1] R. Khorsan and C. Crawford, "External Validity and Model Validity: A Conceptual Approach for Systematic Review Methodology," *Evidence-Based Complementary and Alternative Medicine*, vol. 2014, p. 694804, Mar. 2014, doi: https://doi.org/10.1155/2014/694804.

[2] M. G. Findley, K. Kikuta, and M. Denly, "External Validity," *Annual Review of Political Science*, vol. 24, no. 1, pp. 365–393, 2021, doi: https://doi.org/10.1146/annurev-polisci-041719-102556.

[3] N. Egami and E. Hartman, "Elements of External Validity: Framework, Design, and Analysis," *American Political Science Review*, vol. 117, no. 3, pp. 1070–1088, Aug. 2023, doi: https://doi.org/10.1017/S0003055422000880.

[4] C. M. Loyka, J. Ruscio, A. B. Edelblum, L. Hatch, B. Wetreich, and A. Zabel, "Weighing People Rather Than Food: A Framework for Examining External Validity," *Perspect Psychol Sci*, vol. 15, no. 2, pp. 483–496, Mar. 2020, doi: https://doi.org/10.1177/1745691619876279.

[5] T. Slough and S. A. Tyson, "External Validity and Meta-Analysis," *American Journal of Political Science*, vol. 67, no. 2, pp. 440–455, 2023, doi: https://doi.org/10.1111/ajps.12742.

[6] F. Coelho, D. Gonçalves, and A. M. Abreu, "The Impact of Game-based Interventions in Adult Cognition: A Systematic Review," Manuscript submitted for publication, 2024.

[7] F. Ferreira-Brito *et al.*, "Game-based interventions for neuropsychological assessment, training and rehabilitation: Which game-elements to use? A systematic review," *Journal of biomedical informatics*, vol. 98, p. 103287, 2019, doi: https://doi.org/10.1016/j.jbi.2019.103287.

[8] F. Coelho and A. M. Abreu, "The Corporate (Magic) Circle: Fun Work or Controlled Play?," *TechTrends*, vol. 67, no. 1, pp. 160–177, Jan. 2023, doi: https://doi.org/10.1007/s11528-022-00776-z.

[9] M. T. Soo and H. Aris, “Game-based learning in requirements engineering: an overview,” in *2018 IEEE Conference on e-Learning, e-Management and e-Services (IC3e)*, IEEE, 2018. doi: https://doi.org/10.1109/IC3e.2018.8632650.
[10] J. Krath, L. Schürmann, and H. F. O. von Korflesch, “Revealing the theoretical basis of gamification: A systematic review and analysis of theory in research on gamification, serious games and game-based learning,” *Computers in Human Behavior*, vol. 125, p. 106963, Dec. 2021, doi: https://doi.org/10.1016/j.chb.2021.106963.
[11] I. H. López-Nava *et al.*, “Current state and trends of the research in exergames for the elderly and their impact on health outcomes: a scoping review,” *J Ambient Intell Human Comput*, Aug. 2022, doi: https://doi.org/10.1007/s12652-022-04364-0.
[12] J. L. Plass and S. Kalyuga, “Four Ways of Considering Emotion in Cognitive Load Theory,” *Educ Psychol Rev*, vol. 31, no. 2, pp. 339–359, Jun. 2019, doi: https://doi.org/10.1007/s10648-019-09473-5.
[13] B. Bruya and Y.-Y. Tang, “Is Attention Really Effort? Revisiting Daniel Kahneman’s Influential 1973 Book Attention and Effort,” *Frontiers in Psychology*, vol. 9, 2018, doi: https://doi.org/10.3389/fpsyg.2018.01133.
[14] A. Baddeley, M. W. Eysenck, and M. C. Anderson, *Memory*, 3rd ed. London: Routledge, 2020. doi: https://doi.org/10.4324/9780429449642.
[15] A. Diamond and D. S. Ling, “Review of the evidence on, and fundamental questions about, efforts to improve executive functions, including working memory,” in *Cognitive and working memory training: Perspectives from psychology, neuroscience, and human development*, New York, NY, US: Oxford University Press, 2020, pp. 143–431. doi: https://doi.org/10.1093/oso/9780199974467.003.0008.
[16] R. Gross, *Psychology: The science of mind and behaviour*, 8th ed. Hodder Education, 2020.
[17] J. Higgins *et al.*, *Cochrane Handbook for Systematic Reviews of Interventions version 6.4 (updated August 2023)*. Cochrane, 2023. Accessed: May 23, 2023. [Online]. Available: https://training.cochrane.org/handbook
[18] M. J. Page *et al.*, “The PRISMA 2020 statement: an updated guideline for reporting systematic reviews,” *BMJ*, vol. 372, p. n71, Mar. 2021, doi: https://doi.org/10.1136/bmj.n71.
[19] D. Pati and L. N. Lorusso, “How to Write a Systematic Review of the Literature,” *HERD*, vol. 11, no. 1, pp. 15–30, Jan. 2018, doi: https://doi.org/10.1177/1937586717747384.
[20] F. Coelho, D. Gonçalves, and A. M. Abreu, “External and model validity in a systematic review of game-based interventions (Appendix).” Mar. 21, 2024. doi: https://doi.org/10.17605/OSF.IO/MWEZ2.
[21] S. de Freitas, “Are Games Effective Learning Tools? A Review of Educational Games,” *Journal of Educational Technology & Society*, 2018, pp. 74–84. Accessed: Oct. 23, 2023. [Online]. Available: https://www.jstor.org/stable/26388380
[22] R. N. Landers, E. M. Auer, A. B. Collmus, and M. B. Armstrong, “Gamification science, its history and future: Definitions and a research agenda,” *Simulation & Gaming*, vol. 49, no. 3, pp. 315–337, 2018, doi: https://doi.org/10.1177/10468781187743.
[23] R. N. Landers, “Developing a Theory of Gamified Learning: Linking Serious Games and Gamification of Learning,” *Simulation and Gaming*, vol. 45, no. 6, pp. 752–768, 2014, doi: https://doi.org/10.1177/1046878114563660.
[24] S. Schöbel, M. Saqr, and A. Janson, “Two decades of game concepts in digital learning environments – A bibliometric study and research agenda,” *Computers & Education*, vol. 173, p. 104296, Nov. 2021, doi: https://doi.org/10.1016/j.compedu.2021.104296.
[25] M. Sailer and L. Homner, “The Gamification of Learning: a Meta-analysis,” *Educ Psychol Rev*, vol. 32, no. 1, pp. 77–112, Mar. 2020, doi: https://doi.org/10.1007/s10648-019-09498-w.
[26] N. González-Gálvez, G. M. Gea-García, and P. J. Marcos-Pardo, “Effects of exercise programs on kyphosis and lordosis angle: A systematic review and meta-analysis,” *PLOS ONE*, vol. 14, no. 4, p. e0216180, Apr. 2019, doi: https://doi.org/10.1371/journal.pone.0216180.
[27] C. Crawford *et al.*, “The Impact of Massage Therapy on Function in Pain Populations—A Systematic Review and Meta-Analysis of Randomized Controlled Trials: Part I, Patients Experiencing Pain in the General Population,” *Pain Medicine*, vol. 17, no. 7, pp. 1353–1375, Jul. 2016, doi: https://doi.org/10.1093/pm/pnw099.
[28] I. D. Coulter *et al.*, “Manipulation and mobilization for treating chronic low back pain: a systematic review and meta-analysis,” *The Spine Journal*, vol. 18, no. 5, pp. 866–879, May 2018, doi: https://doi.org/10.1016/j.spinee.2018.01.013.