Check for updates

The use of serious games for psychological education and training: a systematic review

Antonio Pio Facchino[1†], Daniela Marchetti[1†], Marco Colasanti[2*], Lilybeth Fontanesi[1] and Maria Cristina Verrocchio[1]

[1]Department of Psychology, G. d'Annunzio University of Chieti-Pescara, Chieti, Italy, [2]Department of Psychological, Health, and Territorial Sciences, G. d'Annunzio University of Chieti-Pescara, Chieti, Italy

Introduction: The present systematic review aims to synthesize and critically analyze the use of serious games in the professional training and education of psychologists and psychology students.

Methods: Following PRISMA guidelines, database searches from inception to July 2023 (PsycINFO, PubMed, Web of Science, and Scopus) yielded 4,409 records, of which 14 met the eligibility criteria, including 17 studies. Quality assessment was performed using the Newcastle-Ottawa Scale and the Risk of Bias Tool for Randomized Trials.

Results: The review identified three pivotal areas where serious games demonstrated significant educational impact: enhancing psychological traits and attitudes (e.g., prejudice, empathy), promoting theoretical knowledge acquisition (e.g., biopsychology), and developing professional skills (e.g., investigative interview with children). Serious games, particularly those providing feedback and modeling, significantly enhance the quality of learning and training for psychology students and professionals.

Discussion: Key findings revealed that serious games operate by offering realistic, engaging, and flexible learning environments while mitigating risks associated with real-world practice. Methodological limitations, including moderate to high risk of bias in many studies, especially those that relied on cross-sectional data, underscore the need for rigorous designs and long-term evaluations. Practical implications suggest integrating serious games into curricula to address gaps in experiential learning for psychologists, facilitating skill development and knowledge retention. Future research should explore the long-term impact of serious games on professional competencies and assess their applicability across diverse educational contexts.

1 Introduction

In recent years, in parallel with the growth of multimedia technologies, the use of interactive software specifically designed for pedagogical purposes has surged (Girard et al., 2013; Marchetti et al., 2015). In this context, two distinct methodologies can be identified (Landers, 2014): on the one hand, serious games, which incorporate all game elements and are explicitly designed to provide instructional content, typically within a computer-based environment (Gentry et al., 2019; Wouters et al., 2013); on the other, gamification, which refers to the use of game design elements in non-game contexts to enhance user engagement and motivation (Deterding et al., 2011). Gamification often employs discrete game components,

such as points, badges, and leaderboards, to complement pre-existing instructional content. For instance, users completing an electronic learning module may be rewarded with badges or points (Gentry et al., 2019). While both approaches aim to leverage the motivational and experiential benefits of games, focusing on education rather than entertainment (Fraticelli et al., 2016), they differ fundamentally in scope and intent. Serious games are created as fully realized games with an educational purpose, often featuring avatars or interactive environments designed to immerse the learning (Miller et al., 2011). In contrast, gamification integrates specific elements characteristic of games (e.g., feedback systems and goal-setting mechanisms) into otherwise non-game activities to foster a sense of "gamefulness" (Deterding et al., 2011). The distinction lies in the extent to which game structures are applied: Serious games embody complete game ecosystems, whereas gamification adopts selected design principles to enhance existing processes.

Compared to traditional approaches, serious games offer unique advantages by complementing traditional teaching methods and bridging the gap between theoretical knowledge and practical application. Traditional approaches often lack interactivity and experiential learning opportunities, whereas serious games immerse learners in dynamic, simulated environments where they can actively engage with the material and apply their knowledge in meaningful ways (Ke et al., 2016; Wouters et al., 2013). Additional benefits of serious games include enhanced learner engagement and motivation (Breuer and Bente, 2010; Söbke et al., 2020), as well as the opportunity for immediate feedback (Boyle et al., 2011), which promotes iterative learning and allows users to adjust their approaches in real time. Consequently, serious games have been implemented in various domains (Laamarti et al., 2014) with different applications, such as training, skill development, and knowledge acquisition.

Particularly, there has been a growing interest in serious games as a pedagogical tool in the education of healthcare professionals. International literature suggests that they can be an effective way to teach complex practical skills, such as technical and non-technical skills in the surgical field (e.g., Graafland et al., 2012) or diagnosis and treatment in the medical and nursing professions (e.g., Buajeeb et al., 2023; Craig et al., 2023). Various applications have also been developed for mental health, such as behavior change (Hammady and Arnab, 2022), development of metacognitive skills (Amo et al., 2023), treatment of psychopathological symptoms such as depression (Gómez-Cambronero et al., 2023), chronic pain (Stamm et al., 2022), disordered eating behaviors (Tang et al., 2022), and so on.

However, while the usefulness of serious games for training medical students and professionals has been extensively evaluated (Gentry et al., 2019; Maheu-Cadotte et al., 2021; Sipiyaruk et al., 2018; Sitzmann, 2011; Wouters et al., 2013), albeit with somewhat inconsistent results, little is known about the efficacy of serious games for training mental health professionals, specifically psychologists or psychotherapists. This pedagogical methodology could be particularly beneficial, as it would address some of the existing shortcomings in the training of these professional profiles. For example, the international literature repeatedly highlighted that students encounter difficulties in learning different psychological techniques, such as testing (e.g., Viglione et al., 2017) and assessment (e.g., Pompedda et al., 2022), mainly due to the limited availability of practical training provided in university curricula and to the difficulty of gaining practical experience with real patients. Currently, the only way to

ensure quality training for psychology students and psychologists is to provide continuous feedback and supervision over time, but it remains unclear how much supervision and feedback are needed to achieve long-term results (Lamb, 2016; Lamb et al., 2002). In response to these challenges, serious games have some characteristics that could make them particularly attractive for educational programs in the field of psychology: first, they provide a realistic environment in which students can safely learn and make mistakes without real-life consequences (Mikropoulos and Natsis, 2011); second, as an interactive and engaging medium, students are not just passive recipients but play an active role in learning (Bellotti et al., 2013; Greitzer et al., 2007); third, they provide greater flexibility (Heyselaar et al., 2017); and finally, trainers can provide unbiased feedback more efficiently (Pompedda, 2018).

To the authors' knowledge, empirical studies on the use of serious games in psychological education and training have not been systematically reviewed so far. Consequently, the aim of this systematic review is threefold: first, to survey the use of serious games in the professional training and education of psychologists and psychology students; second, to evaluate the effects of serious games on the quality of learning; and third, to assess the characteristics that make serious games effective and the limitations within which they operate.

2 Materials and methods

2.1 Initial search strategy

A preliminary exploration of the literature was conducted to identify keywords related to the effectiveness of serious games in training within psychology and healthcare professions. Through this process, a set of search terms was identified, prioritizing general terms where feasible (such as assessment, testing, and evaluation). When deemed necessary, technology-specific keywords, such as "avatar," were also included in the search strategy to ensure comprehensive coverage of the relevant literature. Likewise, although this review focused on serious games, the term "gamification" was also included because, as will be discussed in greater detail subsequently, there is a certain degree of confusion surrounding the differences between the two terms (Warsinsky et al., 2021).

The present systematic review was carried out following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Page et al., 2021).

The literature search included all publication years until the date of the systematic search (July 2023). For each database (Scopus, PsycINFO, Web of Science, PubMed), search terms related to the psychological profession, psychological activities, and serious games

TABLE 1 Search terms employed.

Category	Search terms
Psychologists	psycholog* OR psychotherap*
Psychological activities	assessment OR test OR interview* OR diagnos* OR evaluation
Serious games	avatar OR serious gam* OR gamification OR applied gam*

(see Table 1 for more detailed information on search terms). No filters or limits were applied, except for the search conducted on Scopus, in which terms were searched within "Article title, Abstract, Keywords."

2.2 Inclusion and exclusion criteria

Studies were considered eligible and included in the systematic review if they: (a) were original research articles published in peer-reviewed journals; (b) were written in English, French, or Italian; (c) used serious games or avatar-based software for training psychologists; (d) clarified that the sample used consisted of students of psychological sciences, psychologists, or psychotherapists.

Studies were excluded if they: (a) were unfinished manuscripts, reviews, conference proceedings, or theoretical articles; (b) were not written in English, French, or Italian; (c) did not use serious games or avatar-based software for training psychologists; (d) did not include a sample of students of psychological sciences, psychologists, or psychotherapists.

2.3 Studies screening and selection

After retrieving results from electronic databases, all findings were collated in Mendeley Desktop (version 1.19.8), where duplicate entries were eliminated. Two reviewers independently examined the titles and abstracts of the studies identified by our search strategy to determine inclusion or exclusion. Any discrepancies between them were addressed through discussion and mutual agreement, and consulting a third researcher if needed. Both reviewers then independently assessed the full text of the eligible articles to determine their final inclusion in the review. A preliminary list of potential articles was compiled, and any disagreements were resolved by consensus, involving a third reviewer if necessary. Finally, both reviewers independently assessed the full texts of the eligible articles to determine their appropriateness for final inclusion in the review. As in the previous steps, any disagreements regarding inclusion or exclusion were resolved through discussion and, if necessary, by consulting a third researcher.

2.4 Data extraction and data synthesis

A synthesis of the evidence from all studies included in this systematic review was conducted by two reviewers independently. The extracted data included the authors' names, the year of publication, country, study design (i.e., cross-sectional studies, case–control studies, or randomized controlled trials), sample characteristics (i.e., sample size, gender, mean age, occupation), questionnaires or tools used to assess variables related to the serious games, information on outcome measures (i.e., effects on learning), and main findings on the effects of serious games in training psychological or psychotherapeutic skills. A quantitative synthesis of the evidence (meta-analysis) was not performed because of incomplete data, substantial variation among the included studies (e.g., in research design, measurements, and statistical analyses performed), and a large amount of qualitative data in the studies included in this review.

2.5 Quality assessment

The methodological quality of the included studies was assessed using two different tools, tailored to the study designs. The quality of the cohort studies was assessed using the Newcastle-Ottawa Scale (NOS; Wells et al., 2023), while for cross-sectional studies, an adapted version of the NOS was used (Herzog et al., 2013). The NOS is one of the most commonly used tools for assessing the quality of observational research (Luchini et al., 2017). The scales provide a checklist of items that assess three domains of potential bias, that are sample selection, comparability, and outcome. Cohort studies are scored between 0 and 9, while cross-sectional studies are scored between 0 and 10. For both scales, a higher score indicates a lower risk of bias. In general, a score ≤ 5 points has been identified as the cut-off for a high risk of bias (Luchini et al., 2017). To assess the quality of randomized controlled trials (RCTs), version 2 of the Risk of Bias Tool for Randomized Trials (RoB-2; Higgins et al., 2019) was used. This tool provides a checklist of items that assess five domains of potential bias, which are: risk of bias due to the randomization process; risk of bias due to deviations from the intended interventions (effect of the assigned intervention); risk of bias due to missing outcome data; risk of bias in the measurement of the outcome; risk of bias in the selection of the reported outcome. The judgments of bias for RCTs were expressed as: "low risk" when the study was assessed to have a low risk of bias across all domains; "some concern" when the study raises some concerns in at least one domain but does not present a high risk of bias in any domain; and "high risk" if the study is assessed to have a high risk of bias in at least one domain or multiple concerns that significantly reduce confidence in the result (Higgins et al., 2019).

The risk of bias was assessed by two independent reviewers for cross-sectional, case–control, and RCT studies. Any disagreement regarding the judgments was resolved by discussion and adjudication by a third reviewer.

3 Results

3.1 Study selection

A comprehensive search of all electronic databases identified 4,409 records. The selection of studies for the final synthesis involved three steps. First, all records were exported to Mendeley Desktop, a reference management tool, to remove duplicates (*n* = 1834). In the second step, two researchers independently screened the titles and abstracts of the remaining 2,575 records to remove material that was clearly not related to the research question (*n* = 2,409). All inconsistencies were discussed until a consensus was obtained, involving a third reviewer if necessary. At the end of this process, 166 records were sought for retrieval, with 13 records not retrieved. Finally, in the third step, both reviewers independently evaluated the full text of the eligible articles to determine their suitability for final inclusion in the review. As in the second step, consensus was reached by discussion on inclusion or exclusion and, if needed, by consulting the third researcher. During this step, 139 records were removed as they did not fulfill the inclusion criteria. Specifically, 32 articles were unfinished manuscripts, reviews, conference proceedings, or theoretical articles; 1 was not written in any of the included languages; 45 did not use serious games for training psychologists; 61 did not include a sample of psychology students, psychologists, or psychotherapists.

At the end of the process, 14 articles met the inclusion criteria. These articles comprised a total of 17 original studies that were included in the present systematic review (Cangas et al., 2017; Conn et al., 2023; Dancey et al., 2011; Haginoya et al., 2021, 2023; Iwamoto et al., 2017; Krach and Hanline, 2018; Olivier et al., 2019; Pompedda et al., 2015, 2020; Redondo-Rodríguez et al., 2023; Rogers et al., 2022; Segal et al., 2023; Sugden et al., 2021). Figure 1 shows the study selection process based on the PRISMA flowchart (Page et al., 2021).

3.2 Study characteristics

The selected studies were published between 2002 and 2023. Most of them ($n = 10$) were conducted in Europe, while others in the United States ($n = 2$), Australia ($n = 2$), Japan ($n = 2$), and South Africa ($n = 1$). Overall sample sizes ranged from 21 (Pompedda et al., 2015) to 102 (Redondo-Rodríguez et al., 2023). The included studies had observational or experimental designs.

Specifically, 10 were cross-sectional studies, 1 was a case–control study, and 6 were RCTs. Of the included studies, 3 investigated the role of serious games in improving psychological characteristics and attitudes of an individual that may vary in stability, defined in the present review as specific psychological traits and attitudes (i.e., aggression, prejudice and empathy, and stigma). In addition, 4 studies focused on the use of serious games in promoting theoretical knowledge and learning university subjects (i.e., biopsychology and social psychology). Finally, 10 studies examined the role of serious games in teaching practical or professional skills (i.e., investigative interviewing with children, counseling, and consulting skills). The main results are organized and described according to these uses of serious games. The main characteristics and results of the 10 cross-sectional studies, 1 case–control study, and 6 RCTs included in the systematic review are reported in Tables 2–4, respectively. A comprehensive summary of the serious games utilized across the studies and their effectiveness across various application contexts is presented in Table 5.
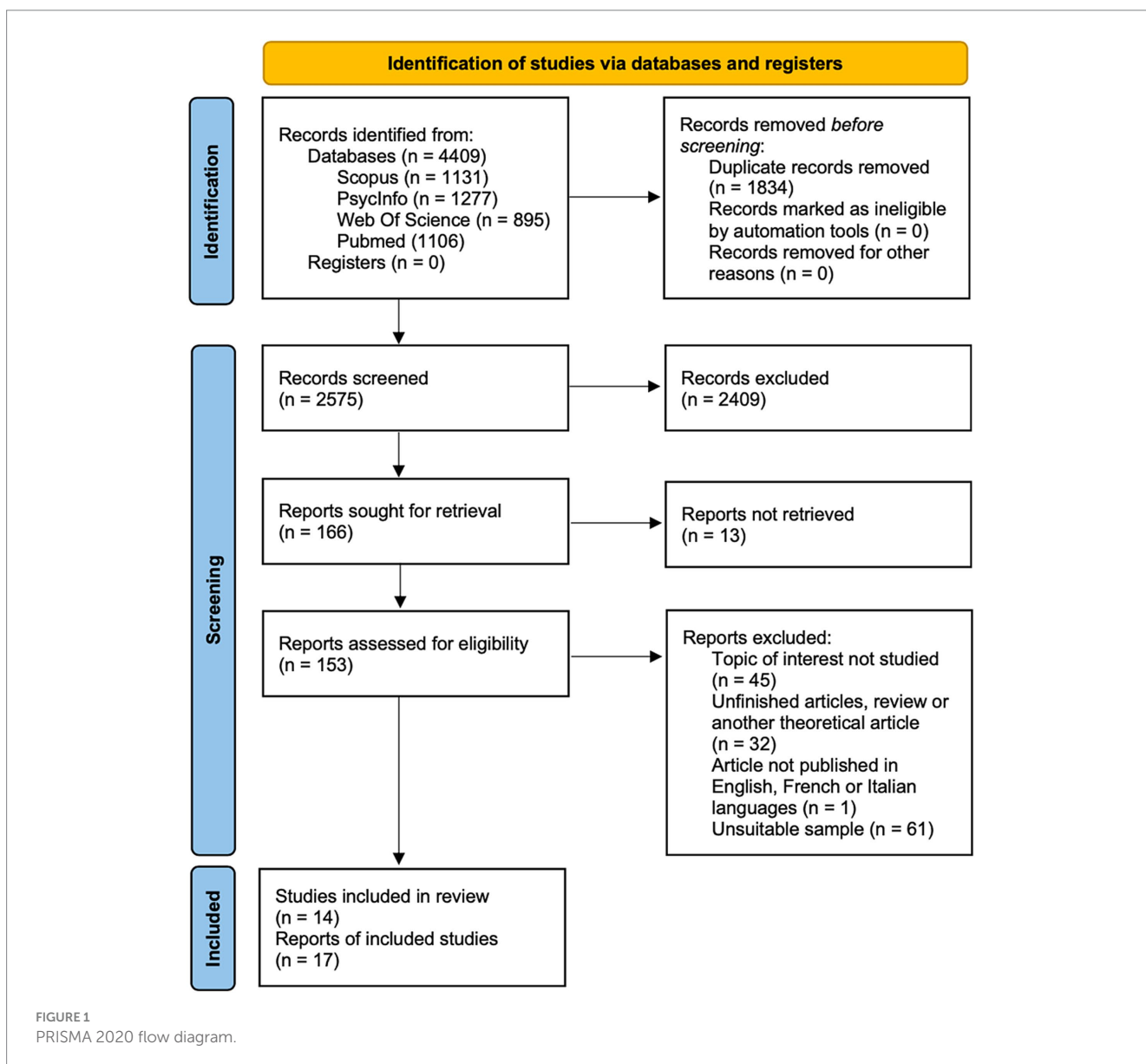

FIGURE 1
PRISMA 2020 flow diagram.

TABLE 2 Cross-sectional studies.

Authors	Country	Sample	Serious game	Assessment of outcomes	Results
Dancey et al. (2011) (Study 1)	UK	Psychology Student - $N = 45$	- SL Virtual Conference Center	- *Ad hoc* list of feedback about the scenario - Open-ended written feedback about the experience of the SL and tutorials	- ANOVA revealed a significant effect of SL tutorial attendance on final exam mark, $F_{(3, 85)} = 10.36$, $p < 0.001$, partial $\eta^2 = 0.27$; - Bonferroni comparisons show that students who attended two or three SL tutorials performed significantly better ($p = 0.008$ and $p < 0.001$, respectively) in the final exam than those with no tutorials or just one ($p = 0.047$). - ANCOVA revealed that coursework ($F_{(1, 83)} = 60.40$, $p < 0.001$) and virtual learning environment ($F_{(1, 57)} = 12.08$, $p < 0.001$) both covaried with exam marks. Yet, SL tutorial attendance effect persisted ($F_{(3, 85)} = 2.85$, $p < 0.05$, partial $\eta^2 = 0.09$), despite considering independent measures of academic ability and engagement. - Students not attending tutorials scored significantly lower in the final exam compared to those attending two ($p = 0.025$) or three ($p = 0.022$) SL tutorials.
Dancey et al. (2011) (Study 2)	UK	Psychology Student - $N = 24$ Academic staff - $N = 3$	- SL Virtual Conference Center	- *Ad hoc* list of feedback about the conference scenario - Open-ended written feedback about the experience of the SL and tutorials	- The vast majority of students liked presenting, enjoyed seeing other students' posters, found it fun, and thought SL should be used more in their degree course. Students also liked the anonymity of SL. - Numerous participants defined the conference as 'interesting', and they would like to use the environment more in the future.
Cangas et al. (2017)	Spain	Second-year psychology students - $N = 26$ - $F = 20$ - $M = 6$ - Age = 20.12 ± 2.85	- Stigma-Stop	- Stigma-Stop assessment questionnaire - *Ad hoc* question about Stigma-Stop experiences	- Participants gave the program an average score of 8.08 (SD = 1.35) and 6.58 (SD = 1.21) for its usefulness and entertainment value. - Participants would recommend Stigma-Stop to a friend and emphasized its educational value, ability in providing intervention strategies, and usefulness in raising awareness against stigma.
Krach and Hanline (2018)	USA	Seniors in a combined undergraduate/M.S. special education teacher training program - $N = 21$ - $F = 21$ - Age = 21–23 Graduate students enrolled in a school psychology combined M.S./Ed.S. program - $N = 11$ - $F = 10$ - $M = 1$ - Age = 22–27	- TeachLive	- TCR - CEF - CEF Adapted - IRP-15 - Evaluation with NASP standard	- Consultants evaluated the TeachLive experience significantly ($t_{(19)} = 4.01$, $p < 0.001$; $d = 1.84$) more positively than consultees ($M = 18.33$, SD = 0.41 vs. $M = 14.93$, SD = 0.74). - F test for equal variances shows a statistically significant difference between the two groups ($F = 0.19$; $p = 0.012$). - Consultees rated consultants ($M = 62.18$, SD = 8.20) significantly higher ($t_{(30)} = 0.002$, $p = 0.998$; $d = 0.72$) than how consultants rated consultees ($M = 56.10$, SD = 8.99). - F test for equal variances for the CEF data found no significant difference in variance ($F = 1.20$; $p = 0.035$). - Statistically significant score differences on the CEF-Consultant and the CEF-Consultee ($t_{(30)} = 0.002$, $p = 0.998$; $d = 0.72$), indicating a medium effect size. - Consultees' average rating for the interventions' quality was $M = 74.05$ (SD = 8.33). - Consultation students were given foundational knowledge and skills in most of the areas described in the standards (66.67%). - Participants enjoyed the experience and required more time with the mixed-reality simulator. But faced challenges implementing intervention strategies due to assignment constraints and technological issues.

(Continued)

TABLE 2 (Continued)

Authors	Country	Sample	Serious game	Assessment of outcomes	Results
Sugden et al. (2021)	Australia	Students enrolled in biopsychology and social psychology - N = 63 - F = 51 - M = 12 - Age = 36.4 ± 10.6	- Online learning activities	- *Ad hoc* question about students' access to online learning activities - *Ad hoc* question about perceived levels of engagement - *Ad hoc* question about perceived levels of deep learning - Qualitative interview - Focus Group	- Each activity being used by at least 75% of survey participants. - Students perceived all of the activities as highly engaging and conducive to deep learning. The Biopsychology interactive games had the highest ratings of engagement, whilst the brain dissection code word activity received the lowest ratings. - Students' access to the learning management system and the time they spent interacting with learning resources within the system were associated with student success in the subject.
Rogers et al. (2022)	Australia	Undergraduate psychology students - N = 73 - M = 22 - F = 51 - Age = 25.34 ± 10.08	- Two scenarios of semi-structured conversation scripts (one via desktop and one via HMD) in which the participant plays the role of a student counsellor interacting with a young female.	- *Ad hoc* question about experience appraisal - *Ad hoc* single question about medium of presentation (HDM vs. Desktop) preference - *Ad hoc* question about the potential of the experience - BEES - BEES modified	- HDM-based experience was preferred by 82% of the students and was rated as significantly more interesting, engaging, and immersive than desktop-based version ($M = 3.38$, SD = 0.64 vs. $M = 3.60$, SD = 0.62, $z = 2.92$, $p < 0.005$; $M = 3.36$, SD = 0.6 vs. $M = 3.05$, SD = 0.76, $z = 2.98$, $p < 0.005$; $M = 3.30$, SD = 0.85 vs. $M = 2.86$, SD = 0.10, z = 3.78, $p < 0.001$, respectively). - Both the experiences were rated positively "interesting" (94%), "engaging" (82%), and "immersive" (71%). - 55% of students considered this experience a valuable addition to their degree course, and 71% evaluate it with a lot of potential as a teaching tool in undergraduate psychology. - The virtual character was perceived by 96% of students as feeling negative, with no significant difference between desktop-based and HMD modes ($M = -1.46$, SD = 0.74; $M = -1.49$, SD = 0.80; $t_{(72)} = 0.32$, $p = 0.75$, $d = 0.04$, respectively). - After engaging with the scenarios, on average the mood of participants significantly improved post scenarios compared to pre-scenarios ($M = 1.81$, SD = 0.70; $M = 1.35$, SD = 0.71, $t_{(69)} = 5.79$, $p < 0.001$, $d = 0.69$, respectively), with 61% of participants' emotions becoming more positive after the scenarios, 24% remaining the same, and 14% becoming more negative.
Redondo-Rodríguez et al. (2023)	Spain	Undergraduate students of psychology and primary education - N = 102 - M = 24 - F = 78 - Age = 18.70 ± 1.7	- Winter is Coming, a serious game based on the novel Game	- Personal Academic Motivation Self-Perception Scale - TMMS-24	- University students who experienced a gamified and cooperative environment in peer, mixed, and interdisciplinary teams increased the mean score in emotional clarity (-1.44 ± 5.07, $p = 0.005$) and emotional repair (-1.37 ± 4.09, $p = 0.001$). No significant changes in the pre-post mean difference were found. - Students increased the use of learning strategies and the vital goals in their motivation to learn (0.40 ± 0.93, $p < 0.001$; 0.13 ± 0.66, $p < 0.005$, respectively).

(Continued)

TABLE 2 (Continued)

Authors	Country	Sample	Serious game	Assessment of outcomes	Results
Segal et al. (2023)	Lithuania	Psychology students - $N = 60$ - $M = 13$ - $F = 47$ - Age = 22.75 ± 6.74	- EIT	- Emotional Reactivity scale of CE-CSA - 1 *ad hoc* question about the avatar's realism - IMotion Biometric Research Platform for recognitions of facial expressions - AU12 Lip Corner Puller for recognition of facial expression of relief - Shimmer3 GSR for monitoring skin conductivity - Adapted PPG for analysis of heart rate	- From the first interview to the second, participants reduced levels of disgust (from $M = 1.107$, SE = 0.436 to $M = 0.359$, SD = 0.216; $\chi^2_{(1)} = 4.200$, $p = 0.040$), surprise (from $M = 3.764$, SE = 0.923 to $M = 1.097$, SD = 0.318; $\chi^2_{(1)} = 11.102$, $p < 0.001$), GSR resistance (from $M = 169.119$, SE = 17.313 to $M = 152.594$, SE = 13.979; $\chi^2_{(1)} = 11.082$, $p < 0.001$), and heart rate (from $M = 85.298$, SE = 2.794 to $M = 79.244$, SE = 2.876; $\chi^2_{(1)} = 13.601$, $p < 0.001$). Participants also reduced their level of sadness (from $M = 0.495$, SE = 0.170 to $M = 1.273$, SD = 0.338; $\chi^2_{(1)} = 11.083$, $p < 0.001$) and lip-corner suppressor activity (relief) (from $M = 0.901$, SE = 0.397 to $M = 3.711$, SD = 1.161; $\chi^2_{(1)} = 7.255$, $p = 0.007$). - Participants expressed more anger in their faces preceding closed as opposed to open questions, $\chi^2 = 2.917$, $p = 0.088$. - Participants had higher GSR resistance, and lower heart rate and preceding closed as opposed to open questions ($\chi^2 = 11.350$, $p < 0.001$, $\chi^2 = 13.496$, $p < 0.001$, respectively).
Conn et al. (2023) (Study 1)	UK	Master's Students in Counseling and Positive Psychology - $N = 24$ - $M = 4$ - $F = 20$ - Age = 34.08 ± 9.75	- Perspective: Counseling Simulator	- SUS - *Ad hoc* questions on Likert's scale about the usability of the serious game - Qualitative feedback about bug - *Ad hoc* question about prior experiences with video games - PCES	- SUS scores yielded an average score of 61.46 (SD = 16.25), with no game-breaking bugs identified. - Participants reported that they agreed or strongly agreed that it was useful in teaching them about risk assessment skills in counseling (62.8%), that what they learned from the game would be useful in their real-world practice (70.9%), and that they would play the game again in the future to practice their risk-assessment skills (66.6%).
Conn et al. (2023) (Study 2)	UK	Master's Students in Counseling and Positive Psychology - $N = 24$ - $M = 2$ - $F = 18$ - NB = 1 - Age = 36.79 ± 12.50	- Perspective: Counseling Simulator	- SUS - *Ad hoc* questions on Likert's scale about usability - Qualitative feedback about experiences - *Ad hoc* question about prior experiences with video games - PCES	- 100% of participants reported that they agreed or strongly agreed that the game was useful in teaching about risk-assessment skills in counseling and that what they learned from the game would be useful in real-world practice. - 96% of participants reported they would play the game again in the future to practice their risk-assessment skills. - SUS scores yielded an average score of 86.25 (SD = 12.25), and no game-breaking bugs were identified.

BEES, Brief Emotional Experience Scale; CE-CSA, Cognitions and Emotions about Child Sexual Abuse; CEF, Consultant Evaluation Form; EIT, Empowerment Interviewer Training; GSR, Galvanic Skin Response; HDM: Head-mounted Display; IRP-15, Intervention Rating Profile for Consultee and Client; NASP, National Association of School Psychologists; PCES, Psychologist and Counsellor Self-Efficacy Scale; PPG, Photoplethysmogram; RMSA, Research Method Skills Assessment; SATS-36, Survey of Attitudes Toward Statistics; SL, Second Life; SUS, System Usability Scale; TCR, TeachLive for Consultation Rating Form; TMMS-24, Trait MetaMood Scale 24.

TABLE 3 Case control study.

Authors	Country	Sample	Serious game	Assessment of outcomes	Results
Iwamoto et al. (2017)	USA	Undergraduate General Psychology students - $N = 49$ - $M = 16$ - $F = 33$	- Kahoot	- Multiple-choice exam score - CAT modified	- T-test shows a significant effect of Kahoot on test scores ($F_{(1, 47)} = 7.801$, $p = 0.008$). - Students reported that Kahoot was helpful in their preparation for the exam.

CAT, Classroom Assessment Technique.

3.3 Risk of bias

To address the diversity of study designs and methodologies, results were analyzed separately based on the type of study. Cohort and cross-sectional studies were grouped for quality assessment and synthesized using the NOS criteria (Wells et al., 2023; Herzog et al., 2013), while RCTs were evaluated independently using RoB-2 (Higgins et al., 2019). This stratification allowed for a clearer comparison of results across methodologies and facilitated the identification of potential sources of bias unique to each design.

The included studies exhibited varying levels of risk of bias, ranging from low to high, contingent upon the study design. The total score for cross-sectional studies ranged from 1 to 4 (out of 10), indicating a high risk of bias, mainly related to sample representativeness, sample size, lack of description of non-respondents, and lack of comparability. The case–control study scored between 3 points (out of 9), indicating a high risk of bias due to lack of adequate definition, lack of representativeness of cases, lack of definition of control, and lack of ascertainment of exposure for case–control studies. Overall, the methodological quality of the included cross-sectional and case–control studies was low.

Similarly, the RCTs raised some concerns regarding the overall risk of bias. Specifically, all studies showed some concerns in the randomization process or in the selection of reported outcomes. Only one RCT study had a low risk of bias (Olivier et al., 2019).

The assessment of risk of bias for all studies is reported in Table 6 and Figures 2, 3.

3.4 Specific traits and attitudes

Three studies were conducted with the objective of enhancing specific traits and attitudes among psychology students and professionals through the implementation of serious games (Cangas et al., 2017; Olivier et al., 2019; Redondo-Rodríguez et al., 2023). More specifically, one study (Cangas et al., 2017) focused on the usefulness of a serious game in reducing the stigma associated with mental illness in a sample of psychology students. This serious game features numerous characters who are dealing with various mental health conditions. The player's goal is to interact with each character, persuade them to contribute their knowledge, and collaborate on the common goal of designing a video game for a contest. The sample consisted of 26 participants, and the activity was conducted in a group

format: 4 participants at a time took turns interacting directly with the video games while the rest of the sample observed on the projector screen. At the end of the program, all students completed the Stigma-Stop assessment questionnaire. Qualitative results of this study revealed that participants found the serious game useful and entertaining, and emphasized the educational value of the game in providing information about the symptoms of various mental disorders. Students also highlighted the usefulness of this serious game in learning intervention strategies for dealing with mental disorders and in promoting helping and prosocial behaviors. They also highlighted the game's usefulness in raising awareness of stigma, dispelling myths, and normalizing mental disorders. In addition, the serious game was useful in helping students identify with the experiences of the characters.

One study (Olivier et al., 2019) investigated the use of a serious game that allowed players to interact with a series of characters with disabilities with the aim of promoting empathy and reducing prejudice toward people with disabilities in a sample of psychology students ($n = 83$). The study included an experimental group and two control groups. The experimental group ($n = 26$) played the serious game "The World of Empa." The first control group ($n = 26$) read case studies and background information related to the theoretical foundation of the serious game. The second control group ($n = 31$) did not receive any intervention. All three groups underwent pre-test and immediate post-test assessments, as well as a follow-up post-test approximately 1 month later, to evaluate differences in empathy, interpersonal reactivity, implicit attitudes toward others, and their changes over time. Results revealed that both control groups showed decreases in perspective-taking, empathic concern, and personal distress between the first and third measurement. However, the first control group scored significantly higher than the other groups at baseline but showed a decline at follow-up. No significant association was found between serious game use and empathy quotient or implicit attitudes toward others. In other words, participants who interacted with the serious game did not show a statistically significant decrease in their empathy scores, unlike the two control groups.

Lastly, Redondo-Rodríguez et al. (2023) explored the impact of a learning approach that incorporated a gamified and cooperative methodology on emotional intelligence, learning strategies, and life goals that serve as motivational factors for a sample of psychology university students ($n = 102$). The serious game was based on the narrative of the series Game of Thrones, which offered a story with

roles for each student, who were divided into small work groups and presented with challenges and cooperative activities. All students completed pre- and post-test assessments of emotional intelligence, learning strategies, and life goals motivating the study of the university subjects. Regarding emotional intelligence, the results show that university students who experienced a gamified and cooperative environment in peer, mixed, and interdisciplinary teams showed an increase in mean scores for emotional clarity and emotional repair. No significant changes were found in emotional attention. Similarly, at the end of the course, participants exhibited enhanced motivation to learn as a result of the implementation of more efficient learning strategies and an increase in the levels of a number of variables defined by the authors as "vital goals" (e.g., assertiveness), compared to the beginning of the course.

3.5 Theoretical knowledge and academic courses

Four studies investigated the use of serious games and gamification as alternative approaches to support students in classroom engagement and exam preparation (Dancey et al., 2011; Iwamoto et al., 2017; Sugden et al., 2021).

In particular, two studies conducted by Dancey et al. (2011) focused on an avatar-based serious game as a research skills acquisition tool that allowed a sample of psychology students ($n = 43$) to attend academic lectures (Study 1); of these, 27 participants presented the findings of their empirical project at a virtual conference (Study 2). This serious game was conducted within the pre-existing video game "Second Life." In the first study, students were divided into 3 groups based on the number of Second Life tutorials they attended. The scores of the participants who attended the academic lectures in Second Life were compared with the exam scores of the students who did not attend the lectures ($n = 46$). The results demonstrated a statistically significant improvement in performance on the final module exam among students who utilized the serious game, as compared to those who did not. Additionally, participants who attended three tutorials demonstrated superior performance compared to those who attended none or only one. The effect of the serious game tutorial participation remained significant even after controlling for variables such as engagement (i.e., additional visits to the virtual learning environment) and academic abilities (i.e., coursework mark). In addition, both studies incorporated a qualitative assessment of student perceptions regarding the efficacy of the serious game. The vast majority of students expressed that the game was enjoyable and more conducive to their learning than traditional in-person tutorials. They also highlighted the flexibility of engaging with the game from their personal computers at home as a key advantage. Furthermore, they perceived the anonymity afforded by the game as a factor that enhanced their engagement compared to real-life interactions. Moreover, students appreciated the interactive environment, which facilitated peer-to-peer interactions, and highlighted the supportive nature of the tutorials, which encouraged collaborative learning among classmates. In addition, students who considered themselves less competent reported feeling more comfortable. Finally, benefits reported by students included access to distance learning and the ability to make up missed lessons. Disadvantages also emerged,

mainly related to technical problems (i.e., sound quality, Internet connection, and computer crashes).

Iwamoto et al. (2017) used the serious game "Kahoot," an online no-stakes quiz, to teach general psychology to a sample of psychology students (n = 49). Prior to the start of the semester, participants were randomly assigned to an experimental group that used Kahoot and traditional learning approaches (e.g., notes, PowerPoint) for exam preparation and a control group that received only traditional materials (e.g., study guide). Results showed that using Kahoot had a significant positive effect on test scores compared to the control group. Participants in the experimental group reported that the serious game helped them to prepare for exams, to understand the material and they were more satisfied with the learning material they received than the control group.

Consistent with these results, Sugden et al. (2021) developed a serious game based on multiple web conferencing tutorials, virtual demonstrations, mind maps, interactive games, and case-based scenarios designed to promote deep learning as conceptualized by Craik and Lockhart (1972) and engagement in biopsychology and social psychology in a sample of psychology students ($n = 63$). Qualitative analysis revealed that psychology students who used the serious game found it to be highly engaging and to promote deep learning. They also reported that the serious game could be a resource for their learning experience and improve their learning approach. Students highlighted that activities involving the practical application of content were highly memorable for exams and remained ingrained even post-course. Additionally, participants emphasized that the serious game provided immediate feedback that compelled them to delve into deeper levels of learning. Moreover, students who spent more time interacting with the learning resources within the serious game received higher scores on the final exam in the Biopsychology and Social Psychology courses.

3.6 Practical or professional skills

Ten studies investigated the use of serious games to teach practical or professional skills to psychology students and psychologists, with promising results (Conn et al., 2023; Haginoya et al., 2021, 2023; Krach and Hanline, 2018; Pompedda et al., 2015, 2020; Rogers et al., 2022; Segal et al., 2023).

In particular, four studies used avatar-based serious games to teach consultation skills to psychology students. Krach and Hanline (2018) used a serious game based on TeachLive to teach early consultation skills to a sample of psychology students ($n = 21$) in a safe environment. Specifically, students first received information about a case, and after deciding with a graduate psychology student on which intervention to implement, students had to attempt to implement the chosen intervention with a virtual avatar. Results showed that students enjoyed the process, considered it a positive learning experience, and would have liked more time with the serious game. Participants also reported learning various methods of effective school counseling.

Similarly, Rogers et al. (2022) developed an avatar-based serious game to teach consultation skills to psychology students ($n = 63$) and evaluated the students' experiences in interacting with two avatars representing young women who were either dealing with the end of a romantic relationship or having difficulties in studying. The consultations were conducted via both a head-mounted display and a

TABLE 4 Randomized controlled trials.

Authors	Country	Sample	Serious game	Assessment of outcomes	Results
Pompedda et al. (2015)	Turin, Italy	Psychology Student - N = 21 - M = 5 - F = 16 - Age = 24.5 ± 1.6	- EIT	CSAABS	- Participants correctly identified the sexual abuse status of the avatar in 22.6% of the cases, in 26.2% for avatars with no-abuse status, and 19.0% for avatars with abuse status ($p < 0.434$). - Participants used twice as many closed ($M = 23.85$) than open-ended ($M = 11.32$) question types. - Participants obtained three ($M = 2.93$) correct (out of the mean total of 7.75 total correct details) and one-and-a-half ($M = 1.48$) neutral details for each ($M = 0.75$) incorrect detail for each scenario. - The interaction between feedback (feedback vs. no feedback) and the phase of the interview (two first interviews vs. two last interviews) was significant for both the number of open-ended (Wald $\chi^2[1] = 5.30$, $p < 0.021$) and closed questions (Wald $\chi^2[1] = 7.02$, $p < 0.008$); - The expected interaction between feedback (feedback vs. no feedback) and the phase of the interview (two first interviews vs. two last interviews) was significant for the number of correct details (Wald $\chi^2[1] = 7.58$, $p < 0.006$) and neutral details (Wald $\chi^2[1] = 14.42$, $p < 0.001$) but not for the number of incorrect details ($p < 0.381$). - The interaction between feedback (feedback vs. no feedback) and the phase of the interview (two first interviews vs. two last interviews) was significant for the proportion of correct conclusions (Wald $\chi^2[1] = 3.99$, $p < 0.046$). - During the first two interviews, the feedback group and the no-feedback group reached 15 and 18% correct conclusions, respectively, in contrast to the last two interviews (45 and 14%, respectively).
Olivier et al. (2019)	South Africa	Senior psychology students - N = 83 - M = 18 - F = 65	- World of Empa serious game; - Exposure to case reports and information with the same theoretical foundation as the serious game	- EQ - IRI - IAT	- Hierarchical linear models show that the Perspective Taking sub-scale of the IRI control group 1 (CL) decreases from the first time ($M = 4.14$) to the third measurement ($M = 3.84$). - ANOVAS indicated a significant main effects of the Time × Condition interaction for Perspective Taking ($F_{(102,8)} = 3.07$, p = 0.02), Empathic concern ($F_{(105,8)} = 3.63$, p = 0.01), personal distress ($F_{(105,7)} = 3.46$, $p = 0.01$), and time ($F_{(100,8)} = 5.0$, p = 0.01), but not for IAT prejudice ($F_{(100,8)} = 3.96$, $p = 0.43$). - The reading group (CL) showed an average decline of empathic concern over the three measurements ($M = 4.38$ at T_1, $M = 4.01$ at T_3, respectively), while the average score for personal distress increased in T_1 and T_2 ($M = 3.72$ and $M = 3.85$) and decreased at T_3 ($M = 3.35$). - Experimental and CL groups showed a medium practical significant decrease ($d = 0.49$ and 0.43, respectively) in the fantasy subscale on the immediate post-test. Control group showed no significant changes over time.

(Continued)

TABLE 4 (Continued)

Authors	Country	Sample	Serious game	Assessment of outcomes	Results
Pompedda et al. (2020) (Study 1)	Italy	Psychologists - $N = 40$ - $F = 37$ - $M = 3$ - Age = 27	- EIT	- Percentage of recommended questions - Number of correct details elicited - Number of incorrect details elicited - Number of correct conclusions	- One-way ANOVAs showed no differences between the two groups at baseline (Interview 1) for incorrect details ($F_{(1, 38)} = 3.86$, $p = 0.057$) or correct conclusions. For number of correct details, a significant difference in favor of the feedback group was found ($F_{(1, 38)} = 7.96$, $p = 0.008$). - Receiving feedback during the avatar training was associated with a higher proportion of recommended questions ($M = 63.19$, SD = 20.18 vs. $M = 30.83$, SD 15.22 for control group, Wald $\chi^2(11, N = 240) = 661.09$, $p < 0.001$); a higher number of correct details ($M = 6.39$, SD = 2.63 vs. $M = 2.75$, SD = 2.25 for control group, Wald $\chi^2(11, N = 240) = 262.04$, $p < 0.001$); a lower number of incorrect details ($M = 0.71$, SD = 1.31 vs. $M = 4.03$, SD = 3.92 for control group, Wald $\chi^2(11, N = 240) = 66.72$, $p < 0.001$) (Group × Time); and a higher proportion of correct conclusions between groups ($M = 0.35$, SD = 0.48 vs. $M = 0.04$, SD = 0.20 for control group, Mann–Whitney $U = 25$, $p < 0.001$) within the avatar interviews. - In contrast to control groups, receiving feedback during the avatar interviews significantly increased the proportion of recommended questions ($M = 13.59$, SD = 27.83, and $M = 14.24$, SD = 31.73, respectively; $E = 14.87$, $t = 5.36$, $p < 0.01$, $d = 0.04$), the proportion of correct details elicited from the children ($E = 31.59$, $t = 2.16$, SE = 14.60, $p < 0.05$, $d = 0.02$), but not for the number of correct conclusions ($M = 0.15$, SD = 0.36 and $M = 0.19$, SD = 0.40, respectively). - Disclosure of the child did not affect the proportion of recommended questions used by the interviewer ($E = 31.59$, $t = 2.16$, SE = 14.60, $p < 0.05$, $d = 0.02$).
Pompedda et al. (2020) (Study 2)	Estonia	Psychology students - $N = 69$ - $M = 22$ - $F = 47$ - Age = 23	- EIT	- Percentage of recommended questions - Number of correct details elicited - Number of incorrect details elicited - Number of correct conclusions	- One-way ANOVAs showed no differences between the two groups at baseline for the proportion of recommended questions ($U = 592$, $p = 0.996$), correct details ($U = 562$, $p = 0.701$), or incorrect details ($U = 485$, $p = 0.158$). A chi-square test showed no differences for correct conclusions ($\chi^2_{(1)} = 0.16$, $p = 0.900$). - During the training, in contrast to the control group, the feedback group showed a higher proportion of recommended questions ($M = 52.63$, SD = 16.55 vs. $M = 69.44$, SD = 17.41, Wald $\chi^2 (11, N = 384) = 222.38$, $p < 0.001$), a higher number of correct details elicited ($M = 3.37$, SD = 2.56 vs. $M = 3.98$, SD = 2.55, Wald $\chi^2 (11, N = 384) = 59.47$, $p < 0.001$), and a lower number of incorrect details ($M = 1.67$, SD = 1.96 vs. $M = 0.89$, SD = 1.30, Wald $\chi^2 (11, N = 384) = 32.23$, $p = 0.001$) (Group × Time). A Mann–Whitney test showed no differences between groups for the proportion of correct conclusions between experimental and control groups ($M = 0.40$, SD = 0.76 vs. $M = 0.32$, SD = 0.67, U = 462, $p = 0.495$). - In contrast to the control group, feedback groups show higher means for the proportion of recommended questions ($M = 20.43$, SD = 18.32 vs. $M = 40.80$, SD = 15.87) but not for the proportion of correct details ($M = 65.20$, SD = 35.01 vs. $M = 48.90$, SD = 35.97) nor the proportion of correct conclusions ($M = 0.28$, SD = 0.45 vs. $M = 0.26$, SD = 0.44). - Receiving feedback in the avatar interviews significantly increased the proportion of recommended questions used compared to the control group in the whole sample ($E = 11.96$, $t = 3.15$, $p < 0.01$, $d = 0.04$), and significantly decreased the proportion of correct details elicited compared to the control group when analyzing the cases in which the child talked about the event ($E = -13.46$, $t = -2.14$, $p < 0.05$, $d = -0.01$). - Disclosure of the child did not affect the proportion of recommended questions used by the interviewer ($E = 11.74$, $t = 3.05$, SE = 3.85, $p < 0.01$).

(Continued)

TABLE 4 (Continued)

Authors	Country	Sample	Serious game	Assessment of outcomes	Results
Haginoya et al. (2021)	Japan	Clinical psychologists - $N = 32$ - $F = 22$ - $M = 10$ - Age = 35.1 ± 8.8	- EIT	- Proportion of recommended questions - Number of correct details - Number of incorrect details	- Significant main effects on the combined dependent variables were found for intervention ($F_{(6, 54)} = 3.37$, $p = 0.007$, ηp^2 = 0.27, $1 - \beta = 0.91$) and time ($F_{(12, 18)} = 11.84$, $p < 0.001$, ηp^2 = 0.89, $1 - \beta > 0.99$). - No significant interaction between intervention and time on the combined dependent variables ($F_{(24, 36)} = 1.41$, $p = 0.171$, ηp^2 = 0.49, $1 - \beta = 0.80$) was found. - At a univariate level, significant effects of intervention emerged for the proportion of recommended questions ($F_{(2, 29)} = 6.59$, $p = 0.004$, ηp^2 = 0.31, $1 - \beta = 0.88$) and the number of correct details ($F_{(2, 29)} = 11.54$, $p < 0.001$, ηp^2 = 0.44, $1 - \beta = 0.99$), but not for the number of incorrect details ($F_{(2, 29)} = 0.62$, $p = 0.547$, ηp^2 = 0.04, $1 - \beta = 0.14$). - A univariate level significant effect of time emerged for the proportion of recommended questions ($F_{(4, 116)} = 66.34$, $p < 0.001$, ηp^2 = 0.70, $1 - \beta > 0.99$), the number of correct details ($F_{(4, 116)} = 49.79$, $p < 0.001$, ηp^2 = 0.63, $1 - \beta > 0.99$), and the number of incorrect details ($F_{(4, 116)} = 7.14$, $p < 0.001$, ηp^2 = 0.20, $1 - \beta = 0.99$). - Planned comparisons revealed that participants in the combination group used a significantly greater proportion of recommended questions compared to the feedback group at all time points except the first ($t = 4.61$, $p = 0.001$, $d = 2.08$, $1 - \beta > 0.99$ for time point 2; $t = 3.90$, $p = 0.004$, $d = 2.00$, $1 - \beta > 0.99$ for time point 3; $t = 3.06$, $p = 0.014$, $d = 1.60$, $1 - \beta = 0.95$ for time point 4; $t = 3.98$, $p = 0.003$, $d = 1.52$, $1 - \beta = 0.89$ for time point 5); elicited a significantly greater number of correct details compared to the feedback group at all time points except the first and fourth ($t = 4.85$, $p = 0.001$, $d = 2.15$, $1 - \beta = 0.59$ for time point 2; $t = 4.16$, $p < 0.001$, $d = 3.15$, $1 - \beta = 0.98$ for time point 3; $t = 4.47$, $p = 0.002$, $d = 1.85$, $1 - \beta = 0.90$ for time point 5); reached a significantly greater proportion of correct conclusions compared to the feedback group ($M = 0.08$, SD = 12, $t_{(9)} = 3.54$, $p = 0.006$, $d = 0.1.56$, $1 - \beta = 0.92$). - Modeling group outperformed the feedback group on the proportion of recommended questions only at time points 2 and 5 ($t = 3.03$, $p = 0.014$, $d = 1.30$, $1 - \beta = 0.81$ and $t = 2.81$, $p = 0.020$, $d = 1.10$, $1 - \beta = 0.68$, respectively), and on the number of correct details elicited only at time point 2 ($t = 3.32$, $p = 0.009$, $d = 1.25$, $1 - \beta = 0.58$). - For the proportion of correct conclusions, a significant main effect was found ($F_{(2, 28)} = 6.13$, $p = 0.006$, ηp^2 = 0.30, $1 - \beta = 0.90$). - The difference between the combined and modeling group ($M = 0.18$, SD = 0.23) and the difference between the modeling and feedback group were not significant.

(Continued)

TABLE 4 (Continued)

Authors	Country	Sample	Serious game	Assessment of outcomes	Results
Haginoya et al. (2023)	Japan	Clinical Psychologist, Police Personnel, Child Guidance Office Staff, Hospital Workers, Educational Facility Workers. - $N = 42$ - $M = 14$ - $F = 25$ - NR = 3 - Age = 37.1 ± 7.1	- AI-based interview simulator	- 2 *ad hoc* questions about their preliminary impression of the case - 3 *ad hoc* questions about their conclusions about the case	- The number of recommended questions had a significant positive correlation with the number of relevant details ($r = 0.63$, $p < 0.001$), whereas the number of not recommended questions had a significant positive correlation with the number of incorrect details ($r = 0.64$, $p < 0.001$). The proportion of recommended questions had a significant positive correlation with the number of relevant details and a significant negative correlation with the number of incorrect details ($r = 0.51$, $p < 0.001$, and $r = -0.39$, $p < 0.001$, respectively). - Significant effects on the combined dependent variables were found for time ($F_{(5, 35)} = 9.90$, $p < 0.001$, $\eta p^2 = 0.59$, $1 - \beta = 1.00$) and for the interaction between intervention and time ($F_{(10,70)} = 2.72$, $p = 0.007$, $\eta p^2 = 0.28$, $1 - \beta = 0.95$). - Univariate level significant effects of time emerged for the number and the proportion of recommended questions ($F_{(1, 39)} = 46.95$, $p < 0.001$, $\eta p^2 = 0.55$, $1 - \beta = 1.00$; $F_{(1, 39)} = 24.45$, $p < 0.001$, $\eta p^2 = 0.39$, $1 - \beta = 1.00$, respectively), and the number of relevant details ($F_{(1, 39)} = 6.65$, $p = 0.014$, $\eta p^2 = 0.15$, $1 - \beta = 0.71$). - Significant effect of intervention emerged for the number of relevant details ($F_{(2, 39)} = 3.50$, $p = 0.040$, $\eta p^2 = 0.15$, $1 - \beta = 0.62$), but not for the number and the proportion of recommended questions ($F_{(2, 39)} = 1.38$, $p = 0.264$, $\eta p^2 = 0.07$, $1 - \beta = 0.28$; $F_{(2, 39)} = 0.71$, $p = 0.499$, $\eta p^2 = 0.04$, $1 - \beta = 0.16$, respectively), the number of not recommended questions ($F_{(2, 39)} = 0.40$, $p = 0.674$, $\eta p^2 = 0.02$, $1 - \beta = 0.11$), and the number of incorrect details ($F_{(2, 39)} = 0.30$, $p = 0.742$, $\eta p^2 = 0.02$, $1 - \beta = 0.09$). - Univariate level significant interaction between time and intervention emerged for the number and the proportion of recommended questions ($F_{(2, 39)} = 9.62$, $p < 0.001$, $\eta p^2 = 0.33$, $1 - \beta = 0.97$; $F_{(2, 39)} = 12.10$, $p < 0.001$, $\eta p^2 = 0.38$, $1 - \beta = 0.99$, respectively), the number of not recommended questions ($F_{(2, 39)} = 12.25$, $p < 0.001$, $\eta p^2 = 0.39$, $1 - \beta = 0.99$), and the number of incorrect details ($F_{(2, 39)} = 3.97$, $p = 0.027$, $\eta p^2 = 0.17$, $1 - \beta = 0.68$).

CSAABS, Child Sexual Abuse Attitude and Belief Scale; EIT, Empowerment Interviewer Training; EQ, Empathy Quotient; IAT, Implicit Association Test; IRI, Interpersonal Reactivity Index; NR, Not Reported.

TABLE 5 A comprehensive summary of the serious games utilized across the studies and their effectiveness across various application contexts.

Serious game use category	Authors	Aim of the serious game	Effectiveness of the serious game
Specific traits and attitudes	Cangas et al. (2017)	Reducing the stigma associated with mental illness	Delivered intervention strategies and raised awareness to combat stigma.
	Olivier et al. (2019)	Fostering empathy and reducing prejudice toward individuals with disabilities	Enhanced perspective taking, empathic concern, and reducing personal distress over time. No effect in reducing prejudice.
	Redondo-Rodríguez et al. (2023)	Enhancing emotional intelligence, learning strategies, life goals, and motivation	Improved emotional clarity, emotional repair, learning strategies, and motivation.
Theoretical Knowledge and Academic Courses	Dancey et al. (2011)	Acquisition of research skills	Higher final exam scores.
	Iwamoto et al. (2017)	Teach general psychology	Higher test scores.
	Sugden et al. (2021)	Promote deep learning and engagement in biopsychology and social psychology	Promoted deep learning and success in the subject.
Practical or professional skills	Krach and Hanline (2018)	Teach consultation skills	Higher ratings for intervention quality. Students gained key knowledge and skills but struggled with time and technology issues.
	Rogers et al. (2022)		Valuable for students' degrees and as a teaching tool. The virtual character was viewed negatively by most students. Participants' mood generally improved after the scenarios.
	Conn et al. (2023)		Highly useful for learning risk-assessment skills in counseling and valuable for real-world practice.
	Pompedda et al. (2015)	Teach to conduct investigative interviews with children	Improved the identification of correct details and conclusions.
	Pompedda et al. (2020)		Higher proportion of recommended questions, a greater number of correct details, and a lower number of incorrect details. Better performance in the avatar interviews, including a higher proportion of correct conclusions.
	Haginoya et al. (2021)		Improved participants' use of recommended questions, the number of correct details, and the proportion of correct conclusions.
	Haginoya et al. (2023)		Improvement in the number and proportion of recommended questions and relevant details, number and proportion of recommended questions. It did not significantly affect the number of not-recommended questions or incorrect details.
	Segal et al. (2023)		Reduced emotional and physiological responses from the first to the second interview, including lower levels of disgust, surprise, sadness, and heart rate, as well as decreased Galvanic Skin Response resistance. They also displayed more facial anger before closed questions and had higher Galvanic Skin Response resistance and lower heart rate before closed versus open questions.

desktop. Results showed that students rated the serious game as very positive, interesting, engaging, and immersive. Students stated that the serious game could be a valuable addition to a psychology course, and it has great potential as a teaching tool. In addition, students generally reported that they perceived the virtual character as having negative emotions, but after interacting with these characters, participants reported that their emotions became more positive. Finally, students reported greater appreciation for the head-mounted display, but no significant differences were found in terms of effectiveness compared to desktop consultation.

In a recent publication, Conn et al. (2023) describe the development of a serious game, entitled "Perspective: Counseling Simulator," to increase self-efficacy in risk assessment skills among counseling students. Two studies were conducted. The first, which included a sample of students from a master's-level course in Counseling and Positive Psychology ($n = 24$), was an initial development and evaluation of the user experience and usability of the software. The second study, which included a sample of students from a master's level course in Counseling and Positive Psychology ($n = 24$), was used to evaluate a further developed version of the game in terms of its ability to elicit meaningful improvements in self-efficacy, risk assessment knowledge, and confidence in making risk assessment judgments. In both studies, participants found the serious game useful and noted that the skills they learned were applicable to real-world scenarios. They expressed a willingness to participate in future sessions to further improve their risk assessment skills.

Six studies were conducted to assess the efficacy of an avatar-based serious game in teaching psychology students and professionals to conduct investigative interviews with children (Haginoya et al., 2021, 2023; Pompedda et al., 2015, 2020; Segal et al., 2023), using the Empowerment Interviewer Training (EIT; Pompedda et al., 2015). In EIT, participants interact with a realistic avatar of a child who may or may not be a victim of abuse or maltreatment, simulating an interview with real children. During this interaction, participants can learn optimal approaches for engaging with children in a forensic setting and improve their questioning techniques through practice and refinement. The software utilizes algorithms based on prior research on children's memory and suggestibility to automatically select responses from pre-defined answer options for various question types, guided by the probabilities associated with each option. The goal of EIT is to improve the quality of the investigative interview by increasing the number of recommended questions identified in the literature (e.g., more open-ended questions and fewer closed and suggestive questions). Overall, the results showed that the serious game improved the quality and accuracy of the investigative interview, especially when combined with outcome feedback, highlighting the importance of the latter in learning with serious games.

In more detail, Pompedda et al. (2015) conducted an RCT to teach the investigative interview to a sample of 21 psychology students within the EIT paradigm. Results revealed that students who received feedback during the simulation increased the number of open-ended questions and decreased the number of closed-ended questions, obtained a greater amount of correct details, and drew more correct conclusions than students who did not receive feedback.

Similarly, Haginoya et al. (2021) utilized EIT with a sample of 32 clinical psychologists, who were divided into 3 groups that received either feedback, modeling (i.e., watching good and bad interviews with the avatar before the training), or both. In each group,

participants directly interacted with the avatar and observed four other participants' interviews. Results showed that the EIT, regardless of the group, was significantly associated with an increase in recommended questions in interviewers and a higher number of correct details elicited by the avatar, but not with the number of incorrect details elicited. Significant effects of time (conceived as the number of interviews conducted and followed by the participants) emerged for the proportion of recommended questions, the number of correct details, and the number of incorrect details. Moreover, students in the group that received both modeling (following the Behavioral Modeling Theory, Bandura, 1977) and feedback had a significantly higher proportion of recommended questions, correct details, and correct conclusions compared to the feedback – only or modeling–only groups.

Two RCTs (Pompedda et al., 2020) conducted in Italy with a sample of psychologists ($n = 40$) and in Estonia with a sample of psychology students ($n = 69$) supported these results, finding that receiving feedback during the serious game training, compared to the control group, was associated with a higher proportion of recommended questions, a higher number of correct details evoked in the avatar, and, for the Italian sample, a lower number of incorrect details and a higher proportion of correct conclusions, whereas for the Estonian sample only a higher proportion of recommended questions was found.

Haginoya et al. (2023) implemented the EIT paradigm with an automated question classification system for avatar interviews while also providing automated intervention in the form of feedback and modeling to improve interview quality. Similar to the previous studies, a sample consisting of clinical psychologists ($n = 16$), police officers ($n = 8$), child welfare workers ($n = 4$), hospital workers ($n = 4$), educational facility workers ($n = 3$), and others who preferred not to specify their affiliation ($n = 7$) was divided into 3 groups (feedback, modeling, no intervention), each of which conducted two interviews with the avatar. Results showed that the no-intervention group elicited a significantly greater number of incorrect details in the second interview compared to the first. The feedback group used a significantly greater number of recommended questions in the second interview than in the first interview. The modeling group used a significantly greater number and proportion of recommended questions, a significantly smaller number of non-recommended questions, and elicited a significantly greater number of relevant details in the second interview than in the first interview. The modeling group also used a significantly greater number and proportion of recommended questions at the second interview compared to the no-intervention group and feedback groups and a significantly smaller number of non-recommended questions at the second interview compared to the no-intervention group.

A further step was taken by Segal et al. (2023), who investigated the association between emotions and psychological parameters with question formulation in a simulated child sexual abuse investigative interview setting in a sample of psychology students. Participants were required to watch an avatar speak, watch someone else interview the avatar, and then conduct two interviews with two avatars. Results show a decrease in participants' levels of disgust and surprise when the avatar tells the story and an increase in levels of sadness from the first to the second interview. Participants' galvanic skin response (GSR)

TABLE 6 Risk of bias assessment.

Reference	Design	Selection (Maximum 5 stars)	Comparability (Maximum 2 stars)	Outcome (Maximum 3 stars)	Score
Dancey et al. (2011) (Study 1)	Cross-sectional	*			1/10
Dancey et al. (2011) (Study 2)	Cross-sectional	*			1/10
Cangas et al. (2017)	Cross-sectional	**		*	3/10
Krach and Hanline (2018)	Cross-sectional	***		**	5/10
Sugden et al. (2021)	Cross-sectional	**		*	3/10
Rogers et al. (2022)	Cross-sectional	**		**	4/10
Redondo-Rodríguez et al. (2023)	Cross-sectional	***		**	5/10
Segal et al. (2023)	Cross-sectional	****		**	6/10
Conn et al. (2023) (Study 1)	Cross-sectional	***		**	5/10
Conn et al. (2023) (Study 2)	Cross-sectional	***		**	5/10

Reference	Design	Selection (Maximum 4 stars)	Comparability (Maximum 2 stars)	Outcome (Maximum 3 stars)	Score
Iwamoto et al. (2017)	Case Control	*		**	3/9

resistance and heart rate decreased from the first to the second interview. Furthermore, facial expressions of anger predicted closed versus open questions in participants, as did higher GSR resistance and lower heart rate.

4 Discussion

Although serious games have become increasingly successful in various fields over the past decade, their use in psychology has almost always been limited to psychological and psychotherapeutic interventions and treatments (e.g., Ahmed et al., 2023; Gómez-Cambronero et al., 2023; Marchetti et al., 2018; Martins et al., 2023; Stamm et al., 2022; Tang et al., 2022). In fact, compared to other health professions, such as nursing (e.g., Baek and Lee, 2024) and medicine (e.g., Kadri et al., 2024), less research has been conducted on the use of serious games in the professional training of psychologists and psychotherapists, although the number of studies on this topic has increased significantly in recent years.

To the best of the authors' knowledge, this is the first systematic review to: (1) summarize the studies investigating the use of serious games in the professional training and education of psychologists, psychotherapists, and psychology students; (2) evaluate the effects of serious games on the quality of learning; and (3) analyze the benefits of using serious games in the training of psychologists, as well as the features that contribute to their utility and the limitations under which they function. Fourteen articles with a total of 17 original studies met the eligibility criteria. Most of these studies used serious games to improve the practical skills of students and professionals; a smaller proportion investigated the use of serious games to teach theoretical knowledge and university subjects in psychology courses, while the smallest number of studies used serious games to support the development of specific traits and attitudes in psychology

students. Overall, the included studies provided evidence to support the usefulness of serious games in promoting learning in various areas of psychology.

From a qualitative perspective, serious games were generally valued by participants in the included studies. In fact, those who experienced serious games in their education reported high levels of engagement, interest, and enjoyment. This seems particularly relevant for the new generation of students, who have shown a high level of comfort in using technology for learning (Iwamoto et al., 2017), and it can serve as an indicative basis for the development of curricula for the activities of students with psychological training profiles (Nagovitsyn et al., 2021).

The present systematic review highlighted several factors that should be considered to make serious games effective tools for teaching basic psychological knowledge, practical techniques and skills, and specific traits and attitudes.

First, several studies underlined the importance of the appropriate duration of the serious game experience. Indeed, short exposures to the tool may compromise the engagement and realism of the experience, resulting in little or no significant effect on learning (Redondo-Rodríguez et al., 2023). In this regard, the specialized literature suggests that serious game experiences should last 10 to 30 minutes to be effective (e.g., Englar, 2019; Gibbs, 2019; Rogers et al., 2022). Similarly, more frequent exposures appear to be associated with better training outcomes (Olivier et al., 2019).

A second crucial element pertains to the significance of aligning the immersive experience of a serious game with the actuality of real-world scenarios (Pompedda et al., 2015). In this sense, serious games should provide a wide range of response options to make the experience more realistic and interactive; in fact, fewer available response options seem to be associated with less immersion and engagement, which in turn reduces the effectiveness of the serious game from a learning perspective (Rogers et al., 2022).

FIGURE 2
Overall risk of bias assessment across all randomized controlled trials.
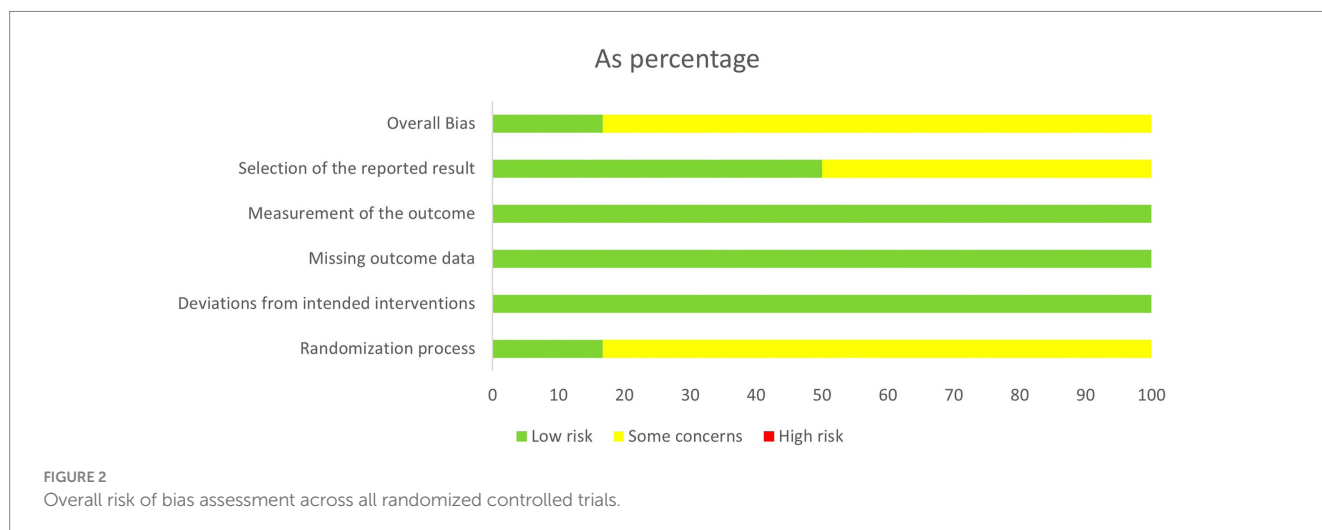
Furthermore, serious games should be designed to be both challenging and engaging, as research has demonstrated that these characteristics are associated with enhanced perceived learning and higher examination scores (Conn et al., 2023; Hamari et al., 2016; Sugden et al., 2021).

Another important factor appears to be the usability and user-friendliness of serious games. A clear and intuitive design, particularly advantageous for individuals lacking technological expertise, and a straightforward approach to navigation have been demonstrated to enhance the efficacy of serious games in promoting learning. In other words, serious games should prioritize a user-centered design approach, recognizing that not all games are suitable candidates for serious adaptation (Conn et al., 2023).

Another crucial aspect to consider is that prior to the serious game training experience, students should receive instruction and education on basic knowledge related to the topic addressed (Krach and Hanline, 2018; Pompedda et al., 2020). In this sense, Haginoya et al. (2021), starting from the Behavior Modeling Training (BMT; Bandura, 1977), suggested that identifying behaviors for learning and presenting models to students before serious games training could potentially optimize learning outcomes. These results appear to be consistent with a meta-analysis by Taylor et al. (2005), which emphasized that presenting detailed rules with explanations, such as providing specific guidance before engaging in behavioral modeling, can maximize its effectiveness. In addition, the use of positive and negative models (in terms of teaching what should or should not be done), known as a mixed model, could increase motivation by showing behavioral outcomes and produce more significant learning effects than the use of positive models alone. Therefore, experiencing serious games after initial basic training can promote long-term knowledge acquisition and deep learning by stimulating connections between theoretical concepts and their practical application (Sugden et al., 2021).

Several studies analyzed in the current systematic review (Haginoya et al., 2021, 2023; Iwamoto et al., 2017; Pompedda et al., 2015, 2020; Segal et al., 2023; Sugden et al., 2021) also emphasized the importance of teacher feedback. These studies demonstrated that teacher-provided feedback enhanced the quality of interaction and learning within serious games, particularly when it was provided in a timely manner rather than with a delay. This finding is consistent with the existing literature (e.g., Johnson et al., 2017), which highlighted that an additional benefit of using serious games is the ability to provide immediate feedback, allowing students to engage in activities at their own pace without having to wait for teacher responses. This is particularly the case when students perceive a sense of emotional involvement and feel supported by their instructors, which may result in increased cognitive and behavioral engagement in the completion of tasks (e.g., Smith, 2017). Consequently, any serious games-based training module should incorporate supervision and feedback as essential components (Stoffregen et al., 2003).

Beyond identifying key design factors, it is crucial to elucidate how these elements interact with user experiences to yield educational outcomes. For example, the serious game experience duration not only influences engagement but also directly impacts cognitive load and memory retention by providing sufficient time for learners to apply and reflect on new knowledge (Redondo-Rodríguez et al., 2023). Similarly, immersive realism achieved through diverse response options fosters critical thinking and decision-making skills by replicating the nuanced challenges of real-world scenarios (Pompedda et al., 2015). Feedback mechanisms act as a critical driver of skill acquisition by reinforcing correct behaviors and enabling learners to correct mistakes in real time, promoting self-efficacy and long-term knowledge retention (Stoffregen et al., 2003). Moreover, user-centered design features, such as intuitive navigation and responsive interfaces, do not merely enhance usability but also minimize frustration, allowing learners to focus cognitive resources on task performance rather than tool mastery (Conn et al., 2023). Finally, preparatory instruction, coupled with behavior modeling, extends the learning impact of serious games by enabling participants to draw explicit connections between theoretical knowledge and applied scenarios, amplifying their ability to generalize learned skills to novel contexts (Haginoya et al., 2021; Taylor et al., 2005).

An important technical consideration in serious games, particularly avatar-based ones, is graphic resolution (Dancey et al., 2011). While high-quality audio and visuals can enhance immersion and engagement (Laffan et al., 2016), the "Uncanny Valley" phenomenon, introduced by Mori (1970) and supported by subsequent studies (e.g., Di Natale et al., 2023), suggests that as artificial entities become more human-like, they may initially evoke comfort, but beyond a certain threshold, imperfections can make

FIGURE 3
Risk of bias assessment for each randomized controlled trial study.
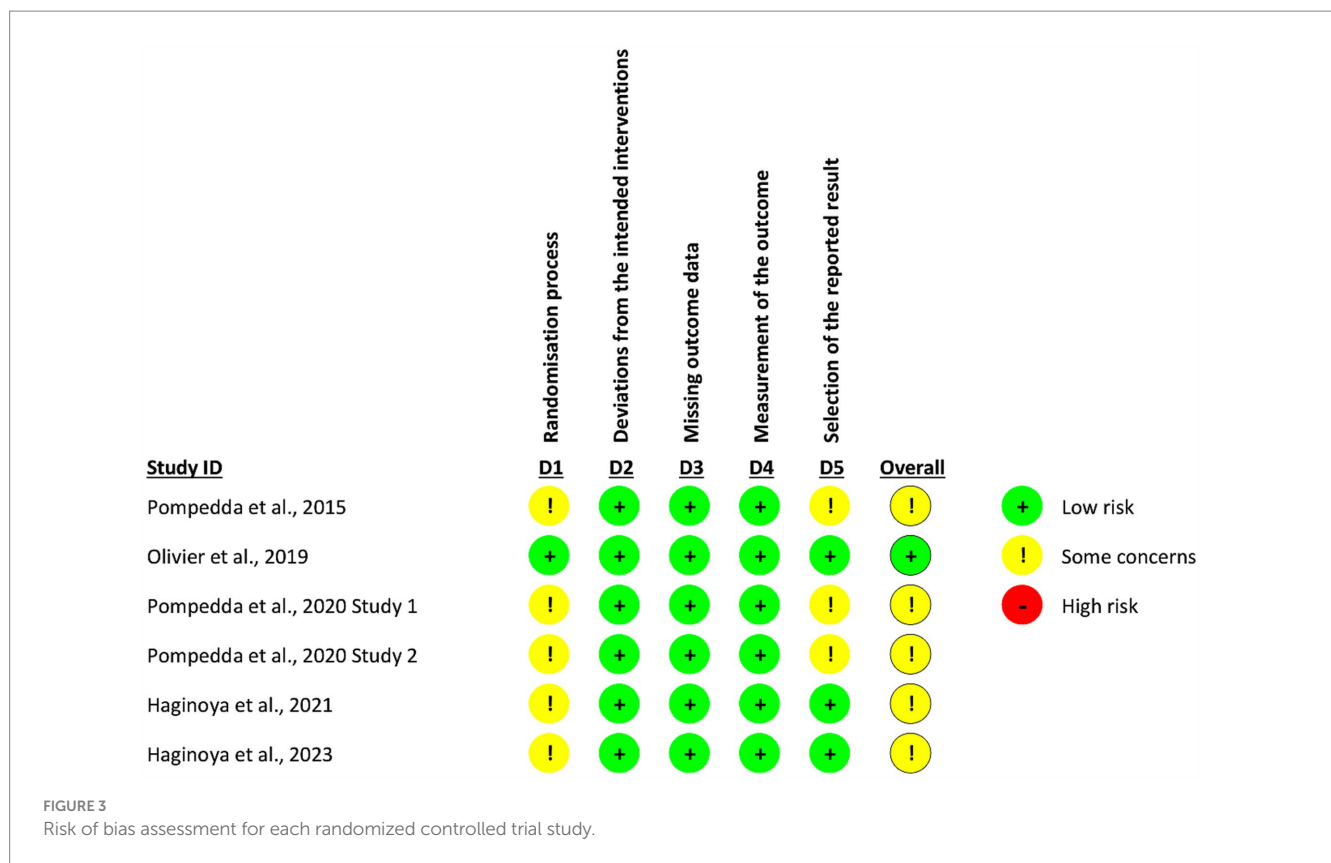
them unsettling or unpleasant. This phenomenon can influence the perceived realism of the experience, which in turn affects learning and memory by increasing cognitive load and emotional responses, though the latter tend to diminish with repeated exposure (Lehtonen et al., 2005; Segal et al., 2023). Additionally, Rogers et al. (2022) emphasized the value of preemptively notifying students about the possibility of encountering sensitive content in the serious game to help mitigate distress, particularly when such content is represented in a realistic manner.

The current systematic review also highlights the numerous advantages offered by serious games over traditional forms of education and professional training. Firstly, in contrast to traditional educational approaches, serious games serve as flexible and dynamic tools that allow students to practice in a safe environment without the potential consequences of errors or the risk of harm to patients due to a lack of experience. It can be argued that serious games facilitate the acquisition of practical knowledge by allowing students to enhance their skills and competencies in a protected environment under expert guidance before engaging with real clients or patients. This approach has been shown to increase confidence and perceived effectiveness and reduce anxiety when transitioning to real-world scenarios (Cangas et al., 2017; Dancey et al., 2011; Iwamoto et al., 2017; Krach and Hanline, 2018).

In general, students found serious games to be engaging and enjoyable due to the interactive nature of such games, which encourages student effort, motivation, and commitment to learning. This ultimately leads to improved learning outcomes (Sugden et al., 2021).

Another benefit of using serious games appears to be the opportunity for psychologists to engage in more frequent hands-on experiences, especially with hard-to-reach populations. For example,

serious games have allowed for the practice of investigative interviews with mistreated or abused children (e.g., Pompedda et al., 2015), counseling interviews with individuals facing specific problems (e.g., Rogers et al., 2022), and strategies for reducing the stigma associated with disabilities (e.g., Cangas et al., 2017). In these and similar scenarios, serious games provide a valuable alternative for students to practice and refine their skills.

It is also noteworthy that serious games appear to be effective in promoting the learning of a skill even when students and psychologists already have practical experience with the subject at hand. It can be reasonably assumed that serious games may be effective regardless of student experience. However, it is important to note that participants with more experience may show lower performance and improvement compared to those with less experience. This is due to the fact that previously acquired habits may be difficult to change and require more time to learn the skill (Haginoya et al., 2021).

Among the studies analyzed, a difference emerged between the use of serious games and avatar-based serious games. Specifically, the latter were used for developing practical and professional skills and enhancing specific traits and attitudes. On the other hand, non-avatar-based serious games were mainly used for theoretical knowledge and academic courses. Although it is not possible to compare the differences between these two types of serious games due to the heterogeneity of the results, it can be assumed that the use of avatar-based serious games is more fruitful in the area of practical training or personal aptitudes. This could be due to the fact that, compared to traditional serious games, they can offer a more realistic and real-world experience in which they are confronted with patients or people (e.g., Mikropoulos and Natsis, 2011). Conversely,

non-avatar-based serious games might be more suitable in the context of theoretical training, where the theoretical notions presented within the serious game might have to take priority, thereby promoting learning and motivation (e.g., Wouters et al., 2013). However, no comparative studies were found regarding the difference between these types of serious games, so further studies are needed.

Despite the benefits identified in this review, it is essential to acknowledge the limitations associated with serious games.

First, there remains confusion surrounding the definitions of serious games and gamification. Although the literature consistently highlighted the distinction between these terms (e.g., Deterding et al., 2011; Miller et al., 2011; Werbach, 2014), researchers frequently report instances where these terms are used ambiguously or interchangeably (e.g., Warsinsky et al., 2021).

Moreover, some studies included in this review (i.e., Cangas et al., 2017; Olivier et al., 2019; Redondo-Rodríguez et al., 2023) demonstrated limited effectiveness in using serious games to promote specific traits and attitudes. This may partly stem from the reliance on self-report instruments for the pre- and post-assessments, which may be less reliable than performance-based measures (Cangas et al., 2017). To enhance reliability and ensure more robust findings, combining self-reports with objective metrics is recommended.

Another significant limitation is the brevity of intervention durations, which may not adequately capture long-term skill acquisition and retention. For instance, many interventions were limited to a single session, raising concerns about their sustainability and the extent to which learning outcomes persist over time. As suggested by Olivier et al. (2019), longer or more frequent exposure periods may yield greater benefits by allowing participants to consolidate their knowledge and skills. Future research should prioritize longitudinal designs with follow-up assessments spanning several months to evaluate the enduring effects of serious games on professional development and learning outcomes.

The methodology limitations observed in the reviewed studies further highlight areas for improvement. The quality assessment revealed that most RCTs (Haginoya et al., 2021, 2023; Pompedda et al., 2015, 2020) had some concerns regarding the risk of bias, mainly in randomization processes and outcome reporting. Only one RCT study was classified as having a low risk of bias (Olivier et al., 2019). Additionally, all cross-sectional studies (Cangas et al., 2017; Conn et al., 2023; Dancey et al., 2011; Krach and Hanline, 2018; Redondo-Rodríguez et al., 2023; Rogers et al., 2022; Segal et al., 2023; Sugden et al., 2021) and case–control studies (Iwamoto et al., 2017) demonstrated a higher risk of bias due to issues such as sample representativeness, sample size, lack of description of nonrespondents, lack of comparability for cross-sectional studies, lack of adequate definition, lack of representativeness of the cases, lack of definition of control, and lack of ascertainment of exposure for case–control studies; thus, the quality assessment highlighted that findings from these studies should be treated with caution. The moderate to high risk of bias observed in many studies suggests the need for methodological improvements: rigorous randomization protocols, validated tools for outcome measurement, and standardized frameworks for participant recruitment and study reporting. Employing blinded assessments and integrating self-reported data with performance-based measures could further enhance the reliability and reproducibility of findings.

The heterogeneity of tools employed in serious games research is another potential limitation, with some tools lacking validation. This variability constrains the ability to compare effectiveness across studies. To overcome this, future research should focus on developing standardized measures for assessing learning outcomes and ensure consistent use of validated tools.

Technical challenges, including software bugs and user interface issues, as well as trainer variability, can hinder the effectiveness of serious games and result in suboptimal learning outcomes by diminishing realism and reducing engagement (Conn et al., 2023). As highlighted by Krach and Hanline (2018), serious games that are not fully automated may exhibit variability in trainer-mediated interactions. Differences in the behavior of trainers managing the game can influence participant experiences and outcomes. Furthermore, the cost of equipment and development remains a practical concern, potentially limiting the widespread adoption of serious games (Krach and Hanline, 2018). Developers should prioritize user-friendly designs to accommodate varying levels of technological proficiency. The integration of artificial intelligence-based feedback systems could provide personalized, real-time guidance, possibly enhancing the learning experience (Tolks et al., 2024). Additionally, routine updates and rigorous testing are essential to maintain the functionality and usability of serious games.

Future research should investigate whether the positive effects of serious games observed in controlled settings can be translated into real-world applications. Expanding this work to other populations, such as trainee psychotherapists, who were not included in the studies reviewed, would enhance the generalizability of findings. Furthermore, it is crucial to elucidate the mechanisms and moderators, including the role of feedback, that contribute to the effectiveness of serious games in psychological education. The integration of advanced technologies such as artificial intelligence, virtual reality, and augmented reality into serious games offers the possibility of creating immersive and adaptive learning experiences (e.g., Arif et al., 2024; Lee et al., 2024). Lastly, future research should explore the development of serious games tailored to diverse educational needs and cultural contexts, ensuring inclusivity, accessibility, and engagement. By addressing these diverse needs, such research could also provide valuable insights into the scalability and adaptability of serious games across different educational and professional settings. Moreover, understanding cultural attitudes toward technology and gamification in education could be critical for further tailoring serious games.

It is also essential to recognize the limitations of this systematic review when interpreting its findings. A first limitation is the exclusion of gray literature, such as conference abstracts, preprint papers, and dissertations. While peer-reviewed studies were prioritized to ensure methodological rigor, this approach may have omitted potentially valuable insights from unpublished work that could have provided a more comprehensive understanding of the topic. Second, the review only included studies published in English, Italian, and French, which may have resulted in data loss and raises concerns about publication bias. Finally, a protocol for this systematic review was not registered in the PROSPERO database.

5 Conclusion

This systematic review's findings underscore several critical practical implications. Educational institutions should explore

integrating serious games into their curricula alongside traditional teaching methods. Curriculum designers should ensure that serious games are appropriately structured and make the most of key characteristics that enhance their effectiveness, particularly in terms of duration, frequency, and level of challenges, to maintain student engagement. Teacher training should emphasize the importance of providing timely and constructive feedback during serious game activities, allowing students or professionals to reflect on their performance and make improvements. Additionally, educators should be equipped to manage the technical aspects of serious games to ensure they are user-friendly and accessible.

The present systematic review expands the understanding of serious games' utility in psychology education, emphasizing their role in enhancing learning quality and professional training. The main advantages of using serious games in the training of psychologists are highlighted, as well as the characteristics that make them effective and the limitations within which they operate. Serious games are emerging as a complementary educational tool rather than a substitute for standard methods. They have the potential to enhance the quality of learning by creating a context in which students perceive the possibility of making mistakes as an opportunity for growth and learning and in which the consequences of possible mistakes are minimal. Finally, the results provide evidence to support the use of serious games to increase the effectiveness of professional training and contribute to the formation of specific cultural and professional competence groups.

Data availability statement

Author contributions

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The authors declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

References

Ahmed, F., Carrión, J. R., Bellotti, F., Barresi, G., Floris, F., and Berta, R. (2023). Applications of serious games as affective disorder therapies in autistic and Neurotypical individuals: a literature review. *Appl. Sci.* 13:4706. doi: 10.3390/app13084706

Amo, V., Prentice, M., and Lieder, F. (2023). A gamified Mobile app that helps people develop the metacognitive skills to cope with stressful situations and difficult emotions: formative assessment of the InsightApp. *JMIR Form. Res.* 7:e44429. doi: 10.2196/44429

Arif, Y. M., Ayunda, N., Diah, N. M., and Garcia, M. B. (2024). A Systematic Review of Serious Games for Health Education: Technology, Challenges, and Future Directions", in Transformative Approaches to Patient Literacy and Healthcare Innovation, ed. M. B. Garcia., R. P. Almeida (Hershey, PA: IGI Global), 20–45.

Baek, G., and Lee, E. (2024). Development and effects of advanced cardiac resuscitation nursing education program using web-based serious game: application of the IPO model. *BMC Nurs.* 23:206. doi: 10.1186/s12912-024-01871-7

Bandura, A. (1977). Social learning theory. Oxford: Prentice-Hall.

Bellotti, F., Kapralos, B., Lee, K., and Moreno-Ger, P. (2013). User assessment in serious games and technology-enhanced learning. *Adv. Hum.-Comput. Interact*, 1–2. doi: 10.1155/2013/120791

Boyle, E., Connolly, T. M., and Hainey, T. (2011). The role of psychology in understanding the impact of computer games. *Entertainment Comput.* 2, 69–74. doi: 10.1016/j.entcom.2010.12.002

Breuer, J., and Bente, G. (2010). Why so serious? On the relation of serious games and learning. *J. Comput. Game Cult.* 4, 7–24. doi: 10.7557/23.6111

Buajeeb, W., Chokpipatkun, J., Achalanan, N., Kriwattanawong, N., and Sipiyaruk, K. (2023). The development of an online serious game for oral diagnosis and treatment planning: evaluation of knowledge acquisition and retention. *BMC Med. Educ.* 23:830. doi: 10.1186/s12909-023-04789-x

Cangas, A. J., Navarro, N., Parra, J. M. A., Ojeda, J. J., Cangas, D., Piedra, J. A., et al. (2017). Stigma-stop: a serious game against the stigma toward mental health in educational settings. *Front. Psychol.* 8:1385. doi: 10.3389/fpsyg.2017.01385

Conn, C., Patel, A., Gavin, J., Granda-Salazar, M., Williams, A., and Barnes, S. (2023). "Development and evaluation of perspective: counselling simulator": a gamified tool for developing risk-assessment skills in trainee counsellors. *Ment. Health Soc. Incl.* 27, 140–153. doi: 10.1108/MHSI-12-2022-0090

Craig, S., Stark, P., Wilson, C. B., Carter, G., Clarke, S., and Mitchell, G. (2023). Evaluation of a dementia awareness game for undergraduate nursing students in Northern Ireland: a pre-/post-test study. *BMC Nurs.* 22:177. doi: 10.1186/s12912-023-01345-2

Craik, F. I. M., and Lockhart, R. S. (1972). Levels of processing: a framework for memory research. *J. Verbal Learn. Verbal Behav.* 11, 671–684. doi: 10.1016/S0022-5371(72)80001-X

Dancey, C. P., Attree, E. A., Painter, J., Arroll, M., Pawson, C., and McLean, G. (2011). Real benefits of a second life: development and evaluation of a virtual psychology conference Centre and tutorial rooms. *Psychol. Learn. Teach.* 10, 107–117. doi: 10.2304/plat.2011.10.2.107

Deterding, S., Dixon, D., Khaled, R., and Nacke, L. (2011). From game design elements to gamefulness: defining "gamification". In Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments. New York, NY, USA: Association for Computing Machinery. (pp. 9–15)

Di Natale, A. F., Simonetti, M. E., La Rocca, S., and Bricolo, E. (2023). Uncanny valley effect: a qualitative synthesis of empirical research to assess the suitability of using virtual faces in psychological research. *Comput. Hum. Behav. Rep.* 10:100288. doi: 10.1016/j.chbr.2023.100288

Englar, R. E. (2019). Using a standardized client encounter to practice death notification after the unexpected death of a feline patient following routine ovariohysterectomy. *J. Vet. Med. Educ.* 46, 489–505. doi: 10.3138/jvme.0817-111r1

Fraticelli, F., Marchetti, D., Polcini, F., Mohn, A. A., Chiarelli, F., Fulcheri, M., et al. (2016). Technology-based intervention for healthy lifestyle promotion in Italian adolescents. *Ann. Ist. Super. Sanita* 52, 123–127. doi: 10.4415/ANN_16_01_20

Gentry, S. V., Gauthier, A., Ehrstrom, B. L., Wortley, D., Lilienthal, A., Car, L. T., et al. (2019). Serious gaming and gamification education in health professions: systematic review. *J. Med. Internet Res.* 21:e12994. doi: 10.2196/12994

Gibbs, S. I. (2019). The thespian client: the benefits of role-playing teaching techniques in psychology. *Am. J. Appl. Psychol.* 8, 98–104. doi: 10.11648/j.ajap.20190805.12

Girard, C., Ecalle, J., and Magnan, A. (2013). Serious games as new educational tools: how effective are they? A meta-analysis of recent studies. *J. Comput. Assist. Learn.* 29, 207–219. doi: 10.1111/j.1365-2729.2012.00489.x

Gómez-Cambronero, Á., Casteleyn, S., Bretón-López, J., García-Palacios, A., and Mira, A. (2023). A smartphone-based serious game for depressive symptoms: protocol for a pilot randomized controlled trial. *Internet Interv.* 32:100624. doi: 10.1016/j.invent.2023.100624

Graafland, M., Schraagen, J. M., and Schijven, M. P. (2012). Systematic review of serious games for medical education and surgical skills training. *Br. J. Surg.* 99, 1322–1330. doi: 10.1002/bjs.8819

Greitzer, F. L., Kuchar, O. A., and Huston, K. (2007). Cognitive science implications for enhancing training effectiveness in a serious gaming context. *J. Educ. Resour. Comput.* 7:2-es. doi: 10.1145/1281320.1281322

Haginoya, S., Ibe, T., Yamamoto, S., Yoshimoto, N., Mizushi, H., and Santtila, P. (2023). AI avatar tells you what happened: the first test of using AI-operated children in simulated interviews to train investigative interviewers. *Front. Psychol.* 14:1133621. doi: 10.3389/fpsyg.2023.1133621

Haginoya, S., Yamamoto, S., and Santtila, P. (2021). The combination of feedback and modeling in online simulation training of child sexual abuse interviews improves interview quality in clinical psychologists. *Child Abuse Negl.* 115:105013. doi: 10.1016/j.chiabu.2021.105013

Hamari, J., Shernoff, D. J., Rowe, E., Coller, B., Asbell-Clarke, J., and Edwards, T. (2016). Challenging games help students learn: an empirical study on engagement, flow and immersion in game-based learning. *Comput. Hum. Behav.* 54, 170–179. doi: 10.1016/j.chb.2015.07.045

Hammady, R., and Arnab, S. (2022). Serious gaming for behaviour change: a systematic review. *Information* 13:142. doi: 10.3390/info13030142

Herzog, R., Álvarez-Pasquin, M. J., Díaz, C., Del Barrio, J. L., Estrada, J. M., and Gil, Á. (2013). Are healthcare workers' intentions to vaccinate related to their knowledge, beliefs and attitudes? A systematic review. *BMC Public Health* 13:154. doi: 10.1186/1471-2458-13-154

Heyselaar, E., Hagoort, P., and Segaert, K. (2017). In dialogue with an avatar, language behavior is identical to dialogue with a human partner. *Behav. Res. Methods* 49, 46–60. doi: 10.3758/s13428-015-0688-7

Higgins, J. P. T., Savović, J., Page, M. J., Elbers, R. G., and Sterne, J. A. C. (2019). "Assessing risk of bias in a randomized trial", in Cochrane Handbook for Systematic Reviews of Interventions, 2nd edition, ed. J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, V. A. Welch (Chichester, UK: John Wiley and Sons), 205–228.

Iwamoto, D. H., Hargis, J., Taitano, E. J., and Vuong, K. (2017). Analyzing the efficacy of the testing effect using Kahoot™ on student performance. *Turk. Online J. Distance Educ.* 18, 80–93. doi: 10.17718/tojde.306561

Johnson, C. I., Bailey, S. K., and Van Buskirk, W. L. (2017). "Designing effective feedback messages in serious games and simulations: a research review". In: Instructional

Techniques to Facilitate Learning and Motivation of Serious Games, ed. P. Wouters, H. van Oostendorp (Springer, Cham), 119–140.

Kadri, M., Boubakri, F.-E., Kaghat, F.-Z., Azough, A., and Zidani, K. A. (2024). IVAL: immersive virtual anatomy laboratory for enhancing medical education based on virtual reality and serious games, design, implementation, and evaluation. *Entertain. Comput.* 49:100624. doi: 10.1016/j.entcom.2023.100624

Ke, F., Xie, K., and Xie, Y. (2016). Game-based learning engagement: a theory- and data-driven exploration. *Br. J. Educ. Technol.* 47, 1183–1201. doi: 10.1111/bjet.12314

Krach, S. K., and Hanline, M. F. (2018). Teaching consultation skills using interdepartmental collaboration and supervision with a mixed-reality simulator. *J. Educ. Psychol. Consult.* 28, 190–218. doi: 10.1080/10474412.2017.1301818

Laamarti, F., Eid, M., and El Saddik, A. (2014). An overview of serious games. *Int. J. Comput. Games Technol.* 2014:358152, 1–15. doi: 10.1155/2014/358152

Laffan, D. A., Greaney, J., Barton, H., and Kaye, L. K. (2016). The relationships between the structural video game characteristics, video game engagement and happiness among individuals who play video games. *Comput. Hum. Behav.* 65, 544–549. doi: 10.1016/j.chb.2016.09.004

Lamb, M. E. (2016). Difficulties translating research on forensic interview practices to practitioners: finding water, leading horses, but can we get them to drink? *Am. Psychol.* 71, 710–718. doi: 10.1037/amp0000039

Lamb, M. E., Sternberg, K. J., Orbach, Y., Esplin, P. W., and Mitchell, S. (2002). Is ongoing feedback necessary to maintain the quality of investigative interviews with allegedly abused children? *Appl. Dev. Sci.* 6, 35–41. doi: 10.1207/S1532480XADS0601_04

Landers, R. N. (2014). Developing a theory of gamified learning: linking serious games and gamification of learning. *Simul. Gaming* 45, 752–768. doi: 10.1177/1046878114563660

Lee, L. K., Wei, X., Chui, K. T., Cheung, S. K., Wang, F. L., Fung, Y. C., et al. (2024). A systematic review of the Design of Serious Games for innovative learning: augmented reality, virtual reality, or mixed reality? *Electronics* 13:890. doi: 10.3390/electronics13050890

Lehtonen, M., Page, T., and Thorsteinsson, G. (2005). "Emotionality considerations in virtual reality and simulation based learning" in Proceedings of the international conference on cognition and exploratory learning in digital age (Porto: Portugal), 26–36.

Luchini, C., Stubbs, B., Solmi, M., and Veronese, N. (2017). Assessing the quality of studies in meta-analyses: advantages and limitations of the Newcastle Ottawa scale. *World J. Meta-Anal.* 5, 80–84. doi: 10.13105/wjma.v5.i4.80

Maheu-Cadotte, M.-A., Cossette, S., Dubé, V., Fontaine, G., Lavallée, A., Lavoie, P., et al. (2021). Efficacy of serious games in healthcare professions education: a systematic review and Meta-analysis. *Simul. Healthc.* 16, 199–212. doi: 10.1097/SIH.0000000000000512

Marchetti, D., Fraticelli, F., Polcini, F., Fulcheri, M., Mohn, A. A., and Vitacolonna, E. (2018). A school educational intervention based on a serious game to promote a healthy lifestyle. *Mediterr. J. Clin. Psychol.* 6, 1–16. doi: 10.6092/2282-1619/2018.6.1877

Marchetti, D., Fraticelli, F., Polcini, F., Lato, R., Pintaudi, B., Nicolucci, A., et al. (2015). Preventing Adolescents' Diabesity: design, development, and first evaluation of "Gustavo in Gnam's planet". *Games Health J.* 4, 344–351. doi: 10.1089/g4h.2014.0107

Martins, I., Perez, J. P. P., Osorio, D., and Mesa, J. (2023). Serious games in entrepreneurship education: a learner satisfaction and theory of planned behaviour approaches. *J. Entrep.* 32, 157–181. doi: 10.1177/09713557231158207

Mikropoulos, T. A., and Natsis, A. (2011). Educational virtual environments: a ten-year review of empirical research (1999–2009). *Comput. Educ.* 56, 769–780. doi: 10.1016/j.compedu.2010.10.020

Miller, L. M., Chang, C.-I., Wang, S., Beier, M. E., and Klisch, Y. (2011). Learning and motivational impacts of a multimedia science game. *Comput. Educ.* 57, 1425–1433. doi: 10.1016/j.compedu.2011.01.016

Mori, M. (1970). Bukimi no tani - the uncanny valley (K. F. MacDorman & T. Minato, trans.). *Energy* 7, 33–35.

Nagovitsyn, R. S., Berezhnykh, E. A., Popovic, A. E., and Srebrodolsky, O. V. (2021). Formation of legal competence of future bachelors of psychological and pedagogical education. *Int. J. Emerg. Technol. Learn. IJET.* 16, 188–204. doi: 10.3991/ijet.v16i01.17591

Olivier, L., Sterkenburg, P., and van Rensburg, E. (2019). The effect of a serious game on empathy and prejudice of psychology students towards persons with disabilities. *Afr. J. Disabil.* 8:10. doi: 10.4102/ajod.v8i0.328

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 372:n71. doi: 10.1136/bmj.n71

Pompedda, F. (2018). Training in investigative interviews of children: Serious gaming paired with feedback improves interview quality [Doctoral Thesis]. [Turku (FL)]: Åbo Akademi University

Pompedda, F., Palu, A., Kask, K., Schiff, K., Soveri, A., Antfolk, J., et al. (2020). Transfer of simulated interview training effects into interviews with children exposed to a mock event. *Nord. Psychol.* 73, 43–67. doi: 10.1080/19012276.2020.1788417

Pompedda, F., Zappalà, A., and Santtila, P. (2015). Simulations of child sexual abuse interviews using avatars paired with feedback improves interview quality. *Psychol. Crime Law* 21, 28–52. doi: 10.1080/1068316X.2014.915323

Pompedda, F., Zhang, Y., Haginoya, S., and Santtila, P. (2022). A mega-analysis of the effects of feedback on the quality of simulated child sexual abuse interviews with avatars. *J. Police Crim. Psychol.* 37, 485–498. doi: 10.1007/s11896-022-09509-7

Redondo-Rodríguez, C., Becerra-Mejías, J. A., Gil-Fernández, G., and Rodríguez-Velasco, F. J. (2023). Influence of gamification and cooperative work in peer, mixed and interdisciplinary teams on emotional intelligence, learning strategies and life goals that motivate university students to study. *Int. J. Environ. Res. Public Health* 20:547. doi: 10.3390/ijerph20010547

Rogers, S. L., Hollett, R., Li, Y. R., and Speelman, C. P. (2022). An evaluation of virtual reality role-play experiences for helping-profession courses. *Teach. Psychol.* 49, 78–84. doi: 10.1177/0098628320983231

Segal, A., Bakaitytė, A., Kaniušonytė, G., Ustinavičiūtė-Klenauskė, L., Haginoya, S., Zhang, Y., et al. (2023). Associations between emotions and psychophysiological states and confirmation bias in question formulation in ongoing simulated investigative interviews of child sexual abuse. *Front. Psychol.* 14:1085567. doi: 10.3389/fpsyg.2023.1085567

Sipiyaruk, K., Gallagher, J. E., Hatzipanagos, S., and Reynolds, P. A. (2018). A rapid review of serious games: from healthcare education to dental education. *Eur. J. Dental Educ.* 22, 243–257. doi: 10.1111/eje.12338

Sitzmann, T. (2011). A Meta-analytic examination of the instructional effectiveness of computer-based simulation games. *Pers. Psychol.* 64, 489–528. doi: 10.1111/j.1744-6570.2011.01190.x

Smith, T. (2017). Gamified modules for an introductory statistics course and their impact on attitudes and learning. *Simul. Gaming* 48, 832–854. doi: 10.1177/1046878117731888

Söbke, H., Arnold, U., and Montag, M. (2020). "Intrinsic motivation in serious gaming: a case study" in Games and learning Alliance: 9th international conference, GALA 2020, Laval, France, December 9–10, 2020 (Berlin, Heidelberg: Springer-Verlag), 362–371.

Stamm, O., Dahms, R., Reithinger, N., Ruß, A., and Müller-Werdan, U. (2022). Virtual reality exergame for supplementing multimodal pain therapy in older adults with chronic back pain: a randomized controlled pilot study. *Virtual Real.* 26, 1291–1305. doi: 10.1007/s10055-022-00629-3

Stoffregen, T. A., Bardy, B. G., Smart, L. J., Pagulayan, R. (2003). "On the Nature and Evaluation of Fidelity in Virtual Environments", in: Virtual and Adaptive Environments: Applications, Implications, and Human Performance Issues, ed. L. J. Hettinger, M. W. Haas (Mahwah, NJ, US: Lawrence Erlbaum Associated Publishers), 111–128.

Sugden, N., Brunton, R., MacDonald, J., Yeo, M., and Hicks, B. (2021). Evaluating student engagement and deep learning in interactive online psychology learning activities. *Australas. J. Educ. Technol.* 37, 45–65. doi: 10.14742/ajet.6632

Tang, W. S. W., Ng, T. J. Y., Wong, J. Z. A., and Ho, C. S. H. (2022). The role of serious video games in the treatment of disordered eating behaviors: systematic review. *J. Med. Internet Res.* 24:e39527. doi: 10.2196/39527

Taylor, P. J., Russ-Eft, D. F., and Chan, D. W. L. (2005). A Meta-analytic review of behavior modeling training. *J. Appl. Psychol.* 90, 692–709. doi: 10.1037/0021-9010.90.4.692

Tolks, D., Schmidt, J. J., and Kuhn, S. (2024). The role of AI in serious games and gamification for health: scoping review. *JMIR Serious Games* 12:e48258. doi: 10.2196/48258

Viglione, D. J., Meyer, G. J., Resende, A. C., and Pignolo, C. (2017). A survey of challenges experienced by new learners coding the Rorschach. *J. Pers. Assess.* 99, 315–323. doi: 10.1080/00223891.2016.1233559

Warsinsky, S., Schmidt-Kraepelin, M., Rank, S., Thiebes, S., and Sunyaev, A. (2021). Conceptual ambiguity surrounding gamification and serious games in health care: literature review and development of game-based intervention reporting guidelines (GAMING). *J. Med. Internet Res.* 23:e30390. doi: 10.2196/30390

Wells, G.A., Shea, B., O'Connell, D., Peterson, J., Welch, V., Losos, M., et al. (2023). The Newcastle-Ottawa scale (NOS) for assessing the quality if nonrandomized studies in meta-analyses. URL: https://www.ohri.ca/programs/clinical_epidemiology/oxford.asp (Accessed November 12, 2023)

Werbach, K. (2014). "(re)defining gamification: a process approach" in PERSUASIVE technology. PERSUASIVE. Lecture notes in computer science, 8462. eds. A. Spagnolli, L. Chittaro and L. Gamberini (New York: Springer, Cham).

Wouters, P., van Nimwegen, C., van Oostendorp, H., and van der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *J. Educ. Psychol.* 105, 249–265. doi: 10.1037/a0031311