Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 03 June 2026
Revised 27 July 2026
Accepted 12 August 2026

Natural Resources for Human Health 2026, Vol. 6 (6): 386–400

KEYWORDS:
gamification, harassment-prevention training, psychometric validation, technology-enhanced learning, natural language processing, Sustainable Development Goal 5

Gamified harassment-prevention training: A framework and protocol for psychometric validation in the financial sector

Anjali Kushwaha[1*], Garima Bhardwaj[2]

[1]Research Scholar, Amity University Uttar Pradesh, Noida, India · ORCID: 0009-0006-5365-7120

[2]Professor, Amity University Uttar Pradesh, India · ORCID: 0000-0001-6839-696X

*Corresponding author: Anjali Kushwaha, Amity University Uttar Pradesh, Sector 125, Noida 201313, Uttar Pradesh, India, E-mail: anjali.kushwaha165@gmail.com

ABSTRACT: Gender-based harassment (GBH) persists in financial-sector workplaces despite legal and compliance frameworks, partly because conventional slide-and-acknowledgement training does little to engage learners, support retention, or change reporting behaviour. This protocol describes a mixed-methods study to design and psychometrically validate a technology-enhanced learning framework and to test its effect on harassment-prevention knowledge, attitudes, and behavioural intentions relative to conventional training. The framework integrates (1) the Gamified GBH Assessment Instrument (GGAI), to be validated for content, construct, criterion, and reliability evidence within an enterprise learning management system; (2) a gamified learning environment (points, badges, leaderboards, and branching scenario simulations) grounded in self-determination theory; and (3) a natural language processing pipeline for sentiment and theme detection in learner feedback. A two-condition, quasi-experimental pre–post design with unit-level allocation will be applied to a stratified sample of full-time employees from Indian non-banking financial companies and information technology firms (target N = 360). Between-condition effects will be tested with mixed ANOVA and ANCOVA, behavioural confidence with logistic regression, gamification predictors with multiple regression, and open-ended feedback with thematic analysis triangulated by validated sentiment classification. The analysis plan is pre-registered on the Open Science Framework, and outcomes are mapped to Sustainable Development Goal 5 (Targets 5.1 and 5.2) as a measurable composite index. As a protocol, no empirical results are reported. The study is expected to yield a validated, theory-driven model for designing and evaluating gamified harassment-prevention training that supports safer and healthier workplaces.

1 INTRODUCTION

Digital technologies have reshaped the way organisations design, deliver, and evaluate learning. Gamification, the use of game-design elements in non-game contexts, has developed as one of the most actively researched approaches in technology-enhanced learning (TEL), motivated by its potential to increase engagement, motivation, and knowledge retention (Deterding et al., 2011; Seaborn and Fels, 2015; Koivisto and Hamari, 2019). Yet the promise of gamification has not been evenly realised across learning domains. One domain where conventional digital training has conspicuously underperformed is workplace harassment prevention, an area of significant social and regulatory importance worldwide.

In India, the Sexual Harassment of Women at Workplace (Prevention, Prohibition and Redressal) Act, 2013 has modified the legislative framework surrounding gender-based harassment (GBH). Organisations are now mandated to constitute Internal Complaints Committees, impart awareness training and maintain institutional accountability. Despite this legal scaffolding, there is a consistent pattern in survey evidence of under-reporting of incidents, difficulty in recognising harassment in ambiguous situations, and a lack of confidence in acting (Antecol and Cobb-Clark, 2003; Willness et al., 2007). We contend that the persistence of under-reporting is not primarily a legal failure but an instructional failure.

The common training paradigm of formal presentations, compliance videos and acknowledgement forms often does not lead to meaningful learning or behavioural change. Employees attend because attendance is mandatory, invest little cognitively, and return to work with little durable change in knowledge, attitudes, or readiness to report (Bingham and Scherer, 2001; Roehling and Huang, 2018). The evidence on conventional harassment-prevention training is discouraging: standard programmes have been found not only to leave attitudes unchanged but, in some cases, to provoke backlash (Bingham and Scherer, 2001). The instructional question, then, is whether training can produce a real shift in learner knowledge, attitude, and behavioural intention, which are the outcomes that well-designed TEL is theorised to influence.

Gamified learning environments that incorporate simulations, leaderboards, points, badges, and progress tracking offer a theoretically grounded alternative, with the potential for greater engagement and more durable knowledge (Hamari et al., 2014; Sailer et al., 2017; Sailer and Homner, 2020). Yet despite a substantial literature on gamified learning in organisational and educational settings, we are not aware of a peer-reviewed study that has psychometrically developed and validated an instrument to measure GBH training outcomes (knowledge, attitudes, and behavioural intentions) within a gamified environment, and certainly not in the Indian financial sector. Validated harassment measures exist, but they are neither embedded in nor designed for gamified delivery.

This paper addresses that gap. The framework is both a learning-design and a measurement contribution, and it makes three contributions in particular. First, it specifies the design of a gamified harassment-prevention learning environment grounded in self-determination theory and the theory of gamified learning (Ryan and Deci, 2000; Landers and Landers, 2014). Second, it sets out a complete protocol for developing and psychometrically validating the Gamified GBH Assessment Instrument (GGAI). Existing validated harassment measures, such as the Sexual Experiences Questionnaire, assess victimisation experiences rather than training-relevant knowledge, attitudes, and behavioural intentions, and were not designed for or validated within gamified delivery; the GGAI is intended to fill that specific measurement gap so that learning outcomes can be assessed with established validity and reliability. Third, it incorporates an artificial-intelligence pipeline, using NLP for sentiment and theme detection, to analyse open-ended learner feedback at scale. Throughout, the framework treats SDG 5 (Gender Equality) as a measurable learning outcome rather than a rhetorical reference point (United Nations, 2015).

Prior gamified-training research has largely targeted onboarding, safety, and general compliance, and reports positive but context-dependent effects on engagement and knowledge; very few studies address harassment prevention, and to our knowledge none combines a psychometrically validated outcome instrument, gamified delivery, and AI-based feedback analysis in a single framework, particularly in the Indian financial sector. The present work is timely because India's POSH Act 2013 has created a regulatory mandate for effective training while evidence shows conventional delivery underperforms, and because validated NLP and IRT tooling now make rigorous, scalable evaluation feasible. These gaps motivate the five research questions and hypotheses set out in Section 3, which test, in turn, the instrument's psychometric qualities (RQ1), pre-to-post knowledge gains (RQ2), behavioural confidence (RQ3), the relative contribution of individual gamification elements (RQ4), and the qualitative learner experience (RQ5).

2 THEORETICAL BACKGROUND

2.1 Gamification and technology-enhanced learning

The theoretical basis of gamification is predominantly derived from self-determination theory (SDT) and its concept of motivational affordances—elements in a learning environment that satisfy the needs for autonomy, competence, and relatedness, and thereby maintain intrinsic motivation (Deci and Ryan, 1985; Ryan and Deci, 2000; Deterding et al., 2011). Most of the studies that Hamari, Koivisto and Sarsa reviewed in their extensive review and meta-analysis confirmed positive motivational and engagement effects, but also pointed out important contextual dependencies (Hamari et al., 2014; Buckley

and Doyle, 2016). More recent meta-analytic work has largely confirmed a positive but attenuated effect of gamification on cognitive, motivational and behavioural learning outcomes (Bai et al., 2020; Huang et al., 2020; Sailer and Homner, 2020).

Research has also shifted from questioning whether gamification works to identifying which elements work and under what conditions. Sailer and colleagues isolated points, badges, and performance graphs as the strongest drivers of psychological need satisfaction (Sailer et al., 2017), while Mekler and colleagues examined the unique effects of individual elements on intrinsic motivation and performance (Mekler et al., 2017). Landers' theory of gamified learning explains how game elements influence learning-relevant behaviours such as time-on-task and goal persistence, which in turn drive learning (Landers and Landers, 2014). Recent systematic reviews continue to refine the theoretical basis of the field (Dichev and Dicheva, 2017; Zainuddin et al., 2020; Krath et al., 2021), and large reviews of education-specific gamification report consistent—if heterogeneous—benefits (Denny, 2013; Dicheva et al., 2015; Looyestyn et al., 2017; Nacke and Deterding, 2017; Majuri et al., 2018).

The broader serious-games and game-based-learning literature reinforces these findings. Meta-analyses of serious games report positive cognitive and motivational effects relative to conventional instruction, including immersive virtual-reality-based instruction (Connolly et al., 2012; Wouters et al., 2013; Merchant et al., 2014; Clark et al., 2016), and game-based learning has been theorised as a multifaceted construct integrating cognitive, motivational, affective, and socio-cultural perspectives (Plass et al., 2015). Simulation games in particular have demonstrated instructional effectiveness in workplace and adult-learning contexts (Hamari et al., 2016). What this literature has not yet delivered is a validated assessment instrument for measuring outcomes of gamified harassment-prevention learning.

2.2 Why conventional compliance training falls short

The limitations of conventional harassment-prevention training are well documented. Educational programmes have failed to shift attitudes and, in some instances, produced backlash among participants—an outcome that underscores the risk of poorly designed training (Bingham and Scherer, 2001). One major reason is motivational. When employees perceive training as a compliance ritual rather than a learning opportunity, cognitive engagement is reduced to the minimum necessary to complete the task (McDonald, 2012). If there is no engagement there is no retention, if there is no retention there is no behavioural change. Explicit calls for theory-grounded, engagement-oriented redesign and rigorous outcome measurement are made in reviews of harassment-training effectiveness (Roehling and Huang, 2018).

These instructional deficiencies are compounded by structural antecedents. Situational antecedents of GBH include the organisational climate for harassment tolerance and the gendered composition of work groups (Fitzgerald et al., 1997). A meta-analysis of almost 70,000 respondents found harassment experiences to be associated with lower job satisfaction, organisational commitment and mental health and increased work withdrawal (Willness et al., 2007). Steep hierarchies and norms of institutional silence in Indian financial-sector organisations mean that harassment can go unrecorded (Fitzgerald et al., 1995; Antecol and Cobb-Clark, 2003; Magley et al., 2010). These dynamics are preserved in the training that does not require building recognition skills and reporting confidence (Cortina and Berdahl, 2008; McDonald, 2012).

2.3 Evaluation of learning: Psychometric standards

The outcome measures of the TEL intervention need to be valid and reliable in order to be evaluated credibly. The Standards for Educational and Psychological Testing offers a comprehensive framework for validity that includes criterion-related, content, and construct evidence (AERA et al., 2014). Content validity ensures the item pool addresses the desired domain and can be measured using Lawshe's Content Validity Ratio (Lawshe, 1975; Polit and Beck, 2006); construct validity—usually validated through confirmatory factor analysis using well-known fit indices—ensures the instrument measures the latent constructs it claims to (Hu and Bentler, 1999; Rosseel, 2012); criterion validity tests whether scores can predict real-world behaviour (Messick, 1995). Cronbach's alpha and McDonald's omega are the usual way to report internal consistency (Cronbach, 1951; McDonald, 1999; Streiner, 2003). No GBH instrument has been validated for this full framework in a gamified environment, thereby creating a scientific gap and practical risk that organisations evaluate training on unmeasured tools.

2.4 Artificial intelligence and natural language processing for learning analytics

Open-ended learner feedback is a rich, but under-exploited source of evaluation data in TEL. Advances in NLP make it feasible to analyse such feedback at scale. Lexicon- and rule-based models such as VADER provide efficient sentiment classification of short texts (Hutto and Gilbert, 2014), while transformer-based language models—introduced by the attention mechanism and operationalised in models such as BERT—enable contextual understanding that substantially improves classification of

nuanced affect (Vaswani et al., 2017; Devlin et al., 2019). Combined with established qualitative methods such as thematic analysis (Braun and Clarke, 2006), these tools allow a mixed quantitative–qualitative reading of learner experience that complements survey-based outcome measures (Creswell and Plano Clark, 2018).

2.5 SDG 5 as a measurable learning outcome

United Nations SDG 5 (Gender Equality) aims at the eradication of discrimination, violence, and harassment against women and girls in all institutional contexts; Targets 5.1 (end all forms of discrimination) and 5.2 (eliminate violence and harassment) directly relate to the workplace (United Nations, 2015; UN Women and UN DESA, 2020). International Labour Organization Violence and Harassment Convention No. 190 offers a complementary international standard (ILO, 2019). We operationalise SDG 5 alignment as a measurable outcome rather than a framing device by mapping specific GGAI items to SDG 5.1/5.2 sub-indicators: the knowledge scale indexes recognition of discrimination and harassment (Target 5.1/5.2), and the behavioural-intention scale indexes willingness to report and intervene (Target 5.2). A composite SDG-5 learning index is computed from these mapped items, and pre-to-post gains on this index constitute the paper's concrete, reportable measure of contribution to SDG 5; the item-to-indicator mapping is provided in the supplementary material.

3 RESEARCH QUESTIONS AND HYPOTHESES

The framework is guided by five research questions (RQ1–RQ5):

RQ1. What are the psychometric qualities—including content, criterion and construct validity, and internal consistency of the GGAI?

RQ2. Does gamified GBH training produce significantly greater pre-to-post knowledge gains than traditional compliance training, as measured by the validated GGAI?

RQ3. Does gamified training produce significantly greater improvements in participants' confidence to report or intervene in harassment incidents than traditional compliance training?

RQ4. Which gamification elements—points, leaderboards, or scenario simulations—are most strongly associated with post-training knowledge and behavioural outcomes?

RQ5. How do participants describe the experience of gamified GBH training, and what emotional and attitudinal themes characterise their open-ended feedback?

Corresponding hypotheses.

H1 (Psychometric): the GGAI will demonstrate satisfactory content validity (CVR ≥ 0.62), adequate construct validity for the reflective attitude and behavioural-intention scales (CFI > 0.90, RMSEA < 0.08, SRMR < 0.08), and acceptable reliability (Cronbach's α and McDonald's $\omega \geq 0.80$ for the attitude and intention scales; KR-20 ≥ 0.70 and adequate item-fit under a 2-parameter logistic IRT model for the dichotomous knowledge scale).

H2 (Knowledge): the gamified condition will show significantly larger pre-to-post knowledge gains than the traditional condition ($p < 0.01$; Cohen's $d > 0.50$).

H3 (Behaviour): a significantly higher proportion of gamified-condition participants will report post-training confidence to report harassment ($p < 0.01$; OR ≥ 2.0).

H4 (Predictors): leaderboard engagement and simulation completion will be the strongest gamification predictors of behavioural confidence (β at $p < 0.05$; model $R^2 \geq 0.30$).

H5 (Sentiment): on a held-out, human-annotated gold-standard subset used to validate the classifier, AI-driven NLP will classify a majority (pre-registered threshold $> 60\%$, benchmarked against the human-coded distribution) of open-ended feedback on gamified training as positive.

We treat sentiment as a secondary, experience-level indicator that triangulates—rather than substitutes for—the learning and behavioural outcomes in H2–H4.

Natural Resources for Human Health
eISSN: 2583-1194

4 MATERIALS AND METHODS

4.1 Research design

This study will use a mixed-methods, quasi-experimental, two-group pre-post design. The comparison (control) condition is a traditional, slide-based programme delivering content equivalent in scope, duration, and learning objectives to the gamified programme but without any game elements; the experimental condition receives the purpose-built gamified programme with the embedded GGAI. Because individual randomisation within a workplace risks cross-condition contamination, allocation is at the organisational-unit level: intact units within participating organisations are matched on size, sector, and prior-training history and then assigned to conditions, approximating a matched-pairs cluster design. Each participant receives only one condition. We acknowledge that the absence of individual randomisation weakens causal inference; to mitigate this, baseline equivalence (including prior harassment training and prior exposure to gamified learning) is tested, group differences are adjusted via ANCOVA, and a propensity-score-matched sensitivity analysis is planned as a robustness check. As described in Section 4.5, three measurement points are used: T1 (baseline, before any content), T2 (automatic process data during training in the gamified condition), and T3 (post-training outcomes). This design is consistent with established practice for the evaluation of complex training interventions (Shadish et al., 2002; Arthur et al., 2003). The analysis plan, hypotheses and coding of outcomes are registered on the Open Science Framework prior to data collection. Ethics approval and consent procedures are described in Section 4.9.

4.2 Design of the gamified learning environment

The learning environment is built on an enterprise LMS and instantiates four game-design elements selected for their documented effects on need satisfaction and engagement: (i) branching scenario simulations that situate decisions in realistic workplace contexts; (ii) an immediate point-based feedback system that rewards correct reasoning, not merely correct answers (Mekler et al., 2017; Sailer et al., 2017); (iii) progressive content unlocking tied to performance thresholds, supporting competence and goal persistence (Landers and Landers, 2014); and (iv) a cohort-visible social leaderboard supporting relatedness and healthy comparison (Tondello et al., 2016; Hamari, 2017). The selection of elements is theory-driven, each mapping onto an SDT need and a learning-relevant behaviour in the gamified-learning model (Ryan and Deci, 2000; Landers and Landers, 2014), while the design guards against documented risks such as over-competition and extrinsic crowding-out (Attali and Arieli-Attali, 2015; Toda et al., 2018).

4.3 Instrument development: GGAI

Phase 1 – Domain mapping and item generation. A multidisciplinary expert panel (n = 8; HR professionals, employment-law specialists, and organisational psychologists) maps the GBH knowledge domain onto three anchoring frameworks: India's POSH Act 2013, ILO Convention No. 190 (ILO, 2019), and SDG 5 indicators (United Nations, 2015). The panel generated an initial pool of 60 candidate items across three constructs, including 25 scenario-based multiple choice items on declarative knowledge of GBH policy and law, 20 Likert items on attitudes to gender bias and harassment, and 15 Likert items on behavioural intentions to report or intervene. The pool is thus reduced to a final 45-item instrument, consisting of a 20-item knowledge scale, a 15-item attitude scale and a 10-item behavioural-intentions scale, as summarised in Table 1, after content validation and pilot item analysis (Phase 2 and the pilot in Section 4.7).

Phase 2 – Content validation. A second independent expert panel (n = 10) rates each item using Lawshe's Content Validity Ratio procedure. At ten judges the critical CVR for significance is 0.62. Items below threshold are removed or revised (Lawshe, 1975; Polit and Beck, 2006). This is consistent with validation procedures used in established harassment-measurement work (Fitzgerald et al., 1995; Cortina and Berdahl, 2008).

Phase 3 – Gamification integration. Items that pass content validation are embedded within the gamified LMS interface described in Section 4.2, such that assessment is woven into the learning experience instead of being appended to it (Vlachopoulos and Makri, 2017).

4.4 Participants and sampling

The target population is full-time employees of NBFC and IT-sector organisations in India. Recruitment uses purposive selection of approximately eight collaborating organisations followed by stratified random sampling of employees within each, stratified by job level, gender identity, and organisational tenure to improve representativeness and support subgroup analysis.

We acknowledge that purposive organisational selection limits external validity; participating organisations are therefore profiled (size, sub-sector, workforce composition) against national sector statistics, and any divergence is reported as a generalisability boundary. We aim for roughly balanced recruitment of about 45 participants per organisation (around 22–23 per condition), summing to the target sample. Sample size is determined by the most demanding planned analysis rather than the simplest. Power analyses were conducted a priori with G*Power 3.1 (Cohen, 1988; Faul et al., 2007). A two-tailed independent-samples t-test ($\alpha = 0.05$, power 0.80, Cohen's $d = 0.50$) requires 128 participants per condition; detecting an odds ratio ≥ 2.0 in binary logistic regression at the same α and power requires a comparable order of magnitude; and detecting $R^2 \geq 0.30$ in a five-predictor multiple regression requires fewer than 60 cases. Confirmatory factor analysis is the binding constraint: applying the conventional $N \geq 200$ minimum and a 10:1 case-to-item ratio for the reflective scales, we require at least 300 valid cases for stable estimation. We therefore set a recruitment target of $N = 360$ ($n = 180$ per condition) to retain at least 300 complete cases after an anticipated 15% attrition across the three data-collection points. Missing and incomplete data are addressed as follows: attrition and item missingness are reported with a CONSORT-style flow diagram; Little's MCAR test screens the missingness mechanism; and, where data are missing at random, full-information maximum likelihood (for CFA/SEM) and multiple imputation ($m = 20$) are used, with complete-case analysis reported as a sensitivity check. The separate validation pilot (Section 4.7) is sized independently for its purpose.

4.5 Data collection procedure

Measurements are taken at three time points. T1 (pre-training baseline): all participants complete the full GGAI and a brief demographic questionnaire before accessing content. T2 (training engagement, gamified condition only): the LMS automatically logs module completion, time-on-task, point accumulation, badge awards, and leaderboard rankings—operationalising each gamification element and capturing learning-relevant behaviours (Landers and Landers, 2014; Koivisto and Hamari, 2019). T3 (post-training): all participants complete the GGAI again within 48 hours of completion, and a structured open-ended form collects reflective narrative data for NLP analysis.

4.6 Measures

Outcome and process measures are summarised in Table 1. The GGAI comprises a 20-item knowledge scale (scenario-based multiple-choice items scored dichotomously as correct/incorrect, 0–20; analysed via item-response theory and KR-20), a 15-item attitude scale and a 10-item behavioural-intentions scale (both 5-point Likert, 1 = strongly disagree to 5 = strongly agree; analysed by confirmatory factor analysis) with scale scores reported as mean (SD). LMS interaction logs (module completion %, time-on-task in minutes) and gamification achievement logs (points accrued, badges earned, leaderboard rank) operationalise the four game elements and serve as moderators and RQ4 predictors. For the qualitative component (RQ5), each participant completes a structured open-ended feedback form of six prompts (covering scenario realism, perceived relevance, psychological safety, motivation, comparison with prior training, and suggested improvements), administered online at T3; responses are exported verbatim with identifiers removed. Sentiment and themes are analysed with a two-step NLP pipeline: VADER for baseline lexicon-based polarity and a fine-tuned transformer (BERT) for contextual sentiment, with classifier performance reported as accuracy, precision, recall, and macro-F1 against a human-annotated gold-standard subset (two independent coders; Cohen's $\varkappa \geq 0.70$ required before automated scoring). Themes from the automated pipeline are cross-validated against reflexive thematic analysis (Braun and Clarke, 2006) coded in NVivo, and discrepancies are reconciled by consensus. SDG 5 adherence is quantified, not asserted: a composite SDG-5 learning index is computed from GGAI items mapped a priori to SDG Targets 5.1 and 5.2 (mapping table in the supplementary material), and pre-to-post change on this index is the reportable SDG-5 outcome.

Table 1. Measures used in the study.

Measure	Description	Role in analysis	Response scale / units
GGAI knowledge subscale	20 scenario-based MCQs on GBH law and policy	Primary outcome (RQ2): mixed ANOVA/ANCOVA; IRT scoring	0–20 correct (dichotomous); reported as mean (SD)

Measure	Description	Role in analysis	Response scale / units
GGAI attitude subscale	15 Likert items on gender-bias/harassment attitudes	CFA; regression predictor	5-point Likert (1–5); mean (SD)
GGAI behavioural-intentions subscale	10 Likert items on confidence to report/intervene	Primary outcome (RQ3): logistic regression	5-point Likert (1–5); mean (SD)
LMS interaction logs	Module completion; time-on-task (auto-generated)	Descriptive; moderation variable	% complete; minutes (auto-logged)
Gamification achievement logs	Points; badges; leaderboard rank	Regression predictors (RQ4)	Count (points/badges); ordinal rank
Open-ended feedback form	Post-training narrative reflections	Thematic + NLP sentiment analysis (RQ5)	Free text; 6 prompts (verbatim)

GGAI = Gamified GBH Assessment Instrument; LMS = learning management system; NLP = natural language processing; RQ = research question. Likert items range from 1 (strongly disagree) to 5 (strongly agree).

4.7 Psychometric validation plan

Validation proceeds in two stages. A content-validation stage uses Lawshe's CVR with 10 expert judges (critical value 0.62) (Lawshe, 1975; Polit and Beck, 2006). The instrument is then administered in a dedicated psychometric pilot whose sample is sized for factor analysis: a minimum of 300 respondents (≥ 10 cases per reflective item; N ≥ 200 floor) (Hu and Bentler, 1999; Rosseel, 2012). Because the three subscales differ in measurement type, they are analysed with appropriate models rather than a single reflective framework. The dichotomous knowledge scale (correct/incorrect items) is evaluated with item-response theory (2 parameter logistic model: item difficulty, item discrimination and fit) and KR-20 reliability, treating knowledge as an attainment/test score rather than a reflective latent trait. The attitude and behavioural-intention scales (Likert items) are evaluated with confirmatory factor analysis in R via lavaan, using robust (WLSMV) estimation appropriate to ordinal data, with acceptable fit defined as CFI ≥ 0.90, RMSEA ≤ 0.08, SRMR ≤ 0.08 (Hu and Bentler, 1999; Rosseel, 2012) and internal consistency by Cronbach's alpha and McDonald's omega (target ≥ 0.80) (McDonald, 1999; Streiner, 2003). Criterion validity is assessed by correlating GGAI behavioural-intention scores with a behaviourally anchored, supervisor-rated index of harassment-awareness conduct collected over a four-week window from supervisors blind to condition; inter-rater reliability is established on a subsample (two raters per participant; target ICC ≥ 0.70) before the index is used, and the target validity coefficient is r ≥ 0.40 (Messick, 1995). Test–retest reliability is estimated on an independent subsample (n = 60) over a two-week interval using an alternate-form or counterbalanced administration to limit practice effects (target ICC ≥ 0.75).

4.8 Data analysis plan

The analytic methods and decision criteria for each research question are summarised in Table 2.

Table 2. Data analysis plan mapped to research questions.

Research question	Analytic method	Decision criterion
RQ1: Psychometric properties	CVR; CFA (attitude/intention, WLSMV); IRT/KR-20 (knowledge); α, ω; ICC	CVR ≥ 0.62; CFI ≥ 0.90, RMSEA ≤ 0.08; α/ω ≥ 0.80; KR-20 ≥ 0.70; ICC ≥ 0.75
RQ2: Knowledge gain	Mixed 2 × 2 ANOVA (condition × time); ANCOVA	Interaction $p < 0.05$; Cohen's d/partial η^2 with 95% CI
RQ3: Behavioural confidence	Chi-square; binary logistic regression	$p < 0.05$; OR ≥ 2.0 with 95% CI

Research question	Analytic method	Decision criterion
RQ4: Gamification predictors	Multiple linear regression	β at $p < 0.05$; $R^2 \geq 0.30$ with 95% CI
RQ5: Qualitative experience	Thematic analysis (Braun and Clarke, 2006); NLP (VADER/BERT) validated on human-annotated gold standard	Classifier accuracy/F1 reported vs. gold standard; > 60% positive (pre-registered, benchmarked); ≥ 3 themes

CVR = content validity ratio; CFA = confirmatory factor analysis; IRT = item response theory; ICC = intraclass correlation coefficient; OR = odds ratio.

Quantitative analyses are conducted in SPSS v27 and R (lavaan, psych, mirt). Analyses are mapped to research questions as follows. RQ1 (psychometrics): content validity via Lawshe's CVR; construct validity via CFA (WLSMV) for the reflective scales and a 2-parameter logistic IRT model for the knowledge scale; reliability by Cronbach's alpha and McDonald's omega (Likert scales) and KR-20 (knowledge); criterion validity via Pearson/point-biserial correlations; test–retest via ICC. RQ2 (knowledge gain): a 2 (condition) × 2 (time) mixed-design ANOVA on knowledge scores, with the condition × time interaction as the key test, followed by ANCOVA controlling for baseline and covariates; partial η^2 and Cohen's d with 95% confidence intervals are reported. RQ3 (behavioural confidence): chi-square and binary logistic regression, reporting odds ratios with 95% CIs. RQ4 (gamification predictors): hierarchical multiple regression of behavioural-confidence change on points, badges, leaderboard engagement, and simulation completion, reporting standardised β, R^2, and CIs. RQ5 (qualitative): thematic analysis triangulated with NLP, reporting inter-coder reliability and classifier metrics. Statistical assumptions are explicitly tested: normality (Shapiro-Wilk and Q-Q plots), homogeneity of variance (Levene's test) and sphericity (Mauchly's test with Greenhouse-Geisser correction when violated). Non-parametric alternatives (Mann-Whitney U, Wilcoxon signed-rank) or robust/bootstrapped estimators and adequate transformations are used where assumptions are not met. Outliers are screened using standardised residuals and Cook's distance, and handled by sensitivity analyses rather than silent deletion. All inferential tests are reported with effect sizes and 95% confidence intervals. The Benjamini–Hochberg false-discovery-rate procedure is used to control the false discovery rate across the confirmatory hypotheses. Confirmatory tests are pre-registered and exploratory analyses are reported separately. Exploratory analyses include whether gender, age, role, or prior gamified-learning experience moderate training effects.

4.9 Ethical considerations

The study involves human participants and will not begin until written approval is obtained from the Institutional Review Board / Research Ethics Committee of Amity University, Uttar Pradesh (approval number to be inserted on grant of approval). The protocol adheres to the Declaration of Helsinki and to APA ethical standards (AERA et al., 2014). Each participant is provided with a plain-language description of purpose, procedures, risks, voluntariness, and right to withdraw at any time without penalty prior to access to any study materials, and written informed consent is obtained. Participation is on a voluntary basis and not linked to employment evaluation. Confidentiality and privacy are protected as follows: responses are pseudonymised at collection; individual responses are never disclosed to employers; organisational reporting uses aggregate data only; and data are stored on encrypted, access-controlled institutional servers and deleted no later than five years after completion. Because open-ended responses are processed by NLP tools, all text is de-identified before analysis and processed on institutionally controlled infrastructure, so automated analysis does not compromise participant privacy. Given the topic's sensitivity, a confidential referral pathway to professional support is available at every stage, and no deception is used.

5 ANTICIPATED OUTCOMES AND DISCUSSION

As a registered protocol and methodological contribution, this paper specifies the framework and analysis plan rather than reporting collected data; no empirical results are presented and all hypotheses await testing. Drawing on the meta-analytic evidence base, we articulate the outcomes the framework is designed to test and the mechanisms through which they are expected to arise.

For RQ1, the staged validation plan is designed to yield an instrument with documented content, construct, and criterion validity and acceptable reliability—filling the measurement gap identified in Section 2.3 (Messick, 1995; AERA et al., 2014). For RQ2

and RQ3, the gamification literature gives strong reason to anticipate larger knowledge gains and greater reporting confidence in the gamified condition: Meta-analyses show positive effects of gamification and serious games on cognitive and behavioural outcomes (Clark et al., 2016; Bai et al., 2020; Huang et al., 2020; Sailer and Homner, 2020), and the theory of gamified learning specifies the behavioural pathway (engagement, time-on-task, persistence) through which those effects operate (Landers and Landers, 2014). The expected effect sizes are consistent with the medium effects reported across the field (Wouters et al., 2013; Hamari et al., 2014).

For RQ4, isolating the contribution of individual elements responds to a recognised need to move beyond whether gamification works to which elements drive outcomes (Garcia-Cabot et al., 2015; Mekler et al., 2017; Sailer et al., 2017). We anticipate that the key drivers will be engagement with the leaderboard and completion of simulations, as found for performance feedback and challenge (Sitzmann, 2011; Hamari, 2017), while we remain cognisant of the well-documented risks of over-competition for certain learner profiles (Attali and Arieli-Attali, 2015; Sung et al., 2016; Toda et al., 2018). For RQ5, the NLP pipeline should identify a mainly positive sentiment and recurring themes (e.g., realism of scenarios, psychological safety, perceived relevance) triangulating the quantitative findings (Braun and Clarke, 2006; Hutto and Gilbert, 2014).

These expectations are dampened by two boundary conditions. First, gamification effects are moderated by context, learner type, and design quality. Null and mixed results are well represented in the literature (Hanus and Fox, 2015; Koivisto and Hamari, 2019; Sailer and Homner, 2020). Second, knowledge and attitude gains do not guarantee transfer to workplace behaviour; the training-transfer literature cautions that transfer depends on design and climate as much as on the intervention itself, and that evaluation should span reaction, learning, behaviour, and results (Baldwin and Ford, 1988; Kirkpatrick, 1994; Arthur et al., 2003; Salas et al., 2012). The framework's criterion-validity component and behavioural-intention subscale are included precisely to probe this transfer question rather than assume it.

All outcome statements in this paper are conditional and forward-looking: where we refer to expected gains, these should be read as 'if the hypotheses are supported by the planned analyses' rather than as demonstrated effects. Several alternative explanations are explicitly anticipated and probed. A novelty effect could inflate short-term engagement and sentiment; this is addressed by including a delayed retention component in future work and by treating sentiment as a secondary indicator. Social-desirability and demand characteristics are likely given the sensitive topic and salient 'correct' answers; these are mitigated by anonymity, the criterion-validity (supervisor-rated, blind) component, and by measuring prior exposure to gamified learning so that effects can be distinguished from pre-existing familiarity. Because self-report dominates the behavioural-intention measure, intention is not equated with behaviour, and the transfer question is treated as open rather than assumed.

6 CONTRIBUTIONS TO THE FIELD

Learning-technology design. The framework specifies a theory-driven gamified learning environment for a high-stakes domain that conventional TEL has under-served, mapping each game element to an SDT need and a learning-relevant behaviour (Ryan and Deci, 2000; Landers and Landers, 2014).

Measurement. Based on a structured search of Scopus, Web of Science, PsycINFO and ERIC (search terms combining 'gamif*'/'serious game' with 'harassment'/'sexual harassment'/'gender-based' and 'psychometric'/'validation'/'scale'; no date limit), we did not identify a peer-reviewed instrument that psychometrically validates GBH knowledge, attitudes, and behavioural intentions for delivery within a gamified environment, particularly in the Indian financial sector. To the best of our knowledge from that search, the GGAI is therefore the first such instrument; we frame this as a bounded claim rather than an absolute one, and the full search strategy is in the supplementary material to allow scrutiny and replication (Lawshe, 1975; Messick, 1995; Hu and Bentler, 1999).

Analytics. The integration of NLP-based sentiment and theme detection demonstrates a scalable, reproducible pipeline for mixed quantitative–qualitative evaluation of learner experience (Hutto and Gilbert, 2014; Creswell and Plano Clark, 2018; Devlin et al., 2019).

Policy. By operationalising SDG 5 adherence as a measurable learning outcome, the framework gives organisations a replicable model for documenting contributions to global gender-equity commitments in rigorous rather than aspirational terms (United Nations, 2015; ILO, 2019).

7 LIMITATIONS AND FUTURE WORK

As a registered protocol, the principal limitation is the absence of collected empirical data; all hypotheses await testing and no efficacy is claimed here. The quasi-experimental, unit-allocated design, while practical in organisational settings, does not afford the causal strength of individual randomisation (Shadish et al., 2002); matched allocation, ANCOVA, and a planned propensity-score sensitivity analysis only partially offset this. Purposive selection of collaborating organisations bounds generalisability, which we report against sector benchmarks. Self-report behavioural-intention measures are susceptible to social-desirability bias, partially offset by the supervisor-rated criterion component (Messick, 1995). Future work should investigate the longer-term retention and transfer of behaviour, adaptive and mobile delivery of the gamified environment (Garcia-Cabot et al., 2015; Subhash and Cudney, 2018), and the comparative performance of lexicon-based versus transformer-based sentiment models on learner feedback (Hutto and Gilbert, 2014; Devlin et al., 2019).

8 CONCLUSIONS

Conventional harassment-prevention training fails not for want of legal mandates but because the instructional model does not engage learners, support retention, or change behaviour. This paper has proposed a technology-enhanced learning framework that addresses these three weaknesses together: a theory-driven gamified learning environment, a psychometrically validated assessment instrument (the GGAI), and an AI-based analytics pipeline for learner feedback, all tied to the measurable outcome of SDG 5. By treating assessment and learning-design rigour as equally important, the framework gives educational-technology researchers and practitioners a replicable model for building and evaluating gamified learning in a socially important domain. The next step is empirical: to instantiate the framework, validate the GGAI, and test the hypotheses set out here.

CONFLICT OF INTEREST

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

ETHICS STATEMENT

This protocol involves human participants. Data collection will begin only after written approval from the Institutional Review Board / Research Ethics Committee of Amity University Uttar Pradesh (approval number to be provided upon grant). The study adheres to the Declaration of Helsinki, and written informed consent will be obtained from all participants. Full procedures are described in Section 4.9. The study involves no animal experiments.

DATA AVAILABILITY

This is a registered protocol; no datasets have yet been generated. Upon completion and publication of the empirical study, the de-identified participant-level dataset (survey responses, scored GGAI items, and LMS interaction logs) will be deposited in a public repository (e.g., the Open Science Framework or Zenodo) with a citable DOI, subject to ethics approval and participant consent. Where organisational data are commercially sensitive or where re-identification risk precludes open sharing, those elements will be made available to qualified researchers on reasonable request and under a data-use agreement, and any such restriction will be stated explicitly with its justification. The pre-registration, study materials, and the GGAI item-to-SDG-5 mapping are available on the Open Science Framework.

CODE AVAILABILITY

All analysis code will be made publicly available to support reproducibility. The GGAI scoring algorithm (IRT and classical scoring), the R/SPSS analysis scripts (CFA, ANOVA/ANCOVA, regression), and the NLP pipeline (VADER and the fine-tuned BERT classifier, including training and evaluation scripts and the human-annotation schema) will be deposited in a version-controlled public repository (GitHub) and archived with a DOI (Zenodo) at publication, consistent with the journal's data and code availability policy. No custom code is withheld; any third-party model weights or licences required to run the pipeline will be documented in the repository.

DECLARATION OF GENERATIVE AI USE

The authors used a generative AI assistant (large language model) during manuscript preparation only to improve readability and language and to check the consistency of in-text citations against the reference list. All scholarly content, arguments, study design, and interpretation are the authors' own; all AI-assisted text was reviewed and revised by the authors, who take full

responsibility for the content. Generative AI was not used to generate or analyse research data and is not listed as an author. The natural language processing tools described in Section 4 are part of the planned research methodology, not writing assistance.

REFERENCES

[1] AERA, APA, NCME, 2014. Standards for Educational and Psychological Testing. American Educational Research Association, Washington, DC.

[2] Antecol, H., Cobb-Clark, D., 2003. Does sexual harassment training change attitudes? A view from the federal level. Social Science Quarterly 84, 826–842. https://doi.org/10.1046/j.0038-4941.2003.08404001.x

[3] Arthur, W., Bennett, W., Edens, P.S., Bell, S.T., 2003. Effectiveness of training in organizations: A meta-analysis of design and evaluation features. Journal of Applied Psychology 88, 234–245. https://doi.org/10.1037/0021-9010.88.2.234

[4] Attali, Y., Arieli-Attali, M., 2015. Gamification in assessment: Do points affect test performance? Computers & Education 83, 57–63. https://doi.org/10.1016/j.compedu.2014.12.012

[5] Bai, S., Hew, K.F., Huang, B., 2020. Does gamification improve student learning outcome? Evidence from a meta-analysis and synthesis of qualitative data in educational contexts. Educational Research Review 30, 100322. https://doi.org/10.1016/j.edurev.2020.100322

[6] Baldwin, T.T., Ford, J.K., 1988. Transfer of training: A review and directions for future research. Personnel Psychology 41, 63–105. https://doi.org/10.1111/j.1744-6570.1988.tb00632.x

[7] Bingham, S.G., Scherer, L.L., 2001. The unexpected effects of a sexual harassment educational program. Journal of Applied Behavioral Science 37, 125–153. https://doi.org/10.1177/0021886301372001

[8] Braun, V., Clarke, V., 2006. Using thematic analysis in psychology. Qualitative Research in Psychology 3, 77–101. https://doi.org/10.1191/1478088706qp063oa

[9] Buckley, P., Doyle, E., 2016. Gamification and student motivation. Interactive Learning Environments 24, 1162–1175. https://doi.org/10.1080/10494820.2014.964263

[10] Clark, D.B., Tanner-Smith, E.E., Killingsworth, S.S., 2016. Digital games, design, and learning: A systematic review and meta-analysis. Review of Educational Research 86, 79–122. https://doi.org/10.3102/0034654315582065

[11] Cohen, J., 1988. Statistical Power Analysis for the Behavioral Sciences, 2nd ed. Lawrence Erlbaum Associates, Hillsdale, NJ. https://doi.org/10.4324/9780203771587

[12] Connolly, T.M., Boyle, E.A., MacArthur, E., Hainey, T., Boyle, J.M., 2012. A systematic literature review of empirical evidence on computer games and serious games. Computers & Education 59, 661–686. https://doi.org/10.1016/j.compedu.2012.03.004

[13] Cortina, L.M., Berdahl, J.L., 2008. Sexual harassment in organizations: A decade of research in review, In: Barling, J., Cooper, C.L. (Eds.), The SAGE Handbook of Organizational Behavior, Vol. 1. SAGE, London, pp. 469–497. https://doi.org/10.4135/9781849200448.n26

[14] Creswell, J.W., Plano Clark, V.L., 2018. Designing and Conducting Mixed Methods Research, 3rd ed. SAGE, Thousand Oaks, CA.

[15] Cronbach, L.J., 1951. Coefficient alpha and the internal structure of tests. Psychometrika 16, 297–334. https://doi.org/10.1007/BF02310555

[16] Deci, E.L., Ryan, R.M., 1985. Intrinsic Motivation and Self-Determination in Human Behavior. Plenum Press, New York. https://doi.org/10.1007/978-1-4899-2271-7

[17] Denny, P., 2013. The effect of virtual achievements on student engagement, In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, pp. 763–772. https://doi.org/10.1145/2470654.2470763

[18] Deterding, S., Dixon, D., Khaled, R., Nacke, L., 2011. From game design elements to gamefulness: Defining "gamification", In: Proceedings of the 15th International Academic MindTrek Conference. ACM, New York, pp. 9–15. https://doi.org/10.1145/2181037.2181040

[19] Devlin, J., Chang, M.-W., Lee, K., Toutanova, K., 2019. BERT: Pre-training of deep bidirectional transformers for language understanding, In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT). ACL, pp. 4171–4186. https://doi.org/10.18653/v1/N19-1423

[20] Dichev, C., Dicheva, D., 2017. Gamifying education: What is known, what is believed and what remains uncertain: A critical review. International Journal of Educational Technology in Higher Education 14, 9. https://doi.org/10.1186/s41239-017-0042-5

[21] Dicheva, D., Dichev, C., Agre, G., Angelova, G., 2015. Gamification in education: A systematic mapping study. Educational Technology & Society 18, 75–88. https://eric.ed.gov/?id=EJ1070047

[22] Faul, F., Erdfelder, E., Lang, A.-G., Buchner, A., 2007. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods 39, 175–191. https://doi.org/10.3758/BF03193146

[23] Fitzgerald, L.F., Drasgow, F., Hulin, C.L., Gelfand, M.J., Magley, V.J., 1997. Antecedents and consequences of sexual harassment in organizations: A test of an integrated model. Journal of Applied Psychology 82, 578–589. https://doi.org/10.1037/0021-9010.82.4.578

[24] Fitzgerald, L.F., Swan, S., Fischer, K., 1995. Why didn't she just report him? The psychological and legal implications of women's responses to sexual harassment. Journal of Social Issues 51, 117–138. https://doi.org/10.1111/j.1540-4560.1995.tb01312.x

[25] Garcia-Cabot, A., de-Marcos, L., Garcia-Lopez, E., 2015. An empirical study on m-learning adaptation: Learning performance and learning contexts. Computers & Education 82, 450–459. https://doi.org/10.1016/j.compedu.2014.12.007

[26] Hamari, J., 2017. Do badges increase user activity? A field experiment on the effects of gamification. Computers in Human Behavior 71, 469–478. https://doi.org/10.1016/j.chb.2015.03.036

[27] Hamari, J., Koivisto, J., Sarsa, H., 2014. Does gamification work? A literature review of empirical studies on gamification, In: Proceedings of the 47th Hawaii International Conference on System Sciences. IEEE, pp. 3025–3034. https://doi.org/10.1109/HICSS.2014.377

[28] Hamari, J., Shernoff, D.J., Rowe, E., Coller, B., Asbell-Clarke, J., Edwards, T., 2016. Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. Computers in Human Behavior 54, 170–179. https://doi.org/10.1016/j.chb.2015.07.045

[29] Hanus, M.D., Fox, J., 2015. Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. Computers & Education 80, 152–161. https://doi.org/10.1016/j.compedu.2014.08.019

[30] Hu, L., Bentler, P.M., 1999. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling 6, 1–55. https://doi.org/10.1080/10705519909540118

[31] Huang, R., Ritzhaupt, A.D., Sommer, M., Zhu, J., Stephen, A., Valle, N., Hampton, J., Li, J., 2020. The impact of gamification in educational settings on student learning outcomes: A meta-analysis. Educational Technology Research and Development 68, 1875–1901. https://doi.org/10.1007/s11423-020-09807-z

[32] Hutto, C.J., Gilbert, E., 2014. VADER: A parsimonious rule-based model for sentiment analysis of social media text, In: Proceedings of the 8th International AAAI Conference on Web and Social Media 8, pp. 216–225. https://doi.org/10.1609/icwsm.v8i1.14550

[33] ILO, 2019. Violence and Harassment Convention, 2019 (No. 190). International Labour Organization, Geneva. https://www.ilo.org/resource/ilc/ilc108/violence-and-harassment-convention-2019-no-190

[34] Kirkpatrick, D.L., 1994. Evaluating Training Programs: The Four Levels. Berrett-Koehler, San Francisco.

[35] Koivisto, J., Hamari, J., 2019. The rise of motivational information systems: A review of gamification research. International Journal of Information Management 45, 191–210. https://doi.org/10.1016/j.ijinfomgt.2018.10.013

[36] Krath, J., Schürmann, L., von Korflesch, H.F.O., 2021. Revealing the theoretical basis of gamification: A systematic review and analysis of theory in research on gamification, serious games and game-based learning. Computers in Human Behavior 125, 106963. https://doi.org/10.1016/j.chb.2021.106963

[37] Landers, R.N., Landers, A.K., 2014. An empirical test of the theory of gamified learning: The effect of leaderboards on time-on-task and academic performance. Simulation & Gaming 45, 769–785. https://doi.org/10.1177/1046878114563662

[38] Lawshe, C.H., 1975. A quantitative approach to content validity. Personnel Psychology 28, 563–575. https://doi.org/10.1111/j.1744-6570.1975.tb01393.x

[39] Looyestyn, J., Kernot, J., Boshoff, K., Ryan, J., Edney, S., Maher, C., 2017. Does gamification increase engagement with online programs? A systematic review. PLOS ONE 12, e0173403. https://doi.org/10.1371/journal.pone.0173403

[40] Magley, V.J., Gallus, J.A., Bunk, J.A., 2010. The gendered nature of workplace mistreatment, In: Chrisler, J.C., McCreary, D.R. (Eds.), Handbook of Gender Research in Psychology. Springer, New York, pp. 423–441. https://doi.org/10.1007/978-1-4419-1467-5_18

[41] Majuri, J., Koivisto, J., Hamari, J., 2018. Gamification of education and learning: A review of empirical literature, In: Proceedings of the 2nd International GamiFIN Conference. CEUR Workshop Proceedings 2186, pp. 11–19. https://ceur-ws.org/Vol-2186/paper2.pdf

[42] McDonald, P., 2012. Workplace sexual harassment 30 years on: A review of the literature. International Journal of Management Reviews 14, 1–17. https://doi.org/10.1111/j.1468-2370.2011.00300.x

[43] McDonald, R.P., 1999. Test Theory: A Unified Treatment. Lawrence Erlbaum Associates, Mahwah, NJ. https://doi.org/10.4324/9781410601087

[44] Mekler, E.D., Brühlmann, F., Tuch, A.N., Opwis, K., 2017. Towards understanding the effects of individual gamification elements on intrinsic motivation and performance. Computers in Human Behavior 71, 525–534. https://doi.org/10.1016/j.chb.2015.08.048

[45] Merchant, Z., Goetz, E.T., Cifuentes, L., Keeney-Kennicutt, W., Davis, T.J., 2014. Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis. Computers & Education 70, 29–40. https://doi.org/10.1016/j.compedu.2013.07.033

[46] Messick, S., 1995. Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. American Psychologist 50, 741–749. https://doi.org/10.1037/0003-066X.50.9.741

[47] Nacke, L.E., Deterding, S., 2017. The maturing of gamification research. Computers in Human Behavior 71, 450–454. https://doi.org/10.1016/j.chb.2016.11.062

[48] Plass, J.L., Homer, B.D., Kinzer, C.K., 2015. Foundations of game-based learning. Educational Psychologist 50, 258–283. https://doi.org/10.1080/00461520.2015.1122533

[49] Polit, D.F., Beck, C.T., 2006. The content validity index: Are you sure you know what's being reported? Critique and recommendations. Research in Nursing & Health 29, 489–497. https://doi.org/10.1002/nur.20147

[50] Roehling, M.V., Huang, J., 2018. Sexual harassment training effectiveness: An interdisciplinary review and call for research. Journal of Organizational Behavior 39, 134–150. https://doi.org/10.1002/job.2257

[51] Rosseel, Y., 2012. lavaan: An R package for structural equation modeling. Journal of Statistical Software 48, 1–36. https://doi.org/10.18637/jss.v048.i02

[52] Ryan, R.M., Deci, E.L., 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist 55, 68–78. https://doi.org/10.1037/0003-066X.55.1.68

[53] Sailer, M., Hense, J.U., Mayr, S.K., Mandl, H., 2017. How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction. Computers in Human Behavior 69, 371–380. https://doi.org/10.1016/j.chb.2016.12.033

[54] Sailer, M., Homner, L., 2020. The gamification of learning: A meta-analysis. Educational Psychology Review 32, 77–112. https://doi.org/10.1007/s10648-019-09498-w

[55] Salas, E., Tannenbaum, S.I., Kraiger, K., Smith-Jentsch, K.A., 2012. The science of training and development in organizations: What matters in practice. Psychological Science in the Public Interest 13, 74–101. https://doi.org/10.1177/1529100612436661

[56] Seaborn, K., Fels, D.I., 2015. Gamification in theory and action: A survey. International Journal of Human-Computer Studies 74, 14–31. https://doi.org/10.1016/j.ijhcs.2014.09.006

[57] Shadish, W.R., Cook, T.D., Campbell, D.T., 2002. Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Houghton Mifflin, Boston.

[58] Sitzmann, T., 2011. A meta-analytic examination of the instructional effectiveness of computer-based simulation games. Personnel Psychology 64, 489–528. https://doi.org/10.1111/j.1744-6570.2011.01190.x

[59] Streiner, D.L., 2003. Starting at the beginning: An introduction to coefficient alpha and internal consistency. Journal of Personality Assessment 80, 99–103. https://doi.org/10.1207/S15327752JPA8001_18

[60] Subhash, S., Cudney, E.A., 2018. Gamified learning in higher education: A systematic review of the literature. Computers in Human Behavior 87, 192–206. https://doi.org/10.1016/j.chb.2018.05.028

[61] Sung, Y.-T., Chang, K.-E., Liu, T.-C., 2016. The effects of integrating mobile devices with teaching and learning on students' learning performance: A meta-analysis and research synthesis. Computers & Education 94, 252–275. https://doi.org/10.1016/j.compedu.2015.11.008

[62] Toda, A.M., Valle, P.H.D., Isotani, S., 2018. The dark side of gamification: An overview of negative effects of gamification in education, In: Cristea, A.I., Bittencourt, I.I., Lima, F. (Eds.), Higher Education for All. From Challenges to Novel Technology-Enhanced Solutions. Communications in Computer and Information Science 832. Springer, Cham, pp. 143–156. https://doi.org/10.1007/978-3-319-97934-2_9

[63] Tondello, G.F., Wehbe, R.R., Diamond, L., Busch, M., Marczewski, A., Nacke, L.E., 2016. The gamification user types Hexad scale, In: Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play. ACM, pp. 229–243. https://doi.org/10.1145/2967934.2968082

[64] UN Women, UN DESA, 2020. Progress on the Sustainable Development Goals: The Gender Snapshot 2020. United Nations, New York. https://www.unwomen.org/en/digital-library/publications/2020/09/progress-on-the-sustainable-development-goals-the-gender-snapshot-2020

[65] United Nations, 2015. Transforming our world: The 2030 Agenda for Sustainable Development. Resolution A/RES/70/1, UN General Assembly. https://sdgs.un.org/2030agenda

[66] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need, In: Advances in Neural Information Processing Systems 30, pp. 5998–6008. https://arxiv.org/abs/1706.03762

[67] Vlachopoulos, D., Makri, A., 2017. The effect of games and simulations on higher education: A systematic literature review. International Journal of Educational Technology in Higher Education 14, 22. https://doi.org/10.1186/s41239-017-0062-1

[68] Willness, C.R., Steel, P., Lee, K., 2007. A meta-analysis of the antecedents and consequences of workplace sexual harassment. Personnel Psychology 60, 127–162. https://doi.org/10.1111/j.1744-6570.2007.00067.x

[69] Wouters, P., van Nimwegen, C., van Oostendorp, H., van der Spek, E.D., 2013. A meta-analysis of the cognitive and motivational effects of serious games. Journal of Educational Psychology 105, 249–265. https://doi.org/10.1037/a0031311

[70] Zainuddin, Z., Chu, S.K.W., Shujahat, M., Perera, C.J., 2020. The impact of gamification on learning and instruction: A systematic review of empirical evidence. Educational Research Review 30, 100326. https://doi.org/10.1016/j.edurev.2020.100326