Received: 9 March 2021 | Revised: 23 March 2022 | Accepted: 26 May 2022

DOI: 10.1111/jcal.12703

ARTICLE

Journal of Computer Assisted Learning WILEY

Satisfaction is insufficient: Insights from a randomized, controlled trial of a marketing simulation game

Siv-Elisabeth Skjelbred | Stephan Daus

Nordic Institute for Studies in Innovation, Research and Education, Oslo, Norway

Correspondence
Siv-Elisabeth Skjelbred, Nordic Institute for Studies in Innovation, Research and Education, Oslo, Norway.
Email: siv-elisabeth.skjelbred@nifu.no

Funding information
Regionale forskningsfond Oslofjordfondet, Grant/Award Number: 296402

Abstract

Background: The development and promotion of educational games are still outpacing knowledge of these games' effects, raising calls for evidence of benefits and challenges. Studies suggest that students and teachers like games, but the payoff of the investment in terms of increased motivation and achievement remains unclear.

Objectives: This study investigates the pure effect of a marketing simulation game on motivation, perceived learning and achievement, above and beyond regular student-active instruction.

Methods: We applied a randomized, controlled experiment in a marketing course in upper-secondary schools ($N_{\text{classes}} = 22$; $N_{\text{students}} = 433$) comparing a collaborative–competitive marketing simulation game with regular, case-based, student-active instruction on three groups of outcome measures: motivation, perceived ability, and achievement. Additionally, students and teachers provided quantitative and qualitative feedback on game experiences.

Results and Conclusions: We showcase the importance of a robust study design with valid compound instruments. Moreover, investigations of the game implementation and experiences reveal insights about intervention timing, differential negative consequences by gender and need for reflection opportunities. We find no clear evidence of positive or negative effects of the game, despite students' and teachers' satisfaction.

Implications: Beyond the effect evaluation, we offer recommendations to researchers and developers of educational games about scaffolding, timing and teacher competence building.

KEYWORDS
digitalization, game-based learning, marketing education, randomized trial, simulation games

1 | INTRODUCTION

Digital learning technologies have gradually attracted increasing attention from educators and researchers. With the drive towards integrating technology into the classroom, it is vital to understand where such tools deliver and where their potential has yet to be explored (Coleman & Money, 2020).

Evaluating the successes and challenges of digital tools requires clearly defined outcomes, suitable measurement instruments and sound design. Despite the ever-increasing literature on educational games, few studies have employed standardized and validated outcome measures for use across classrooms, exemplified by the common use of local assessment activities of unknown or poor quality (Wollscheid & Skjelbred, 2021; Vos, 2015).

To infer that educational games improve outcomes requires a research design that can establish causal effects. Many existing studies employed a before-after analysis or a quasi-experiment, which may produce false positive results because of natural variation in student performance (Boevé et al., 2019). Some lack before-after and control group design features, analysing only post-test data for the treatment group only (Buil et al., 2017). Randomized, controlled trials are considered the gold standard for establishing an intervention's effect (Escueta et al., 2020). When employed correctly, differences in outcome measures between the intervention and comparison groups can be attributed to the intervention, in our case, a game.

A major question with educational games is not only whether students gain from games, but whether the gains justify the investment when the game is tested against viable and less expensive alternative ways to increase motivation, knowledge and skills (Clark, 2007). A compelling argument for digital learning games requires an improvement in outcomes or the generation of outcomes not provided by traditional instruction. As such, effects should be both statistically significant and of a magnitude that is important to practitioners and policymakers to make the costs worthwhile.

Using educational games to teach aims towards, and commonly results in, student-active instruction,[1] yet effect studies of games have tended to compare game instruction with traditional teacher-centred instruction (Gee, 2011). A comparison between games and another form of student-active instruction can provide the game with a tougher competitor and produce more information about whether games are more efficient than other high-quality alternative instruction methods. Such a comparison is useful given the costs of game interventions.

School learning often leads to disinterested and disengaged students, resulting in the introduction of educational games as a solution (Chee, 2015). Educational game popularity rests on the idea that games allow play and learning to merge and that games are inherently motivating (Munkvold & Sigurdardottir, 2018). However, questions remain regarding whether popularity and user satisfaction are sufficient evidence for effectiveness and how game use can be structured to ensure that satisfaction is transformed into measurable improvements in motivation and learning. To investigate these issues, we compare outcomes of students randomized to use the educational game with those using regular student-active learning on three groups of outcomes: motivational measures of interest, perceived ability, and achievement, while also evaluating satisfaction among game users and qualitative interview data.

2 | BACKGROUND

In this paper, we consider educational games to be digital games that combine serious learning and interactive entertainment features (Prensky, 2001, p. 5). They typically have specific learning goals and objectives related to gameplay (Nadolny & Halabi, 2016; Pivec, 2007; Pivec et al., 2003). There exist a large variety of games, including simulation games which are interactive games with realistic representations of reality and clearly defined aims and interaction rules (Ranchhod et al., 2014).[2] Identifying effects of digital games requires specificity, and the highest chance of detecting effects and achieving high internal validity occurs when the game purpose, assessment measures and curriculum objectives approximately align (i.e., statistical power; Martone & Sireci, 2009).

In the case of educational interventions, especially digital games, the projected primary outcomes are typically knowledge, skills, attitudes or behaviour (All et al., 2014). Furthermore, motivation for continued learning is usually a desired additional outcome. Researchers have considered intrinsic motivation for performing the activity to be an important aspect of games that can benefit higher levels of enjoyment, interest and performance, as well as resulting in higher learning quality and heightened self-efficacy (Ryan & Deci, 2000). To identify whether game instruction meets its purpose, evidence is needed to see whether the instrument captures changes in the desired outcome.

Pervasive student disengagement is an international problem (Hamari et al., 2016) that likely results from multiple causes. One possible explanation is a need to introduce theory and terminology before students can apply this content knowledge in practically oriented, realistic and contextualized exercises at the end of the course. This delayed introduction of the latter part challenges the engagement of less motivated and more impatient students who are at risk of falling behind early or disengaging completely. Games are most effective when they build bridges between the domain of the game world and an overlapping domain of professional practice (Delwiche, 2006, p. 160; Plass et al., 2015). Thus, early introduction of an intuitive, hands on and interactive digital game can hypothetically increase student motivation for the course, including motivation to learn the terminology and theory demanded by the game. Teachers might employ this understanding when they cite student motivation as a primary reason for using games in instruction (Munkvold & Sigurdardottir, 2018).

Despite this assumption, students' willingness to embrace this kind of activity in school is not self-evident (Hanghøj, 2011; Nousiainen et al., 2018; Sandford et al., 2006; Squire, 2005). Educational games may have lower entertainment value than traditional video games, making them less interesting than their entertaining counterparts (Bellotti et al., 2013). Even if students are satisfied with the game, effectiveness may be diluted when education and entertainment are combined (Barzilai & Blau, 2014). In a review of 31 studies, Wouters and colleagues found no effect of educational games on motivation compared with regular instruction (Wouters et al., 2013). Wouters et al. included motivation in broad terms, by including motivation, interest, engagement, and attitude towards the topic and further showed that motivation is not merely a predictor of learning but also a key educational outcome. It is thus unclear which measures of motivation should be used to capture the effect of educational games. The literature further indicates that it is difficult to optimize the learning elements consistent with interrelated game principles of challenge, engagement, flow and immersion (Hamari et al., 2016), and the conditions for this balance remain to be identified.

Many studies and reviews of educational games have focused on students' perceptions of outcomes such as satisfaction, motivation, perceived ability or perceived learning. Research that examines participants' affective reactions has found that students tend to like educational games, feel they have benefitted from playing and view them more positively than lectures and case discussions (Anderson & Lawton, 2009; Faria, 2001; Kolić-Vehovec et al., 2019). Whether student perceptions are suited for evaluating the success of educational games depends on the outcome studied and the purpose of using games in education (Bacon, 2016). Regardless, student satisfaction and perceived competence represent only one possible objective with educational games.

Although leveraging entertainment with learning environments may increase motivation and allow learners to gain skills and knowledge, students' satisfaction is distinct from competence improvements (Daus et al., 2021; Bawa, 2020; Schumann et al., 2014). Furthermore, empirical studies have firmly established that student-perceived learning and actual student learning are different constructs (Bacon, 2016; Clayson, 2009). Sitzmann et al. (2010) found zero correlation between self-reported knowledge gain, meaning perceived learning, and actual knowledge. Similarly, a meta-analysis of student-evaluations and objectively measured achievement found only a small positive correlation between the two (Clayson, 2009). Most existing literature on educational games and learning has relied on students' and teachers' perceptions of the game's effectiveness. Perceptions of ability address feelings related to learning and the belief that learning has occurred. If game instruction results in a higher subjective assessment of one's abilities than non-game instruction, the game may have increased the student's confidence, which is related to academic self-efficacy (Bong & Skaalvik, 2003). As such, students' perceptions about their own academic competence are important for motivation, life-long learning and performance (Bong & Skaalvik, 2003).

Conversely, measures of actual knowledge rely on direct evidence of learning (Anderson & Lawton, 2009), occasionally called "objective" evaluations (e.g., Schumann et al., 2014). This term is somewhat imprecise, as any assessment is subjective by being situated in the assessment developer's and interpreter's perspectives. We use the term *achievement* to emphasize that we are interested in measuring domain-specific knowledge and skills rather than domain-generic abilities, which is an elusive research field fraught with obstacles (Tricot & Sweller, 2014). Competence improvements from games potentially cover both lower cognitive levels, such as knowing theory and terminology, and higher levels, such as applying knowledge across contexts and integrating multiple curriculum objectives into a larger whole. Yet, games are generally inefficient for learning terminology and theory compared with lectures (Anderson & Lawton, 2009). This finding contradicts the earlier mentioned game purpose of combatting disengagement among students who find terminology and theory uninspiring.

The benefits of educational games depend on the game's learning strategies and those used by the teacher in the game instruction. Games may incorporate motivational drivers such as competition and collaboration, which are important considerations when studying the effect of games, as different student groups may respond differently to various features. For example, extensive literature has shown that girls respond differently to competition than boys (Gneezy & Rustichini, 2004; Morin, 2015; Niederle & Vesterlund, 2010, 2011; Tobin & Garnett, 2003).

During gameplay, players acquire "intuitive knowledge" (Swaak & de Jong, 1996) about rules and strategies. Prompting students to reflect on what they do and experience can make this knowledge more explicit. Digital learning games are thus typically not stand-alone devices and require instructors to guide the learning process (e.g., Wouters & van Oostendorp, 2017). Gameplay coupled with external teacher-assisted scaffolding increases the opportunity to reflect and abstract relevant information for effective learning to occur (Barzilai & Blau, 2014). This understanding echoes Dewey's early idea about experience and reflection equalling learning (Dewey, 1938).

After examining the literature, we summarize the research gap as follows. The literature has suggested that most students like games, but the extent to which educational games result in measurable increases in desirable outcomes remains unclear. While some studies (e.g., Brom et al., 2016) have exhibited higher-quality designs, a strong need exists for further studies that provide outcome instruments with good validation documentation and a robust research design allowing causal interpretation of the effectiveness of educational games over other types of instruction building on the same pedagogy (Anderson & Lawton, 2009; Wollscheid & Skjelbred, 2021; Escueta et al., 2020; Wouters et al., 2013). O'Neil et al. (2005) noted that the methods used in games are not unique to games but are also used effectively in non-game instruction. A key question is thus whether digital games produce significantly more learning or motivation than other instructional platforms.

The existing literature and the large number of educational games on the market suggest that games are a popular remedy, so we ask the questions: Is the popularity of educational games sufficient evidence for their effectiveness? How can game use be structured to ensure that satisfaction is transformed into measurable improvements in educational outcomes? In this paper, we move beyond satisfaction with the game to investigate whether the incorporation of a widely used digital simulation game leads to measurable improvements in interest in the course and the content, perceived ability, and achievement.

By employing a randomized, controlled trial comparing a digital game to an alternative instruction method and using validated instruments that capture a broad variety of outcomes, we seek to showcase and contribute a path towards building evidence on digital learning tools. Investigating a digital simulation game in marketing education, we identify the effect of an embedded educational game beyond regular student-active instruction.

3 | METHOD

During the autumn of the 2019–2020 school year, we conducted a cluster-randomized, controlled trial to gauge the effectiveness of a simulation game in the elective course Marketing and Leadership during the second year of Norwegian upper-secondary education. We collected background information and outcome measures in November 2019, a few weeks after the intervention group ended the simulation game instruction. Students answered the survey in the classroom with the teacher present. We surveyed teachers and conducted in-depth interviews in early spring 2020.

3.1 | Participants

We conducted the project in collaboration with the Municipal Educational Officer in Oslo, so we invited all 24 public upper-secondary schools in this municipality. However, some schools refused, and some schools left the project after sign-up. Thus, to increase statistical power, we also invited some schools from other regions. Hubro Education, the company providing the simulation game, recruited schools outside the municipality of Oslo. Only schools not already using the game were eligible.

A total of 22 marketing teachers and their upper-secondary school students, aged 16–18, participated in the trial. We conducted a stratified randomization at the teacher level. The stratification ensured that the intervention and control groups were balanced on important characteristics, which increases the study's power when few units are randomized. Consequently, we paired teachers based on geography and average grade of the school's previous marketing classes (based on data from the Norwegian Directorate for Education and Training) prior to randomization. In each pair, an intervention teacher was drawn to teach with the simulation game in fall 2019, while the comparison teacher continued with regular student-active instruction. Teachers who taught in the same school had the same geography and average course grade of previous marketing classes, and two teachers from the same school thus constituted one pair (strata). Separating teachers from the same school into intervention and control allowed us to reduce school-confounding effects.

Most students in the classes signed an informed consent form for participation in the experiment. Students who refused consent received the same instruction as the other students but were exempted from data collection. Of the eight teachers who had some students decline to participate, four agreed and four disagreed that the non-participating students were similar to those participating.

A total of 433 students completed the test measuring enjoyment, motivation, perceived ability, and achievement. Class sizes ranged from 10 to 35, with an average of 20 students per class. Class registers from the beginning of the term indicated that about 15% of the students left their marketing class prior to the test. These changes seemed to be natural attrition, as marketing was an elective subject and school regulations allowed students to switch subjects a month into the semester. We did not find any traces of differential attrition across intervention and comparison groups.

3.2 | Intervention

While the intervention teachers were allowed to use the game in the fall of 2019, the comparison teachers were allowed to use the game in the spring of 2020. Thus, the experiment was a delayed-intervention experiment in which all participants received the intervention in the end, which avoids a common issue of high attrition in the comparison group. Furthermore, it allowed teachers at the same school in different intervention groups to collaborate and learn from each other, as the comparison teachers were also attending instruction sessions and were due to use the game later.

We invited the teachers to two game-instruction sessions and provided individual instruction to those unable to attend either session. Training offered ample opportunities to play the game, and the game developer was available to answer questions throughout the intervention period. Table 1 summarizes the main stages of the trial.

The intervention classes used the game in instruction for a total of seven sessions over two periods in fall 2019. The first period consisted of three sessions at the beginning of the school year. The second period of four sessions was administered in October or November. Each session was expected to last about 90 min, and all teachers received suggestions for how to organize each session. Within each session, teachers were expected to switch between active gameplay, topic introduction and class discussion.

Teachers used the simulation game within the allotted lecture time, substituting some regular activities. In collaboration with the teachers, we determined that the game should be used to teach the

TABLE 1 Sequence of events in the trial

Recruitment of schools/teachers	Spring 2019
First teacher training	June 2019
Second teacher training	August 2019
Survey with collection of background characteristics of students	August 2019
First intervention period	September 2019
Session 1: Introduction to the game	
Session 2: Play game, review results, make plan for further gaming	
Session 3: Game, reflection task, Kahoot!	
Second intervention period	October/November 2019
Session 4: Segmentation, theory and gameplay	
Session 5: SWOT, theory and gameplay	
Session 6: New SWOT, action plan and gameplay	
Session 7: Game play, reflection task	
Achievement test and survey about satisfaction (only intervention), motivation and perceived learning	December 2019

Note: SWOT = strengths, weaknesses, opportunities and threats.

topics segmentation and situational analysis. The comparison classes taught the same topics in parallel.

The comparison classes followed business as usual, including case-based instruction highlighting examples from the business community. The comparison and intervention classes used the same book as the basis for instruction and thus had access to the same case descriptions.

3.3 | Game environment

The intervention group used the online simulation game Hubro Marketing Simulation (HMS; https://hubro.education). HMS is a blended learning supplement to teacher instruction and substitutes parts of regular instruction. It intends to help students obtain an overall practical understanding of the marketing curriculum's basic concepts and how marketing interplays with the rest of the business. HMS is a widely used simulation game in marketing courses in upper-secondary schools and tertiary education in Norway. Its popularity serves as a relevant backdrop to our investigation of popularity against effectiveness.

HMS intends to help students obtain an overall practical understanding of the marketing curriculum's basic concepts and how marketing interplays with the rest of the business. The game addresses three of Griffin et al. (2012) list of 21st century skills, beginning with information literacy, as the students must manage various sources of information and interpret it according to set criteria. However, the information is invariably presented as factual, thus impeding the promotion of critical evaluation skills. Second, the students are organized in groups, fostering collaborative, social and leadership skills. Yet, the game could be played alone and does not strictly qualify as a collaboration-demanding game. Third, the decision-making process requires problem-solving with no obvious correct answers or trial-and-error that would have encouraged a brute-force approach. A core task is to prioritize actions with limited resources and partially incomplete information. Students analyse customer segments, identify preferences and habits, invest in technology and product development, organize promotions and distribution channels, set prices and define a marketing mix to maximize sales. The game gives indirect feedback through company performance, however, there is no direct corrective or explanatory feedback during game play.

HMS integrates a range of functional business areas such as marketing, finance, research and development and production. Students manage a virtual company within a dynamic competitive environment, and it thus resembles other business and marketing simulation games. Furthermore, HMS is consistent with the definition of simulation games offered by Lovelace et al. (2016, p. 101) as it provides a synthetic experiential learning environment that facilitates decision-making within a complex and dynamic setting. The game adds to the case-based instruction by letting students live out the cases.

Students work together in groups and compete with other student groups in the classroom. Figure 1 shows a central task during the intervention where students collaborate to determine a segment's preferences regarding a jet-pack product, which serves as a central product throughout gameplay. Within-group discussion and collaboration on decisions requires the students to apply theory, make decisions and review consequences. Their decisions have important consequences as the students seek to gain profits, considered as overall scores, within each financial quarter and compete against other businesses run by fellow students or simulated by the game. Making decisions within a multifirm industry, students are responsible not only for their own firm's strategic decisions, but additionally for anticipating and reacting to the competitive decisions of their peers (Brooks et al., 2006).

During classroom gameplay, the teacher is expected to explain gameplay rules, connect game content to curriculum objectives, encourage reflection and decide when to move from one financial quarter to the next. Thus, the game is an administrator-driven synchronized computer-controlled simulation with fixed scaling of cycles across groups (Crookall et al., 1986; Thavikulwat, 1996). The game elements of collaboration within a team, game rules, clear goal of selling products, interactivity, feedback as scores, competition between firms and an engaging narrative of jetpack products overlap heavily with Vandercruysse et al. (2012).

3.4 | Instruments

We developed instruments for this project because high-quality, marketing-relevant instruments of interest, perceived ability, and achievement were lacking. All instruments have five categories with symmetric labels around a midpoint. See Table A1 for an item overview. We describe below the psychometric properties of the model fit of a unidimensional model and measurement precision along the latent scale (see Table 2 for fit statistics). We also assess measurement invariance, or to what extent the measure works similarly in both the intervention and comparison groups. Configural invariance indicates that the same items are relevant in the measure of both groups, whereas scalar invariance indicates that the measure is generally functioning similarly in both groups.

The measure of satisfaction with the game consisted of nine agreement items presented to the intervention group only. Items covered cognitive and affective potential outcomes. The latent scale showed acceptable unidimensional fit with precision range at middle to low parts of the scale.

As compound measures for motivation, we constructed on the basis of the framework of Hidi and Renninger (2006), three measures of interest: course interest, content interest and job interest. [Corrections made on 30 June 2022, after first online publication: 'topic interest' in the previous sentence has been corrected to 'content interest' in this version.] The detailed validation process is documented in a separate study (Daus et al., 2021). The first of these three consisted of six agreement items on the students' positive perception of the marketing course. We adapted these items to the marketing context from the student questionnaires in Trends in International

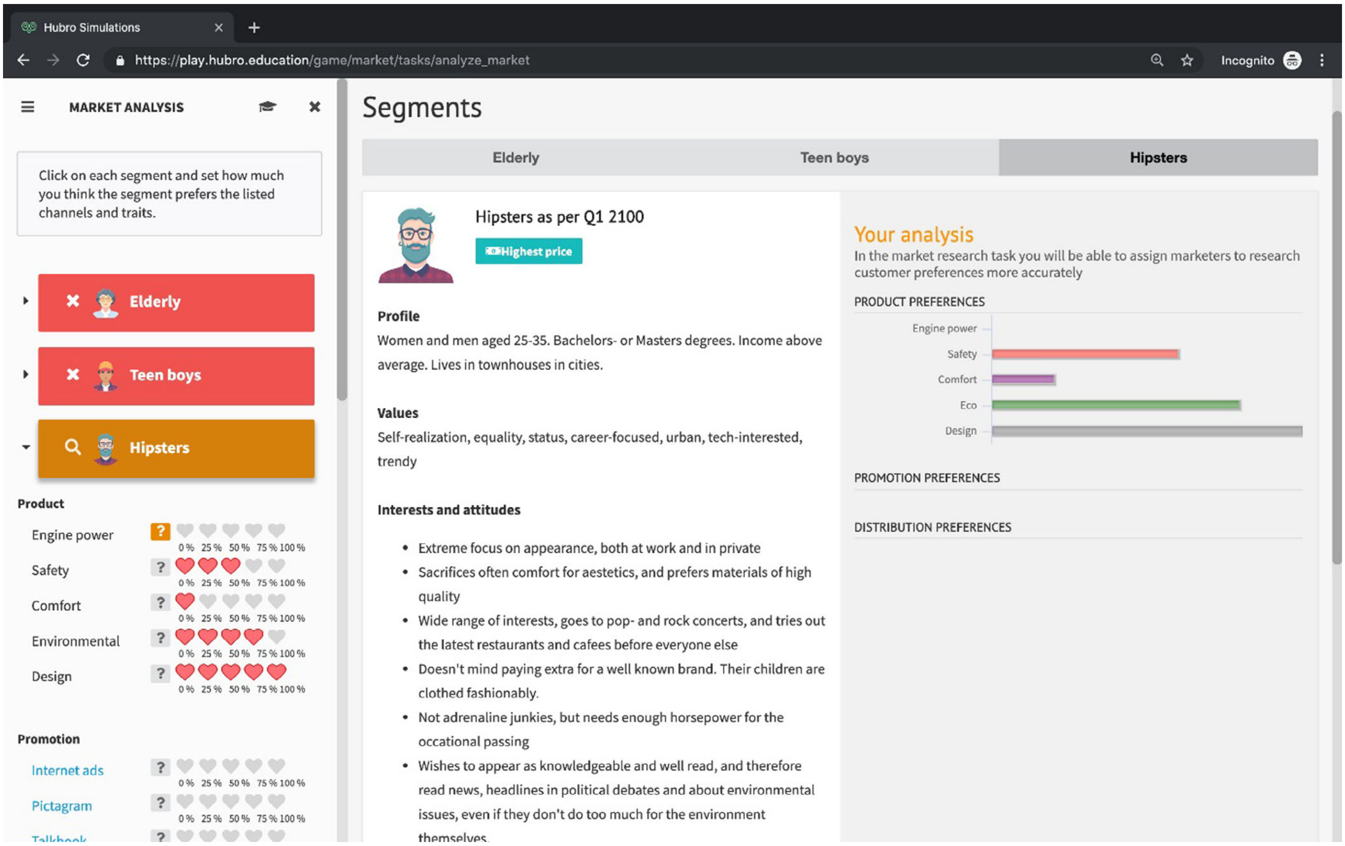

FIGURE 1 Excerpt from the game on a topic of focus during the intervention. Intervention gameplay was in Norwegian

Mathematics and Science Study (Hooper et al., 2017), and from Adams et al. (2000). The measure had an acceptable unidimensional fit with a broad precision range. The second measure consisted of four items on interest in learning four marketing topics, closely aligned with the course syllabus. The measure showed perfect unidimensional fit with high measurement precision, mostly for students at low levels. These first two measures worked similarly across both intervention and comparison groups. The third measure consisted of seven items addressing how interesting it would be to work on seven typical marketing tasks ranging from stereotypically easy and popular to more challenging and tedious tasks. This was inspired by Lemken and Siguaw (2019). This job interest measure showed mediocre unidimensional fit with a broad precision range. This measure was non-invariant across the two groups.

Two perceived ability measures each consisted of four items, with perfect unidimensional fit and a broad measurement precision range. The first measure on the degree of confidence in one's own performance on the assessment showed configural measurement invariance, meaning partial support that the measure functioned similarly in both groups. The second on the perception of one's own learning gains in the course functioned identically across both intervention and comparison groups.

We used two achievement outcome instruments in this study. An achievement test captured different topics and levels of knowledge, including recall of facts and basic concepts, application of knowledge to a new situation, evaluation and analysis of information, and application of course content in decision-making. A central lingering issue has been, and still is, the difficulty in establishing a valid measure along a single dimension that adequately covers the curriculum and game contents. Due to various critiques of the multiple-choice response format for assessing higher-level thinking skills, we included four constructed-response items, scored by two independent teachers. As a result, the final test consisted of 55 multiple-choice items and four short-response items arising from larger case-study-inspired tasks. Due to poor item fit, the measure was post-hoc reduced to a measure of seven items having perfect unidimensional fit and a broad measurement precision range, yet without equal functioning across the two groups (non-invariance). We also investigated the students' mid-term grades as set by their teacher 4–6 weeks after the intervention ended. All measures except the achievement measures correlated positively with medium-to-high strength, suggesting evidence of validity regarding their expected internal relations. For more information about the validation, see Daus, Skjelbred & Pedersen (2021).

3.5 | Statistical analyses

We evaluated the effect sizes of the compound measures using multigroup confirmatory factor analysis for each measure individually

TABLE 2 Summary of measurement validation and reliability analyses

Measure	$X^2(df)$, p	RMSEA	CFI	SRMR	Unidimensional fit judgement	Group invariance	Precision range	α	Sum-score
Experience with Digital Tools	1.51 (2), $p = 0.470$	0 [0, 0.084]	1.000	0.011	Perfect	Scalar	Very low, spike	0.18	Rejected
Satisfaction with The Game	40.84 (27), $p = 0.004$	0.047 [0.009, 0.075]	0.997	0.004	Acceptable	NA	Middle-low	0.94	Rejected
Course Interest	33.5 (9), $p < 0.001$	0.08 [0.05, 0.11]	0.993	0.002	Acceptable	Scalar	Broad, middle	0.68	Rejected
Content Interest	1.38 (2), $p = 0.501$	0 [0, 0.087]	1.000	0.004	Perfect	Scalar	Very low, spike	0.79	Rejected
Job Interest	39.81 (14), $p < 0.001$	0.066 [0.042, 0.091]	0.985	0.004	Acceptable	Non-invariant	Broad	0.87	Rejected
Perceived Test Performance	0.51 (2), $p = 0.776$	0 [0, 0.063]	1.000	0.004	Perfect	Configural	Broad	0.09	Rejected
Perceived Learning Gains in Course	2.98 (2), $p = 0.232$	0.033 [0, 0.108]	0.999	0.011	Perfect	Scalar	Broad	0.86	Rejected
Achievement post-survey	15.0 (14), $p = 0.378$	0.013 [0, 0.049]	0.978	0.054	Perfect	Non-invariant	Broad, middle-low	0.52	Rejected

Note: RMSEA, root mean square error of approximation. CFI, Comparative Fit Index. SRMR, standardized root mean squared residual. α = Cronbach's alpha. [Corrections made on 30 June 2022, after first online publication: CFI values for 'Experience with Digital Tools' and 'Content Interest' have been corrected from '10.000' to '1.000' in this version.]

and with multiple measures simultaneously. As the latter showed the same findings and suffered from convergence issues, we report only the former. We obtained parameter estimates through robust weighted least-squares using a diagonal weight matrix with standard errors and mean- and variance-adjusted chi-square test statistics that use a full-weight matrix (WLSMV) in Mplus 8.4 (Muthén & Muthén, 1998-2020). The benefit of WLSMV is that it makes no assumption about the distribution of the observed categorical variables. Thus, our binary and ordinal indicators, with possibly heavy skew, were treated accordingly. WLSMV does not handle missingness-at-random adequately, unlike full-information maximum likelihood estimation. However, the latter made no difference to conclusions about fit or effects, except for occasionally poorer fit indices. Hence, we only report WLSMV-estimates.

In simple analyses of individual items, we ignored any measurement error and invariance by using bivariate ordinal regression with the indicator as the dependent variable and the group as an independent variable.

In both the multigroup confirmatory factor analyses and the simple bivariate ordinal regression analyses, we accounted for the clustering of students within schools using a sandwich estimator. This approach avoids erroneous inferences in cases where students within a school have correlated characteristics, thereby violating the sampling independence assumption (Snijders & Bosker, 2011). Negatively phrased items were reversed in factor analyses. Robustness analyses that dropped responses from students with extremely long or short survey completion times made no difference to the results.

3.6 | Covariate measures and validation of randomization

Covariates were included to confirm that randomization was successful and to serve as possible control variables in regressions. Table 3 presents the results of Pearson's chi-squared tests which shows that the two groups have similar proportions of male and female students. Table 3 also presents the results from an ordinal regression indicating that the grade-point average for the previous school year was similar between the groups. The analysis of covariates showed that the groups equalled in composition based on gender, grade-point average from the previous school year and parents' education. Furthermore, they also equalled on a four-item measure of experience with digital tools (see Table A1 for items; $b = 0.21$, 95% CI = [−0.0.11, 0.52]).

3.7 | Teacher experience data

We supplemented student survey data with a teacher survey 3 months after the intervention had ended (9 of 11 intervention and 10 of 11 comparison teachers replied) and with semi-structured interviews with five intervention teachers. The gathered data served to confirm that the intervention was implemented as planned and to collect in-depth teacher experiences with the intervention and their evaluation of the experience.

TABLE 3 Characteristics of groups at baseline. Categorical covariates

Variable	Category	Comparison n	Comparison %	Intervention n	Intervention %	Difference p
Gender						
Female		91	46.4	117	49.4	0.32
Male		102	42.0	115	48.5	
Other		3	1.5	5	2.0	
Grade point average previous year						0.71
(Low) 1.0–3.4		14	7.1	18	7.6	
3.5–3.9		32	16.3	38	16.0	
4.0–4.4		66	33.7	78	32.9	
4.5–4.9		55	28.1	79	33.3	
(High) 5.0–6.0		29	14.8	24	10.1	
Educational level of parents[a]						0.44
0 Not relevant/missing		571	58.3	695	59.3	
1 Lower secondary		30	3.1	34	2.9	
2 Upper secondary		65	6.6	81	6.9	
3 Polytechnical		63	6.4	87	7.4	
4 Higher ed., ≤ 4 years		139	14.2	154	13.1	
5 Higher ed., > 4 years		112	11.4	121	10.3	
N		196		237		

[a]The number of occurrences exceeds the sample size because a student typically has more than one parent/guardian, up to a maximum of four such persons.

4 | FINDINGS

We present first a discussion of lessons learned from implementing the game in the intervention group based on survey questions and teacher interviews, followed by measurements of the satisfaction among students and teachers in the intervention group. Then we present an evaluation of the effects on satisfaction, interest, perceived ability, and achievement. Throughout these sections, we consecutively number the main findings.

4.1 | Game implementation and experiences

We used the teachers' and students' written responses to open survey items and the teacher interviews to explore the gaming elements of collaboration and competition. First, the intervention was designed to encourage collaboration within the groups of two or three students, as company decisions were meant to be made after deliberation with some form of consensus. Indeed, most students thought that the group had a good discussion around the decisions (66% responded always/often, 23% occasionally, and 7% never/seldom) and that group work functioned well (80% always/often, 11% occasionally, 4% never/seldom). These findings indicate that the students generally were satisfied with their collaboration.

Second, the intervention and game design also encouraged competition, as the pairs and groups competed against each other and other computer-simulated companies in the same market. Unlike collaboration, competition seemed to motivate some personalities and stress others. Teachers reported that the competition element seemed to motivate some students, particularly boys, but failed to fully engage weaker-performing girls.

Finding 1: The motivating effect of competition in games differs across genders.

The teachers expressed concerns that some students, often the more competitive groups with less-disciplined, all-male students, were too concerned with moving ahead. These groups focused on "gaming the game" to win instead of reflecting on strategy. Perhaps unsurprisingly, these students objected to the teacher cutting off the gameplay to continue a broader discussion of the course's main principles. Teachers thus expressed a need for the game design to induce reflection.

Finding 2: Competition might come at the cost of reflection.

The intervention design involved game use early in the course and teachers reported in the interviews mixed experiences with using the game before covering most of the curriculum. Some teachers experienced that it was difficult to return to more traditional teaching of terminology, and some felt that they had too soon exploited their best "weapon" to increase student motivation and engagement. While other teachers found it attractive to allow the students to play a game while being introduced to new terminology and theory, they expressed that the game was somewhat too complex for such use as it covers large parts of the curriculum.

Finding 3: The timing and duration of gameplay require further investigation.

Nine intervention teachers answered the five survey questions regarding their participation in training before and during the intervention, as well as their collaboration with other teachers. All but one reported at least above average participation in the main training (variables 2 and 3 in Table A2). Additionally, two-thirds had either received individual training/support from the company or collaborated with another teacher using the same game. However, three interviewees revealed that they struggled to grasp how they most effectively could implement it within the curriculum despite finding the gameplay itself to be relatively straightforward. On the positive side, some teachers emphasized the creation of useful collaborative networks between teachers across the schools and explained that such networks are particularly important in smaller courses such as marketing, where there are typically few teachers in each school.

Finding 4: Teacher training lacked a strong connection between the game and the curriculum.

4.2 | Satisfaction with the game

User feedback is typically related to game aspects such as user friendliness, progression, collaboration and suitability between gameplay and regular instruction. More than two-thirds of the intervention students reported that they somewhat or strongly agreed to statements that they were having fun when using the game, would recommend the game to other students and classes, and learned something from the game (Table 4). Moreover, more than half agreed that the game caught their interest from the beginning, that the tasks were good challenges, that the game was easy to use and understand, and that they would consider playing the game in coming weeks.

Although most enjoyed the game, they also recognized that it was not a stand-alone tool. More than half agreed that they learned things from the teacher or textbook that they would not have learned if they only played the game, suggesting the game complements regular instruction rather than replaces it. This complementarity between the game and teacher instruction was embedded in both the game and the intervention design.

The feedback from the students and teachers aligned. Teachers felt that the game increased the students' interest in and understanding of marketing. Additionally, they expressed that the game was a motivational factor for some students, particularly boys. Both teachers and students believed that the game increased students' understanding of the practical application of marketing. Furthermore, teachers thought that the game better communicated the complexity of marketing and the relationship between the course topics than what traditional instruction offers. Even though their responses were generally positive, teachers felt they could improve their effective use of the game.

Finding 5: Students and teachers are satisfied with the game.

4.3 | Effect evaluation

We based the previous finding that students like games on the intervention group's subjective opinions, without comparison to non-users. Below, we compare the students who used the game with similar students who did not use the game, allowing us to identify the causal effect of the game on motivation, perceived ability, and achievement.

4.3.1 | Interest

Comparing the motivation for marketing of non-users to users requires measures that are not related to the gameplay itself, that is, treatment-independent measures. We used measures of three aspects of interest, an aspect of motivation that we assumed would be

TABLE 4 Response distributions in percent on items included in the instrument for satisfaction with the game

ID	How do you think the simulation game Hubro Marketing Simulation worked?	1	2	3	4	5	NA
b29	I had fun when we played the game.	7	6	13	24	48	2
b30	I received the necessary training in using the game so that I could focus on the marketing principles.	9	11	30	24	23	3
b31	The game caught my interest from the beginning.	9	10	23	18	37	3
b32	I would recommend this kind of game to other students and classes.	8	5	18	22	44	3
b33	The tasks in the game were neither too simple nor too hard; they were good challenges.	5	7	26	30	30	3
b34	I am confident that I learned something through the game.	6	5	15	35	35	3
b35	The game was easy to use and understand.	6	12	24	28	26	4
b36	I will consider continuing the game in the next weeks if I get the opportunity to play it.	13	8	20	20	35	4
b37	I learned things from the teacher or textbook that I would not have learned if I only played the game.	6	10	25	31	24	4

Note: Scale (1–5): Strongly disagree—strongly agree. NA, missing.

strongly linked to enjoyment of marketing, namely, course interest, content interest and job interest. Table 5 shows that we did not find significant differences between the intervention group and the comparison group on these measures. This suggests that even though students enjoy the game, the game does not lead to measurable increases in interest in marketing.

Finding 6: The simulation game does not produce measurable increases in interest in marketing.

4.3.2 | Perceived ability

The two perceived ability measures were confidence in assessment performance and perception of learning gains in the course. Overall, we found no significant differences between the intervention group and the comparison group on these two perceived ability measures (see Table 4). The exception was that students who played the game seemed to be more confident in one of the two topics covered in the intervention period. This item is not part of a validity-probed compound instrument. We will delve into the discussion of single-item versus multi-item (compound) analysis later.

Finding 7: The simulation game does not produce measurable increases in confidence in assessment performance or perception of learning, but students using the game expressed greater confidence in the topic covered during the intervention.

4.3.3 | Achievement

We measured student achievement through a selection of multiple-choice and open-response items testing knowledge and reasoning. We did not find an overall difference between the intervention group and the comparison group on the achievement test, irrespective of investigating all 59 items, a selection of well-fitting items, a selection of items deemed to have the closest alignment between the game and the curriculum, or a selection of short-response items representing the most reasoning-demanding items. We also did not find any differences in these sets when including background variables as covariates to increase statistical power.

In the absence of these differences, we investigated a potential effect on students' mid-term grades. Teachers submitted mid-term grades for 331 students, based on the teacher's evaluation of student competence in the autumn semester. We used ordinal and linear regression with grades as a dependent observed variable and the intervention group as an independent variable. However, we found no significant difference, irrespective of the regression method.

Finding 8: The game does not improve overall achievement.

4.3.4 | Single-item versus multi-item compound measures analysis

The common approach of treating observed responses as is, without accounting for measurement error, can lead to oversimplified conclusions. Despite the lack of clear evidence on the overall interest, perceived ability, and achievement measures, we found positive traces of effect on single items in favour of the intervention group (Table 6). This finding illustrates the importance of compound instruments, as our conclusions about the game would have differed if we had used only a subset of single items to study the game's effectiveness.

One out of 17 items addressing interest was significant in favour of those who used the simulation game. This item measured prospective interest in analysing website traffic with quantitative data. In addition, three items related to confidence in their ability on the post-test were significant in favour of the intervention group. Compared with the comparison group, students in the intervention group agreed more to the item “I was motivated to do my very best on these questions” and disagreed more to the items “I was unfocused when working on these tasks” and “I could have exerted more effort on these tasks.” Despite the word *motivated* in the first item, we argue for its

TABLE 5 Standardized effects on each of the outcome measures (individual analyses)

				95% CI		
Dependent latent variable	**MI**	***b***	***SE***	**LL**	**UL**	***N***
Course Interest	Non-invariant	0.08	0.23	−0.38	0.53	427
Content Interest	Scalar	0.00	0.18	−0.36	0.35	416
Job Interest	Configural	0.08	0.14	−0.20	0.37	423
Perceived Test Performance	Non-invariant	0.12	0.19	−0.25	0.48	429
Perceived Learning Gains in Course	Scalar	0.15	0.20	−0.23	0.54	421
Achievement: Post-survey (long)[a]	Scalar	0.26	0.29	−0.31	1.55	433
Achievement: Post-survey (short)[b]	Scalar	1.80	1.45	−1.05	4.64	433
Achievement: Mid-term grade	Not available	0.01	0.28	−0.56	0.54	331

Note: MI = Measurement invariance between control and intervention groups (non-invariance < configural < scalar). *b* = estimate. *N* = effective sample size. *CI* = confidence interval; *LL* = lower limit; *UL* = upper limit.

[a]Long version contains 26 mediocre and acceptable-fitting items.

[b]Short version contains only seven acceptable-fitting items.

TABLE 6 Individual items with a significant effect in favour of the intervention group (positive estimate) or the comparison group (negative estimate)

Motivation and perceived learning items	Est	Sig
How do you feel it went on the test items you just took?		
I was unfocused when I worked on these tasks.[a]	0.28	*
I was motivated to do my very best on these tasks.[a]	0.36	**
I could have exerted a greater effort when I worked on these tasks.	0.39	**
So far in Marketing and Leadership 1 I have gotten a very good understanding for:		
Segments, segmentation and target groups	0.44	**
If in the future you are responsible for the marketing of a larger organization, how interesting would it be for you to work with the following?		
Analysing with quantitative data who visits the website, what they do there and what characterizes them[b]	0.22	*
Achievement test items	**Est**	**Sig**
What does the claim mean? The phone producer Huawei has the largest market share globally in the category mobile phones.	0.37	*
Segmentation criteria for consumer market. Classify each piece of information below as either demographic segmentation, psychographic segmentation or behaviour segmentation – Extrovert.	0.40	*
Segmentation criteria for consumer market. Classify each piece of information below as either demographic segmentation, psychographic segmentation or behaviour segmentation – Will have a religious Confirmation next weekend.	0.24	*
Which of the following claims are correct about the adoption of products in the population? – The early majority and the late majority constitute together 34% of the market.	0.22	**
Which business has the highest sales volume today? [interpretation of graphs case task]	0.28	*
We move time ahead one period (imagine how the points move). If none of the businesses introduce new products, which company do you think will have the highest sales volume? [interpretation of graphs case task]	0.28	*
What does "environmentally friendly trips" constitute in the vacation market? [segmentation case task]	−0.66	***
What do the four person descriptions represent? [segmentation case task]	−0.72	***
Are the following categories of segmentation criteria used in the four person descriptions? – Segmentation by buyer's criteria. [segmentation case task]	−0.31	**
Are the following categories of segmentation criteria used in the four person descriptions? – Segmentation by diffusion. [segmentation case task][c]	0.33	*
Whom of these four do you believe… – …represents the largest segment in Bergen centre on any summer day? [segmentation case task]	0.55	**
Whom of these four do you believe… – …will react most positively to an advertisement campaign for MiniHolidays in the newspaper Bergens Tidene with the Bergen singers Jan Eggum, Rune Larsen, Kurt Nielsen and Sissel Kyrkjebø? [segmentation case task]	0.32	*

Note: a = Item is part of the compound measure "Perceived learning – assessment confidence". b = Item is part of the compound measure "Motivation – work tasks". c = Item is part of the compound measure "Achievement test."
*indicates $p < 0.05$. **indicates $p < 0.01$. ***indicates $p < 0.001$.

placement among perceived ability based on the previously discussed measurement fit during the validation effort. Thus, despite the lack of clear evidence of effect on the overall perceived ability measures, we found significant and consistently positive traces of effect on single items in favour of the intervention group.

For the achievement test, we can also temporarily ignore measurement error and bias to investigate single test items. The intervention group significantly outperformed the comparison group on 9 out of 55 multiple-choice items, whereas the comparison group significantly outperformed the intervention group on three multiple-choice items. The nine items were on various topics and cognitive levels, whereas the three items were exclusively related to a single task assessing students' ability to distinguish the terms *segments* and *target groups* and to apply this understanding in practice. In conclusion, we found, at best, a weak indication that the game might have resulted in higher achievement on certain items.

Finding 9: Single-item investigations offer insights into specific topics that may be influenced by the game, but careful interpretations are needed to avoid overly positive conclusions.

5 | DISCUSSION

In this paper, we evaluated the effectiveness of a digital simulation game in a marketing course. We used multiple outcome measures and qualitative data on user satisfaction, interest in marketing, perceived ability, and achievement with a randomized, controlled trial where only the game differed between intervention and comparison groups. By using treatment-independent measures, in line with best practices (Slavin & Madden, 2011), we ensured that intervention and control groups were compared on fair premises.

We found that students and teachers were positive towards and satisfied with the game. When scrutinizing single items, we also found some sign of higher perceived ability. Yet, when using compound instruments, we found no evidence that the simulation game leads to measurable increases in interest in marketing, confidence in assessment performance, perception of learning gains, achievement tests, or mid-term grades. These findings lead to the main conclusion of this paper: student and teacher satisfaction with gameplay is insufficient evidence that games by themselves improve educational outcomes beyond that of non-game-based forms of student-active learning. Consequently, this study supports previous literature arguing that opinions about learning and motivation are unreliable and often in conflict with direct measures when both are gathered (Clark, 2007).

In comparison with the meta-analyses by Wouters et al. (2013), our conclusion challenges their finding of the positive effects of group work, but supports their finding that effects are less likely with randomized trials. Our conclusion also supports Hattie's finding that one cannot expect more positive findings in recent trials (Hattie, 2009, p. 220). Our study is also in line with the literature summarized in Anderson and Lawton (2009), who argued that games tend to be inefficient for teaching terminology and theory in business simulations. Thus, increasing technological sophistication, or perhaps even research design, is no guarantee of improved outcomes. However, the development of research methods and assessment methodology might offer *stronger* evidence regarding educational games.

Our identified effects function as a sort of lower bound. To ensure complete comparability across intervention and control groups, we recruited only teachers without experience with the game. This was necessary, as it would be hard to recruit participants to a research project that randomly prevents someone from using an already acquired tool. Yet, game instruction is likely to improve with teacher experience. In fact, studies have found that the successful implementation of digital learning tools for effective learning requires creative and pedagogically oriented use of such tools in teacher training beyond basic digital skills (Røkenes & Krumsvik, 2016; Tømte et al., 2009). So, one could theoretically expect to detect greater effects if the intervention teachers get more practice.

The when and how of game use remains an empirical question. Games can be used for educational purposes in almost all academic levels and all subjects. Generalizing findings regarding games' beneficial and detrimental effects on outcomes depends on contextual factors, such as school subject, curriculum objectives, duration of game play, game pedagogy (e.g., collaboration, competition), implementation, teachers' experiences with the game and teachers' investment in game instruction and application. Despite the difficulties in generalizing findings, we offer three key insights from the student and teacher experiences, which we think are widely applicable to games in general.

First, competition may be a motivational factor with unintended consequences. We found that competition might motivate differentially, which is important to consider for both game designers and teachers using the game. Specifically, designers and teachers must ensure that girls are not alienated by the competition and, in our case, public display of performance. These findings are in line with extensive literature showing that boys are likely to thrive with competition more than girls (Gneezy & Rustichini, 2004; Morin, 2015; Niederle & Vesterlund, 2010, 2011; Tobin & Garnett, 2003). Moreover, competition sometimes came at the cost of reflection. Competition and a collaborative spirit can be powerful factors for creating a flow that engages the students in the game. Unfortunately, in the investigated game, the balance between flow for engagement and tranquillity for reflection tilted towards flow.

Second, a planning consideration when using games in education is timing the use of the game, both within the lesson and within the overall course. In our study, teachers used the simulation game in two periods: a series of three lectures in the beginning of the first semester and a series of four lectures in the middle of the first semester. The intention was for students to play the game while being introduced to new terminology and theory; however, the game is complex and refers to large parts of the curriculum, making full alignment unrealistic. Moreover, many games offer limited variation and may feel repetitive if used for an extended period. Using games early in the course, as in this experiment, might come at the cost of having to revert to other instructional modes, which students and teachers may find challenging and perhaps disappointing. If introduced later, the beneficial effects might be lost for students who have already disconnected from the course. Hence, the balance between gameplay and non-game activities requires further investigation and can be fruitful across game specifics. Irrespective of early or late introduction, the total time spent is likely also a factor worth investigating, as the effects are likely to increase with exposure, and basic game training takes at least some time from regular instruction time.

Third, training time leads us to the third lesson, teacher training. Regardless of how well-designed and intuitive a game may be, it will always require the teacher's investment to incorporate a new tool into education. In fact, our teachers requested more competence building on effective pedagogical game use. Pedagogical-didactical decisions are critical, both for the development of the game and for the incorporation of the game into the course (Clark, 2007). This presents a practical challenge, as teachers may fail to see the advantage of such tools beyond increasing student motivation (Wikan & Molster, 2011). Moreover, teachers have reported insufficient technological-pedagogical training, not only in this study but generally (Brooks et al., 2019; Gudmundsdottir & Hatlevik, 2018).

In line with previous studies, we recommend that game designers and educators incorporate time and tasks for reflection using scaffolding strategies within and between game sessions. Likewise, we recommend that more effort be devoted to identifying optimal timing of gameplay and that teacher training and teacher materials include educational game advice to offer teachers a more complete package and increase their technological-pedagogical content knowledge.

5.1 | Limitations

One possible explanation for the lack of significant effects on the otherwise sensitive compound measures is low statistical power. Although we had many student participants compared with several other experimental studies, the statistical power was strongly reduced by the need to randomize at the classroom level. To identify a significant effect at a 5% level with 80-percent power, a sample of 22 classrooms with about 25 students each would give a minimal detectable effect size of Cohen's $d = 0.4$ under ideal circumstances, meaning low intra-cluster correlation and high explanatory power of background variables. As such, it is conceivable that the simulation is effective, but with a lower effect size than we could identify. However, for any intervention, researchers must consider what effect size they desire to identify, given the cost of the intervention. Furthermore, our sample was not particularly small, but would rather constitute a medium-sized sample compared with similar experimental studies of computer-assisted learning (Escueta et al., 2020). Lastly, the recruitment of teachers without experience with the game ensured a clean comparison of the game itself, but that may come at the cost of clarity. With experience, teachers may more effectively use the tool. Taken together, these design decisions hedge on the more conservative side.

We measured the impact on motivation, perceptions and achievement, which are likely reasons for using educational games. However, we inadequately covered all aspects of the constructs that we aimed to measure. For instance, we only included one aspect of motivation, namely interest. Wouters et al. (2013) raise the important question whether different measures related to motivation represent different constructs or not. Moreover, despite our efforts at including achievement test items that demand applying and evaluating information, further test development should cover broader and higher cognitive demands. Moreover, the achievement test, curriculum and game insufficiently capture 21st-century skills. Hence, learning objectives are not merely broadening from knowing to reasoning with information, but also from managing information to finding and evaluating conflicting pieces of information. Several games, instructional strategies and curricular objectives aim to cover 21st-century skills like critical thinking, creativity, problem-solving, collaboration, effective communication, motivation, persistence or learning to learn. How to adequately assess such skills remains challenging. Hence, further studies should challenge game development, assessment methodology and learning sciences to adequately evaluate the effectiveness of learning tools having such learning objectives.

5.2 | Concluding remarks

In this paper we gave a digital learning game harder competition than is common in many studies and investigated the effect of game instruction beyond student-active learning. The paper showcases the methodological rigour needed to provide information on the success of games in education in terms of the choice of adequate measures, experimental design, mixed data collection and pedagogical perspectives.

Despite high satisfaction with the game, we found no evidence that such satisfaction indicates game effectiveness. Hence, just like other materials, an educational game in isolation is no panacea for combating student disengagement. Our results indicate a strong need for further research into how to ensure that popularity of educational games is translated into measurable improvements. In this regard, our recommendations about scaffolding, timing and teacher competence represent concrete next steps. Before evaluating effectiveness, researchers and practitioners must be more explicit about the intention of introducing games and pay attention to the possibility of unintended consequences of game elements such as collaboration and competition. We argue for the increased awareness about potential heterogeneous effects, including how games may alienate some students.

In recent decades, games have undergone enormous technological development and are constantly improving. As such, the quality of educational games is high, and we believe that the key to unleashing the benefits lies in fine-tuning game aspects and implementation. We end by proposing that satisfaction in the classroom regarding a game is insufficient as an indicator of effectiveness, despite the market of educational institutions seeming satisfied with the game.

ACKNOWLEDGEMENTS

The authors thank the involved teachers and students for their patience and meticulousness.

FUNDING INFORMATION

The research is funded by the Oslo Regional Research Fund, Norway.

CONFLICT OF INTEREST

The implementation of the randomized controlled trial was conducted in cooperation with the learning tool provider Hubro Marketing Simulation. The development of the outcome measures and the analyses were conducted independently from the learning tool provider. Research funding was not dependent on the findings. The authors have no financial or other relationship with the learning tool provider.

DATA AVAILABILITY STATEMENT

Strictly anonymized quantitative data and relevant syntax are available at Open Science Foundation (https://osf.io/2ajxz/?view_only=aa38216a0558499183a2ede4ec0edba0).

ETHICS STATEMENT

Participation was voluntary, without any repercussions if non-participation. Appropriate permissions and ethical approval for the participation requested and approved by the Norwegian Centre for Research Data.

ORCID

Siv-Elisabeth Skjelbred https://orcid.org/0000-0002-0459-7022
Stephan Daus https://orcid.org/0000-0003-0230-6997

ENDNOTES

[1] Student-active instruction is characterized by students being "actively or experientially involved in the learning process and where there are different levels of active learning, depending on student involvement"

(Bonwell & Eison, 1991). This implies instruction beyond passively listening and covers a range of specific approaches, from collaborative groupwork to reflection-inducing summary writing of each lesson.

[2] For the purpose of this article, we do not separate between simulations and games due to the blurred boundaries. But apply the broad term simulation games in line with Sitzmann (2011).

REFERENCES

Adams, E., Womble, M. N., & Jones, J. H. (2000). Marketing education students' perceptions toward marketing education courses. *Journal of Career and Technical Education*, *17*(1), 46–63. https://doi.org/10.21061/jcte.v17i1.591

All, A., Castellar, E. P. N., & van Looy, J. (2014). Measuring effectiveness in digital game-based learning: A methodological review. *International Journal of Serious Games*, *1*(2). https://doi.org/10.17083/ijsg.v1i2.18

Anderson, P. H., & Lawton, L. (2009). Business simulations and cognitive learning: Developments, desires, and future directions. *Simulation & Gaming*, *40*(2), 193–216. https://doi.org/10.1177/1046878108321624

Bacon, D. R. (2016). The editor's corner. *Journal of Marketing Education*, *38*(3), 143–144. https://doi.org/10.1177/0273475316677657

Barzilai, S., & Blau, I. (2014). Scaffolding game-based learning: Impact on learning achievements, perceived learning, and game experiences. *Computers & Education*, *70*, 65–79. https://doi.org/10.1016/j.compedu.2013.08.003

Bawa, P. (2020). Game on!: Investigating digital game-based versus gamified learning in higher education. *International Journal of Game-Based Learning*, *10*(3), 16–46. https://doi.org/10.4018/IJGBL.2020070102

Bellotti, F., Kapralos, B., Lee, K., Moreno-Ger, P., & Berta, R. (2013). Assessment in and of serious games: An overview. *Advances in Human-Computer Interaction*, *2013*, 1–11. https://doi.org/10.1155/2013/136864

Boevé, A. J., Meijer, R. R., Beldhuis, H. J. A., Bosker, R. J., & Albers, C. J. (2019). On natural variation in grades in higher education, and its implications for assessing effectiveness of educational innovations. *Educational Measurement: Issues and Practice*, *38*(4), 55–66. https://doi.org/10.1111/emip.12283

Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review*, *15*(1), 1–40. https://doi.org/10.1023/A:1021302408382

Bonwell, C. C., & Eison, J. A. (1991). Active learning: Creating excitement in the classroom. https://files.eric.ed.gov/fulltext/ED336049.pdf

Brom, C., Šisler, V., Slussareff, M., Selmbacherová, T., & Hlávka, Z. (2016). You like it, you learn it: Affectivity and learning in competitive social role play gaming. *International Journal of Computer-Supported Collaborative Learning*, *11*(3), 313–348. https://doi.org/10.1007/s11412-016-9237-3

Brooks, B. W., Burson, T. E., & Rudd, D. V. (2006). Addressing current research gaps and directions in educational marketing simulations. *Journal for Advancement of Marketing Education*, *9*, 43–49.

Brooks, E., Gissurardottir, S., Jonsson, B. T., Kjartansdottir, S., Munkvold, R. I., Nordseth, H., & Sigurdardottir, H. I. (2019). What prevents teachers from using games and gamification tools in Nordic schools? In A. L. Brooks, E. Brooks, & C. Sylla (Eds.), *Interactivity, game creation, design, learning, and innovation* (Vol. 265, pp. 472–484). Springer.

Buil, I., Catalán, S., & Martinez, E. (2017). Exploring students' flow experiences in business simulation games. *Journal of Computer Assisted Learning*, *34*(2), 183–192. https://doi.org/10.1111/jcal.12237

Chee, Y. S. (2015). *Games-to-teach or games-to-learn: Unlocking the power of digital game-based learning through performance*. Springer.

Clark, R. E. (2007). Learning from serious games? Arguments, evidence, and research suggestions. *Educational Technology*, *47*(3), 56–59 https://www.jstor.org/stable/44429512

Clayson, D. E. (2009). Student evaluations of teaching: Are they related to what students learn? A meta-analysis and review of the literature. *Journal of Marketing Education*, *31*(1), 16–30. https://doi.org/10.1177/0273475308324086

Coleman, T. E., & Money, A. G. (2020). Student-centred digital game-based learning: A conceptual framework and survey of the state of the art. *Higher Education*, *79*(3), 415–457. https://doi.org/10.1007/s10734-019-00417-0

Crookall, D., Martin, A., Saunders, D., & Coote, A. (1986). Human and computer involvement in simulation. *Simulation & Gaming*, *17*(3), 345–375. https://doi.org/10.1177/0037550086173005

Daus, S., Skjelbred, S.-E., & Pedersen, C. (2021). *Development and validation of a set of outcome measures in marketing education*.

Delwiche, A. (2006). Massively multiplayer online games (MMOs) in the new media classroom. *Educational Technology and Society*, *9*(3), 160–172 https://www.learntechlib.org/p/75354/

Dewey, J. (1938). *Experience and education*. Kappa Delta Pi.

Escueta, M., Nickow, A. J., Oreopoulos, P., & Quan, V. (2020). Upgrading education with technology: Insights from experimental research. *Journal of Economic Literature*, *58*(4), 897–996. https://doi.org/10.1257/jel.20191507

Faria, A. J. (2001). The changing nature of business simulation/gaming research: A brief history. *Simulation & Gaming*, *32*(1), 97–110. https://doi.org/10.1177/104687810103200108

Gee, J. P. (2011). Reflections on empirical evidence on games and learning. In J. D. Fletcher & S. Tobias (Eds.), *Computer games and instruction* (pp. 223–232). Information Age Publisher.

Gneezy, U., & Rustichini, A. (2004). Gender and competition at a young age. *American Economic Review*, *94*(2), 377–381. https://doi.org/10.1257/0002828041301821

Griffin, P. E., McGaw, B., & Care, E. (2012). *Assessment and teaching of 21st century skills*. Springer.

Gudmundsdottir, G. B., & Hatlevik, O. E. (2018). Newly qualified teachers' professional digital competence: Implications for teacher education. *European Journal of Teacher Education*, *41*(2), 214–231. https://doi.org/10.1080/02619768.2017.1416085

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

Hanghøj, T. (2011). Clashing and emerging genres: The interplay of knowledge forms in educational gaming. *Designs for Learning*, *4*(1), 22–33.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. *Educational Psychologist*, *41*(2), 111–127. https://doi.org/10.1207/s15326985ep4102_4

Hooper, M., Mullis, I. V. S., Martin, M. O., & Fishbein, B. (2017). *TIMSS 2019 context questionnaire framework*. TIMSS and PIRLS International Study Center https://timss2019.org/wp-content/uploads/frameworks/T19-Assessment-Frameworks-Chapter-3.pdf

Kolić-Vehovec, S., Smojver-Ažić, S., Dorčić, T. M., & Zubković, B. R. (2019). Evaluation of serious game for changing students' behaviour in bullying situation. *Journal of Computer Assisted Learning*, *36*(3), 323–334. https://doi.org/10.1111/jcal.12402

Lemken, R., & Siguaw, J. A. (2019). The use of interest-inventory measurements in marketing education: Improving MBA student team effectiveness. *Journal of Marketing Education*, *43*(1), 75–90. https://doi.org/10.1177/0273475319872235

Lovelace, K. J., Eggers, F., & Dyck, L. R. (2016). I do and I understand: Assessing the utility of web-based management simulations to develop critical thinking skills. *Academy of Management Learning & Education*, *15*, 100–121.

Martone, A., & Sireci, S. G. (2009). Evaluating alignment between curriculum, assessment, and instruction. *Review of Educational Research*, *79*(4), 1332–1361. https://www.jstor.org/stable/40469099

Morin, L.-P. (2015). Do men and women respond differently to competition? Evidence from a major education reform. *Journal of Labor Economics*, *33*(2), 443–491. https://doi.org/10.1086/678519

Munkvold, R. I., & Sigurdardottir, H. D. I. (2018). Norwegian game-based learning practices: Age, gender, game-playing and DGBL. In M. Ciussi (Ed.), *Proceedings of the 12th European conference on game-based learning* (pp. 460–468). Academic Conferences and Publishing.

Muthén, L. K., & Muthén, B. O. (1998-2020). *Mplus user's guide* (8th ed.). Muthén & Muthén.

Nadolny, L., & Halabi, A. (2016). Student participation and achievement in a large lecture course with game-based learning. *Simulation & Gaming*, *47*(1), 51–72. https://doi.org/10.1177/1046878115620388

Niederle, M., & Vesterlund, L. (2010). Explaining the gender gap in math test scores: The role of competition. *Journal of Economic Perspectives*, *24*(2), 129–144. https://doi.org/10.1257/jep.24.2.129

Niederle, M., & Vesterlund, L. (2011). Gender and competition. *Annual Review of Economics*, *3*(1), 601–630. https://doi.org/10.1146/annurev-economics-111809-125122

Nousiainen, T., Kangas, M., Rikala, J., & Vesisenaho, M. (2018). Teacher competencies in game-based pedagogy. *Teaching and Teacher Education*, *74*, 85–97. https://doi.org/10.1016/j.tate.2018.04.012

O'Neil, H. F., Wainess, R., & Baker, E. (2005). Classification of learning outcomes: Evidence from the games literature. *The Curriculum Journal*, *75*(4), 455–474. https://doi.org/10.1080/09585170500384529

Pivec, M. (2007). Play and learn: Potentials of game-based learning. *British Journal of Educational Technology*, *38*(3), 387–393. https://doi.org/10.1111/j.1467-8535.2007.00722.x

Pivec, M., Dziabenko, O., & Schinnerl, I. (2003). Aspects of game-based learning. In *3rd International Conference on Knowledge Management* (pp. 216–225). Elsevier.

Plass, J. L., Homer, B. D., & Kinzer, C. K. (2015). Foundations of game-based learning. *Educational Psychologist*, *50*(4), 258–283. https://doi.org/10.1080/00461520.2015.1122533

Prensky, M. (2001). *Digital game-based learning*. McGraw-Hill.

Ranchhod, A., Gurau, C., Loukis, E., & Trivedi, R. (2014). Evaluating the educational effectiveness of simulation games: A value generation model. *Information Sciences*, *264*, 75–90. https://doi.org/10.1016/j.ins.2013.09.008

Røkenes, F. M., & Krumsvik, R. J. (2016). Prepared to teach ESL with ICT? A study of digital competence in Norwegian teacher education. *Computers & Education*, *97*, 1–20. https://doi.org/10.1016/j.compedu.2016.02.014

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Sandford, R., Ulicsak, M., & Facer, K. (2006). *Teaching with games: Using computer games in formal education*. https://www.nfer.ac.uk/publications/FUTL49/FUTL49guidanceforeducators.pdf

Schumann, P. L., Anderson, P. H., Scott, T. W., & Lawton, L. (2014). A framework for evaluating simulations as educational tools. *Developments in Business Simulation and Experiential Learning*, *28*, 215–220 https://absel-ojs-ttu.tdl.org/absel/index.php/absel/article/view/832

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

Sitzmann, T., Ely, K., Brown, K. G., & Bauer, K. N. (2010). Self-assessment of knowledge: A cognitive learning or affective measure? *Academy of Management Learning & Education*, *9*(2), 169–191. https://doi.org/10.5465/amle.9.2.zqr169

Slavin, R., & Madden, N. A. (2011). Measures inherent to treatments in program effectiveness reviews. *Journal of Research on Educational Effectiveness*, *4*(4), 370–380. https://doi.org/10.1080/19345747.2011.558986

Snijders, T. A. B., & Bosker, R. J. (2011). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). Sage.

Squire, K. (2005). Changing the game: What happens when video games enter the classroom? *Innovate Journal of Online Education*, *1*(6), 1-8. https://nsuworks.nova.edu/innovate/vol1/iss6/5

Swaak, J., & de Jong, T. (1996). Measuring intuitive knowledge in science: The development of the what-if test. *Studies in Educational Evaluation*, *22*(4), 341–362. https://doi.org/10.1016/0191-491X(96)00019-3

Thavikulwat, P. (1996). Activity-driven time in computerized gaming simulations. *Simulation & Gaming*, *27*(1), 110–122. https://doi.org/10.1177/1046878196271008

Tobin, K., & Garnett, P. (2003). Gender related differences in science activities. *Science Education*, *71*, 91–103. https://doi.org/10.1002/sce.3730710112

Tømte, C., Hovdhaugen, E., & Solum, N. (2009). *ICT in initial teacher training: Norway country report*. https://www.oecd.org/norway/45128319.pdf

Tricot, A., & Sweller, J. (2014). Domain-specific knowledge and why teaching generic skills does not work. *Educational Psychology Review*, *26*(2), 265–283. https://doi.org/10.1007/s10648-013-9243-1

Vandercruysse, S., Vandewaetere, M., & Clarebout, G. (2012). Game based learning: A review on the effectiveness of educational games. In M. M. Cruz-Cunha (Ed.), *Handbook of research on serious games as educational, business, and research tools* (pp. 628–647). IGI Global.

Vos, L. (2015). Simulation games in business and marketing education: How educators assess student learning from simulations. *The International Journal of Management Education*, *13*(1), 57–74. https://doi.org/10.1016/j.ijme.2015.01.001

Wikan, G., & Molster, T. (2011). Norwegian secondary school teachers and ICT. *European Journal of Teacher Education*, *34*(2), 209–218. https://doi.org/10.1080/02619768.2010.543671

Wollscheid, S., & Skjelbred, S.-E. (2021). Playing your way to knowledge—The impact of digital games on student learning in marketing and business studies. NIFU Working Paper 2021:10.

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*(2), 249–265. https://doi.org/10.1037/a0031311

Wouters, P., & van Oostendorp, H. (2017). Chapter 1. Overview. In P. Wouters & H. van Oostendorp (Eds.), *Instructional techniques to facilitate learning and motivation of serious games* (pp. 1–16). Springer.

How to cite this article: Skjelbred, S.-E., & Daus, S. (2022). Satisfaction is insufficient: Insights from a randomized, controlled trial of a marketing simulation game. *Journal of Computer Assisted Learning*, 1–17. https://doi.org/10.1111/jcal.12703

TABLE A1 Overview of item variables and item labels for each measure

Measure	Item stem	Sub-items
Covariate measures		
Experience with Digital Tools	To what extent do you agree with the following statements?[b]	(a01) I am good at finding information online. (a02) I like to play digital strategy games. (a03) I like to use computers and tablets at school. (a04) I lose concentration when we use computers and tablets at school.[a]
Outcome measures		
Satisfaction with The Game	How do you think the simulation game Hubro Marketing Simulation worked? (only intervention)[b]	(b29) I had fun when we played the game. (b30) I received the necessary training in using the game so that I could focus on the marketing principles. (b31) The game caught my interest from the beginning. (b32) I would recommend this kind of game to other students and classes. (b33) The tasks in the game were neither too simple nor too hard; they were good challenges. (b34) I am confident that I learned something through the game. (b35) The game was easy to play and understand. (b36) I would consider continuing the game in the next weeks if I get the opportunity to it. (b37) I learned things from the teacher or textbook which I would not have learned if I only played the game.[a]
Course Interest	To what extent do you agree or disagree with these statements about marketing?[b]	(b01) Marketing seems like a very interesting subject/course. (b02) Marketing is one of my favourite courses. (b03) Sometimes I get totally immersed in working with a marketing problem. (b04) I get bored in marketing classes.[a] (b05) It is exciting to learn the underlying principles and theory in marketing. (b06) I would like to continue with Marketing and Leadership 2 next year.
Content Interest	I think it has been interesting to learn about...[b]	(b07) Situational analysis and the company's working conditions. (b08) The development of marketing over time (history). (b09) Various kinds of markets. (b10) Market information systems and various forms of data collection.[d]
Job Interest	Should you in the future have responsibility for marketing in a larger organization, how interesting would the following be to you?[c]	(b11) Develop a marketing strategy for the entire organization. (b12) Plan the marketing of a specific product or service. (b13) Report my analysis of potential customer groups' preferences. (b14) Produce advertisement material for a new product that has not met expected sales. (b15) Explain to colleagues the principles of marketing and leadership. (b16) Analyse with quantitative data who visits the website, what they do and what characterise these visitors.[d] (b17) Estimate appropriate prices to put on various products or services.
Perceived Test Performance	How confident are you with the questions you just answered?[b]	(b21) I had to guess the answer to several questions.[a] (b22) I was motivated to do my very best on these questions.[d] (b23) I believe I could have done more difficult questions. (b24) The questions aligned with what the teacher covered during instruction.
Perceived Learning Gains in Course	Do you agree with the claims below regarding what the course taught you?[b]	(b25) What we had in class was completely new to me. I have really learned new things. (b26) I have become better at finding solutions for a company with challenges. (b27) I will likely be able to use what I have learned so far later in other parts of the course.

TABLE A1 (Continued)

Measure	Item stem	Sub-items
		(b28) I have become better at analysing large chunks of information about the market more effectively (sorting, categorizing and evaluating the information).
Achievement		(b100-b106) Various test items

[a]Reverse coded.
[b]Scale: Strongly disagree—strongly agree (five categories, colour aided).
[c]Scale: Not at all interesting—very interesting (five categories, colour aided).
[d]The intervention group agrees significantly more to the item than the comparison group.
[e]The comparison group agrees significantly more to the item than the intervention group (no occurences).

TABLE A2 Individual teacher responses to survey questions on participation in training opportunities

#	… the simple introduction course about HMS before the summer break?	… refresher training session between summer and autumn breaks	… advanced user course after the autumn break	… individual training by Hubro Education, the game provider	… collaboration with a teacher using the game from Hubro Education	*M*
1	Not at all	Not at all	Some	Not at all	Not at all	1.4
2	Not at all	To a great extent	NA	Not at all	Some	2.5
3	Not at all	Not at all	To a great extent	To a great extent	Not at all	2.6
4	Not at all	Quite a bit	Quite a bit	Quite a bit	Not at all	2.8
5	Very little	To a great extent	Quite a bit	Not at all	Some	3.0
6	Quite a bit	Quite a bit	Quite a bit	Very little	Very little	3.2
7	Quite a bit	Quite a bit	Not at all	To a great extent	Some	3.4
8	Not at all	To a great extent	To a great extent	Very little	Quite a bit	3.4
9	To a great extent	To a great extent	To a great extent	Very little	Not at all	3.6
M	2.6	4.0	4.0	3.0	2.1	

Note: "To what extent have you participated in…." Response categories were (Norwegian original in parentheses). Not at all (veldig lite)—Very little (lite)—Some (verken lite eller mye)—Quite a bit (mye)—To a great extent (veldig mye). NA = missing. # = Teacher ID. *M* = mean across teachers or across variables using converted values: Not at all = 1, To a great extent = 5.