Evaluating Methods for Measuring the Effect of Serious Games in Primary and Secondary School

Brage Fosso, Halvor Bakken Smedås, Johannes Weiberg-Aurdal
Department of Computer Science
Norwegian University of Science and Technology
Trondheim, Norway
Email: {bragefo, halvorbs, johanwei}@stud.ntnu.no

Abstract—With serious games becoming ever more popular both for use and as a topic of research, we’ve reached a point where it’s crucial to make sure that we know the effects serious games used in educational contexts actually have on their players. To obtain this knowledge we need to first know and understand the methods used to investigate serious games’ effect.

With the intent to put more focus onto how the target audience might play a central role in how we measure the effect of serious games, we look at how serious games’ effect in primary and secondary school is measured, and what the challenges of doing this are, before we finally evaluate how are they dealt with. We investigate these problems through an expert opinion survey anchored by a supplement of literature and find two core methods of evaluating a serious game’s effect, and the challenges that accompany both.

Index Terms—serious games, educational games, evaluation, assessment, challenges.

I. INTRODUCTION

Serious games are on the rise in education in schools[1]. Chen & Michael [2] define serious games as "Games that do not have entertainment, enjoyment or fun as their primary purpose.” The main purpose is learning or practicing a skill. Serious games have many application areas. Healthcare professionals have a need for practicing emergency situations, and serious games can help them prepare for this [3]. Serious games can also be used in military training [4], and be a tool for teaching primary school students’ mathematics [5].

As the use of serious games increases in school systems [1], it’s more important that the effect of the games is measurable. Can serious games lead to greater academic improvement than traditional learning? Does it increase motivation and engagement in the children, and how can we measure such learning attributes? These are some of the practical problems that lead to this research paper.

The paper will focus on the measurements of serious games’ effect on learning. Our results shed light on how the effects are being measured, as well as challenges that come along with it. In this paper, we first look at related work and how our work differs from it. We follow it up with a brief discussion of the methods we’ve applied in our research. Then, our findings are presented, and later discussed in detail. Lastly, the conclusion finishes our research paper.

Our contribution comes from the fact that we have collected related work to our research topic, narrowed these results down to one aspect of serious games, and focused them on a specific target group. Combined with the qualitative data we have gathered from experts in the field, we present new findings to the research community.

The research questions are defined as follows:

RQ1 How is the effect of serious games in primary and secondary school measured?
RQ2 What are the challenges of measuring the effect of serious games in primary and secondary school?
SRQ How are these challenges dealt with?

II. RELATED WORK

With serious games becoming ever more popular in several disciplines, there is an increasing amount of research being done in the field [1, Fig. 1–2]. As such, there are several papers and books written about the problem of assessment and measuring of serious games’ effect.

The success of a serious game ought to be determined by its ability to present a serious matter to the target group — its effectiveness. Effectiveness, in this context, is a high-level concept built upon several “effect factors”; examples being motivation, engagement, and presentation, etc. While it’s of interest to everyone involved to know whether or not serious games are more or less effective than traditional non-interactive means of teaching, it’s important to note that it’s these underlying, low-level factors and how they are addressed that is the core for a serious game’s success and effectiveness.

Backlund and Hendrix performed a meta-analysis on serious games’ effectiveness [6]. Their work is based on evidence found in research exploring the previously mentioned underlying, low-level factors for effectiveness. Amongst others, they prevalently refer to Tobias and Fletcher [7], a comprehensive review of research on computer games, in which research evidence is examined for the effectiveness of games for instruction, i.e. games for serious purposes.

We consider Tobias and Fletcher’s findings to be a result of “lower-level” research because they examine a multitude of effect factors and several categories of usage of serious games — in other words; “low-level” in the sense that they examine the underlying factors for effectiveness.
In their research, 95 studies were categorized into groups with respect to their intention and knowledge claim based on the underlying factors, and their impact on learning.

As serious games have gained popularity [1], it’s important to remain critical to how they should be evaluated, and not just fall back on research discussing other serious games, or research discussing serious games on a higher level, i.e. a generalization of the field. This exact issue of insufficient, or poor assessment of serious games and knowledge of the impact on the target group is raised by Mitgutsch and Alvarado [8].

Based on this, it’s reasonable to think of the target group itself as an underlying factor of a serious game’s effectiveness when evaluating its effect. Hence, it’s crucial that evaluation is done per target group, and not as a projection onto a generalized, idealized, thought up target audience.

This issue is brought up by Tüzün et al. in [9], where learning outcome and motivational impact of primary school pupils are evaluated by subjecting the target audience (i.e. the pupils) to their game over a period of time.

We’ve focused our research on serious games specifically made for primary and secondary school, rather than for the general population precisely because of this.

While Tüzün et al.’s work does well in considering target audience as an underlying factor of effectiveness, and hence is narrowing their evaluation to said target audience. We find that it’s not fully exploring the remaining connections between design (what is made) and purpose (what’s the intended positive effect) when evaluating its effectiveness.

Mitgutsch and Alvarado identify this issue as one of the flaws of serious games in their initial statements:

"Although initial empirical studies on serious games usage deliver discussable results, numerous questions remain unacknowledged. In particular, questions regarding the quality of their formal conceptual design in relation to their purpose mostly stay uncharted."

As both learning outcome and motivational impact is key for a learning experience, a game development team will need to be able to create said learning experience, which is where game design comes into play: Mitgutsch and Alvarado suggest that there are 5 key design elements that should be assessed as underlying factors for effectiveness when it comes to serious game design:

- **Purpose**
- **Fiction & Narrative**
- **Game Mechanics**
- **Content & Information**
- **Aesthetics & Graphics**

In addition, they present **Framing** as a sixth and partially overlooked point of assessment. The framing of the five mentioned key elements in relation to the target audience, and its play literacy [8], which is the aforementioned factor Tüzün et al. were on to with their research.

This is ultimately where we stand when raising our research questions; with serious games’ success depending on a multitude of factors, how do you go about and measure them? — Doing so seems to be a challenge, so how do serious game developers deal with the potentially challenging situations?

### III. Methods

#### A. Data collection

The collection of data for this study was done primarily by conducting a survey with interviews, with a literature review as anchoring of the findings. The sampling frame for the survey focused exclusively on experts within serious games that had experience or knowledge with measuring the effect of such games in a primary or secondary school setting. This sampling frame was chosen in order to be able to answer the research questions.

Surveys are often accomplished with questionnaires. However, questionnaires are best suited when the goal is to obtain standardized, brief, and uncontroversial data from a large number of people [10, pg. 220]. This proved to be difficult to achieve both because of low sample size in the survey and because non-standardized data was needed to be able to answer the research questions. Hence, interviews were chosen as the data generation method. The main sampling technique we used was snowball sampling, in which the first interviewee gave suggestions about other people that could be relevant to interview.

The interviews were recorded by audio recording with the consent of the interviewee, in addition to taking field notes after the interviews. The interviewees were informed that the data could be anonymized and that any personal data would be destroyed after the paper was published. The first interview was held in person, which were also planned for the remaining interviews. Due to the outbreak of COVID-19, the rest of the interviews were held through video chat.

Since the primary purpose of this study was “discovery” rather than “checking”, semi-structured interviews were used. Here, we were able to have a list of questions in order to get comparable data, in addition to being able to ask follow-up questions and let the interviewee speak freely. The questions made for the interviews were mainly focused on the research questions, but also on getting information that could help to describe the research context and related work. By doing this, the generated data was highly beneficial towards answering the research questions.

#### B. Data analysis methods

Qualitative analysis was used on the data from the interviews. After collecting data from the interviews, the data was thoroughly analyzed, and segments that appeared to be relevant to the research questions were focused on. Each segment was then categorized into measurement methods, i.e. ways of measuring the effect of serious games. This categorization used an inductive approach, which meant that we focused on categorizing the data without it being influenced by existing theories or our own experiences and prejudices.

When all the data were categorized, a table was made in order to analyze the data together. Here, all the measurement methods were listed together with the related challenges and how the challenges could be handled.
IV. FINDINGS

The findings for this study were collected from interviews with two Ph.D. Candidates (A-B), one Associate Professor (C), and one Lead Developer from an E-Learning game-company (D). Combined, they have a solid background within research, design, and development of serious games.

Interview subject A and B brought up two popular ways of measuring the effect of serious games in primary and secondary school; stealth assessment and observation. These were deemed the most important ones by interviewee A and B, which is the reason they became the main focus point in the findings. Subject C added onto these, talking about some of the challenges with both, and subject D added information about how they in their recent project had dealt with challenges with these measurement methods.

The findings are summarized in the table below.

<table>
<thead>
<tr>
<th>Measurement Method</th>
<th>Challenges</th>
<th>Ways to handle</th>
</tr>
</thead>
</table>
| Stealth Assessment | - Time-consuming  
- Lack of technical competence in teachers  
- Capturing context of use | - Use Observation  
- Automatic data generation & stats-tracking  
- Provide analysis tools  
- Iterative development |
| Observation        | - Avoiding the Hawthorne effect  
- Availability of parties involved  
- Costly and time-consuming  
- Need for expert knowledge | - Use Stealth assessment  
- Communication tools and/or -platforms  
- More resources  
- Use experts |

1) Stealth assessment: Interview subject A defines stealth assessment as "...algorithms included in the game that measures what the players do and how well they do it...", and said that "...if the player is able to get past certain challenges in a game, and those challenges correspond with a learning goal, this could be used to say something about what they learned." The goal of stealth assessment is to gather learner performance data from the player throughout the playing time without the player noticing the assessment [7]. Subject C points out the importance of the assessment not being noticeable: "The concealed collection of data is trying to avoid the Hawthorne effect - if people know they are being measured on a particular output, such as score, it changes their behavior, which means that it’s no longer useful in correlation to other behaviors."

Both interview subject A and B explains that a key challenge is to make the stealth assessment unobtrusive, and it’s crucial that this does not break the game flow. To solve this problem, the tracking mechanisms in the game should not draw any attention from the children playing. Subject B says that he wants to generate data that can be used to analyze the player’s behavior and that the teachers should have tools to perform analysis. Subject D also presented this to be a data-generation method sought out for by their peers.

2) Observation: Observation is a popular data generation method in research and can be used in many different ways to measure the effect of serious games. When talking about stealth assessment and observation, interview subject A says "...I prefer to observe how the players play [...] to see if this shows any reflection and/or improved learning performance." Interview subject A also thinks that observation is especially beneficial because "...it’s possible to get the complete picture of how the game works within a specific context, e.g., a classroom with primary students playing a coop-game."

As in stealth assessment, both interview subject A and B agree that a key challenge is to make the observation as unobtrusive as possible while still getting the needed data. If the observer interrupts the player by asking questions, which is normal in order to get data during an observation or playtest, it could break the game experience, which could be negative in the learning progress according to interview subject B. According to interview subject A, it is smart to ask questions both before and after an observation in order to get the complete picture on learning performance, classroom dynamics, and student perception.

3) Costly and time-consuming: Subject A, B, and C report that evaluating performance of serious games is costly and time-consuming. Teachers already have a lot of work and it’s challenging for them to learn the analysis tools and to fit serious games into the curriculum. Subject A also mentions that there are several teachers that lack technical competence needed to oversee classes using serious games in the classroom. Subject D explained that the schools that contracted work to them did not have enough experience and knowledge to define complete requirements before seeing a prototype. This led to multiple iterations and exceeded budgets and time frames.

V. DISCUSSION

Serious games as a topic is broad, and it’s difficult to measure the effect of them. It depends on a lot of factors, including context, target group, and scope. The findings focus on the most important and popular ways of measuring the effect of serious games according to our interviewees. The first research question is "How is the effect of serious games in primary and secondary school measured?". From interviewing experts in the field, we have learned two methods of doing this; stealth assessment and observation. Because of the limited sample size, there might be some undiscovered methods of assessment. However, our literature review supports our findings.

The experts shared their opinions and experience of working with these methods, and provided insights regarding our next research questions; "What are the challenges of measuring the effect of serious games in primary and secondary school?" and "How are these challenges dealt with?".

One of the big challenges with observation is that it should be unobtrusive, that is; the Hawthorne effect should be avoided. Unfortunately, observation is subject to a more subjective interpretation of the game’s effect and depends to a larger degree on experts in for example behavioral psychology to perform the observation. Subjective interpretation can also be considered as one of the strengths of observation as a method: where stealth assessment needs to be pre-built into the game, observation does not - allowing the observer to apply...
their own subjective experiences to spot, and reason about non-anticipated positive or negative behavior.

Ultimately, observation has the side-effect of being more difficult to conduct, due to a multitude of concerns:

- Laws and rules that might dictate when and for how long the observation process may go on.
- Administrative rulings on when and how such an observation process might be conducted.
- The availability of the involved parties, and the overlap of time between them to spend together in a productive manner.
- Requires several experts to perform a proper analysis.

The experts agreed that these limitations can to some degree be handled by effective and productive communication between the parties involved and by cooperative planning.

Stealth assessment solves some of the challenges with observation - while observation is preferred when measuring classroom dynamics and student perception, stealth assessment is best at measuring learning performance [11]. In stealth assessment, the students are unaware of being assessed, thus the Hawthorne effect does not come into play.

Observation can be time-consuming for the teachers involved. A solution to make it easier and reduce the time for teachers to measure the effect of serious game, is by using a serious game with stealth assessment that has implemented a assessment design approach i.e. evidence-centered design[7]. This approach is based on evidentiary reasoning and provides a framework for educational assessment [12]. There are however teachers that lack technical competence, that would find it hard to use analysis tools. Training them to use the tools properly is therefore important but is also time and resource-demanding.

Interviewee subject D said that his company had taken great benefit of involving teachers who were engaged in E-learning.

Throughout our research, we’ve realized that the methods used for evaluation that we present, although framed in the context of primary and secondary school, may just as well be applied on broader audiences - the two general approaches remain the same, though the challenges accompanied by both may differ from challenges presented in this paper.

The contribution of the paper is that we have identified two main methods of assessment that are being actively used in serious games for primary and secondary schools today. We have presented a structured collection of the challenges related to the methods, and how experts in the industry handle them.

VI. CONCLUSION

Serious games is an interesting theme that shows a lot of potential. It’s increasing in popularity fast, which makes it an important and relevant research field. The main message of the paper is that there are several ways of measuring the effect of serious games in primary and secondary schools, and many challenges accompany them. Many of these problems arise because of limited resources and time given to research and investigate the games’ effect. There are solutions to some of these problems, indicating that there is still a lot of research to be done.

For future work, it would be interesting to run a case study to look at a specific serious games being used in practice. This could, for example, be to measure the effect of a serious game for teaching mathematics within primary school.

As the sampling frame was small and we had limited time and resources, the data collection was limited. Ideally, we would have more qualitative data to increase the validity of our findings. This could be achieved by conducting more interviews, performing a questionnaire, or by observing a school class playing a serious game.

ACKNOWLEDGEMENT

We would like to thank the interview subjects for their time and helpful contribution to this research, and our peers for reviewing our paper and giving helpful feedback.

REFERENCES


Behavioral Outcome Effects of Serious Gaming as an Adjunct to Treatment for Children With Attention-Deficit/Hyperactivity Disorder: A Randomized Controlled Trial. Authors of this article: Kim CM Bul 1, 2; Pamela M Kato 3; Saskia Van der Oord 4, 5, 6; Marina Danckaerts 7; Leonie J Vreeke 1, 2; Annik Willems 8; Helga JJ van Oers 9; Ria Van Den Heuvel 10; Derk Birnie 11; Thérèse AMJ Van Amelsvoort 12, 13, 14; Ingmar HA Franken 2; Athanasios Maras 1. Table 1. Description of primary and secondary outcome measures. Measures. Research on the effects of primary and secondary education does not withhold education from children, thereby necessitating either natural experiments or correlational designs employing sophisticated statistical techniques, over time, to. Keywords: School effects, pupil cognitions, pupil attributions, academic self concept. Accepted manuscript received 29 July 1993. Although an initial I.Q. advantage for pre-school graduates disappeared by entry to secondary school, there were startling differences in outcome between the 65 children who attended the half-day educational programme over two years and the control group of 58 children who had remained at home (Berrueta-Clement, Schweinhart, Barnett, Epstein & Weikart, 1984). Applying the right teaching methods and techniques in classroom can improve the student learning. But How to implement it? Have a Look. They should be given some practice tests on a regular basis when they can evaluate or assess themselves on the areas where they lack and need to improve. This can include practice problems, virtual flash cards and more. The self learning phase gives them the independence to take control over their learning and they can have a boost to their confidence level. There are a lot more teaching techniques and strategies that give impressive results in improving the student learning. School authorities should also offer a helping hand in supporting teachers to implement advanced and creative strategies in classrooms. However, evaluation of the usability of this kind of games requires a redefinition of techniques, methods and even terminology. In this paper, we elicit six research questions and conduct a systematic review of the scientific literature, which resulted in the selection of 187 papers that contained the most relevant responses. The conclusions of this systematic review illustrate the general status of current academic usability evaluations of these games and the main trends in the selection of methodologies and how are they applied. This view may be a valuable foundation for future research. This is a preview of subscription content, log in to check access. Serious Games (SGs) indicate positive effects on Preschool and Primary school students and promote a multi sensory style of learning. This review paper explores the integration of Serious Games in the area of Preschool and Primary Education, in the last decade (2006-2016). The studies were carried out on Preschool and Primary school settings, respectively. Research showed that Serious Games are able to keep all students engaged in classroom facilities, scaffolding their learning through increased motivation, independence, autonomy and resultant self-esteem. Serious Games Based Learning (GBL) has proven its added value in almost every aspect of the curriculum. Discover the world’s research. 20+ million members.