Same Game, Different Impact

Diagnosing the Successes and Failures of One Game-Based Intervention Across Four Schools

Andreas Lieberoth (Aarhus University), Thorkild Hanghøj (Aalborg University), and Morten Misfeldt (Aarhus University)

Abstract: A game-based learning approach may work in 1 school, but fall flat in another. This paper details how 1 intervention led to different outcomes and analyzes these differences in terms of the social and cultural characteristics of 4 schools. The intervention for 3rd–5th graders included a mix of commercial-off-the-shelf gameplay, classroom exercises based on the game, and gamification aimed at classroom conduct. General effects were observed for intrinsic and external motivation using the children's locus of causality scales (c-PLOC), plus teacher-assessed learning and well-being. Yet very different practical outcomes were observed in the 4 different schools taking part in the project. This study used an explanatory mixed-methods analysis to dive deep into the qualitative characteristics of 1 classroom where the intervention took a definite hold, concluding that successes were significantly determined by a combination of support for the teachers in the schools, visible positive results, and the specific needs at play in each classroom. Specifically, we find that, although the intervention had several kinds of positive impact, simpler gamification elements worked best with younger students and/or for teachers who struggled to maintain focus in their classrooms.

Introduction

Commercial off-the-shelf games hold promise for curriculum learning, development of noncognitive skills, and fostering collaboration among players (Barab, Gresalfi, & Arici, 2010; Gee, 2003; Hanghøj, 2017; Lieberoth, 2017; Squire, 2006). With the rise of esports as an after-school activity, and emergence of both grade- and high school esports programs, researchers and teachers also are experiencing an increased impetus to identify a place for "real games" in school settings and to develop pedagogical practices around them (Lieberoth, Fiskaali, & Spindler, 2018; Lieberoth & Hanghøj, 2017). Yet a recent mixed-methods intervention study showed us that there can be a notable difference in how a game-based intervention impacts students depending on their social positioning and individual learning needs (Hanghøj, Lieberoth, & Misfeldt, 2018).

Through the lens of between-schools differences, we here combine statistics and ethnographic evidence to exemplify what we believe to be critical characteristics of effective and problematic implementations of games in new classrooms.

From field observations, it became clear that all schools and classrooms had unique characteristics, which led to variations in the implementations. Teachers had different motivations and opportunities. Students had different needs, social backgrounds, and existing foundations to build upon. Even technical practicalities came to determine the interventions. As an example, one school was haunted by practical problems, including an impromptu staff replacement after teacher training had been conducted, which lessened staff ownership and led to a less rigorous implementation of the scheduled activities, both in terms of curriculum integration and classroom elements. In another school, teachers originally felt very frustrated, but their principal supported them admirably, to a point where they became highly motivated

converts to game-based learning. This mélange led us, after the initial data analysis, to ask the question: Did the botched implementation observed early leave a measurable mark on the students as measured though the children's perceived locus of causality) (c-PLOC) scales (Pannekoek, Piek, & Hagger, 2014)? While this initial hypothesis was rejected, a statistical analysis of between-schools differences combined with ethnographical accounts reveals interesting patterns mainly related to intrinsic motivation and external regulation (as per Ryan & Deci, 2000) in creating new spaces for game-based teaching practice.

Here, we zoom in on an underprivileged school, where the implementation appears to have been especially successful in order to identify key routes to impact. We also discuss the methodological challenges to assessing the effects of multilayer interventions, and we hope to have demonstrated that these and other methodological limits can be somewhat mitigated using mixed methods.

The Four Schools

The four schools were spread out across Denmark, but were all situated in or close to larger cities. None of the teachers who participated in the intervention had prior experience with multiplayer games in their classrooms.

The four schools got involved with the project for different reasons, so some differences between them were related to the degree of commitment invested by the school management, support from colleagues who already advocated games, and the overall interest of the project teachers in using the "school at play" method.

The *Inner City School* (37 students, third and fifth grade, 20 female) was situated in a poor area with a high concentration of immigrant students. The school management at Inner City was eager to be involved in order to explore new methods that might help include their high rate of at-risk students. The teachers were at first reluctant to participate, and even experienced severe breakdowns in the process, but after diligent support by management, they experienced positive effects on their students, which led to great enthusiasm. In the end, the teachers continued to use games after the intervention.

The Mixed Urban School district overlapped a high-income area and some poorer immigrant blocks (44 students, fourth and fifth grade, 24 female). Management did not display strong commitment, and participation in this intervention was mainly seen as one project among many. This negatively influenced the teachers' relation to the project.

Suburban Schools 1 and *2* were quite similar in their locations and middle-class demographics. Suburban School 1 (55 students, fifth grade, 26 female) wanted to participate in the project as a way of trying out their new computers, but management did not pay much attention to the aims and results of the project, which made it difficult to continue the project at this school. On the contrary, both management and teachers at Suburban School 2 (55 students, fifth grade, 28 female) were highly interested in the project, mainly because of other teachers who advocated games at the school and could share positive experiences, but also align expectations, in their classes.

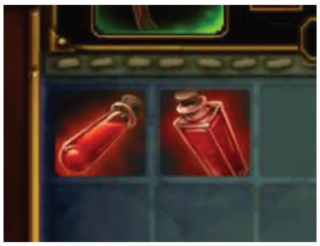
The Intervention

The school at play intervention departs in the core idea of using a selected video game as a linchpin for social activities and themed classwork, while also instituting a gaming "tone" into the overall classroom

experience through metaphors such as quests, progress bars, and levels. This mirrors the ambitions for some high-school esports programs (Lieberoth et al., 2018). Its methodology was conceived by a duo of special education teachers, who would later found a consultancy based on their experiences and tools and apply for funding to expand the technique into normal classrooms, with us as research partners. Sixteen teachers were introduced to the method though two rounds of four one-day courses, with eight teachers in each round. The teachers were given templates to customize classroom tools, and they were actively involved in building curriculum assignments around the themes and mechanisms of a chosen game. Since the intervention was developed in the context of special education, a significant strand of our research, as well as the intervention activities, centered on the promise of creating inclusion opportunities for at-risk students as much as on reframing curriculum activities for everyone.

Torchlight II and Game-Related Assignments

The school at play method uses digital games off the shelf to create contexts for collaboration and learning, as students explore themes, mechanics, and strategies, and connects these to curriculum elements (in this case within math or Danish classes) that may grant advantages in-game. For instance, equations, percentages, and fractals are explored in *Torchlight II* through math assignments related to *time* to swing a weapon and the amounts of *damage* dealt by each swing, relative to an enemy's total life points (for an example, see Figure 1). Through this mapping of math onto gameplay, students may discover more efficient in-game attack combinations. Similarly, students wrote guides to the game for other players in Danish. The hope is that in doing these exercises, students come to identify more with the usefulness of math or clear writing in their real (gaming) lives (see Figure 2). After the *Torchlight II* experience, teachers tried implementing games of their own choosing for a few weeks. In the considerations below, we will focus on an inner-city school where the teachers tried their hand with *Minecraft*.



A Health Potion gives 900 health within 8 seconds. A Big Health Potion gives 1.800 within the same time span.
How much health pr. second do you get from 1 Health Potion?

How much health pr. second do you get from 1 Big Health Potion?

Imagine that you are running away from a group of skeletons and have 200 out of 500 health. Which potion do you drink? Why?

Figure 1. Math assignment for Torchlight II.

Figure 2. The loop between game activities and curricular activities envisioned by the intervention's originators.

Classroom Gamification

To support the overall process, wall charts and gamification activities are used during lessons. Students, for instance, move their names on a *progress bar* as they solve assignments, with the additional option of progressing up to 150%. Higher- and lower-performing students are given differentiated challenges to ensure equal progress for the same amount of effort. Another "classroom game" aims at classroom conduct by making specific social *virtues* visible and rewarding related behaviors with points, of which the cohort can then be traded in for shared benefits (e.g., more game time, finishing up early to play outside, etc.).

Summing up, the goal of the school at play intervention is to create new social spaces and playfully facilitate student identification with social virtues of conduct, teamwork, and inclusion, and the goals of school subjects. As researchers, we did lend support to designing the training process that would allow teachers to adapt the method. This development perspective gave the whole process a twist of design-based research and action research, as we were automatically viewed as representatives of the method when training teachers and doing field observations. A core tenet was, however, to keep the intervention similar across schools.

Previously Reported Results of the Intervention

The intervention was studied using both qualitative and quantitative methods, at various points zooming in on students, teachers, the training of teachers, and subgroups of students. This left a large and rich pool of data that could not easily be collated into a single analysis. The process had the traits of a parallel mixed-methods intervention design (Creswell, 2013; Creswell & Plano Clark, 2011).[1] Still, some strands of the data became separate bracketed components, which have yet to be integrated into the overall evaluation story as of this writing.

After a brief summary of initial results, we report statistical between-schools differences. We measured effects of the intervention using both teacher assessments and the children's perceived locus of causality

1. Maxwell and colleagues (2015) emphasize that descriptive models of mixed-methods designs often become useful only at the end of the process, when the different data strands have fully materialized.

scale (C-PLOC), which assesses five different motivational dimensions, including intrinsic motivation, identification with the value of the learning activity, and external pressure to participate.

Elsewhere (Hanghøj et al., 2018), we have reported on the fairly complex pattern of individual change seen in the 190 students, as well as unique patterns for students identified as being at risk. By way of summary, FDR-corrected ANOVAs (see Figure 3) revealed changes on most variables, which included clear differences between patterns for students' motivation for Danish and math: Of desirable effects, we might mention that general *participation/thriving* rose, while *amotivation* for Danish class and *external* regulation for math fell. On the other hand, intrinsic motivation and identified regulation also dropped. In contrast to simplistic notions that "games make learning fun, and fun makes learning better," the pattern revealed here is complex. Bidirectional findings and differences between c-PLOC variables such as intrinsic and identified regulation for Danish and math show that motivation to participate cannot easily be reduced to one factor, and that it can be tough to pick up such complexities with quantitative measures alone.[2]

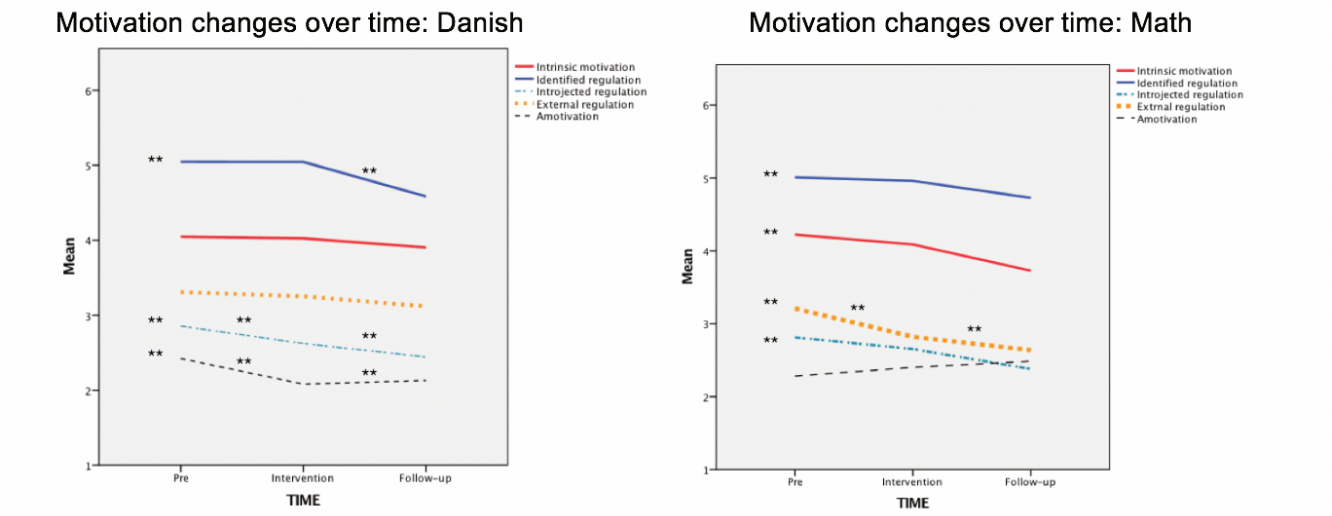


*Figure 3. C-PLOC motivations over time: ** p < .01 change between two time points; ** at the beginning of the line denotes change from pretest to posttest.*

New Hypotheses

One school experienced substantial disruptions, including technical issues and a teacher's being replaced midintervention, which led us to expect that the integration between game and subjects (Danish/math) was weaker, and that (H1) motivational impact toward the school subjects as experienced though the game would probably be less here compared to the other schools. Second, we wanted to know to what extent between-schools differences affected the overall over-time results previously reported (Hanghøj et al., 2018). We expected that significant between-schools differences during the implementation would also affect long-term outcomes (H2).

After this traditional hypothesis testing, we adopt an explanatory analytic approach (Creswell & Plano Clark, 2011), using our field knowledge of practical and social between-schools differences to also understand between-schools differences found in statistical results. Zooming in especially on the Inner

2. The first analysis of the data did not use school as a multilevel factor, as the data set contains too few observations for reliable multilevel analysis, and because our fist publication focused on selective effects on at-risk students compared to their peers.

City School, this brief mixed-methods analysis will be an expansive and strand-challenging reflection on the process as a whole (Greene, 2007; Maxwell, Chmiel, & Rogers, 2015).

Between-Schools Statistics

In order to identify differences between schools as the intervention unfolded, a one-way MANOVA compared the four schools on the c-PLOC scales. The multivariate test showed significant between-school differences: Pillai's Trace = .31, F (30, 474) = 1.83, $p < .01$, $\eta_p^2 = .10$ during the intervention. Univariate analyses revealed that the significant between-schools differences were to be found in two places: on the *intrinsic motivation* subscale for *both* mother-tongue education—F (3, 166) = 8.33, $p < .01$, $\eta_p^2 = .13$—and math—F (3.66) = 5.44, p < .01, $\eta_p^2 = .09$—and on *external regulation* for math—F (3, 166) = 14, 45, $p = .04$, $\eta_p^2 = .10$. No effects were found for the remaining three c-PLOC subscales (see Figure 4).

The Bonferroni post-hoc test revealed that the differences in *intrinsic motivation* were to be found between one school and two others ($p < .01$). We expected that the Mixed City School would fall below the others because of difficulties during the implementation of the intervention. This hypothesis (H1), however, was not confirmed. Instead, the Inner City School scored significantly over ($M = 4.78$, $SD = .19$) the two suburban schools ($M = 3.75$, $SD = .16$ and $M = 3.65$, $SD = .16$). The Bonferroni test did not reveal a significant difference on *external regulation* between any specific schools, however. As can be seen in Figure 4, this might be due to a pattern where pairs of schools (one urban and one suburban) cluster very closely.

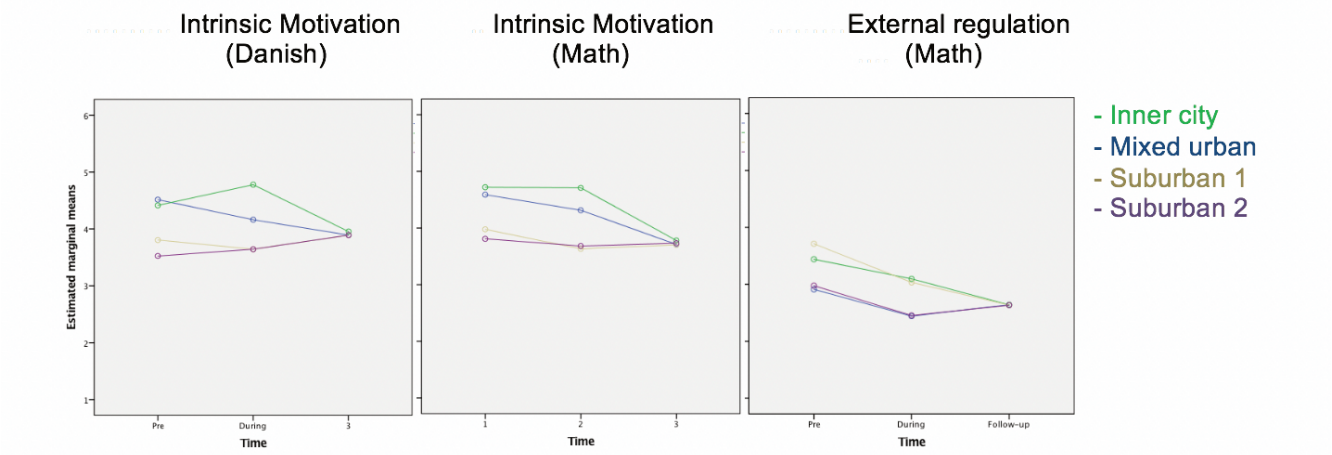


Figure 4. Different change over time for each school on c-PLOC motivation variables.

Next, we examined whether observed between-schools differences were also visible in change from before to after the intervention (see the Previously Reported Results section).[3] For *Danish* the interaction between school and intrinsic motivation, F (6, 270) = 6.29, $p < .001$, $\eta_p^2 = .12$, accounted for a small but significant part of the observed variance, which rendered the change first observed over time nonsignificant: F (2, 270) = 2.02, $p = .14$, $\eta_p^2 = .02$. For *math* the interaction between school and intrinsic motivation was also significant, F (5.88, 264) = 3.91, $p = .001$, $\eta_p^2 = .08$., but the change over time remained significant, F (1.89, 264) = 15.39, $p < .001$, $\eta_p^2 = .10$ (both Hynh-Feldt corrected), with the interaction accounting for a slightly smaller proportion of the variance than the main effect of time.

3. See note 2.

Finally, for extrinsic motivation (math) no interaction effect for school was discovered, F (6, 264) = 1.61, p = .11, η_p^2 = .04, and so the effect over time remained significant: F (6, 264) = 14.59, p < .001, η_p^2 = .11.

In summary, differences between schools were visible in terms of student *intrinsic motivation* and *extrinsic pressure*, confirming H2, and going some way toward explaining the psychological impact of the intervention over time. The most positive general effect of the intervention would seem to be a lessened feeling of being forced into participating through outside pressure, whereas intrinsic motivation for Danish varied significantly by school. Notably, the Inner City School stands out as a success in terms of motivational effects.

Explanatory Ethnography

The process of mixed-methods analysis often involves a conversation between lenses, with one set of initial results, resulting in queries into other types of data. In this case, the statistical analyses left the question: What made the Inner City School special?

Statistical examination of differences between the schools did not confirm the hypothesis that motivation levels for students in the Mixed City School would be significantly worse than in the other locations because of its botched implementation. Indeed, the school did not figure as significantly different from any of the other three schools, revealing that the tumult surrounding the introduction of a new teacher who had not been trained on the intervention method had not rubbed off on the students' engagement with the games and curriculum assignments in a statistically discernable way—at least when looking at c-PLOC scores. If anything, this reveals that the "fun" component of game-infused learning environments can be retained regardless of teacher training, and that any teacher can jump into a game-based teaching regimen if it is clearly structured and he or she has basic skills.

The statistical analysis indicated that students in the Inner City School experienced a significantly higher degree of intrinsic motivation during the intervention than did the other schools. Students also reported high degrees of intrinsic motivation to begin with, which is puzzling, as the teachers report difficulties pertaining to learning and conduct alike in these classrooms. This challenges (as per Greene, 2007) the reliability of the c-PLOC instrument somewhat: Could it be that students simply overreported intrinsic motivation, or was there a genuine surge of engagement? Or could it be that the younger third-grade classrooms or relatively more disadvantaged students simply found the intervention more appealing? Based on teacher interviews as well as observations, it was clear that the Inner City School *did* experience stronger positive effects of the intervention than the other schools. When interviewed, teachers emphasized several positive changes for the at-risk students in relation to using commercial games, gamification, and game-related assignments. As one teacher remarked: "I can feel that the students are much more excited when writing their game guide. Normally, they don't write much, when writing nonfiction. But here they wrote loads and loads in a very short time."

The teachers at the Inner City School emphasized how students benefited from the classroom game. Two teachers of the third grade mentioned several examples of how boys became motivated by being awarded tokens for positive behavior. As an example, one of the at-risk students, who often displayed impatience and aggression, had been so motivated by the tokens that he had told his mother that the project had made him turn from "hating school" to "loving school." The two teachers had at first been very reluctant to use the classroom game, which they viewed as a behavioristic "dog training" (as e.g.,

voiced by Kohn, 1999), but after they began using the tool, they experienced a more positive mood in class, which also meant that they "did not have to shout" at the students.

The teachers also described how the use of *Minecraft* and *Torchlight II* allowed at-risk students to take on new roles in terms of helping others, asking for help, and being seen as "experts." The teachers' observations of positive changes with computer games were mostly related to the boys, but a teacher also mentioned how some of the girls experienced positive effects. As an example, one girl, who was diagnosed with dyslexia, become very engaged not only in playing *Torchlight II*, but also in writing a lengthy guide for the game in Danish, which was addressed at other potential players. This aspect of engaging the at-risk students in curricular activities through game-related assignments is further elaborated elsewhere (Hanghøj, 2017).

In the end, the teachers at the Inner City School became so fond of working with games that they went on to design their own game modules involving *Pokémon GO*, *Hearthstone*, and board game design. This suggests that interventions such as the school at play approach may provide the most value where they can make the most difference for at-risk students, who generally have low intrinsic motivation working with the school subjects.

Discussion

In our first analysis, we found the impact of the school at play intervention to be fairly complex, because it operates on multiple levels: the curricular, the social, and the behavioral. In our first dissemination we unpacked some of this complexity by looking at at-risk students. In the present analysis, we focused on observations of what made each school different. In summary, we see evidence for the same general conclusion in both studies: Reconfiguring school experiences with an array of game elements appears to *most strongly benefit those who otherwise struggle to participate* through a combination of social repositioning and clear new modes of interacting with classmates, teachers, and curriculum. This is the case whether discussing effects at the individual level or the between-schools level. However, we also diagnosed factors that mediate effect:

First, our analysis illustrates how there is no one-size-fits-all solution, and how the successes and problems in each iteration of the same game implementation may depend on factors including age group and whether behavioral, social, or curricular needs are an issue to begin with. Second, the role of leadership backing, as well as the benefits of having other more experienced game teachers available as models and allies, illustrates the benefits of a supportive climate.Third, based on this comparison between schools, we find evidence that different elements of an intervention may become the main springboards for successful implementation in different classrooms. Specifically for our case, this also suggests that the school at play method, which emerged from special needs education, may have had the strongest impact in schools with poorer or younger students who displayed more conduct difficulties but also responded well to simple gamification incentives. Finally, and perhaps most important to the self-determination theory (SDT) framework, the successful case of the inner-city implementation shows that teachers as much as students are motivated by the experience of immediately positive outcomes. As it were, the bar for improvement at this school was low, but after initial pushback, the teachers were among the most dedicated. There were clear signs of learning in all schools (Hanghøj, 2017; Hanghøj et al., 2018), yet changes in terms of conduct and engagement were also most prevalent in the Inner City School—changes that appear to, in turn, have been necessary for *learning* opportunities with games to arise in a positive space for the teachers to ply their craft.

It is puzzling that no effect was found for internalized regulation, which should basically have hinted at the students' identification with the usefulness and meaningfulness of the game-related curriculum assignments. As such, it may be that the school at play intervention's main strength is to motivate and socially reposition students, or simply that the c-PLOC failed to pick up the learning dimension. Indeed, all c-PLOC registers is the students' subjective impression of how the implementation felt, not what it actually did to them. Given how much information was gleaned from observations and interviews, we wish to make the case for more mixed-method work and analysis aimed at understanding *what we study* when we study game-based interventions at scale.

References

Barab, S. A., Gresalfi, M., & Arici, A. (2010). Transformative play: Games as 21st century curriculum. In *Proceedings of ICLS 2010, 2* (pp. 93–100). Retrieved from http://dl.acm.org/citation.cfm?id=1854548

Creswell, J. W. (2013). *Research design: Qualitative, quantitative, and mixed methods approaches*. London, England: Sage. Retrieved from https://books.google.dk/books?id=EbogAQAAQBAJ

Creswell, J., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). London, England: Sage.

Gee, J. P. (2003). What video games have to teach us about learning and literacy. *Computers in Entertainment, 1*(1), 20. https://doi.org/10.1145/950566.950595

Greene, J. C. (2007). *Mixed methods in social inquiry*. San Francisco, CA: Jossey-Bass/Wiley.

Hanghøj, T. (2017). "Read this or die!" Including at-risk students through game-related literacy practices. In *Proceedings of the 7th European Conference on Games-Based Learning, 2017* (pp. 219–226). Sonning Common, England: Academic Conferences and Publishing International.

Hanghøj, T., Lieberoth, A., & Misfeldt, M. (2018). Can cooperative video games encourage social and motivational inclusion of at-risk students? *British Journal of Educational Technology*, *49*(4), 775–799. https://doi.org/10.1111/bjet.12642

Kohn, A. (1999). *Punished by rewards: The trouble with gold stars, incentive plans, A's, praise, and other bribes*. Boston, MA: Houghton Mifflin.

Lieberoth, A. (2017). *Spilpædagogik*. Aarhus, Denmark: Aarhus University Press.

Lieberoth, A., Fiskaali, A., & Spindler, H. (2018, October). *pwning n00bs for highschool credit? A map of modes and expectations for 11 Danish eSports programs.* Poster presented at the UCI Esports Conference, Irvine, CA.

Lieberoth, A., & Hanghøj, T. (2017). Developing professional "game teacher" repertoires: Describing participants and measuring effects in a Danish college course on game based learning. In M. Pivec & J. Gründler (Eds.), *European Conference on Games Based Learning* (pp. 377–386). Graz, Austria: Academic Conferences. https://doi.org/978-1-911218-57-9

Maxwell, J., Chmiel, M., & Rogers, S. E. (2015). Designing integration in multimethod and mixed

methods research. In S. Hesse-Bieber & R. Burke Johnson (Eds.), *The Oxford handbook of multimethod and mixed methods research inquiry* (pp. 223–239). Oxford, England: Oxford University Press.

Pannekoek, L., Piek, J. P., & Hagger, M. S. (2014). The children's perceived locus of causality scale for physical education. *Journal of Teaching in Physical Education*, *33*(2), 162–185. https://doi.org/10.1123/jtpe.2013-0095

Ryan, R., & Deci, E. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, *25*(1), 54–67. https://doi.org/10.1006/ceps.1999.1020

Squire, K. (2006). From content to context: Videogames as designed experience. *Educational Researcher*, *35*(8), 19–29. https://doi.org/10.3102/0013189X035008019