Routledge
Taylor & Francis Group

Journal of Research on Technology in Education

ISSN: 1539-1523 (Print) 1945-0818 (Online) Journal homepage: http://www.tandfonline.com/loi/ujrt20

"No Fun Games": Engagement Effects of Two Gameful Assessment Prototypes

Meg Guerreiro & Chase Nordengren

To cite this article: Meg Guerreiro & Chase Nordengren (2018): "No Fun Games": Engagement Effects of Two Gameful Assessment Prototypes, Journal of Research on Technology in Education, DOI: 10.1080/15391523.2018.1426063

To link to this article: https://doi.org/10.1080/15391523.2018.1426063

JRTE | Vol. 0, No. 0, pp. 1–15 | © 2018 ISTE | iste.org/jrte
DOI: 10.1080/15391523.2018.1426063

"No Fun Games": Engagement Effects of Two Gameful Assessment Prototypes

Meg Guerreiro & Chase Nordengren
NWEA

Abstract

Assessments with features of games propose to address student motivation deficits common in traditional assessments. This study examines the impact of two "gameful assessment" prototypes on student engagement and teacher perceptions among 391 Grades 3–7 students and 14 teachers in one Midwestern and one Northwestern school. Using mixed methods, it finds higher satisfaction for students taking gameful assessments, and conflicting attitudes from teachers regarding the impact of gameful assessments on students' intrinsic motivation and desire to learn. The article concludes by discussing opportunities for continued iteration and innovation in gameful assessment design. (Keywords: gaming, gameful assessment, mixed methodology)

Effective learning relies on students to engage with and participate in a variety of contexts at school, including assessment. Student engagement within an assessment can impact the validity of scores (Swerdzewski, Harmes, & Finney, 2011; Thelk, Sundre, Horst, & Finney, 2009), resulting in scores that may not be a true representation of student skills. Traditional assessments are not explicitly designed to promote student engagement and often provide a sterile environment in which students are asked to demonstrate knowledge and skills. Increasing engagement within an assessment context can therefore impact both student satisfaction and overall assessment accuracy and validity.

In a variety of learning contexts, games have been known to increase engagement (Cheong, Filippou, & Cheong, 2014; Shaffer, Squire, Halverson, & Gee, 2005) by making learning activities more enjoyable and fun. Features implemented within a game context may provide insight on innovative forms of assessment design. Here, we distinguish between using a game itself for a given purpose and using aspects of game design. The idea of "gameful assessment" (also considered "gamified assessment") leverages aspects of games that are germane to the constructs and experiences being assessed. Gameful assessments, for the purposes of this study, utilize game-like features (e.g., feedback systems, student choice, challenge/interest) within a traditional assessment context with the goal of impacting student engagement and overall user experience. This study aims to maintain measurement integrity and features of high-quality formative assessments, and to continue to provide valuable information to teachers and students.

It is advantageous to explore features of games that can be used within an assessment context, and in doing that, the study seeks to better understand teacher and student needs when features of games are included in assessments. Understanding the impacts of gameful assessment on student engagement and teacher perceptions may provide insight on how, when, and what game-like features are important to consider and include in future educational assessments. To study these questions, we use two prototypes of gameful assessments designed to increase formative assessment engagement among students in upper elementary and middle school grades.

The two research questions guiding this study are:

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/ujrt

1. How and in what ways are students engaged in the gameful assessment prototypes?
2. How do teachers currently use and imagine using gameful assessments to impact instruction?

Background

Student Engagement Within Formative Assessment

There is growing awareness that student engagement within an assessment context can impact student outcomes (Swerdzewski et al., 2011; Thelk et al., 2009). In educational assessment, engagement is described as "giving one's best effort to the test, with the goal being to accurately represent what one knows and can do in the content area covered by the test" (Wise & DeMars, 2005, p. 2). Engagement has a strong, positive link with student achievement (Finn, 1989; Miller, Greene, Montalvo, Ravindran, & Nichols, 1996; Saeed & Zyngier, 2012; Smiley & Anderson, 2011; Sundre, 1999; Thelk et al., 2009; Walker, Greene, & Mansell, 2006).

While engagement dynamics are often discussed in the "high-stakes" summative assessment domain, engagement is also critical during "low-stakes" assessment situations, such as classroom and formative assessment (Smiley & Anderson, 2011). During high-stakes situations, students may be aware of associated consequences of outcome performance (Wise & DeMars, 2005) (i.e., placement, college acceptance, graduation); these consequences can increase the perceived importance and the expenditure of effort. However, because formative assessments often do not carry significant or public consequences for students themselves (Great Schools Partnership, 2014), students may not perceive personal benefit in engaging with the text, which can reduce the effort expended (Sundre, 1999; Wise & DeMars, 2005). As a result, low effort and perceived importance may be linked with student engagement on the assessment.

Engagement expended by students during test events may have strong implications for the validity of score inferences (Swerdzewski et al., 2011; Thelk et al., 2009) and therefore may result in scores that are not reflective of students' true ability levels (Swerdzewski et al., 2011). This is a particular concern within the sterile environment of a traditional assessment where there may be little students find exciting, fun, or interesting. As a result, and due to the increased scrutiny of program effectiveness mainly measured by assessment, it is important to understand how design factors that impact student engagement may in turn impact student performance (Wise & Kong, 2005), particularly in a formative assessment context where fun and playfulness may be more relevant and engaging to students. Formative assessments also allow for a lower stakes platform to explore new assessment features and components. In doing this, it is important to explore assessment environments and item designs that may result in a positive link with student engagement.

Gaming and Engagement

The use of digital technology in educational assessments has been examined for decades and is becoming more common due to technological innovations, advanced statistical methods, and the need for the evaluation of more complex skills. Recently, the use of technology to include gamification features has grown in popularity, particularly in education, as a way to increase engagement and motivation. In this context, the term "gamification" refers to the implementation of gaming features in non-game situations (Robson, Plangger, Kietzmann, McCarthy, & Pitt, 2015), such as educational instruction and assessment. These features are responsible for the function of the assessment components and can include points, levels, challenges, trophies, badges/medals, accomplishments, virtual goods, and classification, among others (Seixas, Sandro, & Jos, 2016).

In addition to considering the features of a game within an assessment context, it is also important to include specific game dynamics to help structure the assessment experience and shape the interactions the user has with game mechanics (Zichermann & Cunningham, 2011). The dynamics of a game can be both individually and group based and can include rewards, statuses, accomplishments, self-expression, and competition, to name a few. The inclusion of specific game features and game dynamics can shape the development of a formative assessment experience, as well as the

way students interact with items. This interaction can include shifts in the way students expend effort on the assessment items and in the way students perceive the value of the overall assessment.

The inclusion of game elements within assessment is largely centered on the goal of increasing student motivation and engagement. Although assessment experiences utilizing game features remain largely unexplored, there have been links between instructional game components and outcomes like increased student engagement (Cheong et al., 2014; Shaffer et al., 2005) and student motivation (Banfield & Wilkerson, 2014; De-Marcos, Domínguez, Saenz-De-Navarrete, & Pagés, 2014; Deterding et al., 2011; Papastergiou, 2009; Peng, Lin, Pfeiffer, & Winn, 2012; Seixas et al., 2016). This type of increased engagement is the opposite attitude that most students have toward school and learning (Prensky, 2003).

Beyond engagement and motivation, there are many additional benefits of gaming features within education that have been explored, primarily through a curriculum and instruction lens. Some of these benefits include achievement (Barab et al., 2009; Cheong et al., 2014; Clark, Tanner-Smith, & Killingsworth, 2016; De-Marcos et al., 2014; Dominguez et al., 2013; Hamari, 2015; Hickey, Ingram-Goble, & Jameson, 2009), intrinsic motivation, and self-efficacy (Banfield & Wilkerson, 2014). It is evident that gaming can provide students with the ability to understand complex ideas (Prensky, 2003) and to engage in critical thinking, skills discussed as part of 21st-century learning (Gee, 2007). "Well-designed digital games offer a viable alternative to assessing and developing complex problem solving skills that are needed to succeed in the real world" (Shute, Wang, Greiff, Zhao, & Moore, 2016, p. 106). In doing this, we can better understand how gaming features may have an impact within an educational formative assessment context, specifically on the impact of student engagement and motivation.

Bringing Gamification to Formative Assessment

Current literature has started to connect gamification to education. Only recently, however, have scholars specifically begun to connect gaming elements and assessment. This connection is proposed with the goal of increased student motivation and engagement, as seen by students' expended effort and the perceived value students place on the activity in order to obtain valid outcome scores. Benefits of gaming in education have been explored through the gamification of curricular components such as instruction and intervention support; however, empirical evidence exploring the benefits of gaming within formative assessment remain sparse. Driven by its focus on student feedback and increasing student engagement, this study sits within a cognitive perspective (Greeno, Collins, & Resnick, 1996) and its focus on mental processes related to behavior, particularly within a new assessment practice that include gameful assessment features.

Because the value of formative assessment is derived from how teachers use formative assessment data, it is also important to consider how teachers intend to use data from gameful assessments. An understanding of how teachers view gameful assessments and how they would apply them in practice remains largely absent in this literature. As a result, the current study aims to identify features of gamification within an assessment context that can be attributed with better student engagement and increased relevance for teachers.

Rationale

The worlds of assessment design and game design are not as divergent as they might initially seem. Assessment designers want things similar to those that game designers want: They want to empower students, or players, to fully demonstrate what they know and can do. The traditional forms that assessment designers work in often lack humor, fun, and playfulness. Conversely, games as a genre seem limitless in how they can entice the player to learn incredibly challenging systems. Assessment design is an integrative art. It has long incorporated concepts, tools, and processes from cognitive psychology, learning theory, and content domains. To understand how to motivate students to give their best effort and demonstrate what they know and can do, we might gainfully look to the

world of game design, a field that has mastered the art of user engagement, to create a hybrid form termed gameful assessment.

Gameful assessment is a step toward understanding how to leverage affordances from game design while producing valid and reliable measurement. The emphasis in a gameful assessment is on measurement integrity and sound assessment principles. Teachers and stakeholders need to feel secure that the information provided to them from a gameful assessment is valid, reliable, and actionable. In this paradigm, if there is a tension between the assessment aspect and the game aspect, the assessment aspect drives the decision.

Method

To study the gameful assessment prototypes, we pursued a concurrent mixed methods research design strategy, providing equal priority to quantitative and qualitative data (Creswell & Plano-Clark, 2011). Used together, these methods allowed us to examine gameful assessment within the context of our prototypes, providing students and teachers a point of reference when discussing game-like features, and providing practical application of those features for comparing their effects on student engagement.

Development Process

Development of prototypes began with articulating the assessment's purpose, identifying its theory of action, and developing the evidentiary argument. Proceeding iteratively, we also identified game elements such as the role of the user in the experience, the role of choice or agency in the gaming process, the rewards, goals, and feedback systems to be provided by the game, the game story arc or narrative, and how results should be reported. Each prototype included 36 multiple-choice items, four riddles, and one puzzle. The items for the assessment, which measured elements of spatial reasoning and verbal reasoning and were written to target fourth grade, were developed by content experts, and the content blueprint was equivalent across all three platforms. The content was used as a prototype and was designed independent of school curriculum; however, some school curricula may include spatial and verbal instruction.

These design decisions were provided separately to two game design companies, which pursued two separate development processes. One company produced the "sea" game, which included two-dimensional (2-D) backgrounds and objects with glow and animation effects, the opportunity for student to choose a path, accumulated rewards for correct answers, and a final puzzle based on rewards; a screenshot is shown in Figure 1. We consider this a "gaming lite" experience due to its moderate use of gaming features. The other company produced the "cave" game, which included three-dimensional (3-D) backgrounds and items, earned badges, hidden objects, and a final "boss" to defeat; screenshots are shown in Figures 2 and 3. We consider this a "gaming heavy" experience due to its heavier use of gaming features. Both game prototypes included the same items, and students were provided a report of their results at the conclusion of each game. Notably, the study included a "control group assessment," which used the same items ported on a computer without enhancements, animation, or interactivity.

Participants

Participants for this study include a convenience sample (overall $n = 391$) of students in Grades 4–7 from one K–12 public school in the northwestern United States ($n = 77$) and one Grades 3–5 public school in the midwestern United States ($n = 314$). The sample included 170 fourth graders, 161 fifth graders, 19 sixth graders, and 41 seventh graders. The sample included 29.9% participants identifying as White, with a majority of the participants identifying as non-White (70.8%), including 14.8% Black or African American, 20.2% Hispanic or Latino, 21.0% two or more races, 5.6% Asian, 0.3% Hawaiian or Pacific Islander, 2.3% Native American/Alaskan Native, and 6.6% other. Demographic data for the sample are displayed in Table 1.

Figure 1. Screenshot of the "sea"-themed game, showing the underwater theme where students can select their path.

The schools were identified as an appropriate research site for the program because of their announced commitment to the use of data in instructional improvement and their integration of technology in instruction. In particular, the Midwestern school's 1:1 student-tablet program and use of several curricular games provided a context in which technology use in instruction was understood and routine.

Gameful assessments were administered in the school computer lab on identical computers. Students were randomly assigned to either the cave-themed gameful assessment, the sea-themed gameful assessment, or a control assessment containing the same items without gameful assessment

Figure 2. Screenshot of the "cave"-themed game, showing the cavern with illuminated symbols, words, and pathway.

Figure 3. Screenshot of the "cave"-themed game, showing the final boss of the game, called the Boggler. The Boggler takes the player through riddles that are solved with words on the cave wall as the player progresses through the game.

Table 1. Demographic Data, $N = 391$

Variable	Total
Female	205 (52.4)
Male	178 (45.5)
Unidentified	8 (0.02)
White	114 (29.2)
Non-White	277 (70.8)

Note. Values are *n* (%).

features. After taking the assessments, students participating in the cave- or sea-themed game completed a survey on the game features.

Additionally, students in all three groups (sea-themed game, cave-themed game, control) completed a cognitive engagement survey (Cognitive Engagement Scale—Short—Deep, Shallow, Persistence [CE-S-DSP] & Student Opinion Scale [SOS])—a hybrid of previously used self-reporting tools (Greene & Miller, 1996; Smiley & Anderson, 2011; Sundre, 1997) measuring cognitive engagement across five subscales (deep processing, shallow processing, persistence, importance, and effort) using 21 Likert items (Guerreiro, 2017). On the cognitive engagement measure, the total score was 94. The reliability based on this data set for the overall instrument (Guerreiro, 2017) is $\alpha = 0.84$, which is above the preferred threshold of 0.80 and well above the minimally acceptable threshold of 0.70 (Tavakol & Dennick, 2011). Variance of student scores on the CE-S-DSP & SOS is reported as 78.08 with a standard deviation of 8.84 and a mean of 57.57 (Guerreiro, 2017). This indicates that there were fluctuations in students' self-report of cognitive engagement, which suggests there is enough variance in the measure.

Teachers from the Midwestern school were selected to participate in interviews; teachers from the Northwestern school were unavailable for interviews. Among 15 participating classes, 14 teachers agreed to participate. Teachers completed a consent form and were provided small compensation for their participation. Interviews occurred in private locations throughout the school, and included video walkthroughs of both gameful assessments and comments on the features of both games. Before and after viewing videos, teachers all discussed their general use of games in the classroom and the features they would desire in a gameful assessment system.

Analysis

Results from the engagement survey were analyzed using an analysis of variance (ANOVA) to assess the differences in engagement between platforms. On the user survey, *t*-tests were used to assess differences in student opinion between the two gameful assessments, and one-way ANOVAs were used to understand students' sense of relative item difficulty and overall preference of gaming features between all three platforms.

Qualitative interview data was fully transcribed and coded. The interviewing researcher and two outside researchers separately coded transcripts, using NVivo and evaluation coding techniques in a first round and structural coding in a second round (Saldaña, 2016). These separate coded transcripts were reconciled with each other in sessions including all three researchers, using procedures described in Cornish et al. (2013). In these sessions, researchers used techniques for generating meaning described by Miles, Huberman, and Saldaña (2014), such as noting patterns or themes, subsuming particulars into the general, and making metaphors. These general interpretations were further derived and tested using validity procedures described by Onwuegbuzie and Leech (2005), including triangulation, weighing the evidence, checking the meaning of outliers, and assessing rival explanations. Together, this corpus was reconciled into one master set of coded interviews and interpretations from which qualitative findings were written.

Limitations

This study is primarily limited by the nature of the schools in which it was conducted. For relatively technologically literate students and teachers, gameful assessments may be more accessible and

usable than for other, less technologically literate students and teachers. Self-report measures have a tendency to present differences between groups; students of marginalized groups (e.g., racial/ethnic minorities, language minorities, females) may self-report lower than peers from majority groups, which may have impacted self-report scores of engagement. Methodologically, the analysis does not account for the nesting of data between grades or classrooms. Additionally, no single approach to measurement of a construct is considered universally acceptable; therefore, it is possible researchers will select different behaviors to measure the same construct (Khairani & Razak, 2013), in this case, constructs of cognitive engagement.

This study can speak only to the use of gameful assessment elements within the context of the two gameful assessment prototypes examined here; other implementations of these principles may produce different results. However, the study's design provides the opportunity to examine gameful assessment principles in practice, in a school selected for its interest in and engagement with the use of technology in learning and assessment.

Results

This section summarizes results from presentation of both gameful assessment prototypes to both students and teachers, in comparison to the control group. It describes the results of quantitative student data and qualitative data collected from teachers, with quantitative data answering Research Question 1 and qualitative data answering Research Question 2. The results of both data streams are discussed together in the article's conclusion.

Student Engagement

At the completion of the sea and cave assessments, students were asked questions about the game and item components using a 4-point Likert scale ranging from *strongly disagree* to *strongly agree*. Overall, students had higher enjoyment across the game and item components of the cave-themed gameful assessment. A *t*-test was used to examine mean score differences of item components between the two games; see Table 2. Out of 12 questions, all but two were significantly different between the two platforms, with the cave theme showing higher favorability for questions asking about the games' path, level of interest, shapes and lines, final puzzle, reading passages, sounds, the use of a progress bar, and the final report.

Students were also asked questions about item difficulty between the three platforms (sea, cave, control) using a 5-point Likert scale ranging from very difficult to very easy. Overall, students who took the sea game found the items easier in comparison to the cave assessment and the control assessment. A *t*-test was used to examine mean score differences of reported item difficulty between gaming platforms (see Table 3), as well as between gaming and the control assessment (see Table 4). Out of seven questions, only one item was significantly different, with students

Table 2. User *t*-Test Results of Component for Gaming Platform Items (Sea Theme and Cave Theme)

Item	*t*	*df*
Item 1	−.64	251
Item 2	−2.34*	251
Item 3	−5.20***	251
Item 4	−2.10*	251
Item 5	−1.70	251
Item 6	−5.65***	251
Item 7	−2.41*	222.83
Item 8	−3.16**	251
Item 9	−4.56***	251
Item 10	−6.77***	251
Item 11	−2.41*	251
Item 12	−2.69**	251

Note. Item 7 violated the assumption of homogeneity of variance and therefore equal variances are not assumed.
*$p < 0.05$. **$p < 0.01$. ***$p < .0.001$.

Table 3. User Survey *t*-Test Results of Difficulty Level for Gaming Platform Items (Sea and Cave)

Item	*t*	*df*
Item 1	1.44	192
Item 2	−4.31***	188.17
Item 3	−1.30	186
Item 4	1.61	210
Item 5	1.33	214.18
Item 6	−.34	167
Item 7	.43	169

Note. Items 2 and 5 violated the assumption of homogeneity of variance and therefore equal variances are not assumed.
*$p < 0.05$. **$p < 0.01$. ***$p < .0.001$.

participating in the cave assessment reporting the final puzzle more difficult. A one-way ANOVA was used to examine mean score differences of reported item difficulty between all three platforms; results are provided Table 5. Again, this analysis shows a significant difference in the difficulty of the final puzzle between platforms, $F(2,244) = 3.37$, $p = 0.036$.

Platform Description

Mean scores were found for engagement on the cave-themed ($M = 56.63$, $SD = 6.95$), sea-themed ($M = 55.10$, $SD = 7.63$), and control ($M = 55.72$, $SD = 7.03$) versions. Descriptive statistics for the platforms are displayed in Table 6.

Cognitive Engagement

A between-subjects main effects ANOVA evaluated gaming platform on student cognitive engagement. The main effect of type of platform on cognitive engagement was not significant. Results are presented in Table 7. There was not a significant difference between self-report of cognitive engagement for students who completed the cave-themed gameful assessment ($M = 56.63$), students who completed the sea-themed gameful assessment ($M = 55.10$), and students who completed the control assessment ($M = 55.72$).

Table 4. User Survey *t*-Test Results of Difficulty Level for Gaming Platform Items (Sea and Cave) in Comparison to the TAP

Item	*t*	*df*
Item 1	−1.25	183.75
Item 3	.82	262
Item 4	.59	304
Item 5	.50	316
Item 6	.14	251
Item 7	.14	243

Note. Item 2 is missing because it is not included in the TAP. Item 1 violated the assumption of homogeneity of variance and therefore equal variances are not assumed.
*$p < 0.05$. **$p < 0.01$. ***$p < .0.001$.

Table 5. One-Way Analysis of Variance Summary Table for the Effect of Gaming Platform on Student Rated Item Difficulty

Source	*df*	*SS*	*MS*	*F*	*p*
Item 1	2, 277	2.05	1.03	1.83	0.163
Item 3	2, 263	1.57	0.78	0.99	0.375
Item 4	2, 305	2.01	1.00	1.43	0.240
Item 5	2, 317	3.93	1.96	1.33	0.267
Item 6	2, 252	0.30	0.15	0.21	0.810
Item 7	2, 244	4.77	2.39	3.37	0.036*

*$p < 0.05$.

Table 6. Descriptive Statistics of Gaming Platform

Platform	*n*	*M (SE)*	*SD*	Skew	Kurtosis	Shapiro–Wilk (*df*)
TAP	108	55.72 (0.68)	7.03	−0.52	−0.31	.966 (108)*
Cave	138	55.63 (0.59)	6.95	−0.34	1.04	.977 (138)*
Sea	145	55.10 (0.63)	7.63	−0.70	0.88	.969 (145)*

Note. *df* = degrees of freedom.
*$p < .05$, ** $p < .01$.

Table 7. One-Way ANOVA Summary Table for the Effect of Gaming Platform on Cognitive Engagement

Source	*df*	*SS*	*MS*	*F*	*p*
Gaming platform	2	167.53	83.76	1.60	.20
Error	388	20,292.47	52.30		
Total	390	20,459.99			

An additional between-subjects main effects ANOVA further evaluated gameful assessment on student cognitive engagement, specifically the impact of a game on cognitive engagement, by comparing the cave and sea games together with the control group. The main effect of a gameful assessment on cognitive engagement was not significant. There was not a significant difference between self-report of cognitive engagement for students who completed the assessment on a gaming platform ($M = 55.85$) and students who completed the control assessment ($M = 55.72$). Results are presented in Table 8.

The effect of platform was further investigated by evaluating the two extreme platforms (the gaming-heavy cave assessment in comparison to the control assessment) on cognitive engagement. Results from the one-way, between-subjects ANOVA (see Table 9) indicate a nonsignificant difference between self-report of cognitive engagement for students who completed the assessment on the gaming-heavy cave theme assessment ($M = 56.63$) and the control assessment ($M = 55.72$).

The additional factors of sex and race/ethnicity were added to explore the impact of gaming platform on cognitive engagement. Results from the three-way, between-subjects ANOVA indicated a nonsignificant main effect of gaming platform on cognitive engagement; these results are provided in Table 10. There was not a significant difference between self-report of cognitive engagement for students who completed the cave-themed assessment ($M = 55.31$), students who completed the sea-themed assessment ($M = 57.41$), and students who completed the control assessment ($M = 56.52$). There was a significant main effect of race/ethnicity on cognitive engagement, $F(7, 344) = 2.29$, $p = .03$, $\eta^2_{\text{partial}} = .044$. Lastly, there was also a nonsignificant main effect of sex on cognitive engagement. Female students did not have significant cognitive engagement outcomes ($M = 57.44$) compared to male students ($M = 56.12$) and students reported as neither male nor female ($M = 52.90$). In addition to the main effects, the interaction

Table 8. One-Way ANOVA Summary Table for the Effect of a Gameful Assessment on Cognitive Engagement

Source	*df*	*SS*	*MS*	*F*	*p*
Between groups	1	1.17	1.17	0.02	.88
Within groups	389	20,458.83	52.59		
Total	390	20,459.99			

Table 9. One-Way ANOVA Summary Table for the Effect of Extreme Platform (Cave Theme Versus Control Assessment) on Cognitive Engagement

Source	*df*	*SS*	*MS*	*F*	*p*
Between groups	1	49.97	49.97	1.02	.31
Within groups	244	11,903.82	48.79		
Total	245	11,953.79			

Table 10. Three-Way ANOVA Summary Table for the Effect of Gaming Platform, Sex, and Race/Ethnicity on Cognitive Engagement

Source	*df*	*SS*	*MS*	*F*	*p*
Gaming platform	2	57.95	28.97	0.58	.56
Race/ethnicity	7	796.87	113.84	2.29	.03*
Sex	2	127.21	63.61	1.28	.28
Gaming platform × Race/ethnicity	12	773.37	64.45	1.30	.22
Gaming platform × Sex	3	51.32	17.11	.34	.79
Race/ethnicity × Sex	8	399.87	49.98	1.00	.43
Gaming platform × Race/ethnicity × Sex	11	623.15	56.65	1.14	.33
Error	344	17,121.55	49.77		
Total	390	1,238,362.00			

Note. Platform × Race/ethnicity gives the results of the interaction effect between variables.
*$p < .05$.

effects were also nonsignificant, as shown in Table 10. This indicates that effects on cognitive engagement were the same regardless of gaming platform, sex, and/or race/ethnicity.

Use of Gameful Assessments in Instruction

Teachers focused on a fundamental tension within the design of games in assessment: balancing the intrinsic motivation of a student's desire to learn and grow with the extrinsic motivations provided by the new stimuli games offer. While they are intricately tied to the notion of formative assessments as a part of learning, teachers also recognized the potential benefits and drawbacks of gaming elements for measurement; teachers disagreed, at times with themselves, on whether gaming in assessment improved the accuracy of assessments by improving motivation and engagement, or whether gaming might decrease the accuracy of assessments by serving as a major distraction.

This section identifies four major themes. First, teachers disagreed regarding the impact of gameful assessment on intrinsic motivation. Second, some teachers identified games as potential enhancements for student persistence. Third, teachers discussed the role of what they called "serious" assessments, and how gameful assessments departed from that role. Finally, teachers identified and acknowledged the opportunity for gameful assessments to serve as opportunities for students to learn.

Intrinsic Motivation. Substantial belief among teachers that the use of games in the classroom improved student motivation was tempered for some teachers by questions regarding the ultimate source of that motivation. All participants cited several features of games they had observed motivating students, including the use of reward systems, the ability to move and manipulate a character, the student's ability to choose a path, elements of a journey or quest for students, and increased use of colors and sound. These elements were mostly recognized by teachers based on their experience with using other technology-based games in the classroom; in addition, a few teachers acknowledged that these elements were motivating because they were similar to students' other experiences, like games played at home.

However, participants also focused on the origin of increased motivation and whether motivation was linked to learning and growth. Some participants felt the reward mechanisms enhanced intrinsic motivation by tying rewards to learning. Others thought they were a distraction from the intrinsic motivation to learn they felt was promoted by other assessments or classroom activities. Like other participants, Amy identified intrinsic rewards as more important:

> I'm trying to teach them intrinsic rewards, and feeling proud of yourself, and looking at what you've accomplished and seeing how much you've grown. I'm trying to stay away from "Here is a prize, you did well" rewards.

Participant Corinne in particular identified concern around how the use of games positioned the act of learning:

> In my mind, I'm just wondering if—this is the old-fashioned me talking—we're giving kids the message that the learning, in and of itself, is not exciting enough and we have to jazz it up. Because some kids get immense satisfaction of taking the test and doing a good job and getting their score and that's that.

In this instance, Corinne directly links the seriousness of an activity with the intrinsic motivation it is believed to promote; games by design appear to her to run contrary to that educational model.

For teachers in this mind set, substantial concern was expressed that games might create rewards for students perceived as separate from what they learned, ultimately serving as a deterrent to the assessment's purpose.

Persistence. Alongside increased motivation, participants also recognized the opportunity for games to allow students to better persist through complicated academic material. Several participants identified gaming as an opportunity to reduce the frustration associated with an assessment environment or with a complex academic task: "If you're playing a game, you could die 20 times and you don't get mad" (Jane). For students easily overwhelmed by content, games "would get them to want to keep going" (Kellee) or provide opportunities for redemption (Pamela). Similarly, participant Danielle suggested that the integration of topic areas in games circumvented thought processes where students identify themselves as poor in particular subjects, preventing early sources of frustration.

Other teachers, however, expressed some concern that elements of games could overwhelm students to a greater degree than traditional assessments. While at least one teacher identified competitiveness as a motivator created by games, participant Carol noted concern that providing students their results or a class "ranking" could demotivate students struggling with the content. Participant Carla noted that the intermixed content present in the gameful assessment prototypes was different from how content is traditionally learned in classrooms, and that this could prove overwhelming for students.

"Serious" Assessment. Third, teachers identified the need for certain classroom activities, particularly assessments, to represent "serious" activities, and argued that gaming actively worked against that end. For some participants, entertaining games were conceptualized as barriers to classroom behavior management: "Kids can take advantage of the gaming aspect. They get excited; they start socializing" (Joy). Participant Jennifer noted this concern extended to classroom rule-making: "In fourth grade, we are very strict on no fun games, everything has to be something educational." In a game, argued Amy, students feel "'I can kind of do whatever.' Where, if it's an assessment, they're like 'Oh, I'm graded on this, I need to try my best.'"

Similarly, some participants also expressed concern about games serving as a distraction for students. As participant Tasha described, "There's too much play at times." However, even these teachers acknowledged potential benefits in providing students alternatives to the "monotony" (Alice) of traditional assessment. Participant Pamela described the use of distraction in games as a learning aid to, for example, help motivate students through difficult or lengthy reading passages.

These participants, and others, underscored the importance of rule and expectation setting in averting these potential downsides. Others, like participant Carla, also emphasized the importance of communicating the purpose of an assessment to underscore its seriousness. For Corinne, however, serious assessments are ultimately necessary to accurately represent how students will learn and work in future grades. "When they leave us, not everything is going to be bells and whistles . . . a test should feel like a test" (Corinne).

The notion of a "serious assessment" is ultimately inseparable from both motivation and persistence. For proponents of "serious assessment," traditional assessments motivate students in the correct ways (e.g., learning or achieving for its own sake) and reward students who persist in that effort. By contrast, the "fun assessment" offered by games may promote misbehavior, distance classroom activities from learning, or even produce less accurate assessment results by minimizing student effort.

Opportunity for Learning. Finally, teachers linked the use of gameful formative assessments with opportunities for students to learn. Three major elements of the prototypes were identified with opportunities to learn. First, participant Tasha recognized the importance of self-pacing in the assessment, noting that students "don't move on until they are ready." Second, several participants noted the opportunity games provide for more complex assessment items that could sample a variety of learning. In this view, games provide the ability to do more than "just answer questions" (Bonnie) by presenting items involving higher level thinking (Corinne), promoting thinking "out of the box" (Alice), or engaging in metacognition (Carla). Games, Carla noted, involve elements of strategy, requiring students to think multiple moves ahead; she connected this process with several forms of multistep thinking in mathematics. Finally, nearly all participants acknowledged the importance of immediate feedback on student performance, and requested more time and information for students to review their individual answers and correct those answers they got wrong.

Discussion

Our study asks two main questions:

1. How and in what ways are students engaged in the gameful assessment prototypes?
2. How do teachers currently use and imagine using gameful assessments to impact instruction?

At the heart of both questions is whether, in participant Corinne's terms, "a test should feel like a test." Although results on engagement were not significant, results on student satisfaction suggest the ability to enhance engagement by focusing work on the elements students found most appealing, particularly the act of journeying through a game, the use of sound, and frequent reporting on student progress. This list parallels the list of features cited most frequently by our teacher participants: In many ways, they make gameful assessment truly gameful. In doing this, our work joins the focus of the gameful assessment design literature on the inclusion of appropriate gaming features (Robson et al., 2015; Seixas et al., 2016) that help ensure appropriate and well-designed games (Shute et al., 2016). These design considerations also help to promote specific mechanics and features that impact the interactions students have with the game (Zichermann & Cunningham, 2011).

However, our qualitative findings suggest the need to tread carefully when introducing new assessment designs based on gameful principles. Teachers brought a host of existing conceptions regarding using games in the classroom to their viewing the same prototype assessments, leading to very different and sometimes competing conceptions. Participants often recognized divided thought within themselves, between "old-fashioned me" with traditional views of assessment and learning, and a more pragmatic acknowledgment that using games in the classroom had clear positive effects on students.

Our work demonstrates that gameful assessments can indeed measure achievement (Barab et al., 2009; Cheong et al., 2014; Clark et al., 2016; De-Marcos et al., 2014; Dominguez et al., 2013; Hamari, 2015; Hickey et al., 2009; Swerdzewski et al., 2011; Thelk et al., 2009) while providing a different experience for students; it leaves unanswered whether those assessments can overcome ideas of what tests are and should be. Future research should continue to explore this conflicted identity, seeking new and unique forms of gameful assessment that are more engaging for students without losing instructional relevance for teachers. This may include gathering additional qualitative student feedback or using student focus groups.

Interest in the area of gameful assessment is driven by the intersection of the curricular and the accountability policies of contemporary schooling; simultaneously, assessments are being encouraged to include technology enhancements in order to reduce the negative impacts of low student motivation (Guerreiro, 2017; Wise & Kong, 2005) and increase engagement (Cheong et al., 2014; Shaffer et al., 2005). Both sets of policies come with legitimate desires to make schooling itself more engaging for students. Gameful assessments purport to address both needs simultaneously; however, they do not address them in the same ways and to the same extent. Gameful assessments

may provide a new approach to measurement, including developing twenty-first-century knowledge and skills (Gee, 2007), navigating virtual environments, understanding complex ideas (Prensky, 2003), and leveraging technology and features to promote engagement. Our study underlines the need for continued iteration and improvement in the gameful assessment arena to produce the best possible combination of motivation, higher order content, and complex assessment in order to maximize the relevance of new assessments for all students.

Received: 24 July 2017
Revised 30 November 2017
Accepted 5 December 2017

Author Notes

Meg Guerreiro is a Research Scientist with NWEA in Portland, Oregon, U.S.A, where she works with prototyping technology initiatives in educational assessment. Her research interests include mixed methodology within the context of e-learning and educational measurement and assessment, the impact of STEM initiatives on K-12 teaching and learning, and self-efficacy and cognitive engagement for marginalized populations. Her primary work includes performance-based tasks, measurement of non-academic skills, and the use of innovative technology within assessment. Please address correspondence to Meg Guerreiro, NWEA, 121 NW Everett St, Portland, OR 97209, USA. E-mail: meg.guerreiro@nwea.org

Chase Nordengren is a Research Scientist with NWEA in Portland, Oregon, U.S.A, where he directs qualitative and mixed methods research at the intersection of assessment, curriculum and instruction. His work includes primary research, program evaluation, and methodological design on NWEA assessments, related services, and assessments in active development.

ORCID

Meg Guerreiro http://orcid.org/0000-0003-4722-6376

References

Banfield, J., & Wilkerson, B. (2014). Increasing student intrinsic motivation and self-efficacy through gamification pedagogy. *Contemporary Issues in Education Research*, *7*, 291–298. doi:10.19030/cier.v7i4.8843

Barab, S. A., Scott, B., Siyahhan, S., Goldstone, R., Ingram-Goble, A., Zuiker, S. J., & Warren, S. (2009). Transformational play as a curricular scaffold: Using videogames to support science education. *Journal of Science Education and Technology*, *18*(4), 305–320. doi:10.1007/s10956-009-9171-5 doi:10.1007/s10956-009-9171-5

Cheong, C., Flilippou, J., & Cheong, F. (2014). Towards the gamification of learning: Investigating student perceptions of game elements. *Journal of Information Systems Education*, *25*(3), 233–245.

Clark, D. B., Tanner-Smith, E. E., & Killingsworth, S. S. (2016). Digital games, design, and learning: A systematic review and meta-analysis. *Review of Educational Research*, *86*(1), 79–122. doi:10.3102/0034654315582065 doi:10.3102/0034654315582065

Cornish, F., Gillespie, A., Zittoun, T., Flick, U., Cornish, F., Gillespie, A., & Zittoun, T. (2013). Collaborative analysis of qualitative data. In U. Flick (Eds.), *The SAGE handbook of qualitative data analysis* (pp. 79–93). Thousand Oaks, CA: Sage. Retrieved from http://srmo.sagepub.com/view/the-sage-handbook-of-qualitative-data-analysis/n6.xml

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

De-Marcos, L., Domínguez, A., Saenz-De-Navarrete, J., & Pagés, C. (2014). An empirical study comparing gamification and social networking on e-learning. *Computers and Education*, *75*, 82–91. doi:10.1016/j.compedu.2014.01.012 doi:10.1016/j.compedu.2014.01.012

Deterding, S., Dixon, D., Khaled, R., Nacke, L., Sicart, M., & O'Hara, K. (2011). Gamification: Using game design elements in non-game contexts. Proceedings of the 2011 Annual Conference Extended Abstracts on Human Factors in Computing Systems (CHI 2011), 2425–2428. doi:10.1145/1979742.1979575

Dominguez, A., Saenz-De-Navarrete, J., De-Marcos, L., Fernanndez-Sanz, L., Pages, C., & Martinez-Herraiz, J. J. (2013). Gamifying learning experiences: Practical implications and outcomes. *Computers and Education*, *63*, 380–392. doi:10.1016/j.compedu.2012.12.020 doi:10.1016/j.compedu.2012.12.020

Finn, J. D. (1989). Withdrawing from school. *Review of Educational Research*, *59*(2), 117–142. doi:10.3102/00346543059002117 doi:10.3102/00346543059002117

Gee, J. P. (2007). *Good Video Games and Good Learning*. Bern, Switzerland: Peter Lang Publishing.

Great Schools Partnership. (2014). High-stakes test. In *The glossary of education reform*. Retrieved from http://edglossary.org/high-stakes-testing

Greene, B. A., & Miller, R. B. (1996). Influences on achievement: Goals, perceived ability, and cognitive engagement. *Contemporary Educational Psychology*, *21*, 181–192. doi:10.1006/ceps.1996.0015

Greeno, J., Collins, A., & Resnick, L. (1996). Cognition and learning In D. Berliner, & R. Calfee (Eds.), *Handbook of educational psychology* (pp. 15–46). New York, NY: Macmillan.

Guerreiro, M. (2017). The impact of a technology-enhanced math performance task on student cognitive engagement in mathematics (Unpublished doctoral dissertation).

Hamari, J. (2015). Do badges increase user activity? A field experiment on the effects of gamification. *Computers in Human Behavior*, *71*, 469–478. doi:10.1016/j.chb.2015.03.036

Hickey, D. T., Ingram-Goble, A. A., & Jameson, E. M. (2009). Designing assessments and assessing designs in virtual educational environments. *Journal of Science Education and Technology*, *18*(2), 187–208. doi:10.1007/s10956-008-9143-1 doi:10.1007/s10956-008-9143-1

Khairani, A. Z. Bin, & Razak, N. B. A. (2013). Advance in educational measurement: Rasch model analysis of mathematics proficiency test. *International Journal of Social Science and Humanity*, *2*(3), 248–251. doi:10.7763/IJSSH.2012.V2.104 doi:10.7763/IJSSH.2012.V2.104

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). Los Angeles, CA: Sage.

Miller, R. B., Greene, B. A., Montalvo, G. P., Ravindran, B., & Nichols, J. D. (1996). Engagement in academic work: The role of learning goals, future consequences, pleasing others, and perceived ability. *Contemporary Educational Psychology*, *21*(4), 388–422. doi:10.1006/ceps.1996.0028 doi:10.1006/ceps.1996.0028

Onwuegbuzie, A. J., & Leech, N. L. (2005). On becoming a pragmatic researcher: The importance of combining quantitative and qualitative research methodologies. *International Journal of Social Research Methodology*, *8*(5), 375–387. doi:10.1080/13645570500402447

Papastergiou, M. (2009). Digital game-based learning in high school computer science education: Impact on educational effectiveness and student motivation. *Computers & Education*, *52*(1), 1–12. doi:10.1016/j.compedu.2008.06.004 doi:10.1016/j.compedu.2008.06.004

Peng, W., Lin, J., Pfeiffer, K. A., & Winn, B. (2012). Need satisfaction supportive game features as motivational determinants: An experimental study of a self-determination theory guided exergame. *Media Psychology*, *15*, 175–196. doi:10.1080/15213269.2012.673850 doi:10.1080/15213269.2012.673850

Prensky, M. (2003). Digital game-based learning. *Computers in Entertainment*, *1*(1), 21. doi:10.1145/950566.950596 doi:10.1145/950566.950596

Robson, K., Plangger, K., Kietzmann, J. H., McCarthy, I., & Pitt, L. (2015). Is it all a game? Understanding the principles of gamification. *Business Horizons*, *58*(4), 411–420. doi:10.1016/j.bushor.2015.03.006 doi:10.1016/j.bushor.2015.03.006

Saeed, S., & Zyngier, D. (2012). How motivation influences student engagement: A qualitative case study. *Journal of Education and Learning*, *1*(2), 252–267. doi:10.5539/jel.v1n2p252 doi:10.5539/jel.v1n2p252

Saldaña, J. (2016). *The coding manual for qualitative researchers* (3rd ed.). Thousand Oaks, CA: Sage.

Seixas, R., Sandro, A., & Jos, I. (2016). Computers in human behavior effectiveness of gamification in the engagement of students. *Computer in Human Behavior*, *58*, 48–63. doi:10.1016/j.chb.2015.11.021 doi:10.1016/j.chb.2015.11.021

Shute, V. J., Wang, L., Greiff, S., Zhao, W., & Moore, G. (2016). Measuring problem solving skills via stealth assessment in an engaging video game. *Computers in Human Behavior*, *63*, 106–117. doi:10.1016/j.chb.2016.05.047 doi:10.1016/j.chb.2016.05.047

Shaffer, D. W., Squire, K. R., Halverson, R., & Gee, J. P. (2005). Video games and the future of learning. *Phi Delta Kappan*, *87*(2), 104–111. doi:10.1002/piq.20020 doi:10.1177/003172170508700205

Smiley, W., & Anderson, R. (2011). Measuring students' cognitive engagement on assessment tests: A confirmatory factor analysis of the short form of the Cognitive Engagement Scale. *Research & Practice in Assessment*, *6*, 12.

Sundre, D. L. (1997). Differential examinee motivation and validity: A dangerous combination. *Annual Meeting of the American Educational Research Association*. Retrieved from http://works.bepress.com/donna_l_sundre/193/

Sundre, D. L. (1999). Does examinee motivation moderate the relationship between test consequences and test performance? *Annual Meeting of the American Educational Research Association*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.890.3967&rep=rep1&type=pdf

Swerdzewski, P. J., Harmes, J. C., & Finney, S. J. (2011). Two approaches for identifying low-motivated students in a low-stakes assessment context. *Applied Measurement in Education*, *24*(2), 162–188. doi:10.1080/08957347.2011.555217 doi:10.1080/08957347.2011.555217

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53–55. doi:10.5116/ijme.4dfb.8dfd doi:10.5116/ijme.4dfb.8dfd

Thelk, A. D., Sundre, D. L., Horst, S. J., & Finney, S. J. (2009). Examining inferences about test-taking motivation: The Student Opinion Scale (SOS). Northeastern Educational Research Association. Retrieved from http://works.bepress.com/donna_l_sundre/93/

Walker, C. O., Greene, B. A., & Mansell, R. A. (2006). Identification with academics, intrinsic/extrinsic motivation, and self-efficacy as predictors of cognitive engagement. *Learning and Individual Differences*, *16*(1), 1–12. doi:10.1016/j.lindif.2005.06.004 doi:10.1016/j.lindif.2005.06.004

Wise, S. L., & DeMars, C. E. (2005). Low examinee effort in low-stakes assessment: Problems and potential solutions. *Educational Assessment*, *10*(1), 1–17. doi:10.1207/s15326977ea1001_1 doi:10.1207/s15326977ea1001_1

Wise, S. L., & Kong, X. (2005). Response time effort: A new measure of examinee motivation inf computer-based tests. *Applied Measurement in Education*, *18*(2), 163–183. doi:10.1207/s15324818ame1802_2 doi:10.1207/s15324818ame1802_2

Zichermann, G., & Cunningham, C. (2011). *Gamification by design: Implementing game mechanics in Web and mobile apps*. Sebastopol, CA: O'Reilly Media.