Journal of Special Education Technology
2020, Vol. 35(4) 191-203

DOI: 10.1177/0162643419856272
journals.sagepub.com/home/jst
SAGE

Students' Perceptions of a Gamified Reading Assessment

Deborah K. Reed, PhD[1], Emily Martin, MSW[1], Eliot Hazeltine, PhD[2], and Bob McMurray, PhD[2]

Abstract

To inform the development of gamified assessments, this study explored how students with or at risk for reading difficulties in Grades 6–8 ($N = 202$) perceived and interacted with a decoding assessment designed with gamification characteristics. Three data sources enhanced the methodological triangulation: observations and scores from testing, surveys of students' perceptions, and focus group discussions with a stratified random sample of students ($n = 25$). Findings suggest students became immersed in the gamified reading assessment and were motivated by tasks that were challenging but not frustratingly difficult. However, they were dissatisfied with some design features and reported focusing on identifying patterns and gaming strategies rather than on the reading skills being assessed. This suggests students' expectations of gamified assessments might contribute construct irrelevant variance to the instruments.

Keywords

gamification, reading assessment, middle school

The development of gamified pedagogical and assessment tools has been encouraged by the Federation of American Scientists ([FAS], 2006) and others (e.g., Hines, Jasny, & Mervis, 2009; Watson, Mong, & Harris, 2011), as a way to increase students' engagement and time on task. This may be particularly important for adolescents with and at risk for reading disabilities who have experienced repeated struggles and tend to adopt a passive role during instruction (Zimmerman & Schunk, 2006). According to Morford, Witts, Killingsworth, and Alavosius (2014), elements of game playing (e.g., a player's direct impact on the game's results, clear end goals, rules for play, and development of strategies to complete the tasks) can serve to motivate students in three ways. First, these gaming elements can be socially motivating by offering opportunities for collaboration with team members or competition with other players. Second, academic games are emotionally motivating when students earn rewards and receive feedback because they become immersed in the game's tasks. Finally, to instantiate these properties, the game must be sufficiently challenging to engage students without frustrating them.

A game's motivational components can affect its educational and entertainment value differently. For example, competition among players was found to be less effective than noncompetitive configurations because the requirement that someone loses can reduce students' self-efficacy (Clark, Tanner-Smith, & Killingsworth, 2016). Test takers in previous research admitted that having a low position on a leaderboard made them lose motivation and want to give up on a gamified assessment (Ferrell, Carpenter, Vaughn, Dudely, & Goodman, 2015; Kocadere & Caglar, 2015). Middle school students in special education were observed becoming frustrated and demotivated when they did not attain the scores they wanted to earn (Ke & Abras, 2013). Similarly, incorporating points or rewards in a computerized math assessment was found to increase middle schoolers' satisfaction but have no discernable effect on accuracy (Attali & Arieli-Attali, 2015).

Because computer games used for educational purposes may engender unanticipated expectations or reactions among students, the present study investigated what middle school students (over half of whom were served in special education) thought about a gamified assessment of reading and how those perceptions altered the approaches they were taking to complete the tasks. The assessment used in the present study was designed without the competitive elements of points and leaderboards but with elements conceptualized as contributing to internal motivation such as choice and positive reinforcement.

[1] Iowa Reading Research Center, University of Iowa, Iowa City, IA, USA
[2] Department of Psychological and Brain Sciences, University of Iowa, Iowa, Iowa City, IA, USA

Corresponding Author:
Deborah K. Reed, PhD, Iowa Reading Research Center, University of Iowa, 103 Lindquist Center, Iowa City, IA 52242, USA.
Email: deborah-reed@uiowa.edu

Students' Expectations of Academic Games

The increasing use of computer games at school (Gray, Thomas, & Lewis, 2010) and in special education specifically (Bouck & Flanagan, 2009) risks conditioning students to focus on developing strategies for conquering the digital format rather than demonstrating the targeted academic skills. The overuse of extrinsic rewards may contribute to this (Abramovich, Schunn, & Higashi, 2013), especially the use of participatory rewards (e.g., those earned for completing a specific number of tasks; Deci, Koestner, & Ryan, 2001). Thus, the availability of rewards not related to the educational goal may lead students to become preoccupied with finding patterns or shortcuts in the games that would improve their scores, regardless of whether the patterns were relevant to the academic content.

The combination of gamified pedagogical and assessment tools also may blur the distinction between learning tasks and tasks intended to measure students' academic performance. In fact, a systematic review of gamified assessments found that many included training games with the same types of tasks as appeared on the test (Lumsden, Edwards, Lawrence, Coyle, & Munafo, 2016). The perceived purpose of the gamified tasks (i.e., to learn or to assess skills) could alter students' approaches and, thus, their scores or teachers' subsequent interpretations of students' abilities.

Moreover, it is possible that gaming elements provide a performance boost to some students by lessening the pressure placed on certain skills or abilities, such as by increasing attention during a cognitive test intended to detect attention deficits (Lumsden et al., 2016). The primary purpose for adding game elements is to better engage the test taker (Armstrong, Ferrell, Collmus, & Landers, 2016; Flowers, Kim, Lewis, & Davis, 2011), but by changing students' attitudes and behaviors, gamification could affect assessment outcomes in ways not well understood by the test developers (Landers, 2014). This may be why gamified assessments tend to be simpler and use basic game elements such as 2-D graphics and sound effects; whereas, training games are more likely to resemble high-tech, commercial video games such as 3-D graphics and role-playing (Armstrong et al., 2016; Lumsden et al., 2016).

There is "constant tension between creating an engaging task and the risk of undermining the task's scientific validity" (Lumsden et al., 2016, p. e11). In assessment parlance, the game elements might threaten the construct validity of the measure by contributing to construct irrelevant variance (Bachman, 1990; Messick, 1989). Because the current study was conducted as part of the iterative development and validation of a gamified assessment, we were interested in exploring how students described their behaviors while taking the test and their reactions to certain game elements. This information would be useful in making any necessary improvements to protect the validity of the gamified reading assessment.

Gamified Assessments

Assessments may be computer-delivered but not gamified unless they incorporate entertaining or interactive features (Boyle et al., 2016). Even without gamification, computer delivery can be useful for administering more items more rapidly, offering items that cannot feasibly be presented in other ways, delivering accommodations to students in special education, or for adapting the difficulty as the test taker responds (Almond et al., 2010). Once an assessment exists in digital form, it is relatively easy to add some game elements (Armstrong, Landers, & Collmus, 2015). Such elements might include graphics, avatars, levels, feedback, challenge, content locking, sound effects, and progress bars (Werbach & Hunter, 2012).

Researchers have advised test developers to add game elements iteratively, starting with a few and gradually increasing or refining them (Landers, 2014; Larman & Basili, 2003). This affords the opportunity to investigate how the elements are affecting test takers—including those with disabilities—and the test's validity (Murray, Silver-Pacuilla, & Helsel, 2007). Previous studies have focused on quantitative factors that mediate or moderate students' performance on computerized assessments such as grade or age, gender, special education and free or reduced-price lunch (FRL) status, attitudinal ratings, and categorical designations for computer features (Borgonovi, 2016; Flowers et al., 2011; Wang, Jiao, Young, Brooks, & Olson, 2008). However, there is a paucity of qualitative research that can better elucidate students' perceptions of and approaches to gamified assessments. Extant research on the perceptions of middle school students in special education has examined a computer-based but not gamified assessment (Flowers et al., 2011) or a gamified learning program (Ke & Abras, 2013). The former study administered surveys to collect data, and the latter study recorded anecdotal data while observing the students using the program.

To more systematically and thoroughly gather data in the present study, we surveyed the opinions of all middle school students participating in a validity study of a gamified reading assessment and then held a series of focus groups with randomly selected students to probe students' thinking more deeply. We were interested in how the few initial game elements incorporated in the test might be influencing students' attitudes and performance.

Purpose and Research Question

The purpose of the current study was to understand how middle school students with or at risk for reading difficulties view and approach taking gamified reading assessments. Students experiencing reading difficulties typically are assessed multiple times throughout a school year to monitor their progress toward grade-level skills, and they participate in computer-administered accountability tests annually (Cuevas, Russell, & Irving, 2012; Kieffer, 2010). To inform broader test development for these students, the study used one measure as an example of the kinds of gamification characteristics that are

Table 1. Demographic Characteristics of Full Sample (Survey Respondents) and Focus Group Members.

	School A				School B				School C				School D				Total				Final Sample			
	Survey ($n = 57$)		FocGrp ($n = 9$)		Survey ($n = 64$)		FocGrp ($n = 7$)		Survey ($n = 32$)		FocGrp ($n = 9$)		Survey ($n = 49$)		FocGrp ($n = 7$)		Survey ($n = 202$)		FocGrp ($N = 32$)		Survey ($n = 202$)		FocGrp ($n = 25$)	
Charac.	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*N*	%
Grade																								
6	18	32	4	44	23	36	2	29	17	53	4	44	10	20	1	14	68	34	11	34	68	34	9	36
7	19	33	3	33	26	41	4	57	9	28	3	33	22	45	4	57	76	38	14	44	76	38	10	40
8	20	35	2	22	15	23	1	14	6	19	2	22	17	35	2	29	58	29	7	22	58	29	6	24
Males	26	46	4	44	24	38	4	57	15	47	4	44	18	37	3	43	83	41	15	47	83	41	11	44
IEP	4	7	3	33	10	1	6	86	6	19	4	44	7	14	6	86	27	13	19	59	27	13	13	52
FRL	29	51	8	89	33	52	4	57	13	41	7	78	32	65	5	71	107	53	24	75	107	53	20	80

Note. Charac. = characteristics; FocGrp = focus group participants; IEP = identified with a disability and served with an individualized education program; FRL = receiving free or reduced-price lunch; total = students in all four schools combined; final sample = students after removal of School B.

currently being incorporated in testing tools. Specifically, the assessment was a single-player game without competition that included an avatar, ways to track progress through tasks, multiple game play sessions, clear rules, sounds, and established constraints (see description of measure in methods). The research question that guided our inquiry was: How do students with or at risk for reading difficulties in Grades 6–8 perceive and interact with the game elements incorporated in a reading test?

Method

Setting and Participants

The present investigation was conducted with a subsample of a larger study conducted to establish the technical adequacy of a gamified reading assessment. The 202 participants were from four middle schools in an urban Midwestern U.S. school district (Table 1). Because the test targeted decoding knowledge and automaticity—skills on which proficient readers would demonstrate a ceiling effect—the sample was composed of students in Grades 6 through 8 with or at risk for reading difficulties based on their statewide assessment performance. Those who were English-language learners were excluded to prevent confounding a reading difficulty with limited English-language proficiency. Although only 13% of the students currently were receiving special education services, it is estimated that 60% of adolescents not reading proficiently on state assessments have difficulty with word identification (Cirino et al., 2013), which requires intensive and prolonged intervention (Vaughn et al., 2012).

Observation and survey data were collected on all 202 students, but a maximum of 12 students at each of the four schools were selected for focus groups through stratified random sampling. The strata included, in hierarchical order grade level, gender, disability status (whether or not students were served with an Individualized Education Program or 504 plan), and FRL. This sampling procedure was used to increase representation of students with disabilities and those receiving FRL who exhibit reading difficulties at higher rates than their nondisabled and monolingual English peers (National Center for Education Statistics, 2018). In addition, these populations are more likely to participate in diagnostic reading assessments and interventions in middle school (Kieffer, 2010). As anticipated, the focus groups had larger percentages of students with disabilities and receiving FRL than were represented in the full sample (Table 1).

At School B, the group dynamics discouraged students from speaking freely. That is, students looked to one dominant participant after each question and then responded mostly in shrugs. Scripted follow-up prompting and individual student querying were not successful in eliciting information. Therefore, the data reported here are from the three campuses where students were willing to respond openly and verbally to questions. As shown in Table 1, removing School B did not appreciably change the representation of different student characteristics.

Measure

The instrument assessed students' decoding (i.e., their knowledge of letter–sound correspondences) and their automaticity in mapping sounds to letters. These foundational reading skills and abilities typically are mastered by third grade, so the test was designed to identify the particular areas with which adolescent students needed instructional intervention (e.g., short vowels, consonant digraphs, silent-e syllables, etc.). Most words and all the pseudowords (i.e., a nonword that looks or sounds like an English word but has no real meaning) used in the test were single syllable with regular letter–sound correspondences. Only four variants of tasks included multisyllable words, and these were all concrete words (i.e., those representing tangible things). The experimental version completed by students in the present study had nine tasks, some with different variations (see Table 2), which were delivered via an Internet-based program.

Each task was introduced by an avatar (Figure 1) with a prerecorded human voice that described the rules and constraints for that task and demonstrated an example of how to

Table 2. Summary of Tasks.

Task Name	Variants	Description
Fill in the blank (vowel)	Monosyllabic and multisyllabic	Student hears a spoken word and sees the spelling with the vowel(s) replaced by a _. Selects from among eight choices of vowel and vowel digraphs. For multisyllabic words, only the vowel(s) of one syllable is/are missing.
Fill in the blank (consonant)	Monosyllabic and multisyllabic	Same as *fill in the blank (vowel)* but with either initial or final consonant missing.
Change the (vowel)	N/A	Subject hears "Change X to make Y." Complete spelling of X is shown, and participant must select letter (of eight choices) to make Y.
Change the (consonant)	N/A	Same as *Change the (Vowel)*.
Verify	Speeded (fast) and unspeeded Monosyllabic and multisyllabic	Subject hears one word and sees the spelling of a word that either matches or mismatches the heard word. Selects yes or no.
Rhyme identification	Speeded (fast) and unspeeded	Subject sees a word and selects a rhyming word from among eight visually displayed options. On half the trials, the spelling of the rhyme does not match the target (e.g., *bare/pear*).
Find the picture	Speeded (fast) and unspeeded Monosyllabic and multisyllabic	Subject sees a word and selects which of four pictures match. Pictures represent phonologically similar competitors.
Syllable identification	N/A	Subject hears a multisyllabic word and is cued to find the Xth syllable. Selects the spelling of the syllable from an array of eight options.
Syllable order	N/A	Subject hears a multisyllabic word and assembles it from a series of short (syllable length) strings.

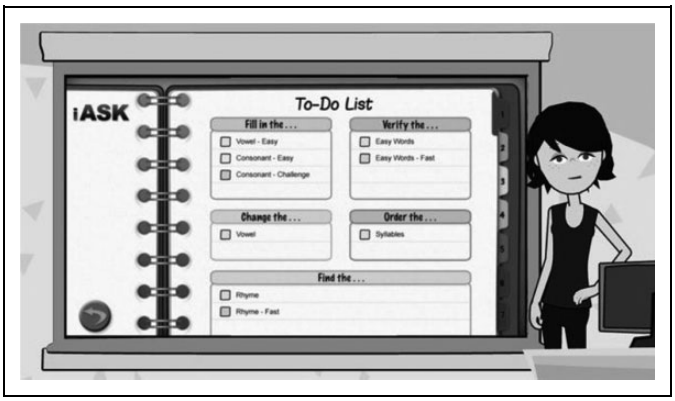

Figure 1. Screenshot of avatar explaining the progress tracker tasks (shown in gray scale).

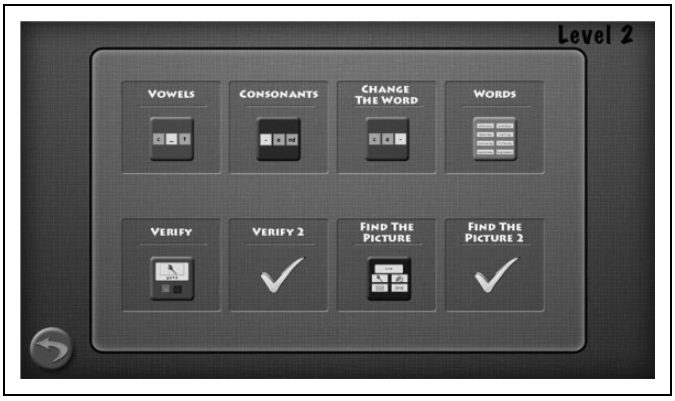

Figure 2. Screenshot of a sample task selection menu (shown in gray scale).

Figure 3. Screenshot of the levels with locked content in Levels 2 and 3 (shown in gray scale).

respond to the items. Because previous research found middle school students in special education were motivated by computer-based tests that allowed them some choice or control (Flowers et al., 2011; Ke & Abras, 2013), students were free to choose a starting task from the menu (Figure 2) and how they wanted to progress through the other tasks. After making a selection, students had to answer all the items before moving on to another task. Similarly, students had to complete all available tasks before progressing to another game play session, or *level* (Figure 3). Levels did not increase in difficulty because the items and tasks were randomly assigned to maintain a constant level of difficulty across sessions as part of the larger study determining the assessment's technical adequacy. Werbach and Hunter (2012) consider it a developer's purview to determine how *levels*, or the components that show a player's position in the game, will be defined. In the experimental study, the levels were paired with content unlocking, which serves as another form of challenge, feedback, and reward (Werbach & Hunter, 2012). Students had to complete all tasks in a given level to unlock the next level.

Other gamified features of the assessment intended to be motivational included encouraging but not evaluative feedback

Figure 4. Screenshot of a "fast" task where the word was replaced with hash marks after 90 ms (shown in gray scale). Progress keys are displayed along the left side of the screen.

embedded within tasks (e.g., the avatar voiced "Good job!"), a progress bar of keys that lit up and made a sound as students progressed through the items within a task, and checking off tasks on the To-Do List (Figure 1) as they were completed within a level. Tasks in the menu were indicated by a unique icon with visual cues to distinguish task variants (Figure 2). For example, some tasks included a speeded processing component in which the stimulus word was concealed or masked after 90 ms (Figure 4). The masking tasks allow for separating students' decoding knowledge from their speed in processing this knowledge (Roembke, Hazeltine, Reed, & McMurray, 2019). To help students distinguish speeded and unspeeded variants, the tasks with masking were labeled *fast*.

Per the advice of other researchers (e.g., Armstrong et al., 2016; Landers, 2014), the game elements of the experimental measure were limited in number and not as complex as commercial video games. Reducing complexity supports students in special education appropriately using the game because focusing too much on features and game play rules would overload cognitive resources needed to concentrate on the tasks (Murray et al., 2007). The assessment is a type of *serious game*, as opposed to an *entertaining game*, and is intended to evaluate skills and abilities (Boyle et al., 2016), so incorporating game elements should not come at the cost of test quality. Through iterative design and validation studies, the intention was to gradually increase or refine game elements that improved the test takers' experience without threatening the technical adequacy of the measure.

Testing was conducted in school classrooms, with 10–15 students per session. Students were told they would be participating in the development of a new test and that they would complete a series of tasks to test their abilities to match sounds and letters. To complete the tasks, each student used a laptop and made selections using either a touch screen or mouse. Because directions and items were delivered through digitized audio, students wore headphones throughout testing. At least two members of the research team were present at all times to assist students, answer questions, troubleshoot technical difficulties, and record observational data.

Study Procedures

To support methodological triangulation (Denzin, 1970), we gathered three sources of information on students' approaches and reactions to the computerized assessment. First, we gathered both students' scores on the experimental measure and observational data on student behaviors while testing (e.g., posture, distractibility, number of requested breaks, complaints expressed, clarifying questions asked, visible actions on the computer interface). Members of the research team who were not involved in the focus groups made these observational notes of the overall classroom context not of targeted students.

To gauge consensus among the broader population of student participants, the second data source was a 12-item survey about students' perceptions of the gamified assessment features, particular tasks, and general testing experience (Table 3; $\alpha = .755$). All students completed the electronically delivered survey at the end of their testing session. Four items asked students to indicate whether certain features made the tasks easier, more difficult, or neither. The final 8 items presented a traditional 0–5 Likert-type scale for students to report their level of agreement with aspects of the testing experience.

Survey items were aligned with questions posed to focus groups of students, the third data source. We used focus groups rather than individual interviews because we believed peers might be able to stimulate each other's thinking, provide a context for expressing a disconfirming opinion, and co-construct responses by contributing further elaboration (Kitzinger, 1995). Holding separate focus groups at each school also allowed differences to emerge from school to school, but each session followed a standardized protocol and posed the same seven questions targeting what students thought about the interface, different tasks, and amount of time they had to complete tasks. In addition, students were asked to describe what made the tasks easier or more difficult, how much they liked different tasks, what helped them to concentrate, and what strategies they might have used to do their best.

The focus groups were held after all testing concluded and lasted 35–40 min, depending on how much the students wanted to say. Although they were audio recorded for later transcription, the researchers made field notes throughout the sessions. During the discussion, the moderator repeated and summarized statements to allow participants an opportunity to make clarifications, additions, or changes. Conducting member checking in real time is recommended whenever there might be difficulty reconvening participants (Kidd & Parshall, 2000).

Data Analytic Procedures

Qualitative data. Guided by Strauss and Corbin's (1998) grounded theory approach, we sought to discover emerging patterns in the transcripts, field notes, and observational data. Accordingly, we conducted microanalysis with NVivo-11 qualitative data analysis software, in which we coded relevant words, phrases, and sentences. These were examined to identify categories and their associated properties and dimensions

Table 3. Survey Results.

Indicate Your Opinion About how Each Feature of the Test Made it (1) *More Difficult*, (2) *Neither More Difficult or Easier*, or (3) *Easier*.	1. *More Difficult*		2. *Neither More Difficult nor Easier*		3. *Easier*	
	n	%	*n*	%	*n*	%
1. The number of items I had to complete for each task made the test.	18	9	136	67	48	24
2. The audio recordings of the words and nonwords made the test.	16	8	66	33	120	59
3. Having word and nonword items together in the same task made the test.	53	26	101	50	48	24
4. Having items covered up so that I had to answer based on my memory made the test.	99	49	66	33	37	18

Indicate Your Personal Opinion About Whether You (1) *Strongly Disagree*, (2) *Disagree*, (3) *Neither Agree nor Disagree*, (4) *Agree*, or (5) *Strongly Agree* With Each Statement.	1. *Strongly Disagree*		2. *Disagree*		3. *Neither Agree nor Disagree*		4. *Agree*		5. *Strongly Agree*	
	n	%	*n*	%	*n*	%	*n*	%	*n*	%
1. I concentrated while completing all tasks	1	0	6	3	28	14	121	60	46	23
2. I gave my best effort while completing all tasks	2	1	2	1	18	9	91	45	89	44
3. The assessment tasks were easy	1	0	17	8	62	31	81	40	41	20
4. The assessment tasks were frustrating	52	26	80	40	45	22	19	9	6	3
5. Thirty min was enough time to work on the test each day	4	2	12	6	36	18	74	37	76	38
6. I got tired during the test	22	11	38	19	44	22	57	28	41	20
7. I think this test will help my teacher plan reading lessons for me	7	4	21	10	79	39	62	31	33	16
8. The directions for each task were easy to follow	1	0	4	2	26	13	60	30	111	55

by making theoretical comparisons among phenomena (Silverman & Marvasti, 2008; Strauss & Corbin, 1998). We refined and clarified the relationships between categories as we moved from within-site to cross-site analyses (Miles & Huberman, 1994). Throughout the iterative process of analysis, the researchers who conducted the focus groups discussed their work with the team members not involved in either observing the tests or the focus groups to aid in examining potential biases and ensure reliability to the coding procedures (Brantlinger, Jimenez, Klingner, Pugach, & Richardson, 2005). The four categories that emerged in the final analysis of the perceptual data were perceptions of task difficulty, making the tasks fun, strategies, and competition.

Quantitative data. We first evaluated students' test data for normality. After confirming a Gaussian distribution, we analyzed scores from each task for central tendency and compared the means of task variants. Next, we analyzed survey data descriptively, using frequency of responses and dispersion, as appropriate. Because items varied in format (i.e., 1–3 rating of easy to difficult and 1–5 rating of agreement), we treated data for each section separately.

Results

Based on the four categories that emerged in our qualitative analysis, we report on the students' perceptions of gaming features that they believed made the tasks of the illustrative instrument more or less difficult and more or less fun to complete. We then report on students' approaches to the tasks that revealed an imposition of strategic responding and competitiveness. Where applicable, we address how the observation and survey data aligned with the comments of focus group members.

Perceptions of Task Difficulty

Focus group participants at all schools commented on how the tasks that included masking were difficult because the stimulus word did not appear on the screen for very long (90 ms) before being covered. This feature is only possible in a digital assessment and was intended as a way to assess automaticity in word recognition, but some students felt these *fast* items disadvantaged their performance. As a School A participant explained, "When it goes away too fast, it just surprises you and you mess up." Similarly, a School C participant stated, "You might see only two of the first letters and then you kinda gotta think about what the last ones are because it seems like they don't even let you see the last four letters." Other members of the School C focus group felt that the masking increased their concentration because "... you had to try to see what it actually was."

Students at School A made impromptu suggestions for how the test might be redesigned to facilitate using gaming strategies for mastering the masked items. They wanted a way to reveal the word again, "like a replay button." Others suggested making the number of hashtags that covered a word match the number of letters, as a clue. At School D, a student thought the timing could stay the same, "... if you could push a button to make the word show, instead of waiting for it to flash for you." The suggestions indicate that students were interested in having mechanisms for improving their game playing, even if the changes would have detracted from the accurate assessment of their reading skills.

Survey responses revealed that just under half of the students (49%) agreed the masking feature made the test more difficult (see masking task example in Figure 4). However, 18% believed it made the test easier and 33% thought the masking made the test neither easier nor harder. Students responded correctly on 86.2% of unmasked variants (SD = 7.8%), compared to 78.3% (SD = 9.6%) on masked variants, $t(201) = 26.3$, $p < .0001$. The significant difference in student performance favoring unmasked words suggests the fast items were more challenging, but the lack of floor effects combined with the absence of a consensus on the survey results suggest students generally were not frustrated by the challenge. Moreover, students at School A complained about tasks that "A lot of them were too easy," and "a couple were too basic." This was echoed at School D where a participant commented, "It was too easy 'cause we . . . learned all the stuff already, and it's like repeating."

Among the 202 survey respondents, 60% agreed or strongly agreed that the assessment tasks were easy. Only about 8% disagreed with that statement, and a relatively small percentage (12%) believed the tasks were frustrating. This suggests that—despite any perceived difficulty of certain items or certain task variants—students' overall impression was that the gamified assessment was feasible for them to master. On average, students responded to 75.9% of all the items correctly (SD = 10.6; range: 37.9–94.0%). There was variation in test performance, but for most students, the items were appropriately challenging, an important design consideration for academic games (Morford, Witts, Killingsworth, & Alavosius, 2014; Watson et al., 2011).

Students made their own judgments about what tasks were more or less difficult, and 64% of the focus group participants reported using this self-imposed leveling to guide the order in which they chose the tasks to complete. Most often, students said they worked from tasks they felt were harder to the ones they thought were easier ($n = 5$ at School A; $n = 4$ at School C; $n = 3$ at School D). A few reported working in the opposite direction, from the easiest to the hardest tasks ($n = 2$ at School A; $n = 0$ at School C; $n = 2$ at School D). The observers' notes from the testing sessions confirmed that students were moving through tasks in different orders. Completed tasks had a check mark next to them on the computer screen, so observers monitoring the sessions could see not only a student's active selection of a task but also that task in relation to the others already completed in the menu. The selection sequences varied from student to student, suggesting the choices were based on personal preferences as was intended to improve motivation (Morford et al., 2014).

Making the Tasks Fun

Several gaming features of the assessment were intended to motivate the group of adolescents with or at risk for reading difficulties, but focus group participants did not always agree that these augmented their experience. For example, the avatar's voice was used to offer praise (e.g., "Good job!" or "Keep going!") at random intervals during tasks. Although focus group participants thought she sounded robotic when giving instructions, they commented on her having "too much enthusiasm" when delivering the praise. Rather than being encouraging, students perceived the comments as disruptive. A student at School A remarked, "Yeah, that was annoying . . . 'cause it interrupts you right at the game." Students at all three schools thought the encouragement actually hampered their performance as described by a student at School D, " . . . it's kinda distracting. It's like, you're going through it and you're trying to figure out the next word, but then she says something and you gotta wait."

Students at Schools C and D described the praise as coming from "out of nowhere," and at all three schools several students believed it caused them to "mess up." In addition, students at School C commented on how disingenuous or even misleading the praise was:

Participant 1: . . . it says, "Good job" and all that, but even if you don't even get the answers right, it still says, "Good job."

Participant 2: Yeah, 'cause one minute if you knew that you got it wrong, and then it says, "Good job," it seems like they were just saying it to make you feel good.

Students were not told which of their answers were correct, but they expressed a desire for feedback that would indicate their accuracy. They described expecting certain gaming elements to provide this information. For example, a student at School D suggested the keys that tracked completion should only "show if you got it right. You could have an X if you get it wrong and a star if you get it right." Another member of the focus group suggested that the progress bar items could be different colors to reflect answers he got right or wrong, "By putting a red star on one, and then put a yellow star." Similarly, a student at School C reflected on the dinging sound that accompanied the key, " . . . it had a beeping noise . . . , so do things like that's the noise that you got right. But then it should have a noise for what you got wrong."

Once the students figured out the keys and noises were unrelated to the accuracy of their responses, they acknowledged the feature was helpful for knowing, " . . . when I was almost done" (student at School A). However, some students at School A found the accompanying ding bothersome because it played "after each question you do." Although one student thought it could occur more intermittently, two others remarked, "Just don't do it at all." One participant "really hated the ding" so much that he "took off the headphones every single time . . . but you could still hear it." As the stimuli were stated orally through digital audio recordings, removing headphones after each item would potentially be detrimental to student performance.

Researchers generally did not observe students removing headphones, but a number of students did readjust their headphones frequently and place them around their necks or hold

them to their ears rather than wear them over their heads. Therefore, it is possible that students' efforts to manipulate the equipment impeded their ability to hear the stimuli completely or accurately. This may have contributed to students' perceptions that the audio recordings of the words made the test more difficult (8%) or neither more difficult nor easier (33%).

Students in the focus groups were interested in having more sounds at the end of a testing session. A student at School A who described the current ending as "kinda boring" suggested, "... give it some confetti. Give it some music." Three other participants in that focus group concurred it should "... play a song or something" and have confetti appear on the screen. A student at School C agreed the encouragement should be reserved for the end, "... when you finished it, say, 'Good job, you finished this!'" Students at School D wanted not only congratulations at the end of a task but also some type of reward. "Or if you get an amount right, you get points. And then if you get enough points, you could play a game."

These additions were considered as a way to make the test more enjoyable. During the testing, some students were observed slumping on their desks more by the third game play session. A student at School C remarked that the repetition of tasks in each level made it boring, "It's like, really, this again? ... If it was something different and not the same [task] over and over again." Another student added, "I felt it didn't improve of how hard or how easy it got. It was just the same." Other focus group participants made suggestions for how to increase interest in completing the tasks multiple times such as by playing loud music or changing the sounds and voices "so then people would actually wanna hear it again." A different student offered, "It should change the background every once in a while to keep you motivated."

Among the 202 survey respondents, 83% indicated they agreed or strongly agreed that they concentrated while completing all tasks and only 3% disagreed or strongly disagreed with the statement. Even more students (89%) agreed or strongly agreed that they gave their best effort while completing all tasks, with only 2% disagreeing or strongly disagreeing. Although they may not have enjoyed certain parts of the test, most believed they applied themselves to the tasks and looked for ways to perform well.

Strategies

Students in the focus groups made impromptu comments that revealed they were searching for patterns in the assessment tasks that were independent of the skills those tasks were targeting. At School A, this was the topic of a discussion among several participants, describing the syllable identification task:

Participant 1: They said ... the first syllable, and then you pick the first one. They switched to three, and people that usually like to rush will click [the first syllable]. So if you realize that you heard third, but you picked the first one, it would do both.

Participant 2: Yeah, 'cause ... that kinda happened. Like, first syllable, first syllable, first syllable, first, first, first, and then, like third.

Participant 3: Second, third—oh!

Participant 1: It's like, "Argh, it's changing!"

Participant 4: To me, I found a pattern. It's—the syllables will either be on top or bottom, or it'll be angled in some way. So ... all I had to do was find the word in the pattern and then, there you go!

Participant 5: Yeah, if I had pieces of those words, I could find exactly where they match up—where they overlap.

Participant 4: Here's the thing ... I'm just saying don't make the patterns too predictable, for me, was that. ... Just mix them up because [it was] too easy for me. ... I could *[snaps fingers]* finish those.

Similar comments were made at School C where students expressed looking for patterns in other tasks as well:

Participant 1: ... It seems like [the verify task items] were hard 'cause it would be in a row. Like-you would have maybe five different mess ups and then you would have two of the good ones. And you would think it has a pattern. Then ... if you guess a pattern for, I think it was six of them at the same pattern, then you try and figure it out. But then it messes you all up.

Moderator: So instead of thinking about the words, you were trying to figure out patterns.

Participant 1: Mm-hmm.

Participant 2: If someone was trying to figure out a pattern and then they think that the pattern is the same, and it could mess you up too.

When referring both to patterns and to tasks that were difficult, focus groups members would describe aspects of the test design that would cause them to "mess up." Hence, students might have been attributing at least some of their performance on the tasks to how skillful they were at gaming strategies or approaching the tasks with some sense of being in competition with the task design elements.

Competition

The decoding assessment was a single-player game without competition or overt scoring. However, focus group participants indicated they were evaluating their own performance and desired opportunities to improve their playing. As students in School C explained:

Participant 1: ... What'd be really good for it is at the end of you doing stuff, it should show up how much in total did you get right or wrong. So

you could do it over again to get really good at it.

Participant 2: Yeah. . . . It don't got to say which answer you actually got right and which answer you got wrong. So if you wanted to redo the score, you didn't know which ones you got right. So, you had to, like, redo every single one and retry everything . . . like if you wanted to improve . . .

Participant 1: . . . when you're done with the picture [task], if you got one wrong, it would say, "Hey, you got one wrong. Try it again." . . . And then when you did it again, you got it all right and you go on to something else.

Five of the nine members of that focus group concurred they wanted the opportunity to redo tasks, even if they thought they only missed 1 item. Although students only could have perceived mistakes or wrong answers because the program never indicated this, the concern with accuracy was echoed by a student at School A who said, "After you do one question, it goes directly to the next one right after you do it. And say that you wanted to go back and fix your mistake, you can't go back and fix it."

For some students, accuracy was associated with flow and becoming immersed in the gaming environment to the point that other noises or movements in the room were unwelcome intrusions on their experience. A student at School A explained, "We can get into the motion, then we'll immediately lose it, and then we get into that, and then we lose it again." A peer added, "I had a perfect streak going until she said something, and then I tried to do something else." Students at School D referred to "getting on a roll" that would be disrupted by talking or having other students make them "lose focus."

Given students' concerns with strategizing and competing against the game, we asked them what they thought the purpose of the test was. Reponses included "to see how students learn," "how [we] think," "I'm helping future students learn," and "show what you need more help with that [game play] session." No students made comments that suggested they recalled that the test was about their reading abilities. In addition to being told this in the assent process, a survey item specifically asked students whether they thought the test would help their teachers plan reading lessons. Nearly half of the 202 respondents (47%) agreed or strongly agreed that it would and only 14% disagreed or strongly disagreed. The remaining 39% neither agreed nor disagreed with the statement, suggesting they were unsure how teachers would use the results.

Discussion

This study explored how middle school students with or at risk for reading difficulties perceived and interacted with a gamified test. These types of tools are rapidly emerging in educational settings (Boyle et al., 2016) and are commonly used with adolescents exhibiting reading difficulties (Cuevas et al., 2012). The results of analyzing students' perceptual comments indicated there are potential benefits and pitfalls of applying game design elements for the purposes of evaluating students' reading performance.

An impetus for using gaming characteristics is to motivate students to complete educational activities such as tests that are not typically enjoyable (Armstrong et al., 2016). Findings of the present study suggest the gamified reading assessment was successful at stimulating students in two ways identified by Morford et al. (2014). First, students generally seemed engaged in the test and willing to persist in completing it because the tasks offered a challenge but were not overly frustrating. Second, the assessment seemed emotionally motivating because students' comments revealed they became immersed in the tasks. The gamified test was not designed to be socially motivating because students needed to be assessed on their individual reading abilities, and based on previous findings (Clark et al., 2016; Ferrell et al., 2015; Kocadere & Caglar, 2015), we did not want to make the test competitive such that it might reduce the self-efficacy of students who were average to below-average readers (Zimmerman & Schunk, 2006).

Nevertheless, students self-imposed a sense of competitiveness and strove to identify patterns or strategies they could use to master the assessment. Applying gaming strategies rather than decoding strategies threatens to introduce construct irrelevant variance into the measurement of students' reading abilities, thus threatening the construct validity of the assessment (Bachman, 1990; Messick, 1989). Because the patterns students sought to identify were not the kinds of patterns that existed in the assessment, strategies applied to master the game could not lead to improved performance. Hence, students blamed the program for making them "mess up" and perceived their own efforts and approaches as not contributing to their success. Although success was a self-determined judgment in the absence of receiving a score, students still seemed to lose self-efficacy.

Others have cautioned that pursuing the development of skills for conquering the digital tasks is more likely when extrinsic rewards are offered (Abramovich et al., 2013; Deci et al., 2001). However, the assessment investigated here did not include extrinsic rewards other than verbal praise, which was presented independently of performance and foster internal motivation to persist. The assessment tracked students' progress through the items, tasks, and successively unlocked levels to reveal their "position" in the game and encourage persistence. It did not display their scores on each item or after each block of items for a task because previous research found middle school students in special education lost enthusiasm when they did not attain a desired score, despite receiving specific feedback on their improvement (Ke & Abras, 2013). Students in the present study complained about both the presence of the random, generic praise and the absence of scores. They may have desired accuracy feedback, in part, to inform their gaming strategies because students wanted to be able to redo items they thought they missed and replay stimuli

they did not process. These may be behaviors related to test taking in a gamified environment, but they are not related to improved decoding ability.

Notably absent from focus group members' remarks was an articulated understanding that the assessment was about reading. This was despite all students being told before beginning the tasks that they were part of an assessment. Moreover, fewer than half of the survey respondents thought the results would be helpful for their teachers to plan reading instruction. If other students have similarly misaligned perceptions of and approaches to taking gamified measures, it may help explain why Attali and Arieli-Attali (2015) found that points and rewards did not improve students' accuracy on computerized math assessments. The test takers were not trying to get better at the academic constructs; rather, they were trying to get better at playing the game. Superfluous strategies likely are not unique to gamified tests because students have been known to try guessing patterns of answers on paper-based multiple-choice tests (Foley, 2016), and the search for clues or "test wiseness" has been studied for over half a century—long before gamification (e.g., Millman, Bishop, & Ebel, 1965). It may be that adding gaming elements exacerbates the problem, but it is not known whether students' abilities are more reliably determined when they are cognizant of being tested on a particular topic than when they are not. Future research could explore these issues by comparing participants' approaches and performance when taking a gamified versus computer-delivered assessment, and when they are or are not reminded about the focus of the test they are taking.

Implications

Findings from the present study suggest students' expectations of the gamified reading assessment shaped their experiences and how they interacted with the tasks and features. Besides wanting the kind of accuracy feedback that is common in digital games, they seemed surprised and annoyed when unlocking a *level* meant they had to complete essentially the same tasks multiple times. Students' preconceived notion of the label *level* was that there would be a progression in difficulty or some type of change in the gaming environment. Werbach and Hunter (2012) have described *levels* as anything that show a player's position in the game, but our findings suggest that developers' choice of how to define game element should be driven by how the test takers expect the component will work. Hence, we echo those who have suggested that game designers and education professionals work together to make gaming features not only appealing but also reinforcing of the intended objective (Clark et al., 2016; Meyen, 2015).

It is possible students' expectations of the gamified features arose from previous experiences with how digital games work, but some of the students' comments could inform superficial aspects of a gamified assessment's design. For example, having different environments, backgrounds, and sounds were other surface-level design elements that students suggested would enhance their experience, especially when completing multiple game play sessions (Boo & Vispoel, 2012). Nevertheless, others investigating gamified assessments caution that even seemingly superficial elements could threaten the psychometric properties of the instruments (Landers, 2014; Lumsden et al., 2016). For example, introducing feedback after items or opportunities to redo a response raises the possibility that students could learn or improve their skills during the assessment. The goals of gamification, including the testers' manipulation of difficulty levels, simply may be incompatible with an assessment in which some quantity of incorrect answers is expected or psychometrically necessary.

Gaming features also may encourage students to explore response strategies that are unrelated to the construct being tested. It is important to note that detecting when or how gamification might be contributing to measurement error is not straightforward in purely quantitative statistical analyses of reliability and validity. It was only by talking with students about their approaches and reactions to the test that it became apparent they were attempting to hone their gaming skills rather than their decoding skills. This reflects the measure's face validity: How the users are judging the purpose and reasonableness of the test (Fink, 1995).

Limitations and Directions for Future Research

Our study was not designed to identify effective solutions to overcoming the inappropriate pursuit of patterns and gaming strategies in the gamified assessment. Despite avoiding features identified in previous research as encouraging those behaviors, such as extrinsic rewards and competitive configurations (Abramovich et al., 2013; Clark et al., 2016; Deci et al., 2001), students still expressed a preoccupation with them. Future research might investigate whether an introduction to an assessment can counteract those tendencies. The introduction could explicitly inform test takers of the purpose of the test, the construct on which they should focus, and the types of abilities that are intended to improve performance.

Given that the current study tested middle school students' decoding abilities, the sample was composed of those with or at risk for reading difficulties. We also oversampled students with disabilities and receiving FRL because these types of characteristics have been associated with a higher likelihood of participation in diagnostic assessment and reading intervention (Kieffer, 2010). Hence, the findings reported here may not represent the perceptions of students at different grade and ability levels or those who take gamified assessments of other reading skills (e.g., comprehension) or subjects such as math. Additional research is needed to understand the extent of individual differences in reactions to single-player gamified assessments without competition.

Finally, students' perceptions of and approaches to the assessment in this study might change if the test were redesigned in some of the surface-level ways the students recommended (e.g., different environments, backgrounds, sounds, and labels). It may be tempting to fix every flaw at once, but software developers recommend iterative and incremental

development (Armstrong et al., 2016; Landers, 2014; Larman & Basili, 2003). Researching each iteration of a gamified test would better identify how small design features impact the users and might be manipulated in productive ways to create a better overall assessment.

Conclusions

Findings from this study suggest students with or at risk for reading difficulties react positively to some motivational aspects of gamified assessments such as offering challenging items that were not frustratingly difficult and incorporating interactive features that allowed them to become immersed in completing the tasks. However, students expressed annoyance with alterable surface-level features such as too much repetition (e.g., using the same sounds, displays, and tasks) and praise or labels that they believed did not accurately convey what they were doing. Although the test studied was noncompetitive, students seemed to impose competition with the gaming aspects. Specifically, they blamed the program for making them "mess up," and they sought to identify patterns and strategies that were not related to construct being assessed. This suggests students' misaligned test-taking behaviors might be reducing their self-efficacy and increasing measurement error.

Authors' Note

The content is solely the responsibility of the authors and does not necessarily represent the official views of the Institute of Education Sciences.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was funded by Contract No. ED-IES-15-C-0023 from the Institute of Education Sciences in the U.S. Department of Education.

ORCID iD

Deborah K. Reed, PhD https://orcid.org/0000-0003-0874-1412

References

Abramovich, S., Schunn, C., & Higashi, R. M. (2013). Are badges useful in education? It depends upon the type of badge and expertise of learner. *Educational Technology Research and Development*, *61*, 217–232. doi:10.1007/s1142-013-9289-2

Almond, P., Winter, P., Cameto, R., Russell, M., Sato, E., Clarke-Midura, J., & Lazarus, S. (2010). Technology-enabled and universally designed assessment: Considering access in measuring the achievement of students with disabilities—A foundation for research. *Journal of Technology, Learning, and Assessment*, *10*, 1–52. Retrieved from http://ejournals.bc.edu/ojs/index.php/jtla/article/view/1605

Armstrong, M. B., Ferrell, J. Z., Collmus, A. B., & Landers, R. N. (2016). Correcting misconceptions about gamification of assessment: More than SJTs and badges. *Industrial and Organizational Psychology*, *9*, 671–677. doi:10.1017/iop.2016.69

Armstrong, M. B., Landers, R. N., & Collmus, A. B. (2015). Gamifying recruitment, selection, training, and performance management: Game-thinking in human resource management. In D. Davis & H. Gangadharbatla (Eds.), *Emerging research and trends in gamification* (pp. 140–165). Hershey, PA: IGI Global.

Attali, Y., & Arieli-Attali, M. (2015). Gamification in assessment: Do points affect test performance? *Computers & Education*, *83*, 57–63. doi:10.1016/j.compedu.2014.12.012

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford, England: Oxford University Press.

Boo, J., & Vispoel, W. (2012). Computer versus paper-and-pencil assessment of educational development: A comparison of psychometric features and examinee preferences. *Psychological Reports*, *111*, 443–460. doi:10.2466/10.03.11.pr0.111.5.443-460

Borgonovi, F. (2016). Video gaming and gender differences in digital and printed reading performance among 15-year-old students in 26 countries. *Journal of Adolescence*, *48*, 45–61. doi:10.1016/j.adolescence.2016.01.004

Bouck, E. C., & Flanagan, S. (2009). Assistive technology and mathematics: What is there and where can we go in special education. *Journal of Special Education Technology*, *24*, 17–29. doi:10.1177/016264340902400202

Boyle, E. A., Hainey, T., Connolly, T. M., Gray, G., Earp, J., Ott, M., & Pereira, J. (2016). An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games. *Computers & Education*, *94*, 178–192. doi:10.1016/j.compedu.2015.11.003

Brantlinger, E., Jimenez, R., Klingner, J., Pugach, M., & Richardson, V. (2005). Qualitative studies in special education. *Exceptional Children*, *71*, 195–207. doi:10.1111/1467-8527.t01-1-00151

Cirino, P. T., Romain, M. A., Barth, A. E., Tolar, T. D., Fletcher, J. M., & Vaughn, S. (2013). Reading skill components and impairments in middle school struggling readers. *Reading and Writing*, *26*, 1059–1086. doi:10.1007/11145-012-9406-3

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

Cuevas, J., Russell, R., & Irving, M. (2012). An examination of the effect of customized reading modules on diverse secondary students' reading comprehension and motivation. *Educational Technology Research and Development*, *60*, 445–467. doi:10.1007/s11423-012-9244-7

Deci, E. L., Koestner, R., & Ryan, R. M. (2001). Extrinsic rewards and intrinsic motivation in education: Reconsidered once again. *Review of Educational Research*, *71*, 1–27. doi:10.3102/00346543071001001

Denzin, N. K. (1970). *The research act in sociology*. Chicago, IL: Aldine.

Federation of American Scientists. (2006). *Summit on educational games: Harnessing the power of video games for learning*. Washington, DC: Author. Retrieved from https://fas.org

Ferrell, J. Z., Carpenter, J. E., Vaughn, E. D., Dudley, N. M., & Goodman, S. A. (2015). Gamification of human resource processes. In

D. Davis & H. Gangadharbatla (Eds.), *Emerging research and trends in gamification* (pp. 108–139). Hershey, PA: IGI Global.

Fink, A. (1995). *How to measure survey reliability and validity* (Vol. 7). Thousand Oaks, CA: Sage.

Flowers, C., Kim, D. H., Lewis, P., & Davis, V. C. (2011). A comparison of computer-based testing and pencil-and-paper testing for students with a read-aloud accommodation. *Journal of Special Education Technology*, *26*, 1–12. doi:10.1177/016264341102600102

Foley, B. P. (2016). Getting lucky: How guessing threatens the validity of performance classifications. *Practical Assessment, Research & Evaluation*, *21*, 1–23.

Gray, L., Thomas, N., & Lewis, L. (2010). *Teachers' use of educational technology in U.S. public schools: 2009* (NCES 2010-040). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://nces.ed.gov/pubs2010/2010040.pdf

Hines, P. J., Jasny, B. R., & Mervis, J. (2009). Adding a T to the three R's. *Science*, *323*, 53. doi:10.1126/science.323.5910.53a

Ke, F., & Abras, T. (2013). Games for engaged learning of middle school children with special learning needs. *British Journal of Educational Technology*, *44*, 225–242. doi:10.1111/j.1467-8535.2012.01326.x

Kidd, P. S., & Parshall, M. B. (2000). Getting the focus and the group: Enhancing analytical rigor in focus group research. *Qualitative Health Research*, *10*, 293–308. doi:10.1177/104973200129118453

Kieffer, M. J. (2010). Socioeconomic status, English proficiency, and late-emerging reading difficulties. *Educational Researcher*, *39*, 484–486. doi:10.3102/0013189X10378400

Kitzinger, J. (1995, July 29). Qualitative research: Introducing focus groups. *British Medical Journal*, *311*, 299–302. PMCID: PMC2550365

Kocadere, S. A., & Caglar, S. (2015). The design and implementation of a gamified assessment. *Journal of e-Learning and Knowledge Society*, *11*, 85–99.

Landers, R. N. (2014). Developing a theory of gamified learning: Linking serious games and gamification of learning. *Simulation & Gaming*, *45*, 752–768. doi:10.1177/1046878114563660

Larman, C., & Basili, V. R. (2003). Iterative and incremental development: A brief history. *Computer*, *36*, 47–56. doi:10.1109/mc.2003.1204375

Lumsden, J., Edwards, E. A., Lawrence, N. S., Coyle, D., & Munafo, M. R. (2016). Gamification of cognitive assessment and cognitive training: A systematic review of application and efficacy. *JMIR Serious Games*, *4*, e11. doi:10.2196/games.5888 Retrieved from http://games.jmir.org/2016/2/e11

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York, NY: Macmillan.

Meyen, E. (2015). Significant advancements in technology to improve instruction for all students: Including those with disabilities. *Remedial and Special Education*, *36*, 67–71. doi:10.1177/0741932514554103

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis* (2nd ed.). Thousand Oaks, CA: Sage.

Millman, J., Bishop, H. I., & Ebel, R. (1965). An analysis of test wiseness. *Educational and Psychological Measurement*, *25*, 707–726. doi:10.1177/001316446502500304

Morford, Z. H., Witts, B. N., Killingsworth, K. J., & Alavosius, M. P. (2014). Gamification: The intersection between behavior analysis and game design technologies. *The Behavior Analyst*, *37*, 25–40. doi:10.1007/s40614-014-0006-1

Murray, B., Silver-Pacuilla, H, & Helsel, I. F. (2007). Improving basic mathematics instruction promising technology resources for students with special needs. *Technology in Action*, *2*, 1–8.

National Center for Education Statistics. (2018). *The nations report card. 2017 reading results*. Washington, DC: Institute of Education Sciences, U.S. Department of Education. Retrieved from www.nationsreportcard.org

Roembke, T. C., Hazeltine, E., Reed, D. K., & McMurray, B. (2019). Automaticity of word recognition is a unique predictor of reading fluency in middle-school students. *Journal of Educational Psychology*, *111*, 314–330.

Silverman, D., & Marvasti, A. (2008). *Doing qualitative research*. Thousand Oaks, CA: Sage.

Strauss, A., & Corbin, J. (1998). Basics of qualitative research: Techniques and procedures for developing grounded theory (2nd ed.). Thousand Oaks, CA: Sage.

Vaughn, S., Wexler, J., Leroux, A., Roberts, G., Denton, C. A., Barth, A., & Fletcher, J. (2012). Effects of intensive reading intervention for eighth-grade students with persistently inadequate response to intervention. *Journal of Learning Disabilities*, *45*, 515–525. doi:10.1177/0022219411402692

Wang, S., Jiao, H., Young, M.J., Brooks, T., & Olson, J. (2008). Comparability of computer-based and paper-and-pencil testing in K-12 reading assessments: A meta-analysis of testing mode effects. *Educational and Psychological Measurement*, *68*, 5–24. doi:10.1177/0013164407305592

Watson, W. R., Mong, C. J., & Harris, C. A. (2011). A case study of the in-class use of a video game for teaching high school history. *Computers & Education*, *56*, 466–474. doi:10.1016/j.compedu.2010.09.007

Werbach, K., & Hunter, D. (2012). *For the win: How game thinking can revolutionize your business*. Philadelphia, PA: Wharton Digital Press.

Zimmerman, B., & Schunk, D. H. (2006). Competence and control beliefs: Distinguishing the means and ends. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology* (pp. 349–367). Mahwah, NJ: Lawrence Erlbaum Associates.

Author Biographies

Deborah K. Reed, PhD, is the director of the Iowa Reading Research Center (IRRC) and an associate professor with the University of Iowa College of Education. As an applied researcher, her work focuses on problems of practice in the areas of reading instruction, intervention, and assessment as well as the use of data-based decision making within reading programs.

Emily Martin, MSW, is a doctoral candidate in the School of Social Work at the University of Iowa. Her research interests include educational equality, the effect of school culture and climate on educational outcomes, how to better train educators to use a trauma-informed care approach, and ways to increase cultural competency among school personnel.

Eliot Hazeltine, PhD, is a professor of Psychological and Brain Sciences at the University of Iowa. His research focuses on how people learn to match external stimuli with internal states to choose responses and engage in flexible, goal-directed behaviors. His approaches have included dual-task interference, bimanual coordination, cognitive control, and reading using a wide array of behavioral procedures and neuroimaging techniques.

Bob McMurray, PhD, is a professor of Psychological and Brain Sciences at the University of Iowa and the director of the DeLTA Center (Development, Learning, Theory and Application). His lab researches how people process spoken and written language, and how these skills develop in both typical and atypical individuals using eye-tracking and other psychological methods, cognitive neuroscience, and computational modeling.