Practical Assessment, Research, and Evaluation

Volume 26 Article 27

2021

Game Design Considerations for Screening, Interim, and Diagnostic Educational Assessments

David Dockterman

Yaacov Petscher
Florida State University

Alison McAfee
Harvard University

Eric Klopfer
Massachusetts Institute of Technology

Scot Osterweil
Massachusetts of Institute Technology

See next page for additional authors

Game Design Considerations for Screening, Interim, and Diagnostic Educational Assessments

Authors

David Dockterman, Yaacov Petscher, Alison McAfee, Eric Klopfer, Scot Osterweil, and Cody Diefenthaler

Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Volume 26 Number 27, December 2021 ISSN 1531-7714

Game Design Considerations for Screening, Interim, and Diagnostic Educational Assessments

David Dockterman, *Harvard University*
Yaacov Petscher, *Florida State University*
Alison McAfee, *Harvard University*
Eric Klopfer, *Massachusetts Institute of Technology*
Scot Osterweil, *Massachusetts of Institute Technology*
Cody Diefenthaler *Florida State University*

To inform instruction, screening and diagnostic assessments must collect accurate data about the current state of the learner. Unfortunately, students may find assessments unengaging, intimidating, or irrelevant, undermining the quality of their effort and the quality of the data. The application of gaming to assessments may provide a way to boost and sustain effortful test-taker engagement, an integration that has thus far yielded mixed results, at best. Our interdisciplinary team reviewed and evaluated existing gamification research to consolidate a set of guiding principles for effectively merging diagnostic assessment tasks and protocols with a motivating game-like context in ways that specifically foster high levels of test-taker effort. We share our work in this paper to help inform ongoing research and development leading to more efficient and effective assessments of children.

Gaming Considerations for Educational Assessments

Assessments in pre-K-12, whether screening, diagnostic, benchmark, formative, or summative, are most accurate and useful at the intersection of psychometrically sound tools with individuals putting forth their best effort. Unfortunately, children may perceive assessments to be stressful (Alexander, et al., 2009; Segool, et al., 2013; von der Embse et al., 2018) or, as with any task, unengaging, undermining high quality effort (Lumsden, et al., 2016). How can assessment designers, specifically those seeking to screen students at risk of not meeting grade level learning objectives or to diagnose specific learning gaps, increase the engagement of assessment-takers in order to obtain the most accurate and useful information about individual abilities and needs? Games offer a tempting solution. People, not just children, enjoy playing games. Video game players, in particular, exhibit sustained engagement and embrace, and often seek, challenges that stretch their abilities (Yee, 2006; Koster, 2013; Hunicke et al., 2004; Ke, Xie, & Xie, 2016). The existing evidence related to the application of games to assessment, however, offer conflicting results and unclear guidance. We seek to identify the elements of gamification, game-based learning, and game-based assessment that can be used

to further student engagement in the screening and diagnostic assessment contexts.

The goal of assessments in education is, quite simply, to collect information about the current state of a learner and to make the data of that assessment palatable and usable by parents, teachers, school administrators, and other educational practitioners. Capturing a kindergarten student's current ability, for example, to decode words, can inform teachers as to what instruction is most appropriate to further develop or master those skills. Screening assessments that are predictive of future struggles with reading words, language development, or comprehending text can prompt preventative interventions (Jenkins et al., 2007). The quality of screening assessments matters if the results are to reliably match need with appropriate instruction and intervention (Catts & Hogan, 2020). Students taking these assessments are typically oriented to the assessment through practice items, but do not get help on the items while they are assessed, and they do not find out if their responses were correct.

It is often the case that when assessment developers are creating items and validating the scores from data collected in field studies, the form of the items often looks very different during a data collection compared to the final, finished product. For example, pilot and calibration data may be collected through functional paper-pencil or computerized means lacking aesthetic or product labeling. Once the assessment is ready for dissemination, the overall functionality of the items remains the same but aesthetic for presentation has likely changed (e.g., adding branding or color schemes to record sheets). In this manner, assessment developers seek to preserve the fidelity of the assessment's core functionality while ensuring the newly applied aesthetic form does not distract from its function. As assessment developers may increasingly pursue gaming elements as a form of aesthetic design to improve student engagement and effort in the assessment, they must do so in a manner that preserves the core functionality and characteristics of the screening and diagnostic assessments.

From Gamification to Gaming Elements

The terminology related to gaming, learning and assessment can be confusing. The same terms have been used to describe different applications. Since the early analyses of the intersection of gaming and learning, "gamification" has referred to the use of game elements - such as points, badges, progress bars, feedback, or avatars - in non-game environments (Deterding et al., 2011; Plass, Homer, & Kinzer, 2015; Martí-Parreño, Méndez-Ibañez, & Alonso-Arroyo, 2016) with the goal of improving engagement and enhancing the user experience. Game-based learning, on the other hand, embeds explicit learning objectives into game play. Playing a game inevitably involves learning. The player must learn the rules of the game, the mechanics of playing the game, and strategies to win the game (Koster, 2013). Weaving learning content into compelling game play and vice versa can foster desirable learning outcomes (Lumsden, et al., 2016; Abdul Jabbar & Felicia, 2015). Game-based assessment has the potential for embedding opportunities to collect data about a player's knowledge, skill, or other desired construct directly into a game (Shaffer, et al., 2005; Klopfer et al., 2015; Shute, Ke, & Wang, 2017; Shute & Sun, 2019).

The categories described above, however, have been challenged as overly broad and ambiguous to support rigorous evaluation of their efficacy in educational contexts (Bedwell et al., 2012; Landers, 2014). Rather than comparing, for instance, gamified features to game-based learning approaches, we should consider the value of specific game design elements under different conditions and contexts (Sailer et. al., 2017), including, when appropriate, whether they occur within a game or outside of it. While recent research has offered different taxonomies of game attributes and design elements, those studies have shared a common focus: identifying the appropriate application of specific gaming features to targeted educational purposes and conditions. Since diagnostic assessments have critical and distinctive features, we are particularly attuned to the intersection of those features and game design elements most relevant to the requirements of that context.

Diagnostic Assessment x Gamified Elements

Traditional diagnostic assessments are not games. Scores from developed and tested items are rigorously evaluated to ensure they provide reliable and valid evidence of targeted learner competencies. Those tasks are then delivered in controlled and consistent conditions across administrations, without immediate feedback or support, so that results are reliable and comparable across populations and over time. In

addition, educational assessments target data collection for the purpose of understanding student knowledge or progress, not new learning (American Educational Research Association et al., 2014). One's performance on an assessment may be informed, influenced, or explained by a number of factors including their effort on the task or their attitudes toward the topic being measured (Petscher, 2010). Can these carefully designed assessment tasks be surrounded by some of the motivating elements of games - autonomy, sense of progress, and rewards, for instance - to encourage more effortful engagement?

The results of these types of gamification experiments have been mixed (Hanus & Fox, 2015; Domínguez, et al., 2013; Attali & Attali, 2015). Highlighting progress toward an endpoint can enhance sustained engagement (Siemens, et al., 2015), but there's a risk that goal attainment - finishing - becomes more important than high effort on each individual task (Dockterman & Weber, 2017; Deright & Jorgensen, 2015). The fastest way to make progress is to complete each task quickly. Since diagnostic assessments don't include performance feedback, putting in little effort has no obvious negative consequences. Similarly, when game elements - points, badges, or even the opportunity to play a "fun" game - are used as rewards for completing assessment tasks, the implicit message is that the learning/assessment activities are so unpleasant that they need external rewards to entice completion. Again, to the child, completion rather than high effort might become the goal, since finishing the task quickly is the fastest way to reach the reward. Research on motivation (for instance, Deci, Koestner, and Ryan, 2001; Lepper & Henderlong, 2000) warns that extrinsic rewards can undermine intrinsic drives. In the context of diagnostic assessment, some common motivation-related game design elements may not be the most effective direction for engaging children in sustained, high effort.

Adding rules and goals, elements of game-based learning and assessment environments, can also potentially undermine task-focused effort. The demands of learning the game rules, for example, may interfere with the goal of assessing the targeted learner traits within that game. The design of a game-based diagnostic assessment must be done in a way to minimize potential distractions while maximizing data quality including the reliability and validity of scores.

Several standards do exist to guide the development of new game-based assessments. Evidence Centered Design (ECD) principles, in particular, can support the creation of tasks that can generate the desired relevant, measurable data (Mislevy, Almond, & Lukas, 2003; Klopfer et al., 2018; Shute & Sun, 2019). For instance, if you want to measure a child's decoding ability, you will need to describe the kinds of observations and behaviors that would be reflective of that ability. You then need to design the kinds of tasks that would expose that evidence. Finally, you need to embed versions of those tasks into a game in a way that allows for the collection of data that may produce reliable and valid scores. As well, the Standards for Educational and Psychological Testing (American Educational Research Association, 2014) provide standards for establishing and documenting the reliability and validity of scores, fairness in testing, test design and development, scaling and norming, test administration, and the rights and responsibilities of test takers and test users. Test developers often use these standards to tether the ECD to then better plan for and validate scores from new assessments.

Although game-based assessments can be readily created to generate helpful formative (lower stakes and lower fidelity) information about a student (see Groff, et al., 2015 for a description of this approach to balanced design), high fidelity screening and diagnostics assessments require a more rigorous application of ECD, which can be complex and lengthy. A body of validity evidence must be curated - Does the content for the task map to empirically supported methods for measuring the skill? Do the scores that these new tasks elicit psychometrically conform to previous evidence in terms of reliability, local fit of items, and global fit of constructs? Do participant-level scores correlate to similar constructs in the magnitude and direction that is consistent with the literature? In the case of decoding, children are typically asked to read words. The game might also require children to "decode" nonsense words, or it might include other ways to capture evidence of decoding capacity. Either way, the performance of a child in the game's decoding tasks should match the child's measured decoding capability in an existing set

of assessment tasks. Designing new tasks is challenging by itself. Designing them in a way that leverages the engaging elements of a non-distracting game is very difficult.

Game Design Elements for Effort

Rather than creating new assessment tasks within a new game environment, a more efficient approach would be to incorporate game design elements that foster task-focused effort on existing items from already validated scores from an assessment. The goal is to amplify the value of the task. Completing each task to the best of one's ability is the goal of the game. What can we leverage from games to motivate effort rather than completion or performance (since no performance feedback is provided)?

Consider this general example. A screening assessment includes a battery of tasks designed to discern relevant traits about the learner. A game introduces a storyline where doing the best possible work on those tasks is critical for moving the narrative forward. It could be that characters depend on the student trying hard to help them achieve some goal, or it could be that effort and focus feed some machine that needs the fuel to overcome obstacles. Effortful completion of each battery, regardless of performance, charts a student's progress through the game. The tasks are intrinsically important within this game structure.

A Features Analysis for Games and Screening/ Diagnostic Assessments

In this section, we identify six features of a game environment that have been shown to foster sustained focus and effort that can be complementary to current principles of ECD for researchers and testing companies who create assessments. Each of these gaming features is accompanied by a description of how those elements can be applied responsibly to screening and diagnostic assessments.

Create a sustained challenge. Overcoming challenges can be very satisfying, leading into a flow state. Tasks are neither too easy (boring) nor too difficult (frustrating) (Martí-Parreño, Méndez-Ibañez, & Alonso-Arroyo, 2016, Csikszentmihalyi, 1975; Vygotsky, 1978) as they are targeted to a player or user's zone of proximal development (ZPD). Applied to screening and diagnostic assessments, use computer adaptive assessments (CAA). CAAs are tasks that dynamically adjust the items that are presented based on the difficulty of the item and the ability of the person. Given this, CAAs are more likely to remain appropriately challenging for the child. Banks of fixed items do not always follow a progression of difficulty. It may be that the adaptivity in screening and diagnostic assessment moves from one construct to the next as soon as a result can be determined from the data. The child will not have to suffer through a barrage of overly difficult items nor be bored with items that are too easy. Computer adaptive assessments in early education and beyond provide a means for item and task adaptivity targeted to the user (Mitchell et al., 2015) and can leverage information in the gaming and assessment environment that move the assessment more quickly without sacrificing the reliability or validity of scores (Petscher et al., 2015; Petscher et al., 2017).

Create low stakes outcomes, at least from the child's perspective. Just as challenges need to exist in the zone between boredom and frustration, anxiety too has a sweet spot. When the perceived consequences of failure are high, players can become overly anxious, a state that can impede performance. On the other hand, when there are no stakes, the cortisol and adrenaline that fuel attention and reflexes may not kick in (Klopfer, Haas, Osterweil & Rosenheck, 2018). Applied to screening and diagnostic assessments, frame the assessment positively. Assessments in general are often viewed as high stakes activities. Unlike in games, failure on an assessment has perceived negative consequences. Screening and diagnostic assessments are designed to inform better instruction, not to judge the test-taker. Build productive language into the assessment itself as appropriate. In addition, provide guidance on positive framing to teachers or others who will be supporting the administration of the assessment. Context matters.

Include feedback and scaffolding. Feedback and scaffolds facilitate learning, helping players overcome challenges, particularly in low-stakes contexts that allow graceful failure and multiple retries (Henderlong & Lepper, 2002; Clark et al., 2016). Applied to screening and diagnostic assessments, focus on between-task effort-related feedback. An adaptive delivery system can monitor elements of effort, like time per item and patterned responses (e.g., always picking the first option), and offer personalized

messages to encourage more concentration and thoughtfulness, as appropriate. Be aware that different audiences will respond differently to different kinds of effort-related feedback. Depending on the social context, young children, for instance, may be more likely to embrace effort messages than adolescents, who may value natural talent over hard work (Amemiya & Wang, 2018). As with the assessment items themselves, feedback information needs to be tested and validated before incorporation into a final assessment system.

Include a system that shows progress. A sense of growing competence and movement can be compelling. Progress can be made visible when players see their scores and ranking improve. Progress can also be seen in movement through a journey or narrative (Siemens, et al., 2015). Applied to screening and diagnostic assessments, show progress through the journey and not the performance. Screening and diagnostic assessments do not keep score in a way that the user can see, and variable-length adaptive assessments are unable to determine how many items a child will be administered beforehand, so it cannot and should not show the number of questions left to complete. Progress, though, can be portrayed episodically. For example, when multiple constructs are being measured, each construct can be attached to locations on a map or chapters in a narrative or characters to meet. Each assessment experience can be pristine within a larger, finite structure. As children complete each part of the structure, they get a sense of movement to completion.

Create a sense of autonomy. A sense of locus of control and agency can feed motivation (Anderson & Gray, 2015; Lepper & Henderlong, 2000). Choice can take many forms, from selecting levels and game strategies to choosing which characters or parts of a narrative to engage. Applied to screening and diagnostic assessments, provide choice where possible. Sometimes the constructs being assessed are progressive; sometimes they can be completed in any order. When possible, consider letting the child choose what to do next. Properly structured narratives can provide opportunities for choice by letting children choose a story path or character to follow and support. Make sure you keep choices centered on the themes of focus and high-quality effort.

Create a sense of relatedness. When a player feels related to the content or has a social connection to other players, effort and attention may be enhanced (Ke, Xie, & Xie, 2016). Applied to screening and diagnostic assessments, make trying hard important. If children get no reward or feedback for performance on a diagnostic assessment, why should they care? The answer is unlikely the same for all children, if only for developmental reasons, but research provides some clues about how to increase the value of high-quality effort. For older children who have self-regulation and meta-cognitive capabilities, appealing to the possibility of improved future academic performance may help (Yeager, et al., 2014). Younger children may be eager to please those they care about (Walton & Brady, 2017), whether family, teachers, or characters in a narrative.

Discussion

Cautions with Feature Analysis in Screening/ Diagnostic Assessments

The lack of consistent results in prior assessment and gaming research may well reflect variations in implementation. While the evidence points to potentially productive features, as described above, it also reveals areas of caution. If not applied appropriately, the use of game design elements in a diagnostic assessment context can actually undermine the goals of the assessment. Here are some key warnings we distilled from the research.

First, avoid including activities that might compete for limited available time to complete the diagnostic assessment. Concern about the amount of testing in schools and how it might crowd out time for instruction has grown in recent years (Hart, et al, 2015). Assessment needs to be time-sensitive. Time spent customizing a character's appearance or buying cosmetic accessories takes away time from the diagnostic tasks. The game overhead, in this context, needs to be very efficient.

Second, as mentioned previously, avoid the use of rewards, such as accruing points or mini games in the midst of assessment, that may encourage children to rush through the assessment to get to the games. The goal is to elevate the value of the diagnostic activities,

not send the subliminal message that they are only worth doing for an extrinsic reward.

Third, avoid *distracting* graphics. Engaging visuals may help enroll children in the narrative context, but attention is a limited resource (Sweller, 1994; Sweller, Aures, & Kalyuga, 2011). Designing instructional graphics and animations effectively is a challenge by itself (Mayer & Moreno, 2003; Ng, Kalyuga, & Sweller, 2013). Extraneous visuals related to the game context risk drawing student attention away from an assessment task.

Fourth, be wary of the cost of narrative overhead. Stories can be motivating but they can also take time to set up. In addition, the elements of the narrative could compete for the limited attentional resources just mentioned. Keep narratives simple, quick, and accessible.

Games offer a tantalizing vehicle for engaging children in assessments. Game players, after all, typically exhibit the kinds of focused attention, effort, and persistence assessment designers hope children will display while completing screening or diagnostic tasks. Existing research suggests game features that should induce sustained, high-quality effort among test-takers. That same research also reveals how the implementation of gaming in assessment might lead to the opposite result. Thoughtful, iterative assessment design should be coupled with rigorous formative evaluation to continue to clarify effective from ineffective practices.

References

Abdul Jabbar, A., & Felicia, P. (2015). Gameplay Engagement and Learning in Game-Based Learning: A Systematic Review. *Review of Educational Research, 85(4),* 740-779.

Alexander, R, Doddington, C, Gray, J, Hargreaves, L, & Kershner, R (eds) 2009, *The Cambridge Primary Review Research Surveys*, Taylor & Francis Group, London. Available from: ProQuest Ebook Central. [14 April 2020].

Amemiya, J., & Wang, M. (2018). Why Effort Praise Can Backfire in Adolescence. *Child Development Perspectives, 12(3),* 199-203.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (Eds.). (2014). *Standards for educational and psychological testing.* American Educational Research Association.

Anderman, Eric M., & Gray, DeLeon. (2015). Motivation, Learning, and Instruction. In *International Encyclopedia of the Social & Behavioral Sciences* (pp. 928-935).

Attali, Y., & Arieli-Attali, M. (2015). Gamification in assessment: Do points affect test performance? *Computers & Education, 83(2),* 57-63.

Bedwell, Wendy L, Pavlas, Davin, Heyne, Kyle, Lazzara, Elizabeth H, & Salas, Eduardo. (2012). Toward a Taxonomy Linking Game Attributes to Learning. *Simulation & Gaming*, 43(6), 729–760.

Catts, H.W, & Hogan, T.P. (2020). Dyslexia: An ounce of prevention is better than a pound of diagnosis and treatment. https://psyarxiv.com/nvgje/

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.

Csikszentmihalyi, M. (1975). Beyond boredom and anxiety (1st ed., The Jossey-Bass behavioral science series). San Francisco: Jossey-Bass.

Deci, E., Koestner, R., & Ryan, R. (2001). Extrinsic Rewards and Intrinsic Motivation in Education: Reconsidered Once Again. *Review of Educational Research, 71(1),* 1-27.

Deright, J., & Jorgensen, R. (2015). I Just Want My Research Credit: Frequency of Suboptimal Effort in a Non-Clinical Healthy Undergraduate Sample. *The Clinical Neuropsychologist*, 29(1), 101-117.

Dockterman, D., & Weber, C. (2017). Does stressing performance goals lead to too much, well, stress? *Phi Delta Kappan, 98(6),* 31–34.

Domínguez, A., Saenz-de-Navarrete, J., de-Marcos, L., Fernandez-Sanz, L., Pages, C., & Martínez-Herraiz, J. J. (2013). Gamifying learning experiences: practical implications and outcomes. *Computers & Education, 63*, 380-392.

Groff, J., Clarke-Midura, J., Owen, V.E., Rosenheck, L., & Beall, M. (2015). *Better Learning in Games: A balanced design lens for a new generation of learning games* [white paper]. Cambridge, MA: MIT Education Arcade and Learning Games Network.

Hanus, M. D., & Fox, J. (2015). Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction,

effort, and academic performance. *Computers & Education, 80,* 152-161.

Hart, R., Casserly, Uzzell, Palacios, Corcoran, Spurgeon, & Council of the Great City Schools. (2015). *Council of the Great City Schools,* Council of the Great City Schools.

Henderlong, J., & Lepper, M. (2002). The Effects of Praise on Children's Intrinsic Motivation: A Review and Synthesis. *Psychological Bulletin, 128(5),* 774-795.

Jim, L., Elizabeth, E., & Marcus, M. (2016). Gamification of cognitive assessment and cognitive training: A systematic review of applications, approaches and efficacy. *Frontiers in Public Health, 4, V*ol.4.

Ke, Fengfeng, Xie, Kui, & Xie, Ying. (2016). Game-based learning engagement: A theory- and data-driven exploration. *British Journal of Educational Technology, 47(6),* 1183-1201.

Klopfer, E., Haas, J., Osterweil, S., & Rosenheck, L. (2018). *Resonant Games: Design Principles for Learning Games that Connect Hearts, Minds, and the Everyday*. MIT Press.

Koster, R. (2013). *A theory of fun for game design* (Second ed.). North Sebastopol, California: O'Reilly Media.

Landers, Richard N. (2014). Developing a Theory of Gamified Learning. *Simulation & Gaming*, 45(6), 752–768. https://doi.org/10.1177/1046878114563660

Lepper, Mark R., & Henderlong, Jennifer. (2000). Turning "play" into "work" and "work" into "play"-Chapter 10:25 Years of research on intrinsic versus extrinsic motivation. In Intrinsic and Extrinsic Motivation (pp. 257-307).

Lumsden, J., Edwards, E., Lawrence, N., Coyle, D., & Munafo, M. (2016). Gamification of Cognitive Assessment and Cognitive Training: A Systematic Review of Applications and Efficacy. *Jmir Serious Games*, 4(2), E11.

Martí-Parreño, J., Méndez-Ibáñez, E., & Alonso-Arroyo, A. (2016). The use of gamification in education: A bibliometric and text mining analysis. *Journal of Computer Assisted Learning, 32(6),* 663-676.

Mayer, R., & Moreno, R. (2003). Nine Ways to Reduce Cognitive Load in Multimedia Learning. *Educational Psychologist*, 38(1), 43-52.

Mitchell, A. M., Truckenmiller, A., & Petscher, Y. (2015). Computer-Adaptive Assessments: Fundamentals and Considerations. *Communique, 43(8),* 1-22.

Ng, H., Kalyuga, S., & Sweller, J. (2013). Reducing transience during animation: A cognitive load perspective. *Educational Psychology*, 33(7), 755-772.

Petscher, Y., Foorman, B. R., & Truckenmiller, A. J. (2017). The impact of item dependency on the efficiency of testing and reliability of student scores from a computer adaptive assessment of reading comprehension. *Journal of Research on Educational Effectiveness, 10(2)*, 408-423.

Petscher, Y., Mitchell, A. M., & Foorman, B. R. (2015). Improving the reliability of student scores from speeded assessments: an illustration of conditional item response theory using a computer-administered measure of vocabulary. *Reading and writing, 28(1),* 31-56.

Petscher, Y., Stanley, C., & Pentimonti, J. (2019). Overall screening and assessment. Washington, DC: U.S. Department of Education, Office of Elementary and Secondary Education, Office of Special Education Programs, National Center on Improving Literacy. Retrieved from improvingliteracy.org.

Sailer, Michael, Hense, Jan Ulrich, Mayr, Sarah Katharina, & Mandl, Heinz. (2017). How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction. *Computers in Human Behavior*, 69, 371–380. https://doi.org/10.1016/j.chb.2016.12.033

Segool, N.K., Carlson, J.S., Goforth, A.N., von der Embse, N. and Barterian, J.A. (2013), Heightened Test Anxiety Among Young Children: Elementary School Students' Anxious Responses To High-stakes Testing. *Psychol. Schs., 50:* 489-499. doi:10.1002/pits.21689

Siemens, J., Smith, S., Fisher, D., Thyroff, A., & Killian, G. (2015). Level Up! The Role of Progress Feedback Type for Encouraging Intrinsic Motivation and Positive Brand Attitudes in Public Versus Private Gaming Contexts. *Journal of Interactive Marketing, 32(C),* 1-12.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, *4*, 295-312.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York, NY: Springer.

Von Der Embse, N., Jester, D., Roy, D., & Post, J. (2018). Test anxiety effects, predictors, and correlates: A 30-year meta-analytic review. *Journal of Affective Disorders, 227*, 483-493.

Vygotskiĭ, L., & Cole, M. (1978). *Mind in society: The development of higher psychological processes*. Cambridge: Harvard University Press.

Walton, G. M., & Brady, S. T. (2017). The many questions of belonging. In A. J. Elliot, C. S. Dweck, & D. S. Yeager (Eds.), *Handbook of competence and motivation: Theory and application* (p. 272–293). The Guilford Press.

Yeager, D.S., Henderson, M., Paunesku, D., Walton, G., Spitzer, B., D'Mello, S., & Duckworth, A.L. (2014). Boring but important: A self-transcendent purpose for learning fosters academic self-regulation. *Journal of Personality and Social Psychology, 107,* 559-580.

Citation:

Dockterman, D., Petscher, Y., McAfee, A., Klopfer, E., Osterwell, S., & Diefenthaler, C. (2021). Game design considerations for screening, interim, and diagnostic educational assessments. *Practical Assessment, Research & Evaluation*, 26(27). Available online: https://scholarwoarks.umass.edu/pare/vol26/iss1/27/

Corresponding Author

David Dockterman
Harvard Graduate School of Education
Harvard University
Cambridge, Massachusetts, USA

Email: david_dockterman [at] gse.harvard.edu