WHITE PAPER

Technology-Enhanced Items: Moving Beyond Multiple-Choice

ClearSight

INTRODUCTION

Researchers and experts in the field of K–12 assessment have expressed concern about overreliance on multiple choice questions. Recent assessment systems incorporate technology-enhanced items (TEIs) that can better engage and motivate students, improve assessment validity, capture higher-order cognitive skills, and provide for greater accessibility for students with disabilities. Such systems are capable of automated scoring that provide informative feedback about student writing with the potential to significantly enhance writing outcomes.

ClearSight is a robust K–12 assessment system that offers standards-aligned Checkpoint and Interim Assessments for Reading, Writing, Editing, Listening, Math, and Science. *ClearSight* incorporates rigorous, validated TEIs and auto-scored essay items, accommodations and accessibility features for students with disabilities, and reporting that helps identify learning gaps and needed interventions.

This paper, researched and prepared by Interactive Educational Systems Design (IESD) of New York, summarizes research and expert opinion about impactful K–12 assessment, and explains how *ClearSight* aligns with the research literature.

Research supports shifting away from an overreliance on traditional multiple-choice items on assessments.

Computer-based assessments have relied predominantly on multiple-choice (MC) items because they have been quick and inexpensive to score (Darling-Hammond & Adamson, 2010; Scalise & Gifford, 2006). Researchers and assessment experts have expressed concern that an overreliance on MC questions fails to take advantage of the potential for rich and embedded assessments provided by online testing (Scalise & Gifford, 2006).

Research indicates that an overreliance on MC questions has several significant disadvantages.

- MC questions tend to focus on measuring lower-level skills such as recall or recognition of information, but they are less effective in measuring deeper learning. MC questions cannot measure communication skills—e.g., whether students can write a logically structured essay (Livingston, 2009), express a point of view and support it with evidence (Darling-Hammond & Adamson, 2010), or communicate scientific understanding (Federer et al., 2014, citing multiple sources). MC questions are also inadequate for measuring the depth of students' abilities to organize and synthesize knowledge (Federer et al., 2014, citing multiple sources). Given their limitations, overreliance on MC questions can give inaccurate information about what students actually understand (Darling-Hammond & Adamson, 2010, citing Sheperd, 2008).
- MC assessments do not measure authentic real-world skills and abilities that require students to go beyond selecting one of several answer choices (Stecher, 2010). Thus, they are less able to predict real-world performance (Federer et al., 2014, citing multiple sources).
- Assessments that rely heavily on MC questions can "drive" instruction that mimics this format (Darling-Hammond & Adamson, 2010, citing Madaus et al., 1992), resulting in students being unable to transfer their knowledge to assessments that measure deeper understanding, e.g., through an open-ended format (Darling-Hammond & Adamson, 2010, citing Sheperd, 2008).
- Cognitive theorists argue that MC questions suggest to students that complex skills can be overly simplified and broken into component parts—implying that "knowledge is additive rather than integrative of developing knowledge structures" (Scalise & Gifford, 2006, p. 5, citing multiple sources).

Research supports incorporating TEIs

Technology-enhanced items (TEIs) are computer-based question types that use formats requiring responses that differ from conventional multiple-choice and short-answer questions (Bryant, 2017). Examples of TEIs include drag and drop, hot spot, simulation, table input, and extended response. There has been increased usage of TEIs to counter criticisms leveled at MC-only assessments, and in response to school districts seeking machine-scorable items that are more authentic, engaging, and demanding (Bryant, 2017, citing multiple sources).

Researchers and assessment experts hypothesize several advantages of TEIs.

- TEIs can capture hard-to-measure constructs and evidence of deeper learning. A recent review of research about TEIs concludes their greatest potential is their capacity to measure higher-order cognitive skills such as reasoning and real-world problem solving—thus broadening the types of learning that are assessed and increasing assessment validity (Bryant, 2017). Examples provided in this review include highlighting and annotating texts (similar to when reading in authentic contexts), and math equation editors that allow construction rather than selection. (Bryant, 2017, citing multiple sources). An investigation by the Institute for Credentialing Excellence and the Association of Test Publishers (ATP) concluded TEIs are superior to MC questions in measuring higher-order cognitive skills (ATP, 2017).
- Researchers postulate that TEIs can increase student engagement—by providing more authentic assessment activities such as real-world challenges and encouraging use of problem-solving abilities and higher-order thinking skills—and by providing novel tasks due to some TEI's dynamic properties (Cayton-Hodges et al., 2012). Several researchers assert that when student engagement during assessment increases, assessment accuracy and validity also increases (Bryant, 2017, citing SCOPE SCALE, 2015; Cayton-Hodges et al., 2012). More specifically in literacy assessment, evidence suggests that less engaging MC questions tend to underestimate student skills (Cayton-Hodges et al., 2012, citing multiple sources).
- TEIs provide for greater accessibility. Research indicates that computer-based assessments have the capacity to improve access for students with disabilities (Almond et al., 2010, citing multiple sources)—in particular by incorporating assistive technologies into the assessment platforms (Cayton-Hodges et al., 2012).

ETS researchers recommend an evidence-based design to address the needs of most students, additional features to improve accessibility for some students, and alternative assessment strategies for students who need greater support (Cayton-Hodges et al., 2012). Other researchers recommend applying Universal Design for Learning (UDL) to TEIs to provide for greater accessibility, including flexible presentation of content, flexible engagement opportunities with content, integration of assistive technologies, and alternative versions of text-based content (Almond et al., 2010).

Research supports including open-response essay items in computer-based systems to assess writing.

It is critical that students develop strong writing skills to succeed in post-secondary education and the workplace, and providing students with informative feedback is crucial to the process of learning to write (Allen, Jacovina, & McNamara, 2016). Feedback gives students support for altering metacognitions to improve their writing skills (Shermis et al., 2008).

Recent assessment systems have the potential to significantly improve writing outcomes. Teachers are unable to give substantive feedback due to time constraints and large class sizes (Allen et al., 2016). Computer-based assessments can enhance writing outcomes by providing automated scoring to give nonjudgmental, informative feedback to help students improve their learning outcomes and keep them engaged (Allen et al., 2016, citing Gikandi, Morrow, & Davis, 2011; Shermis et al., 2008).

In addition to feedback, research indicates that extended writing practice is essential for developing effective writing skills (Allen et al., 2016, citing multiple sources). Automated Essay Scoring—technology to evaluate the content, structure, and quality of student writing—provides many more opportunities for writing practice and is an efficient way to allow teachers to assign more writing practice without increasing their workload (Allen et al., 2016, citing multiple sources). Automated scoring can greatly decrease the time and cost of scoring student assessments (Livingston, 2009).

More essay questions on assessments shifts instructional practice toward a focus on extended writing. Standards-based assessments send a message to educators about the types of content, learning, and performances that are of most value. A robust body of research suggests that teachers change their teaching practices accordingly.

With greater use of performance assessments—including essay writing—and less reliance on multiple choice questions, teachers tend to emphasize more extended writing in language arts teaching (Stecher, 2010, citing Stecher & Mitchell, 1995).

Research confirms that Automated Essay Scoring provides reliable and valid scores on essays. Automated Essay Scoring tends to be highly reliable and accurate (Allen et al., 2016, citing multiple sources). The correlation between machine scoring and human scoring of essays is high[1] (Allen et al., 2016, citing multiple sources; Shermis et al., 2008, citing multiple sources; Stecher, 2010).

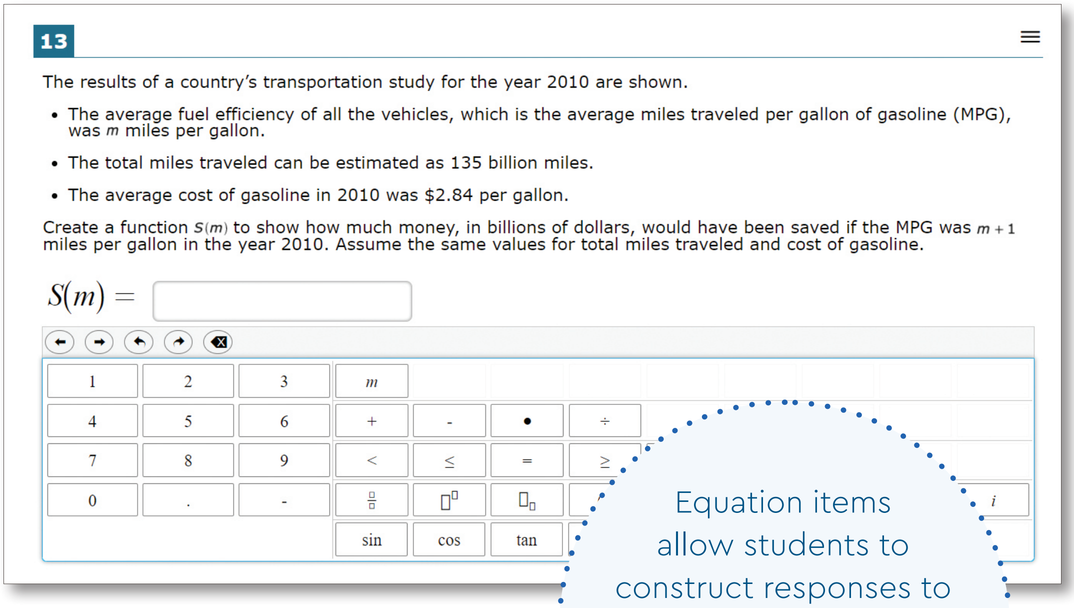
[1] ". . . expert human and computer scores tend to correlate between r = .80 and .85, which is similar to the range found between two human raters" (Allen et al., 2016, p. 318, citing multiple sources).

How *ClearSight* Aligns with the Research

ClearSight takes advantage of recent assessment technologies to provide Checkpoint and Interim Assessments that incorporate a variety of engaging, meaningful item types—with reporting that helps educators discern what students know and can do throughout the depth and breadth of college- and career-ready curriculum standards. *ClearSight* offers the efficiency of automated assessment without overreliance on traditional multiple-choice items. With assessment items created using an evidence-based design process, *ClearSight* provides strong, valid evidence of student learning.

ClearSight **provides an array of technology-enhanced item types (TEIs).**

ClearSight TEIs include drag and drop, hot spot, hot text, editing tasks, table input, equation response (for Math), and other interactive item types to elicit evidence of deeper learning than traditional multiple-choice items. Each TEI type can be machine-scored. Examples include the following:



Equation items allow students to construct responses to real-world Mathematics problems using a keypad with numbers and symbols.

The Creator of Oz
by Shawn Hoffelt

1 In 1900, a group of children listened to an unusual tale with characters made of straw, tin, and fur. L. Frank Baum would often entertain his sons and their friends with his creations. This one, about a girl from Kansas who travels to a wonderful enchanted land, was their favorite.

2 In the middle of the tale, a child asked, "Mr. Baum, where did the Scarecrow and the Tin Woodman live?"

3 Baum looked around the room until he noticed a filing cabinet with the letters A-N on the top drawer and O-Z on the bottom. Putting the last two letters together, he smiled.

4 "Well," Baum said, "the Scarecrow, the Tin Woodman, the Cowardly Lion, and the Great Wizard all lived in the marvelous land of Oz."

Growing Up

5 Lyman Frank Baum was born to a wealthy family in Chittenango, New York, in 1856. He spent most of his childhood living at a beautiful country estate called Rose Lawn. A serious heart condition made it necessary for him to be tutored at home, and

11

Select **two** sentences from the section "Growing Up" that the author uses to support the idea that L. Frank Baum was interested in writing as a child.

Growing Up

5 Lyman Frank Baum was born to a wealthy family in Chittenango, New York, in 1856. He spent most of his childhood living at a beautiful country estate called Rose Lawn. A serious heart condition made it necessary for him to be tutored at home, and one of his favorite pastimes was reading and creating stories. At a young age, Frank, as he preferred to be called, showed a talent for writing, and when he was fifteen, his father bought him a printing press. Frank and his brother, Harry, printed a popular newspaper entitled the *Rose Lawn Home Journal*.

6 By the time Frank turned eighteen, he knew that he wanted to become an actor. With money from his father, Frank formed a Shakespearean acting troupe, and in 1880 became the manager of a string of opera houses owned by his father. In 1881, his first play, *The Maid of Arran,* in which he played the lead, became an instant success.

In hot text and hot spot items, students select text, shapes, or other response elements to answer a question.

21

Sami has 6 times as many books as Jeff.

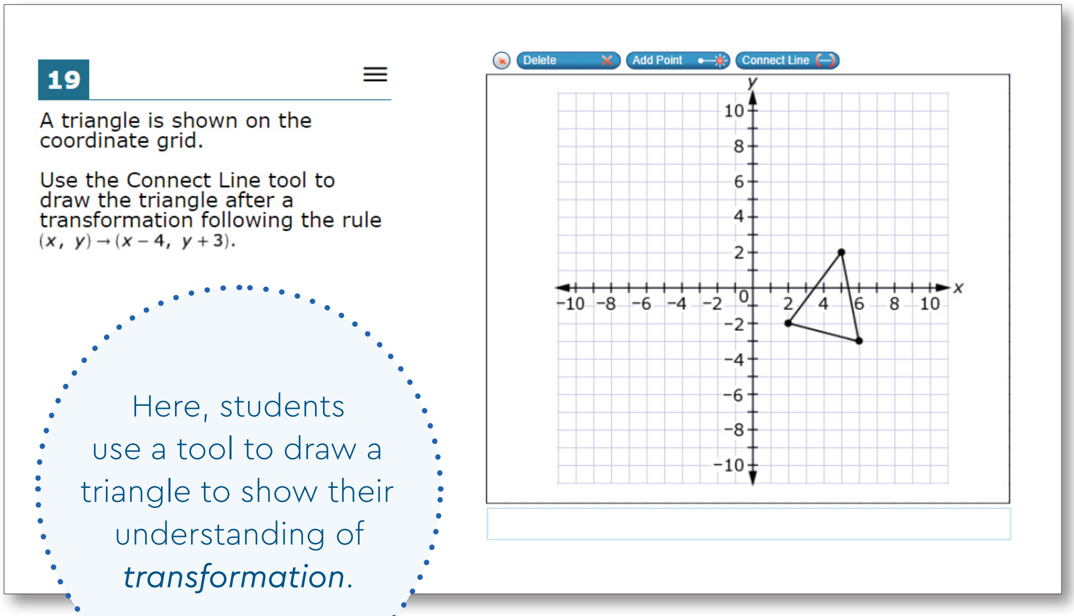
Complete the table to show three different possible amounts of books Sami and Jeff could have.

Sami's Books	Jeff's Books
12	
	4

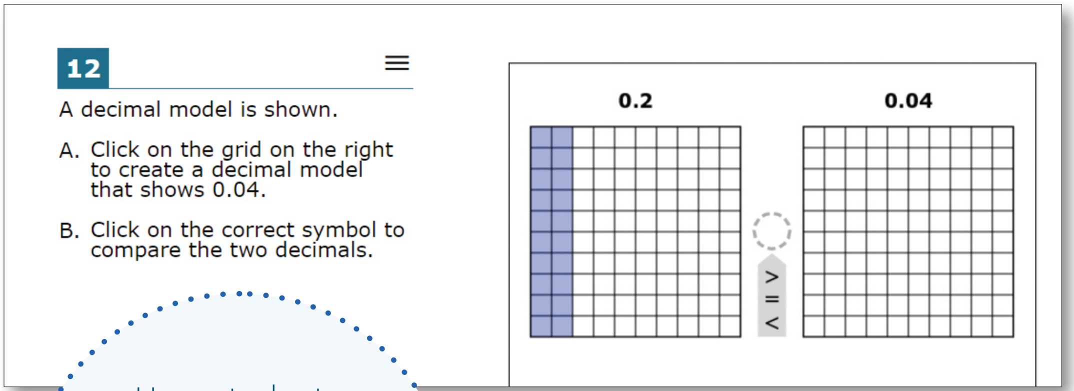
In table interaction items, students fill in a table to represent their thinking and their ability to apply what they have learned.

ClearSight TEIs engage students in active problem solving.

- In some items, students are prompted to drag steps in a process to indicate the proper sequence.
- In other items, students manipulate charts to demonstrate their understanding of data.



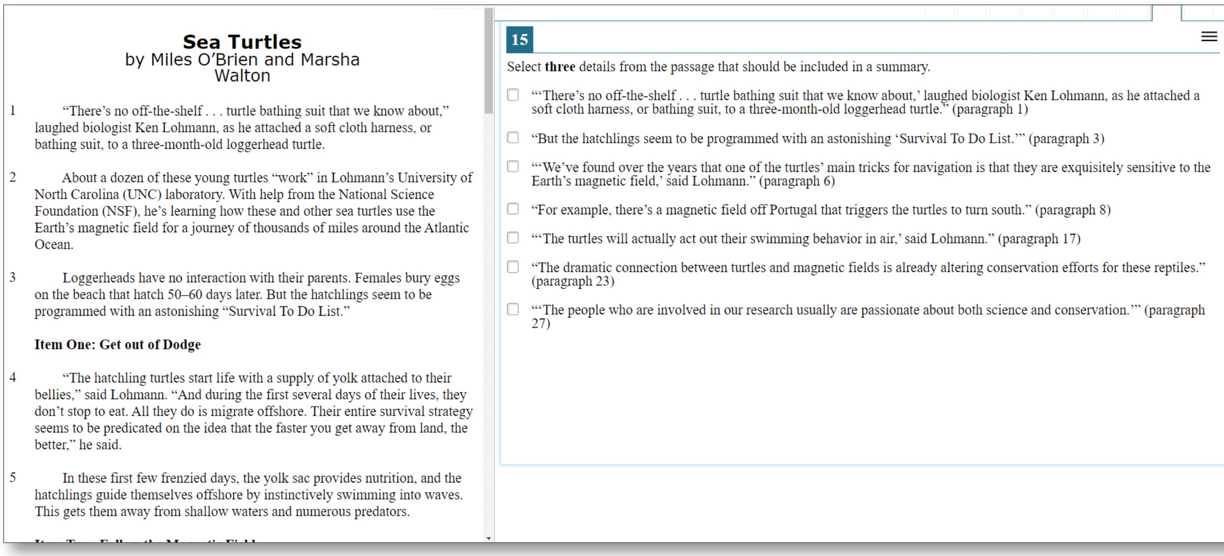
Here, students use a tool to draw a triangle to show their understanding of *transformation*.

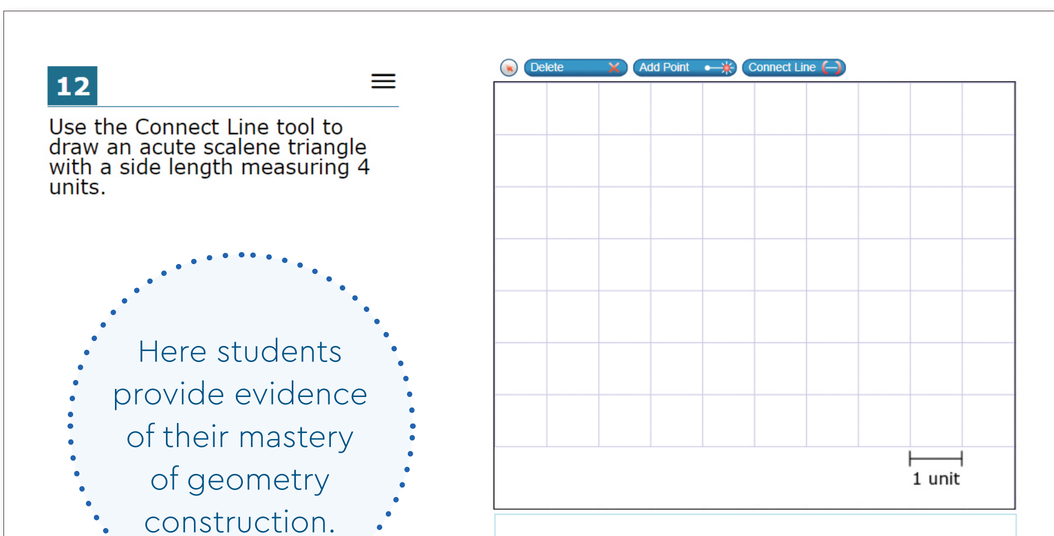


Here students click on squares in a grid to demonstrate their understanding of decimals, then compare decimals.

ClearSight TEIs can address difficult-to-measure skills and understandings.

In typical assessments of reading comprehension, multiple-choice items can be manipulated by students using test-taking tricks for eliminating options and guessing. In the *ClearSight* example below, students are asked to choose multiple details to be included in a summary, thus measuring whether they can identify the key points of the reading passage.

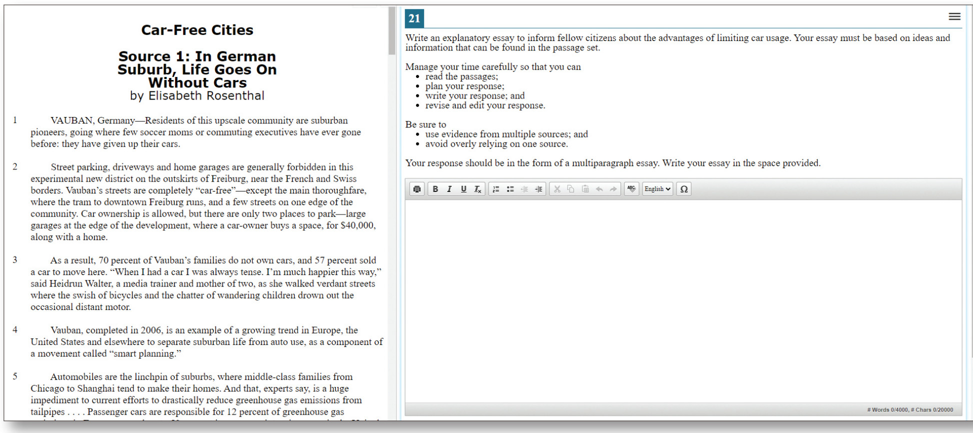




Here students provide evidence of their mastery of geometry construction.

ClearSight Interim Assessments assess writing through automated scoring of extended response essays.

ClearSight Interim Assessment essay items typically engage students in reading multiple sources on a topic and then writing an essay of a specified type (e.g., explanatory, persuasive). Each essay item is designed to be age- and grade-appropriate.



ClearSight's Automated Essay Scoring uses artificial intelligence (AI) to model how human raters would assign scores to essays. The *ClearSight* AI "engine" is trained on each specific essay question. The engine is "taught" how to predict human responses on a specific prompt by exposing the engine to example essays and scores provided by experienced and trained human scorers. After this initial calibration, the engine goes through an extensive quality-control process conducted by professional psychometricians to ensure the level of agreement between the engine and human raters is similar to the level between two human raters.

ClearSight's Automated Essay Scoring provides informative feedback about important aspects of writing. Rather than just providing a single rating of student writing, *ClearSight* analyzes and rates several key features of each essay: organization, language used, sentence variety and complexity, grammar, and spelling. Feature-specific ratings serve as informative feedback to teachers and students to identify which aspects of writing to focus on.

ClearSight includes extensive accommodation and accessibility features for students with special needs.

Some tools are available to all students, including:

- Help
- Calculator
- Dictionary
- Line reader (enabling students to highlight an individual line of text)
- Masking (enabling students to cover any part of the page they want to mask)
- Notes (enabling students to take notes about the assessment)
- Zoom
- American Sign Language videos
- Expand (enabling magnification of a section of a page for greater readability)
- Highlighter
- Strikethrough
- Text-to-speech (if turned on)
- Notepad (enabling note-taking about the current item and drafting written essay responses)

Additional accommodation tools are available such as:

- Color choices (to support a student's visual needs)
- Scribe (enabling students to dictate responses to a qualified person who records verbatim what they dictate)

ClearSight also offers interoperability with most accessibility technologies.

Besides these features, the development team follows Universal Design for Learning (UDL) guidelines to ensure assessment items are appropriate for a wide range of students. Item developers ensure the reading load and graphic complexities are appropriate to the item and do not introduce factors that are irrelevant to what is being assessed. Experts then review for bias, sensitivity, and accessibility, and revisions are made as necessary.

References:

Allen, L.K., Jacovina, M.E., & McNamara, D.S. (2016). Computer-based writing instruction. In C. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research*. New York: Guilford Press. https://files.eric.ed.gov/fulltext/ED586512.pdf

Almond, P., Winter, P., Cameto, R., Russell, M., Sato, E., Clarke-Midura, J., Torres, C., Haertel, G., Dolan, R., Beddow, P., & 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(5). https://ejournals.bc.edu/index.php/jtla/article/view/1605

Association of Test Publishers (ATP) and the Institute for Credentialing Excellence (ICE) (2017). Innovative item types: A white paper & portfolio. https://www.testpublishers.org/assets/innovative%20item%20types%20w.%20appendix%20copy.pdf

Bryant, W. (2017). Developing a strategy for using technology-enhanced items in large-scale standardized tests. Practical Assessment, Research & Evaluation, 22(1). http://pareonline.net/getvn.asp?v=22&n=1

Cayton-Hodges, G., Marquez, E., Keehner, M., Laitusis, C., van Rijn, P., Zapata-Rivera, D., Bauer, M.I. & Hakkinen, M.T. (2012). Technology enhanced assessments in mathematics and beyond: Strengths, challenges, and future directions. Educational Testing Service. International Research on Technology Enhanced Assessments. https://www.ets.org/research/policy_research_reports/publications/paper/2012/jfme

Darling-Hammond, L. & Adamson, F. (2010). Beyond basic skills: The role of performance assessment in achieving 21st century standards of learning. Stanford, CA: Stanford University, Stanford Center for Opportunity Policy in Education.

Federer, M.R., Nehm, R.H., Opfer, J.E. et al. (2015). Using a constructed-response instrument to explore the effects of item position and item features on the assessment of students' written scientific explanations. Research in Science Education, 45, 527–553. https://doi-org.ezp-prod1.hul.harvard.edu/10.1007/s11165–014–9435–9

Livingston, S. (2009). Constructed-response test questions: Why we use them; how we score them. R&D Connections, 11. Educational Testing Service (ETS). https://www.ets.org/Media/Research/pdf/RD_Connections11.pdf

Scalise, K. & Gifford, B. (2006). Computer-based assessment in e-learning: A framework for constructing "intermediate constraint" questions and tasks for technology platforms. Journal of Technology, Learning, and Assessment, 4(6). https://ejournals.bc.edu/index.php/jtla/article/view/1653/1495

Shermis, M. D., Wilson, C., Diao, G.Y. (2008, March 25–27). The impact of automated essay scoring on writing outcomes. Paper presented at the Annual Meetings of the National Council on Measurement in Education, New York, NY. https://files.eric.ed.gov/fulltext/ED501148.pdf

Stecher, B. (2010). Performance assessment in an era of standards-based educational accountability. Stanford, CA: Stanford University, Stanford Center for Opportunity Policy in Education.

ClearSight

LEARN MORE

voyagersopris.com/clearsight

800.956.2860

V5 07.14.2020