

# Online Performance Assessment of Creativity Skills: Findings from International Pilot Study

Yigal Rosen ([yigal.rosen@pearson.com](mailto:yigal.rosen@pearson.com))

Pearson Research & Innovation Network, USA

and

Maryam Mosharraf ([maryam.tager@pearson.com](mailto:maryam.tager@pearson.com))

Pearson Research & Innovation Network, USA

## Abstract

Many educational programs in school settings place great emphasis on promoting student ability to create original ideas, products, and solutions that will be both novel and valuable. However, there is lack research on potential tools that can be used for creativity assessment in the K-12 arena. The aim of this study was to explore patterns in student performance and motivation in creativity computer-based prototype assessment. Eighty seven 14-year-old students from the United States, United Kingdom, Turkey, and South Africa participated in the study. Students were asked to write different short stories based on given images or a video, create titles for a story, and propose captions for a cartoon. The findings showed that, while students were relatively successful in creating cognitively complex creative writing, they struggled with writing in an original manner, expressing emotions in their writing, descriptiveness, and humor. Male students outperformed female students in originality and use of humor. The findings indicated mostly significantly positive relationships between different dimensions of creativity. However, no significant correlations were found between student creativity and factors such as motivation and time-on-task. Directions for future research will be discussed in terms of their implications to teaching, learning and assessment.

*Keywords:* creativity, performance assessment, online assessment

## **Introduction**

Much of the research on creativity and creativity assessment involved studies of adults, rather than school age children. However, yet our review found no examples of widely used and credible methods of assessing creativity in schools. The purpose of this study was to provide empirical evidence of what can be achieved by intertwining computer-based tools in a performance assessment of student creativity in terms of student performance, motivation and time-on-task. This paper provides findings from an empirical pilot study conducted in four counties, and discusses implications of the findings on further research and development.

## **Defining Creativity Skills**

Although most researchers agree that creativity involves the development of a novel product, idea, or problem solution that is of value to the individual and/or the larger social group, researchers have had great difficulty finding consensus on a definition beyond these two criteria of novelty and value (Kaufman, 2003; 2004; Sawyer, 2012). Creativity may range from a low level, as in, for instance, solving a typical insight problem, to the very high level involved in the shift of paradigms or genres involved in science and art. To build a foundation for college and career readiness students need to learn to use creative writing and creative expression. For example, the United States Common Core State Standards in English Language Arts (CCSSO, & NGA, 2010; CCSSO, 2012) require student's competency of writing narratives to develop real or imagined experiences or events, writing informative or explanatory texts to examine a topic and convey ideas, and strengthen writing by planning, revising, or trying a new approach. The students are expected to "comprehend oral and written classroom discourse about the content and craft of expository writing, narration, and other forms of creative writing" (CCSSO, 2012, p. 13). Partnership for 21st Century Skills (2009, 2011) argues that student success in college and careers requires creative writing that can be incorporated as an element of almost any performance task related to English Language Arts. At the international level, creativity was part of problem solving assessment in Program for International Student Assessment (PISA) 2012, with major emphasis on cognitive activities that result in finding solutions to a novel problem (OECD, 2010). PISA 2015 extends the need for creative thinking into interpersonal situations in which collaborative problem solving is assessed (OECD, 2013). According to Assessment and Teaching of 21st Century Skills (Binkley et al., 2012), communicating new ideas to others effectively, being open and responsive to new and diverse perspectives, and the ability to implement creative ideas to make a significant and useful contribution are among the skills required for college and career readiness.

In order to encourage the decision to be creative, one should believe that he or she will be awarded for the attempt to be more creative rather than punished (O'Hara & Sternberg, 2000–2001). In order to promote creativity there is a need to construct opportunities to engage in it, encourage, and reward when people respond to such opportunities. However, most of the conventional assessments penalize students if they try being creative (Beghetto, 2010; Sternberg, 1997). Student answers are often analyzed against prototype responses, while answers that reflect novel perspectives are discouraged. Thus, an educational and social atmosphere in which students feel free to play with ideas is essential in establishing optimal settings for creativity assessment, as well for teaching and learning processes.

There is a notable lack of consensus regarding the generalizability of creativity versus whether creativity is domain-specific. Research literature indicates that large portions of creative competency are domain-specific (e.g., Kaufman, & Baer, 2005; Sawyer, 2012; Treffinger et al.,

2002). However, there are other perspectives, such as intermediate approaches (e.g., only some traits are domain-general) and developmental approaches (e.g., domain-general skills translate into domain-specific accomplishments). In this study we adopt the intermediate approach (e.g. Lubart, & Guignard, 2004;), according to which some creativity skills apply to multiple domains (e.g., creating different ideas or taking risks in introducing new ideas), whereas others are unique to specific subject areas (e.g., the ability to create different solutions in science or to write a conceptually different essay in Language Arts). Although a certain level of domain knowledge is essential for creativity, too ingrained, traditional domain-specific thinking may prevent the individual from manipulating the concepts within a particular field in novel ways (Sternberg, 2006). Furthermore, according to Plucker and Beghetto (2004), the level of specificity-generality changes with the social context and as one develops through childhood into adulthood. This approach suggests involving both domain-general and domain-specific dimensions of creativity in creativity assessment.

In our research, an operational definition of creativity refers to *the capacity of an individual to effectively engage in a process of developing a novel product, idea, or problem solution that is of value to the individual and/or the larger social group.*

### **Assessing Creativity**

Creativity assessments go beyond tests of analytical thinking in measuring performance on tasks that require individuals to deal with relatively novel situations. Creative performance can be observed and measured in ways that creative thinking cannot. Thus, creativity of products is typically the focus of assessments that vary the conditions under which individual's creativity is measured (Runco, 2004). Participants are often asked to write stories, make collages, and do other tasks that result in some tangible product. Assessment of little c product creativity mainly relies on the Consensual Assessment Technique (CAT) because of its relative simplicity and the consistently high levels of inter-rater agreements reached (e.g., Amabile, 1996; Kaufman et al., 2007).

The most widely used tests of creativity are Torrance Tests of Creative Thinking (Torrance, 1974, 2008). The tests were designed to identify children with high creativity potential, and to transform education to fully realize the creative potential of every student. In these tests, Verbal and Figural forms include various subtests. The Verbal subtest consists of picture construction, picture completion, and sketching different objects by using lines and circles. The Figural battery includes creating as many questions as possible about a picture and a common object, proposing unusual uses for a common object, guessing causes, guessing consequences, product improvement, and listing possible ramifications for an improbable situation. The revised tests report scores of fluency, originality, elaboration, resistance to premature closure, and abstractness of titles (Ball, & Torrance, 1984; Torrance, 2008).

Performance tasks are expected to tap into an important part of creativity that might not be measured using multiple-choice items alone because open-ended measures require more authentic and free-form responses. For example, in the Rainbow Project study, creativity was measured using open-ended, performance-based tasks among first-year college students in the United States (Sternberg, & the Rainbow Project Collaborators, 2006). For each of the tasks, participants were given a choice of topic or stimuli on which to base their creative written stories, oral stories, or cartoon captions. Each of the creativity performance tasks was rated based on rubrics that were pre-determined as measures of creativity. The score for creativity in cartoons captions was formed by summing the ratings given by judges on originality, cleverness,

and humor on 5-point scales. Both written and oral stories were rated for originality, complexity, emotional evocativeness, and descriptiveness on 5-point scales. The study found that student creative performance in these tasks was separated from the other more conventional tests. It was also found that adding the creative measures to other measures of analytical and practical measures roughly doubled the predictive value of the SAT for the sample in predicting grades for first-year college students (Sternberg, & the Rainbow Collaborators, 2006). The measures also served to decrease ethnic differences between groups.

The current study adopted the assessment approach developed by Sternberg and the Rainbow Collaborators (2006) and extended it to the computer-based interactive performance tasks. As shown in the methodology section of this paper, school students were asked to write different short stories based on given images or a video, create different titles for a story, and propose different captions for a cartoon. Thinking tools, such as capturing bookmarks from the movie and using them to create the story, were embedded into the assessment. In addition, differently from the traditional approach of creativity measurement, student-created stories, titles, and captions were given a rating by simulated classmates. This technique was used in order to potentially increase students' motivation and provide a reason for asking the student to create a different story, title or a caption later based on the same content; the submitted response was followed by a simulated rating given by classmates. This idea follows the hybrid method proposed by Silvia and colleagues (2008), with expansion to an external source of real-time rating. According to the hybrid method creativity assessment should take into account each participant's own definition of what is highly creative, by asking them to indicate the top two ideas from their own completed responses. These top ideas are then rated for creativity using the CAT. In this study the rating was pre-determined to indicate 5 out of 10 stars and was given a few seconds after the first submission of the story, title or a caption. Then the student was asked to write a different and more creative story, title or a caption based on the same images or video. The second submission was followed by a simulated rating given by classmates and was pre-determined to show 9 out of 10 stars. The implementation of the simplified real-time rating technique and the use of other interactive features within the creativity task were examined in order to inform further research and development of interactive tools that can potentially leverage technology added-value in an assessment context.

### **Research Questions**

The study addressed empirically the following questions regarding student performance, motivation and feedback in a computer-based creativity assessment:

1. What are the differences in the student performance between different creativity measures, as reflected by the computer-based creativity task?
2. How are students' abilities in different measures of creativity related to each other?
3. How is students' performance in creativity task related to students' motivation and time-on-task?
4. What is the students' feedback on various task items and interactive features that were introduced in the task, as reflected in student open-response?

### **Method**

The study participants included 87 students, who were all 14 years old and from the United States, United Kingdom, Turkey, and South Africa. The results presented in the current paper came from a larger study in which students from six countries were recruited to participate

in a 21st Century Skills Assessment project study investigating the innovative ways of developing computer-based assessment in critical thinking, creativity, and collaborative problem solving (see Rosen, & Tager, 2013, for study of collaborative problem solving, and Rosen, & Tager, 2014, for study of critical thinking). The researchers collected data from November 2012 through January 2013. Of the total students who participated, 47 were boys (54%) and 40 were girls (46%). Table 1 summarizes the country and gender distribution of participating students.

### Creativity Assessment

In this ‘You make the story’ computer-based assessment task, the student was asked: (a) to create two different short stories by using given images; (b) write three different captions for a single cartoon; and (c) write a story based on a video and create three different titles for a story. The student’s writing was accompanied by a rating given by simulated classmates. Due to the exploratory nature of the study, the students were not limited in time-on-task. The task was checked by two teachers from each of the four participating countries to ensure that students would be able to work on the task, and that the task could differentiate between high and low levels of creativity ability. Think-aloud interviews were conducted with eight students representing the target population to initially examine the assessment approach.

Figure 1 shows a sample screenshot of the creativity task.

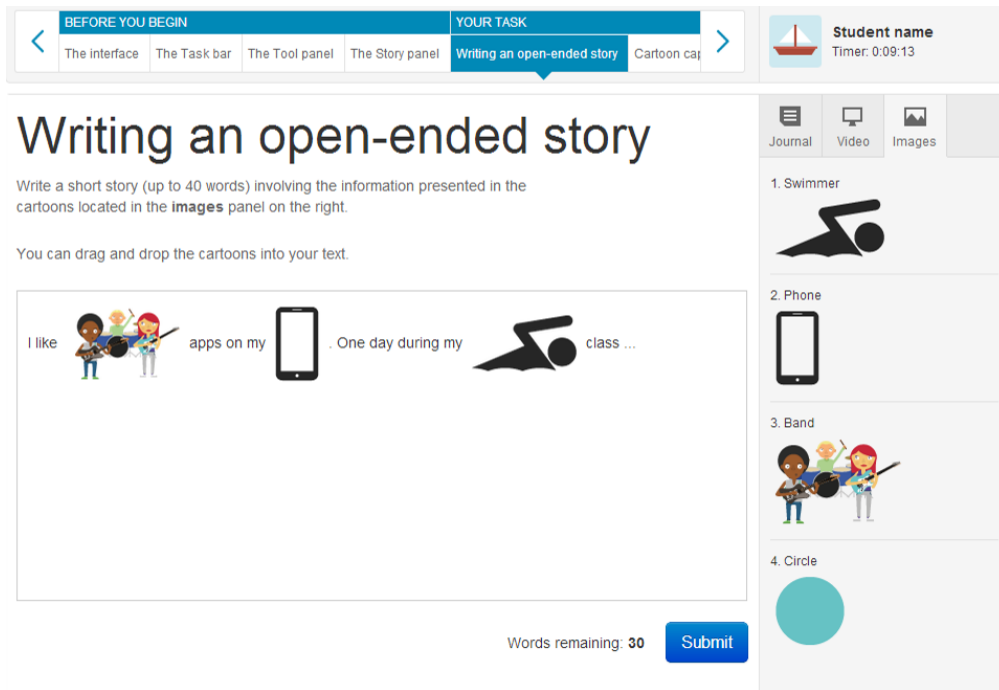


Figure 1. Writing a short story using given images

### Creativity Scoring Criteria

CAT scoring of the student responses was provided independently by four teachers from participating schools in the United States. Each judge rated the stories for originality, cognitive complexity, emotional expressiveness, and descriptiveness on 3-point scales. The captions for the cartoon and the titles for the written story were rated for originality on 3-point scales. For the

purposes of more meaningful interpretation of student scores for instructional purposes, the 0-3 scale was later converted into 0-100% scale. Inter-coded agreement of scoring was 87% for originality, 92% for cognitive complexity, 90% for emotional expressiveness, and 94% for descriptiveness. It should be noted that student responses were scored based on the criteria presented above, while spelling and grammar issues did not affect the student score.

### **Feedback Questionnaire**

Due to the exploratory nature of this pilot study, the students were asked to share their general feedback on the computer-based creativity assessment task. They were asked to respond to the following question: “This is your chance to help us build assessment tasks that you would like to take. Give us your feedback on the task you just went through. What did you like the most about this task? What did you least like about this task?” All the participating students completed the feedback questionnaire.

### **Results**

All results are presented on an aggregative level beyond the countries because no interaction with country was found. First, the descriptive results of student performance in a creativity assessment are presented to determine whether there is a difference in student scores in different measures of creativity. Next, the results regarding the relationship between different measures of creativity are shown. A Bonferroni correction was included to adjust the type one error for the number of correlations examined. Then, the relationships between student performance in creativity and student motivation, as well as time-on-task, are demonstrated. Last, findings from qualitative data analysis of students’ general feedback on the creativity task are presented.

The results of the creativity scores indicated that while students were relatively successful in cognitively complex writing ( $M=70.0$ ,  $SD=26.0$ ), they struggled with writing in an original manner ( $M=13.9$ ,  $SD=14.8$ ), expressing emotions in their writing ( $M=21.0$ ,  $SD=26.0$ ), and descriptiveness ( $M=25.0$ ,  $SD=20.2$ ). To better understand the relationship between the dimensions of creativity, analysis of intercorrelations between the variables was conducted. The findings showed significantly positive relationships between student originality score and the ability to express emotions, and descriptiveness ( $r=.53$ ,  $p = .000$ , for both measures). Emotional expression was also positively correlated with descriptiveness ( $r=.77$ ,  $p = .000$ ). No additional statistically significant correlations were found.

Data was analyzed to determine the possible relationship between student motivation to participate in creativity assessment and student performance. The results demonstrated no significant relationship between student motivation and student performance in creativity task ( $r=.11$ ,  $p=.74$ ). On average, student motivation was 3.6 on scale 1-4 ( $SD=.7$ ), indicating high level of motivation to work on the creativity task. In addition, no significant correlation was found between time-on-task and student performance in creativity ( $r=.19$ ,  $p=.08$ ). On average, time-on-task in the assessment was 20.8 minutes ( $SD=6.0$ ), while the maximum and the minimum time-on-task were 48.0 and 8.1, respectively.

Qualitative analysis of the major trends in students' open-ended feedback has revealed three major themes. It was found that the primary two themes in students' feedback were related to the video-based story episode. Students' feedback was associated with the engaging content of the video itself (82.8%) and the bookmark capturing interactive tool (58.6%), as represented in the following students' quotations: *"I liked the video a lot because it showed me how different the future will be and the music was really good too."*; *"The movie was the best part because it didn't feel like I was working on a test and I was having fun."*; *"I would have made even more entertaining and maybe more visual with some more videos or talking."*; *"I liked how you can drag in the pictures to help whoever is reading the story to get a better visual of it."*; *"It was very helpful to be able to bookmark photos from the video. It helped my visual because I'm a visual person."*

The third major theme was the automated simulated rating provided in each episode of the task in order to increase student engagement (48.3%). While most of the comments were positive in terms of the added-value of this feature in the task, some students reported that they were disappointed due to the fact that it was a static automated rating without human intervention. The following students' quotations represent students' feedback on the simulated rating aspect of the creativity task: *"I liked how we could add captions to the pictures and have them rated. It helps you improve your own caption."*; *"What I liked most about this test is that it gave me many tries to do better and showed the rating for each try. It made me think of more ideas."*; *"I did like that you made me work hard and didn't let me pass if I didn't wrote a good story or a caption."*; *"Make the rating a little more realistic. I didn't feel they could grade my title or story in 5 seconds."*; *"I would have a person grade the work and not the computer system."*; *"I would write next to the score what was wrong with the answer so the student will know how to correct it."*

Among the additional less frequently reported topics by students were the suggestion to allow more words in the written stories (*"I would change the number of limited words in every story we had to write in to a bigger number so we can give more details about the things that we write about."*), and the ability to write imaginative stories (*"I thought that this was a great way of getting kids to really use their imagination while doing work."*).

## **Discussion**

The 'You create a story' task that was developed is innovative in three aspects: it not only serves as task model to generate tasks with different content complexity and difficulty in school context, but also proposes digital tools that can be embedded into the tasks and facilitates the collection of data from participants in various populations, including participants across cultures. Use of the online task also streamlines the scoring process, as all responses are collected in digital format in a common database. While one task model was used in this study, the paper's aim is to offer empirically-studied strategies for evaluating data from computer-based creativity assessment tasks generally.

The goal of this study was to explore patterns in student creativity performance, motivation and feedback in computer-based interactive prototype of assessment. Students assessed in creativity computer-based performance task showed relatively high performance level in cognitive component of creativity, but their performance on all other components of creativity was significantly lower. Findings showed low performance levels in terms of originality and emotional expressiveness. It should be noted that the assessment was introduced to students as a general assessment task of 21st Century skills, emphasizing that there are no

right or wrong answers. The assessment criteria (e.g. “be original in your ideas”) or any other information about the assessment were not included in the introduction (see Method section for more details). Similarly to other types of assessment, providing more information about the assessment measures could potentially affect student performance in a creativity test (e.g., Barron, & Harrington, 1981, in the context of Torrance Test of Creative Thinking). Thus, further investigations are needed to examine the effect of introducing the assessment measures, as well as the scoring rubrics on creativity performance of school students.

The requests to create different stories based on a given theme or propose different titles for a cartoon were new to many students, but allowed each of them to show creativity regardless of his or her reading, writing. Creating different ideas under identical conditions is a cognitively challenging task that requires divergent thinking and novelty, which are often discouraged in school climate and curriculum (Andilou & Murphy, 2010; Benavides, Dumont & Istance, 2008). More often a culture of rewarding uncreative thinking is cultivated, rather than encouraging and rewarding creative answers. According to Partnership for 21st Century Skills (2011), creativity skills are central components in K-12 college and career ready goals of the teaching, learning and assessment, but are often implied in the content standards, rather than explicitly stated. Creativity can be embedded as an element of almost any performance task, and teachers should consider how to engage students in variety of tasks and contexts.

The results from motivation and feedback questionnaire suggest that the students were highly engaged in working on the creativity task. The interesting content, the ability of free writing of imaginary stories and captions and the use interactive tools, such as video bookmarking have promoted students’ motivation. Although, the students were positive about the possibility to have a real-time automated feedback on their creative writing, an authentic real-time rating with recommendations for improvement will make students’ experience even more engaging and productive. Human creativity is generally considered as one of the most distinguishable human capacities that cannot be easily scored automatically by computers. Automated scoring algorithms tend to use features that are easily computable, such as structure, word complexity, and average word length and combine them in ways that best predict the scores awarded by human judges. Despite a few attempts to develop automated techniques for creativity automated scoring by employing language technologies and computational–statistical machine learning methods to grade students’ natural language responses automatically (e.g., Shermis, et al., 2002; Wang, Li, & Chang, 2008), currently there are no examples of widely used and credible methods. Future studies could consider exploring ways to support creativity assessments with a real-time meaningful feedback. Students need an opportunity to use feedback on creativity to brainstorm new ideas and approaches they might have (Brookhart, 2013).

The current study had several limitations. First, it is based on a small and non-representative sample of 14-year-old students in four countries. However, due to a lack of empirical research in the field of computer-based interactive assessment of creativity skills in schools, it is necessary to conduct small-scale pilot studies in order to inform more comprehensive approaches of creativity assessment. Further studies could consider including a representative sample of students with a wider range of ages and backgrounds. Second, the study operationalized the interactive tools in critical thinking assessment through a simulated rating and video-capturing functionality, while other approaches could be considered, including semantic organization tools, dynamic modeling, conversational agents and collaboration tools (Jonassen, 2006; Hyerle, 2009; Rosen, 2014). Finally, it is important to conduct comparability studies between different modes of creativity assessment with and without interactive tools to



explore similarities and differences in student performance. Moreover, future studies could consider exploring differences in student performance in a wide range of problems and situations that require creativity.

### References

- Amabile, T. M. (1996). *Creativity in context*. Boulder, CO: Westview Pres.
- Ball, O. E., & Torrance, E. P. (1984). *Streamlines scoring workbook: Figural A*. Bensenville, IL: Scholastic Testing Service, Inc.
- Barron, F., & Harrington, D. M. (1981). Creativity, intelligence, and personality. *Annual Review of Psychology*, 32, 439-476.
- Beghetto, R. A. (2010). Creativity in the classroom. In J. C. Kaufman, & R. J. Sternberg (Eds.), *Cambridge handbook of creativity* (pp. 447-463). New York, NY: Cambridge University Press.
- Benavides, F., Dumont, H., & Istance, D. (2008). The search for innovative learning environments. In OECD (Ed.), *Innovating to learn, learning to innovate* (pp. 21-44). Paris: OECD.
- Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley M., Miller-Ricci, M. & Rumble, M. (2012). Defining twenty first century skills. In P. Griffin, B. McGaw, & E. Care (Eds.), *Assessment and Teaching of 21st Century Skills* (pp. 17-66). Dordrecht: Springer.
- Brookhart, S. (2013). Assessing creativity. *Educational Leadership*, 70(5), 28-34.
- Council of Chief State School Officers. (2012). *Framework for English Language Proficiency Development Standards corresponding to the Common Core State Standards and the Next Generation Science Standards*. Washington, DC: CCSSO.
- Council of Chief State School Officers and National Governors Association Center for Best Practices (2010). *Common Core State Standards*. Washington, DC: Council of Chief State School Officers and National Governors Association Center for Best Practices.
- Hyerle, D. (2009). *Visual tools for transformation into knowledge* (Second Edition). Thousand Oaks, CA: Corwin Press.
- Jonassen, D. H. (2006). *Modeling with technology: Mindtools for conceptual change*. Columbus, OH: Merrill/Prentice-Hall.
- Lubart, T., & Guignard, J. H. (2004). The generality-specificity of creativity: A multivariate approach. In R. J. Sternberg, E. L. Grigorenko, & J. L. Singer (Eds.), *Creativity: From potential to realization* (pp. 43-56). Washington, D.C.: American Psychological Association.
- O'Hara, L. A., & Sternberg, R. J. (2000-2001). It doesn't hurt to ask: Effects of instructions to be creative, practical, or analytical on essay—writing performance and their interaction with students' thinking styles. *Creativity Research Journal*, 13, 197-210.
- OECD (2010). *PISA 2012 Field Trial Problem Solving Framework*. OECD Publishing.
- OECD (2013). *PISA 2015 Collaborative Problem Solving Framework*. OECD Publishing.
- Partnership for 21st Century Skills. (2009). *P21 framework definitions*. Washington, D.C.: P21.
- Partnership for 21st Century Skills. (2011). *P21 Common Core toolkit: A guide to aligning the Common Core State Standards with the framework for 21st century skills*. Washington, D.C.: P21.
- Plucker, J. A., & Beghetto, R. A. (2004). Why creativity is domain general, why it looks domain specific, and why the distinction does not matter. In R. J. Sternberg, E. L. Grigorenko, & J.

- L. Singer (Eds.), *Creativity: From potential to realization* (pp. 153-167). Washington, D.C.: American Psychological Association.
- Rosen, Y. (2014). Thinking tools in computer-based assessment: Technology enhancements in assessing for learning. *Educational Technology*, 54(1), 30-34.
- Rosen, Y., & Tager, M. (2014). Making student thinking visible through a concept map in computer-based assessment of critical thinking. *Journal of Educational Computing Research*, 50(2).
- Rosen, Y., & Tager, M. (2013). *Computer-based assessment of collaborative problem-solving skills: Human-to-agent versus human-to-human approach*. Research & Innovation Network. Pearson Education.
- Runco, M. A. (2004). Creativity. *Annual Review of Psychology*, 55, 657-687.
- Sawyer, R. K. (2012). *Explaining creativity: The Science of human innovation* (Second Edition). Oxford University Press: Oxford.
- Shermis, M. D., Koch, C. M., Page, E. B., Keith, T. Z., & Harrington, S. (2002). Trait ratings for automated essay grading. *Educational and Psychological Measurement*, 62(1), 5–18.
- Sternberg, R. J. (1997). *Successful intelligence*. New York, NY: Plume.
- Sternberg, R. J. (2006). The nature of creativity. *Creativity Research Journal*, 18(1), 87-98.
- Sternberg, R. J., & The Rainbow Project Collaborators (2006). The Rainbow Project: Enhancing the SAT through assessments of analytical, practical, and creative skills. *Intelligence*, 34, 321–350.
- Torrance, E. P. (1974). *Torrance tests of creative thinking: Norms-technical manual*. Princeton, NJ: Personnel Press/Ginn.
- Torrance, E. P. (2008). *Torrance tests of creative thinking: Norms-technical manual*. Bensenville, IL: Scholastic Testing Service.
- Treffinger, D. J., Young, G. C., Selby, E. C., & Shepardson, C. (2002). *Assessing creativity: A guide for educators*. Sarasota, FL: National Research Center on the Gifted and Talented.
- Wang, H-C., Chang, C-Y., & Li, T-Y. (2008). Assessing creative problem-solving with automated text grading. *Computers & Education*, 51(4), 1450-1466.