

12-12-2016

The Development of An Online Divergent Thinking Test

Jiajun Guo

University of Connecticut, jiajun.guo@uconn.edu

Follow this and additional works at: <https://opencommons.uconn.edu/dissertations>

Recommended Citation

Guo, Jiajun, "The Development of An Online Divergent Thinking Test" (2016). *Doctoral Dissertations*. 1304.
<https://opencommons.uconn.edu/dissertations/1304>

The Development of An Online Divergent Thinking Test

Jiajun Guo, PhD

University of Connecticut, 2016

Divergent thinking (DT) tests are the most frequently used types of creativity assessment and have been administered in traditional paper and pencil format for more than a half century. With the prevalence of computer-based testing and increasing demands for large-scale, faster, and more flexible testing procedures, it is necessary to explore and test the usability of computer-based divergent thinking tests. Yet few studies have focused on the use of technologies in the assessment of creativity, including divergent thinking tests.

The purpose of the present study was to design and test the feasibility of an online divergent thinking (DT) test. The following three aims were addressed: (1) evaluate reliability evidence of DT test scores, (2) explore relationships between technology use and online DT performance, and (3) compare the online test with its paper version regarding DT scores. One hundred and sixty-four participants were recruited from the University of Connecticut and randomly assigned into three groups: online-basic (OB), online-advanced (OA), and paper-and-pencil (PP). Based on the results of reliability analysis, six DT scores were selected for analysis of variance and multiple regression modeling. The findings indicated that, despite of the possible link between technology use and online DT performance, no differences were found between different modes (online vs. paper) or different interfaces (simple tools vs. advanced tools) in terms of either DT scores or reliability evidence. Additionally, males were found to produce overall significantly higher originality scores than females did in the line meaning test and the real-world problem test. The implications of these findings are further discussed in the paper.

The Development of An Online Divergent Thinking Test

Jiajun Guo

B.A., East China Normal University, 2009

M.Ed., East China Normal University, 2012

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

at the

University of Connecticut

2016

Copyright by

Jiajun Guo

2016

APPROVAL PAGE

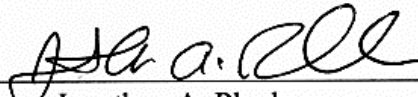
Doctor of Philosophy Dissertation

The Development of An Online Divergent Thinking Test

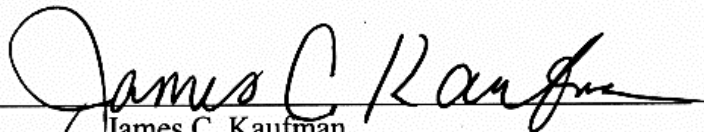
Presented by

Jiajun Guo, B.A., M.Ed.


Major Advisor


Jonathan A. Plucker

Associate Advisor


James C. Kaufman

Associate Advisor


Ronald A. Beghetto

University of Connecticut
2016

This dissertation is dedicated to my beautiful wife Lingjia.

ACKNOWLEDGEMENTS

First and foremost, I am indebted to Jonathan Plucker for his years of patient guidance, support, and encouragement. His many positive contributions to my professional development will always be remembered.

I acknowledge the valued support, suggestions and overall assistance that James Kaufman, Ronald Beghetto, Michael Young, and Jae-Eun Joo provided during the development of this manuscript.

I wish to thank Mark Runco for his permission to use his measure as well as his advice for research design.

I also express my sincere appreciation to Anna, Cody, BK, PJ, Rasis, Shiyu, and Simon for their friendship and support.

Table of Contents

LIST OF TABLES	viii
LIST OF FIGURES	x
INTRODUCTION	1
STATEMENT OF THE PROBLEM	2
RESEARCH QUESTIONS	4
REVIEW OF LITERATURE	5
Divergent Thinking Test	5
Computerized Divergent Thinking Assessment.....	8
RESEARCH HYPOTHESES	11
METHODS	11
Design and Participants	11
Instruments	12
Online Test Interface	16
Procedures	18
RESULTS	20
Reliability Analysis	20
Subjective Ratings of Originality	21
Objective Scoring of Originality	23
Scoring of Fluency.....	24
Summary of Reliability Analysis	24
Selection of Divergent Thinking Scores for Subsequent Analyses	26
Summary of Reliability Evidence for DT Scores from the Three Groups	27
Multiple Regression Analysis	28
Data Transformation.....	28

Simple Multiple Regression	29
Stepwise Multiple Regression	34
ANOVA, MANOVA, and MANCOVA	38
Descriptive Statistics	38
Two-Way Factorial Analysis of Variance (ANOVA)	38
One-Way Multivariate Analysis of Variance (MANOVA)	42
One-Way Multivariate Analysis of Covariance (MANCOVA)	42
DISCUSSION	43
Reliability of Online Divergent Thinking Test Scores	43
Theme 1: Different scoring methods produced different evidence of reliability	43
Theme 2: Different types of responses produced different evidence of reliability	46
The Relationship between Technology Use and Divergent Thinking Abilities.....	48
Theme 3: Technology Use Might Influence Online Divergent Thinking Performance.....	48
Mode Effect and Gender Effect on Divergent Thinking Scores	50
Theme 4: No mode effect was found for DT scores or their reliability estimates	50
Theme 5: Gender differences in DT scores were inconsistent	52
Summary of the Feasibility of an Online Divergent Thinking Test.....	53
Limitations and Disadvantages of the Current Online Test.	53
Advantages and Potential of Online Creativity Testing	55
Future Directions	56
References	58
Appendix 1: Online Divergent Thinking Test (ODTT)	68
Appendix 2: The Creativity with Technology Self-efficacy Scale (CTSS).....	75
Appendix 3: Technology Experience Survey (TES)	76

LIST OF TABLES

Table 1. Inter-rater Reliability Estimates of DT Tests for Originality	22
Table 2. Internal Consistency Estimates of DT Tests for Subjective Rated Originality	23
Table 3. Internal Consistency Estimates of the DT Test Using Objective Scoring	23
Table 4. Internal Consistency Estimates of Fluency Scores	24
Table 5. Frequency of Fluency Scores on the Square Drawing Test	26
Table 6. Summary of Reliability Information of the Six DT Scores for Different Groups	27
Table 7. Descriptive Statistics for the Selected DT Scores	29
Table 8. Technology with Creativity Self-efficacy Scales Regressed on Line Meaning Fluency	30
Table 9. Technology with Creativity Self-efficacy Scales Regressed on Line Meaning Originality	30
Table 10. Technology with Creativity Self-efficacy Scales Regressed on Story Drawing Originality	31
Table 11. Technology with Creativity Self-efficacy Scales Regressed on Square Drawing Originality	32
Table 12. Technology with Creativity Self-efficacy Scales Regressed on Real-world Problem Fluency	32
Table 13. Technology with Creativity Self-efficacy Scales Regressed on Real-world Problem Originality	33
Table 14. Technology Experience Items Regressed on Line Meaning Fluency	35
Table 15. Technology Experience Items Regressed on Line Meaning Originality	35
Table 16. Technology Experience Items Regressed on Square Drawing Originality	36
Table 17. Technology Experience Items Regressed on Real-world Problem Fluency	36

Table 18. Technology Experience Items Regressed on Real-world Problem Originality	37
Table 19. Descriptive Statistics by Group and Gender for DT Indices	38
Table 20. Analysis of the Differences in Group and Gender on Line Meaning Fluency	39
Table 21. Analysis of the Differences in Group and Gender on Line Meaning Originality.....	39
Table 22. Analysis of the Differences in Group and Gender on Story Drawing Originality	40
Table 23. Analysis of the Differences in Group and Gender on Square Drawing Originality	40
Table 24. Analysis of the Differences in Group and Gender on Real-world Problem Fluency ...	41
Table 25. Analysis of the Differences in Group and Gender on Real-world Problem Originality	41

LIST OF FIGURES

Figure 1. Frequency of Fluency Scores on the Story Drawing Test	25
Figure 2. Frequency of Fluency Scores on the Square Drawing Test	26

INTRODUCTION

Creativity has been recognized by many as the most important quality for success in the 21st century workplace (George, 2008; Mumford & Licuanan, 2004; Partnership for 21st Century Skills, 2014; Tierney, Farmer, & Graen, 1999). At the same time, our current era is also marked by exponential growth in technological innovation. Over the last few years, continuous calls have been made for more research examining the intersection between these two trends and more practices that fuse creativity with technology in education (Mishra, 2012). Numerous studies have been conducted to explore the use of various technologies in fostering original and curious minds (Andersen, 2014; Dunleavy, Dede, & Mitchell, 2009; Halsey et al., 2006; Magnussen, Hansen, Planke, & Sherson, 2013; Pignéguy, 2004; Walsh, 2007; Yee, 2015). One of the potential benefits technology brings to creativity is its capability to provide immediate external representation of complex information that can help people review and modify, improve, or enhance their creative action. For example, by using motion tracking and real-time animation techniques, teachers can help children enhance their ability to do creative storytelling (Yee, 2015). With the help of computer visualization tools, such as geographic information systems (GIS), high-spatial thinkers may have more opportunities to develop their creative talents (Andersen, 2014). Even digital games can highly motivate learners at different levels to create new knowledge, thanks to these games' multimedia features (Magnussen, Hansen, Planke, & Sherson, 2013).

Another potential benefit brought by the use of technology to the classroom is the enhancement of communication and collaboration in creative activities. Some technologies may not necessarily be interactive in nature, but in the classroom environment, where activities are often cooperative and collaborative, technology can serve as a platform or means by which those

activities can take place. For example, Pigneguy (2004) found that creating a video as part of a group project could enable students with emotional and behavioral difficulties to learn how to both creatively use technologies and effectively interact with other people. This shows that, for disadvantaged students, technology can be especially useful in promoting creativity (Halsey et al., 2006). Using handheld devices, students can engage in interactive and collaborative problem solving as part of complex and authentic classroom practices, such as in the simulated problem context (Dunleavy, Dede, & Mitchell, 2009). In addition, students' out-of-school literacy with using digital technologies also offers an opportunity for schools to harness students' creativity (Walsh, 2007).

STATEMENT OF THE PROBLEM

Although recent years have witnessed the impact of technology on creativity teaching practices in classrooms, few studies have looked at the use of technology in the assessment of creativity. With the development of digital technologies, many standardized tests can now be administered online or on computers, including TOEFL, GRE, SAT, Smarter Balanced Assessments, and even intelligence tests. However, most assessments of creativity are still administered in the same way as they were 50-60 years ago, in the form of traditional paper and pencil tests, making it difficult for them to serve new and emerging purposes like large-scale testing, flexible scheduling, and faster data collection.

Taking the divergent thinking (DT) test, a popular assessment of creativity, as an example, the test has been administered on paper since its emergence in 1960s, with only a few attempts to computerize it in recent years (Hass, 2015; Kwon, 1996; Pretz & Link, 2008; Zabramski, 2014). Pretz and Link (2008) designed a web-based Java program for the DT test so that they could collect responses from multiple sites and score the tests more easily. Although

this study, along with several other similar ones (Lau & Cheung, 2010; Palaniappan, 2012; Pásztor, Molnár, & Csapó, 2015; Rosen & Tager, 2013), utilized technology in creativity assessments, a closer examination of the tools these studies used reveals an important piece missing in the measurements – a figural form of the divergent thinking test.

A complete DT test, such as the Torrance Test of Creative Thinking (TTCT) (Torrance, 1974), usually consists of two formats: a verbal form and a figural form. The verbal form asks participants to come up with verbal responses to test prompts, while the figural form asks participants to draw their responses on the given figures and shapes. However, most computer-based creativity assessments have only used the verbal form of the DT test and have avoided the figural form, which requires drawing responses to test prompts. Only a few studies have used the figural form for computerized DT tests, and this yielded mixed results regarding the differences between paper and computerized versions of DT tests. For example, Kwon (1996) designed a computerized version of the TTCT figural form and examined the equivalence of the paper-and-pencil and computerized versions. He found significant differences between the two versions in terms of overall DT scores. Zabramski and Neelakannan (2011), on the other hand, found no differences between paper-and-pen and stylus in figural tasks, or between mouse, stylus and touch-input in another study (Zabramski, Gkouskos, & Lind, 2011). Despite these efforts, DT tests are still largely administered on paper instead of computers.

A main reason why the drawing task has been avoided in computerized creativity tests is because it might involve technological issues that could complicate research procedures and findings, including the availability of related tools and the influence of people's technology skills and experiences.

Speaking of the availability of technology tools, unlike the paper-and-pencil DT test, a consensus is yet to be reached regarding which tool to use and how the testing procedure should be carried out in computerized divergent thinking assessments, even when there is no drawing involved. What can a computer-based DT test accomplish that a paper version test cannot, and vice versa? Is it possible to make a computerized creativity test that can automatically calculate scores and then adapt the test to the test-taker's ability level? Does the change of platform (or format) affect the test's reliability estimates and validity? These questions still lack definite answers and cannot be answered by only one or two studies. Thus, the computerized divergent thinking test, including both figural and verbal forms, is a topic that warrants more research.

Regarding the influence of technology experience and skills in computer-based creativity tests, if a test taker is not proficient at using computers to draw or to complete other basic computer tasks, it is likely that his/her creative performance may be hindered by this inability to use the computer's tools, even if these tools have powerful functions. In addition, the skills we learn are always influenced by our daily experience with technology, such as the consumption of visual contents online or verbal editing done on computers and smartphones. In order to understand the influence of such technology-related factors (so that we can improve the computerized environment of the DT test), it is necessary to employ measures to disentangle the technology aspect from the creativity aspect for an online divergent thinking test. However, this has yet to be done in the literature.

RESEARCH QUESTIONS

Taking the above issues into consideration, the present study aims to test the feasibility of an online divergent thinking test protocol, and to investigate the psychometric properties of this online DT test compared to the paper-and-pencil version. Specifically, using a newly designed

online divergent thinking test, a tech-related self-efficacy scale previously developed at the University of Connecticut, and a technology experience survey (created for the current study), we will address the following questions:

- Research Question 1: What are the advantages and limitations of a computerized DT test with regard to the quality of assessment?
- Research Question 2: To what degree do technology-related skills and experience predict performance in online divergent thinking tasks?
- Research Question 3: How do the results obtained from an online DT test differ from the results of the paper-and-pencil version?

REVIEW OF LITERATURE

Divergent Thinking Test

Although there are many forms of creativity assessments, none have received as much attention as divergent thinking tests. The concept of divergent thinking originated with Guilford (1967), who proposed a multi-dimensional model of intelligence called the Structure of Intellect (SOI) that classified mental abilities by operation, content, and product. The most far-reaching impact of SOI was the distinction between two types of operation – convergent thinking and divergent thinking (Clapham, 2011). Basically, convergent thinking is the ability to find a single correct answer to a problem, while divergent thinking is the ability to come up with varied solutions to a problem. Most tests used in education today, like achievement tests and intelligence tests, measure convergent and conforming thinking, whereas very few involve questions asking students to come up with unusual solutions/ideas. This could represent a serious problem with our educational systems, since many important problems we encounter in our life don't have single best answers. Rather, many problem situations require the ability to think

outside the box (e.g. in the face of adverse events), to weigh different values and benefits, and to invent rules/knowledge when no rules/knowledge are to be found in certain area.

An interesting aspect of divergent thinking is its relationship with intelligence. As mentioned previously, Guilford originally identified divergent thinking as one of the components of intelligence. However, this notion of divergent thinking as a subset of intelligence is outdated. In fact, psychologists have been unsuccessfully attempting to include creativity in intelligence tests since the earliest days of IQ tests (Brown, 1989). This led to another view of the relationship: intelligence and divergent thinking are unrelated or only remotely related. This view, again, fell out of favor due to a lack of empirical support. Evidence from more contemporary research shows that divergent thinking and intelligence are indeed overlapping, if not identical, in several respects (Plucker & Esping, 2015). Cognitive control may be one of the functions shared between the two. As an executive ability, cognitive control helps filter out salient but non-relevant information from the external world, which can both facilitate fluent generation of new ideas (divergent thinking) and ensure that those new ideas are of high quality (intelligence) (Benedek et al., 2012). This view stands in contrast with the traditional notion that creative ideation is associationistic and structural rather than executive, thus providing insights into how these two distinct but overlapping constructs interact with one another (Nusbaum & Silvia, 2011).

Perhaps a more important issue is the relationship between divergent thinking and creativity. The appearance and frequent use of DT tests made them so popular in education and research that people – especially those who didn't know much about creativity – began to equate DT with creativity, rather than seeing DT as one component of it. However, this notion that DT=creativity is simply incorrect because it takes for granted other important factors in the

making of real-world creativity, such as contextual influence (e.g. from peers and mentors), disposition (e.g. tolerance of ambiguity), domain knowledge (e.g. science, music, or leadership), and even convergent thinking (e.g. IQ). Research has been conducted to examine how well divergent thinking predicts creative achievement, with mixed evidence and conflicting opinions found in those studies. On the one hand, DT tests show evidence of reliability and concurrent validity (e.g. Torrance, 1968, 2008). Plucker (1999), for example, reanalyzed data from Torrance's longitudinal study and found that divergent thinking was much more predictive of how people differed in terms of creative achievement than traditional IQ was. On the other hand, many scholars have questioned DT tests' predictive validity (Baer, 1993, 1994; Gardner, 1993; Kogan & Pankove, 1974; Weisberg, 1993). This has led many researchers and educators to avoid using DT tests and, broadly, had led to criticism of the psychometric study of creativity (Plucker & Renzulli, 1999).

Given such criticism, it is not surprising that DT tests have fallen out of favor over the last decades. But as Kaufman, Plucker, and Baer (2008) note, many of these criticisms have been overstated, and the psychometric quality of DT test has significantly improved over time. For example, numerous new developments have emerged in terms of DT task scoring, and scores are no longer limited to the traditional four components (fluency, originality, flexibility, and elaboration). Some studies have suggested the use of alternatives to increase reliability and external validity, including summative scores (adding up all four component scores), highly uncommon scores (top 5 percent in originality), weighted fluency scores, and percentage scores (dividing originality scores by fluency scores) (see Benedek et al., 2013; Plucker, Qian, & Wang, 2011).

In addition to the development of new objective scoring methods, subjective assessment has also been recommended by researchers such as Silvia, Martin, and Nusbaum (2009), who proposed that a holistic rating can be given by neutral raters, therefore providing a “snapshot” of creativity. A combination of both objective and subjective methods has also been recommended (Plucker, Qian, & Schmalensee, 2014).

Computerized Divergent Thinking Assessment

Numerous attempts have been made to assess divergent thinking using web-based or computer-based methods. Some researchers have chosen to devise an application or program specifically for the DT tasks, such as Kwon’s (1996) computerized TTCT figural forms, Pretz and Link’s (2008) *Creative Task Creator* (CTC), Cheung and Lau’s (2010) *electronic Wallach-Kogan Creativity Tests* (e-WKCT), Palaniappan’s (2012) *Creativity Assessment System* (CAS), and most recently, Zabramski’s (2014) computerized-multi-input TTCT figural forms. Creative Task Creator (CTC) is a Java-based program that can generate HTMLs through which participants can complete an alternative uses task. The program can also collect data from different sites and store them on a server for researchers to manage (Pretz & Link, 2008). Similarly, the Creativity Assessment System (CAS) is a web application developed on the basis of the Torrance Test of Creative Thinking (TTCT). CAS is able to automatically calculate fluency, originality, and flexibility scores. Interestingly, an elaboration score was intentionally left out due to issues with the complex programming required. Another computer program for the divergent thinking test was designed by Cheung and Lau (2010), who computerized all the Wallach-Kogan Creativity Tests (WKCT) and administered them in computer labs. The authors found no differences in evidence of reliability between the results from the electronic and paper

versions of the WKCT. The study was also the first to provide reliability evidence for the comparability of the two versions of DT tasks (Lau, & Cheung, 2010).

As building a computer application is often time-consuming and expensive, other researchers have turned to alternative ways to put creativity tests on computers. With the recent development of online survey services, administering computer-based DT tests has become much more convenient. Researchers and educators can take advantage of some powerful functions offered by those cost-effective survey tools. The biggest advantages are their cost-effectiveness and convenience. With online survey technologies, test makers are now able to produce a variety of test items, allowing them to adapt almost all of the paper-based assessment content to online platforms. Since test distribution using a survey system is much simpler, test administrators can now carry out large-scale data collection, create flexible testing schedules, and analyze data at any time and from anywhere. For example, Pásztor, Molnár, and Csapó (2015) were able to collect nearly 2,000 responses from 97 classes at 78 schools with the help of an online survey system. Hass (2015) designed and delivered his survey-like DT tests to online participants using the Qualtrics system so that the testing data could be collected through the internet. He also compared scores obtained from online participants with those obtained in person and found that online participants gave fewer responses than participants taking the test in person, a different result from that of Lau and Cheung (2010).

The difference between online and in-person DT test results found in Hass's (2015) study indicate the possibility of a mode effect on test scores. Although few studies have looked at the mode effect in creativity testing, the mode effect in other types of performance testing has been studied by researchers since the inception of computerized testing. The previous literature has identified several factors that may be responsible for differences in some performance tests,

including participant factors (such as computer familiarity and gender) and technological factors (such as screen size and resolution, and item presentation) (Leeson, 2006).

What makes things more complicated is that mixed results have been found among the few studies that included computerized figural forms of the DT test. Kwon's (1996) study was among the first efforts to put the paper DT test onto computers. Using Hypercard, a programming tool provided by Apple Inc., Kwon was able to develop a computer system for the DT test and compare the results between paper figural forms and computerized figural forms. He found significant differences between the two modes. Zabramski and his colleagues (Zabramski, Gkouskos, & Lind, 2011; Zabramski & Neelakannan, 2011), on the other hand, found no significant differences in DT scores between various kinds of input modes, including paper-and-pencil, stylus, mouse, and touch-input.

According to Leeson's (2006) framework, there are at least two reasons for the inconsistencies between Zabramski's (2014) and Kwon's (1996) studies. The first reason, which is technological in nature, concerns the availability of relevant tools for drawing tasks. Even for verbal tasks, no consensus exists regarding which technology tool should be used and how it should be used. For drawing tasks, various kinds of input devices (such as mouse, hand-touch, and stylus) and software (such as computer applications and browser-based interfaces) can be used to collect responses. The differences in tools might contribute to the differences in DT scores between the two studies. The second reason, which is more related to participant characteristics, concerns the confounding influences of participants' technology skills and experience. For example, drawing requires fine motor skills, and the motor skills applied in hand drawing and computer drawing might be different. It is possible that people who often use computers to draw or edit visual contents may possess better skills due to frequent practice, thus

making them more fluent in creating computer drawings. Other technology experience not related to drawing may also influence creativity performance, such as playing video games (Jackson et al., 2012).

All of the abovementioned issues may affect the psychometric properties of an online test, including evidence of reliability and validity. These issues will be addressed in the present study.

RESEARCH HYPOTHESES

On the basis of both my research questions and the previous literature, three general hypotheses were proposed, as follows:

- Hypothesis 1: An online DT test can produce the same reliable DT scores as the paper test.
- Hypothesis 2: Technology-related skills and experience positively predict performance on online divergent thinking tasks.
- Hypothesis 3: Results obtained in an online DT test may be different from the results from the paper-and-pencil version.

METHODS

Design and Participants

A major purpose of this study was to investigate whether there are any differences in terms of psychometric properties between paper-and-pencil and computerized versions of the DT test. Therefore, a three-group, between-subjects experimental design was used in the present

study, which involved a paper-and-pencil group (PP), an online basic group (OB) and an online advanced group (OA).

The present study was carried out at the University of Connecticut in the summer of 2016. Participants were recruited through UConn public announcement systems such as UConn Daily Digest, Neag Student News, and listserv. A survey link was distributed via these systems to recruit potential participants (18 years of age or older) to sign up for the study. The link directed them to an interface where an information sheet/consent form was presented.

Two hundred and ninety-five adults from the University of Connecticut responded to the survey, of which 164 responded to all the survey questions and completed all the divergent thinking tasks. The data collected from these 164 participants were used for analysis. Of these participants, 109 were females, and 55 were males. In addition, 72 were undergraduate students, 85 were graduate students, and 7 were college staff or faculty members.

Instruments

Online Divergent Thinking Test (ODTT). To answer the question regarding the difference between computerized and paper DT tests, this study employed a DT test designed specifically for online use, but which could also be administered on paper so that the researcher could compare the results between the online and paper versions. The test contained three types of tasks: *the Line Meaning test*, *the Drawing test*, and *the Real-world Problem test*.

The Line Meaning test was first used by Wallach and Kogan (1965) as a figural (or visual) test of creativity. Usually, several incomplete and irregular figures were presented and test takers were asked to come up with as many meanings as they could for each figure. However, the original figures in the Line Meaning test were actually designed to measure

personal perception rather than creativity. Tagiuri (1960), who designed the original figures, presented each figure to the participants and asked them to consider the line as a path made by a person moving in space from left to right. That is why all the figures in Wallach and Kogan's Line Meaning test were continuous lines. In the present study, the figures were modified or created to allow for more variation. Three newly designed figures were used, including a curve and point figure, a wave-like figure, and an angle-like figure. First, an example (adapted from Wallach and Kogan's original test) was presented explaining how to respond to the test prompt. Then the participants were asked to imagine and write down all the things each figure might be (see instructions and items of the Line Meaning test in **Appendix 1**).

The Drawing test had two parts: the *Story Drawing* test and the *Square Drawing* test. The first part – the Story Drawing test – was a combination of Urban's (2005) ideas and Wallach and Kogan's figural test. Urban (2005) has argued that traditional creativity tests such as the divergent thinking test focus too much on the quantitative aspects instead of the aspects of quality, so he proposed a drawing production test that asks participants to complete a drawing on the basis of some given figural fragments. This drawing production test was deeply rooted in the Gestalt tradition, which holds that a diagnostic for creativity should be holistic and gestalt-oriented – the whole is more than the sum of the parts. The Story Drawing test used in our study took the three figures presented in the previous task (the Line Meaning test) and asked participants to combine these figures to create an interesting story (see Drawing test 1 in **Appendix 1**). The second part of the Drawing test was the Square Drawing test, as adapted from Guilford's (1967) Sketches test. Here, participants were shown three identical squares and then were asked to add in details to the squares to create original and interesting pictures (see Drawing test 2 in **Appendix 1**).

The Real-world Problem test had two questions, both of which were selected from Runco's *Realistic Presented Problems*, a part of the *Runco Creativity Assessment Battery* (*rCAB*). (Note: The use of this test was permitted by Runco before the start of this study). This test was designed to address the concern that divergent thinking has a low correlation with creative achievement in the natural environment as opposed to the classroom (Okuda, Runco, & Berger, 1991). In the Real-world Problem test used in the current study, participants were first given an example explaining how to respond to the questions. Then they were presented with two problematic situations regarding school and home, and were asked to come up with as many solutions as they could. (see Real-world Problem test 1 and test 2 in **Appendix 1**).

Scoring of Divergent Thinking. Two types of DT scores, *fluency* and *originality*, were calculated for the purpose of the current study. While fluency is an objective score reflecting the number of responses to each question, scoring of originality can be conducted using two different methods, according to the previous literature: objective rating, as carried out by the investigator, and subjective rating, as conducted by outside raters. To find out which method was better for scoring originality in terms of reliability, the present study employed both methods and then compared their reliability evidence to decide which originality score would be used for each sub-test.

For *subjective rating of originality*, the average creativity method was used (Silvia et al., 2008). Three blind raters who were graduate students at the University of Connecticut were recruited and trained to rate all the responses to each question of the DT test. The raters were instructed to use a 1-6 scale to rate each response for creativity, with "1" meaning "not creative at all" and "6" meaning very creative. To obtain the final score, the participant's ratings were summed and divided by the number of responses.

For *objective rating of originality*, the researcher used the top 20% scoring method (Plucker, Qian, & Schmalensee, 2014), counting the number of the responses given by less than 20 percent of the sample to each question. The Drawing test, which included Story Drawing and Square Drawing, did not yield objective ratings of originality, because the researcher could not pool the drawings in order to find uncommon responses – each drawing was unique from an objective perspective.

Creativity with Technology Self-efficacy Scale (CTSS). As mentioned previously, a newly constructed measurement called *the Creativity with Technology Self-efficacy Scale (CTSS)* was employed to examine the influence of technology-related creativity skills on the performance on an online DT test. The scale was constructed to measure self-perceived creativity in technology use. It is a scale containing 18 items that measure three aspects of technology-related creativity (see **Appendix 2**): *General Creativity Self-efficacy (GCS)*, *General Technology Self-efficacy (GTS)*, and *Technology Enabled Creativity Self-efficacy (TECS)*. GCS was adapted from Beghetto, Kaufman, and Baxter's (2006) study, measuring the self-judgement of one's imaginative ability and perceived competence in generating novel and adaptive ideas, solutions, and behaviors. TECS, on the other hand, focuses on one's perceived ability to use computers and other smart devices in daily life. The third sub-scale, TECS, is a combination of the previous two aspects and reflects the self-judgement of one's competence in using technologies to achieve creative goals.

The data collection process for CTSS ended just before the proposal of the current study, with 854 responses collected at the University of Connecticut. Of these respondents, approximately 69% were female and 31% were male. Respondents varied in age, with approximately 53% being 18-25 years old, 24% being 26-34 years old, 16% being 35-54 years

old, and 6% being 55 or older. In terms of education level, about 38% of the participants were undergraduates, 30% had just completed college degrees, and 32% held graduate degrees.

Further analysis showed strong evidence of inter-item reliability for the entire scale (CTSS) (Cronbach's $\alpha = .86$). Evidence of reliability was also found for each subscale: Cronbach's α was .91 for GSC, .88 for GTS, and .75 for TECS. Additionally, the correlations between the three subscales were .21 between GSC and GTS, .46 between GCS and TECS, and .53 between GTS and GCS.

Technology Experience Survey (TES). Another instrument created for the current study measured participants' experience with verbal and figural technologies. TES contains 13 questions/items asking how often people do certain things with technologies (see **Appendix 3**). Item generation was based on both the previous literature and current trends in technology development. The survey questions can be divided into two general categories, including items that focus on visually-focused technologies and items that focus on non-visually-focused technologies. The purpose of including this measurement was to investigate whether people's experience with various technologies influences their performance on the verbal and figural forms of an online DT test. A 5-point Likert scale was used, including *never*, *rarely*, *sometimes*, *very often*, and *always*.

Online Test Interface

To build an online test that allows participants to draw pictures, the researcher experimented with different web tools. The final solution involved several online services, some of which were free and others that charged a reasonable monthly fee.

Web drawing tool. The most important tool utilized in this study was a web drawing tool embedded in an online survey system. The tool, called A Web Whiteboard (AWW), is a web drawing interface designed for simple drawing and communication on computers and tablets. Two features of this tool are worth noting. First, AWW can be used as a plugin that can be embedded in most webpages. Therefore, the researcher was able to combine questions requiring verbal responses and those requiring drawing responses into one streamlined survey system. Second, AWW allows the customization of tools and other interface elements to be presented to participants. Thus, the study could utilize different testing materials for different tasks and create conditions for different online groups. With these features, AWW served as a powerful tool for the current study. Both the Story Drawing and Square Drawing tests used AWW as their drawing interfaces.

Online Survey System. One critical feature that an online test should possess is the capability to streamline the testing procedure, helping test takers to complete all the questions without distraction or disruptions. An online survey service, as provided by Qualtrics through the link from the University of Connecticut, enabled the collection of responses through the internet. Several features were used to meet the needs of the current study. First, a simple link distributed via different email systems enabled the researcher to collect the required amount of data within a relatively short period of time. Second, a system randomizer helped the researcher assign participants to different groups based on the research design. Third, simple coding could be used to embed the drawing tool, as described above, into the survey, and the system also provided ways for participants to upload the pictures they had drawn. The only disadvantage of this service was the lack of an advanced coding environment, so that the drawing tool could not be

directly embedded into the survey. Because of this, another service needed to be used, which will be described in the following section.

Web hosting service. AWW – the drawing tool – allows users to customize its interface to accommodate different needs; this was both an advantage and a disadvantage in the context of present study. While the advantage is obvious, the disadvantage is that it involves coding that is not recognized by the Qualtrics survey system. Thus, the researcher could not customize the drawing interface to create experimental conditions, including the presentation of test materials (instructions and figures) and control over what tools were shown to participants. To compensate for this issue, the researcher used a “bridge” that not only allowed the researcher to change the drawing interface but also allowed the drawing interface to be embedded into the streamlined Qualtrics survey system. Weebly, a web hosting service, played this role.

Two important features of Weebly helped the researcher connect the drawing tool with the survey system. First, Weebly allows and recognizes the HTML coding used to manipulate the web drawing interface. Second, upgrading to a business account, Weebly enables an SSL (Secure Sockets Layer) function that uses an encrypted link between the web server and browser, which in turn is recognized and accepted by the Qualtrics survey system. Another feature of Weebly was that it allowed the researcher to build a website as a demo for people who are interested in the online divergent thinking test, so that people do not have to take the survey in order to see the test. After a few simple steps to connect AWW, Weebly, and Qualtrics, the entire online test system was successfully set up for the present study.

Procedures

After interested participants received an email or saw an advertisement posted on a school announcement board, the test/survey link brought them to a page giving information

about the present study. After reading the information sheet and clicking the “agree” button, they were randomly assigned into three groups by the system randomizer and received different instructions accordingly.

Paper and pencil (PP) group. For the paper and pencil (PP) group, participants were directed to an online ticket platform (Eventbrite), where they were asked to select the times they could attend lab sessions to complete the paper version of the DT test and the two surveys. The sessions were held in a lab located in the Gentry building (Neag School of Education). The elements of informed consent were explained when participants arrived at the lab for their sessions and the participants were provided an informed consent form. After they had signed the form, they received the instructions for the tasks. To maximize each participant’s performance and reduce potential anxiety associated with test taking, the investigator attempted to create a “game-like” environment by saying that this is a thinking game. The general instructions were as follows:

“This is a thinking game. All the instructions are on the sheets. Please follow these instructions and complete all the tasks and questions, taking your time. If you have any questions, let me know.”

The participants were then provided with a pencil with an eraser and started the test. After finishing the tasks, each participant received 10 dollars as compensation, and the procedure was over for the group. All paper-and-pencil responses were transcribed into electronic versions.

Online Groups. For the online basic (OB) group and the online advanced (OA) group, all the tasks, including the divergent thinking test and the two technology surveys, were administered online. The participants followed the instructions posted on the (Qualtrics) survey

that was designed specifically for the current study. The difference between the online basic (OB) group and the online advanced (OA) group was that, in the figural tasks of the DT test, the OB group was asked to use only two drawing tools in the editor (pencil and eraser) to complete the figural tasks, while OA group was allowed to use more colors and brush sizes for the tasks. Upon finishing the final task, the procedure ended for both online groups. The participants were then automatically directed to the online ticket platform (Eventbrite), where they could select times to pick up their 5 dollar compensation or choose to receive their compensation in the form of an Amazon e-gift card.

Raters. Three raters were recruited from the university to rate the creativity of each response. The investigator held an approximately 30-minute video meeting with each rater to explain the scoring procedure. After the meeting, the raters received an email re-iterating the general instructions explained in the meeting, along with several Excel files (containing the verbal responses) and PowerPoint files (containing the drawing responses), with specific instructions for each subtest. After scoring was completed, the raters sent the files back. Each rater received \$100 for completing the scoring.

RESULTS

Reliability Analysis

Research Question 1: How reliable are the scores obtained from the divergent thinking tests?

The current study employed two types of ratings to obtain originality scores: the subjective rating, as conducted by three blinded raters, and the objective rating, which was conducted by the researcher. The subjective ratings were obtained from the Line Meaning test,

the Drawing test, and the Real-world Problem test, while objective ratings were only obtained from the Line Meaning test and the Real-world Problem test. The Drawing test did not produce objective ratings of originality due to the fact that the researcher could not pool all the drawings to find uncommon responses – each drawing was unique from an objective perspective. Additionally, fluency scores were calculated for all the three tests by the researcher, also falling into the objective category.

Two aspects of reliability were assessed in the present study. First, the inter-rater reliability of each item was assessed using the Intraclass Correlation Coefficient (ICC) to determine the level of agreement between raters. Second, the internal consistency of each test was assessed using Cronbach's alpha and Spearman-Brown coefficient analysis.

Subjective Ratings of Originality

Research Question 1a: How reliable are the originality scores produced by the three raters?

Inter-rater Reliability. In the current study, all the items in the divergent thinking (DT) tests were rated for originality by three blind raters. The inter-rater reliability of each item was assessed by interclass correlation coefficients (ICCs) (see Table 1). According to the guidelines recommended by Cicchetti (1994), when the reliability coefficient is between .60 and .74, the level of clinical significance is good, and when it is larger than .75, the evidence of reliability is considered excellent. Following these standards, one DT item had an inter-rater reliability estimate within the excellent range ($ICC = .83$) for originality scores, and the other six DT items were considered to have good evidence of inter-rater reliability ($ICCs = .60$ to $.74$).

With regard to the differences between the three sub-tests, subjective ratings of the Drawing test had the highest inter-rater reliability coefficients ($ICC = .74$ and $.83$), whereas the

subjective ratings of the Line Meaning test and the Real-world Problem test, both of which collected verbal responses, had lower inter-rater reliability coefficients (ICCs = .60 to .67).

Table 1. Inter-rater Reliability Estimates of DT Tests for Originality

<i>Items</i>	<i>ICC</i>
Line meaning 1 (Figure 1)	.61
Line meaning 2 (Figure 2)	.61
Line meaning 3 (Figure 3)	.62
Drawing 1 (<i>Story</i>)	.74
Drawing 2 (<i>Squares</i>)	.83
Real-world problem 1 (<i>Pat</i>)	.67
Real-world problem 2 (<i>Memory</i>)	.60

Note. N=164.

Internal Consistency of DT sub-tests. As mentioned previously, there were three sub-tests of divergent thinking. The Line Meaning test included three items, the Drawing test included two items, and the Real-world Problem test included two items. After collecting subjective ratings of originality from the three raters, scores were averaged for each of the seven items. An internal consistency analysis was then conducted for all three sub-tests using the average scores.

To assess internal consistency, the DT tests were evaluated using Cronbach's alpha and the Spearman-Brown (S-B) coefficient (see Table 2). Cronbach's alpha is often calculated for a test that has three or more items, while the S-B coefficient is most appropriate when a test has only two items (Eisinga, Grotenhuis, & Pelzer, 2012). The results indicated that all three sub-tests had low internal consistency ($\alpha = .34$; *S-B coefficient* = .55 and .59), with the Line Meaning test having the lowest inter-item reliability estimate ($\alpha = .34$).

Table 2. Internal Consistency Estimates of DT Tests for Subjective Rated Originality

<i>Test</i>	<i>Items</i>	<i>alpha</i>
Line meaning	3	.34
<i>Spearman-Brown coefficient</i>		
Drawing	2	.55
Real-world problem	2	.59

Objective Scoring of Originality

Research Question 1b: How reliable are the originality scores obtained by counting the number of top 20% responses?

Internal Consistency of the DT sub-tests. Two of the three DT sub-tests employed the traditional scoring of originality – counting the number of responses provided by less than 20% of the sample. Therefore, no inter-rater reliability was involved in the objective scoring.

Reliability information only comes from the internal consistency analysis.

To assess internal consistency, the DT test was evaluated using Cronbach's alpha and the Spearman-Brown (S-B) coefficient (see Table 3). The results indicated that the Line Meaning test ($\alpha = .88$) reached the benchmark of excellent reliability estimate, and the Real-world Problem test (S-B coefficient = .73) fell just outside the excellent range. Compared to the reliability estimates of the subjective scoring ($\alpha = .34$; S-B coefficient = .59), the objective scoring produced substantially higher internal consistency.

Table 3. Internal Consistency Estimates of the DT Test Using Objective Scoring

<i>Test</i>	<i>Items</i>	<i>alpha</i>
Line meaning	3	.88
<i>Spearman-Brown coefficient</i>		
Real-world problem	2	.73

Scoring of Fluency

Research Question 1c: How reliable are the fluency scores obtained by counting the number of responses?

Internal Consistency of the DT sub-tests. In the present study, fluency scores were obtained by counting the number of responses to each item. To assess the internal consistency of the DT test, these scores were evaluated using Cronbach's alpha and the Spearman-Brown (S-B) coefficient (see Table 4). The results indicated that fluency scores from the Line Meaning test ($\alpha = .93$) and Real-world Problem test (S-B coefficient = .81) reached excellent inter-item reliability, whereas the two items in the Drawing test (S-B coefficient = .14) failed to produce similar fluency scores.

Table 4. Internal Consistency Estimates of Fluency Scores

<i>Test</i>	<i>Items</i>	<i>alpha</i>
Line meaning	3	.93
		<i>Spearman-Brown coefficient</i>
Drawing	2	.14
Real-world problem	2	.81

Summary of Reliability Analysis

According to the reliability analysis conducted above, the traditional scoring method for originality outperforms the subjective rating method in terms of internal consistency for the Line Meaning test and the Real-world Problem test. Furthermore, inter-rater disagreement adds more error to the originality score, making the subjective rating method even worse in terms of overall reliability.

However, the objective scoring method is not without problems, which mainly manifested in the Drawing test. Low inter-item reliability in both originality scores and fluency scores means that the two drawing tasks should not be treated as interchangeable items of a test. In addition, histograms of the fluency scores (Figure 1 and Figure 2) from the two drawing tasks show that most fluency scores clustered on a single number - three. This means that there is not much variability in fluency scores for the drawing tasks, and the drawings should be evaluated on their qualities rather than quantity.

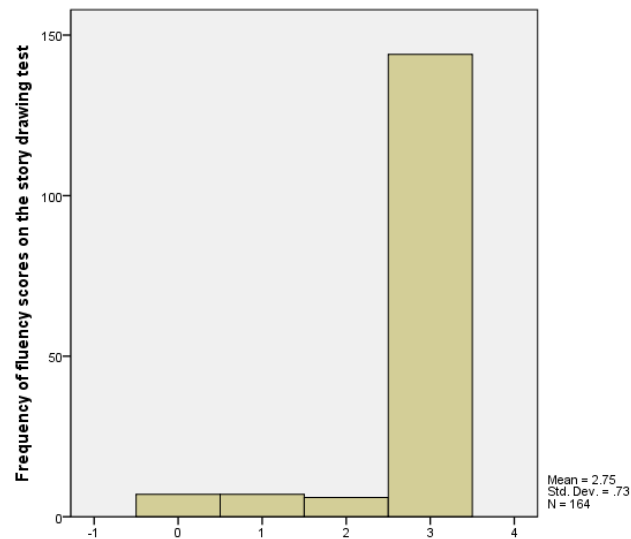


Figure 1. Frequency of Fluency Scores on the Story Drawing Test

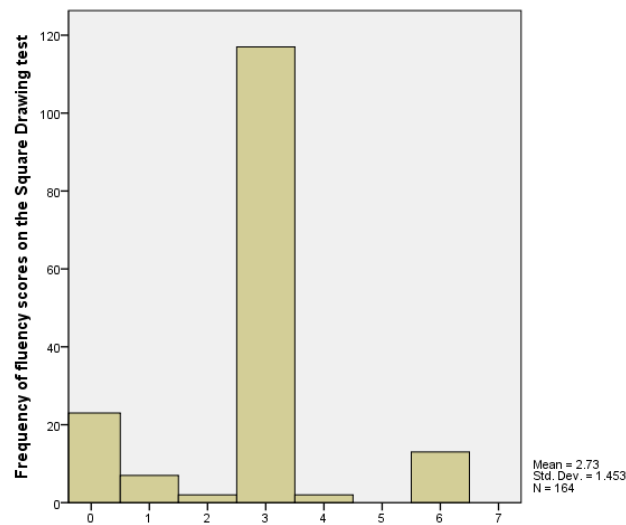


Figure 2. Frequency of Fluency Scores on the Square Drawing Test

Selection of Divergent Thinking Scores for Subsequent Analyses

Taking into consideration the reliability information shown above, the researcher decided to use six creativity scores for subsequent analyses (see Table 5). Two were fluency scores obtained from the Line Meaning test and the Real-world Problem test. The fluency scores from the Drawing test were dropped due to their lack of variability. The other four scores were originality scores obtained from all the three sub-tests, including objective scores from the Line Meaning test and the Real-world Problem test and two subjective scores from the Story Drawing test and the Square Drawing test.

Table 5. Frequency of Fluency Scores on the Square Drawing Test

<i>Index</i>	<i>Scoring Method</i>
Line meaning fluency	Objective
Line meaning originality	Objective
Story drawing originality	Subjective
Square drawing originality	Subjective
Real-world problem fluency	Objective
Real-world problem originality	Objective

Summary of Reliability Evidence for DT Scores from the Three Groups

Another important issue to consider before further analysis is the comparability of the online and paper-and-pencil versions of the DT test in terms of reliability estimates. If large differences existed between different groups in terms of reliability coefficients, there would be no need to compare the DT scores between groups. Therefore, a reliability analysis of the six selected DT scores was conducted for each group, as summarized in Table 6. According to the results, the differences between groups in terms of reliability coefficients were acceptable, indicating that scores obtained from the online groups were generally as reliable as the scores obtained from the paper-and-pencil group.

Table 6. Summary of Reliability Information of the Six DT Scores for Different Groups

	<i>Online-Basic Group</i>	<i>Online-Advanced Group</i>	<i>Paper-and-Pencil Group</i>
<i>Index</i>	<i>alpha</i>		
Line meaning fluency	.91	.93	.94
Line meaning originality	.81	.90	.89
	<i>ICC</i>		
Story drawing originality	.67	.78	.74
Square drawing originality	.81	.84	.83
	<i>Spearman-Brown coefficient</i>		
Real-world problem fluency	.78	.79	.86
Real-world problem originality	.69	.78	.63

Multiple Regression Analysis

Research Question 2: To what degree do technology-related skills and experience predict performance on an online divergent thinking test?

The present study employed two measures, the Creativity with Technology Self-efficacy Scale (CTSS) and the Technology Experience Questionnaire (TEQ), to investigate factors that contribute to performance on an online/computerized divergent thinking test. Six multiple regression analyses were conducted with three CTSS sub-scales as independent variables, and each of the six DT scores obtained from the online test as the dependent variables. Then, a series of forward stepwise multiple regressions using the same dependent variables were conducted with technology experience factors as independent variables to further explore daily technology experience that may influence online DT performance. Since the purpose of regression analysis was to examine the link between technology experience and DT performance in online environment, only data collected from online participants were used for the analysis.

Data Transformation

Before exploring evidence related to the research questions, the descriptive statistics related to each online DT variable are reported (see Table 7). As shown in Table 7, four of the six variables were moderately positively skewed. Therefore, square root transformation of the four variables was performed to establish normality for subsequent analyses.

Table 7. Descriptive Statistics for the Selected DT Scores

	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Skewness</i>	<i>Kurtosis</i>
Line meaning fluency	114	3.83	2.31	1.44	3.64
Line meaning originality	114	2.38	1.79	2.04	6.82
Story drawing originality	114	2.53	1.15	0.85	-0.05
Square drawing originality	114	2.76	1.21	0.45	-0.25
Real-world problem fluency	114	5.77	3.25	1.41	2.49
Real-world problem originality	114	3.06	2.77	1.96	4.70

Simple Multiple Regression

Research Question 2a: To what degree do technology-related and creativity-related self-efficacy predict performance on an online divergent thinking test?

Six multiple regression analyses were used to answer Research Question 2a, with creative self-efficacy, technology self-efficacy, and technology-enabled creative self-efficacy entered into the regression model as independent variables, and each DT score as a dependent variable.

How well do the three self-efficacy scores predict the Line Meaning fluency scores?

The results from the simple multiple regression testing Line Meaning fluency as the outcome variable indicated that the three sub-scales combined explained 12% of the variance ($R^2 = .12$, $F(3, 110) = 4.88$, $p = .003$; see Table 8). The model produced a multiple correlation coefficient that was weak when Line Meaning fluency was the outcome, but was still statistically significant ($R = .34$, $p = .003$). Looking at the regression coefficients in Table 8, creativity self-efficacy ($p = .015$) and technology-enabled creativity self-efficacy ($p = .001$) were strong predictors of Line Meaning fluency, whereas technology self-efficacy ($p = .568$) was not. Surprisingly, the regression coefficient for the technology-enabled creativity self-efficacy (TCS) was negative,

indicating that Line Meaning fluency scores tended to decrease when TCS scores increased. In contrast, this fluency scores tended to increase when creativity self-efficacy scores increased.

Table 8. Technology with Creativity Self-efficacy Scales Regressed on Line Meaning Fluency

<i>Scale</i>	<i>b</i>	<i>SE</i>	β	<i>t</i>	<i>p</i>
Creativity self-efficacy	.03	.01	.25	2.48	.015
Technology self-efficacy	.01	.02	.06	.57	.568
Technology-enabled creativity self-efficacy	-.08	.02	-.34	-3.47	.001

Note. N=114. $R=.34$, $R^2=.12$.

How well do the three self-efficacy scores predict the Line Meaning originality scores?

The results from the simple multiple regression that tested Line Meaning originality as the outcome variable indicated that the three sub-scales combined explained 7% of the variance ($R^2 = .07$, $F(3, 110) = 2.80$, $p = .044$; see Table 9). The model produced a multiple correlation coefficient that was weak when Line Meaning originality was the outcome, but again it was still statistically significant ($R = .27$, $p = .044$). Looking at the regression coefficients in Table 9, technology-enabled creativity self-efficacy efficacy ($p = .010$) were strong predictors of Line Meaning originality, whereas creativity self-efficacy ($p = .083$) and technology self-efficacy efficacy ($p = .412$) were not. Again, the regression coefficient for the technology-enabled creativity self-efficacy (TCS) was negative, suggesting that Line Meaning originality scores, like the previously-tested fluency scores, tend to decrease when TCS scores increase.

Table 9. Technology with Creativity Self-efficacy Scales Regressed on Line Meaning Originality

<i>Scale</i>	<i>b</i>	<i>SE</i>	β	<i>t</i>	<i>p</i>
Creativity self-efficacy	.02	.01	.18	1.75	.083
Technology self-efficacy	.01	.01	.08	.82	.412
Technology-enabled creativity self-efficacy	-.06	.02	-.26	-2.61	.010

Note. N=114. $R=.27$, $R^2=.07$.

How well do the three self-efficacy scores predict the Story Drawing originality scores?

The results from the simple multiple regression that tested Story Drawing originality as the outcome variable indicated that the three sub-scales combined explained only 3% of the variance ($R^2 = .03$, $F(3, 110) = .999$, $p = .396$; see Table 10). The model produced a multiple correlation coefficient that was both weak and statistically non-significant when Story Drawing originality was the outcome ($R = .16$, $p = .396$). As shown in Table 10, none of the three self-efficacy scores were strong predictors of Story Drawing originality ($p = .227$ to $.559$).

Table 10. Technology with Creativity Self-efficacy Scales Regressed on Story Drawing Originality

<i>Scale</i>	<i>b</i>	<i>SE</i>	β	<i>t</i>	<i>p</i>
Creativity self-efficacy	.03	.03	.13	1.22	.227
Technology self-efficacy	.03	.03	.09	.89	.377
Technology-enabled creativity self-efficacy	-.03	.05	-.06	-.59	.559

Note. $N=114$. $R=.16$, $R^2=.03$.

How well do the three self-efficacy scores predict the Square Drawing originality scores?

The results from the simple multiple regression that tested Square Drawing originality as the outcome variable indicated that the three sub-scales combined explained 7% of the variance ($R^2 = .07$, $F(3, 110) = 2.89$, $p = .039$; see Table 11). The model produced a multiple correlation coefficient that was weak but still statistically significant when Story Drawing originality was the outcome ($R = .27$, $p = .039$). As shown in Table 11, only creativity self-efficacy was a strong predictor of Square Drawing originality ($p = .004$), whereas technology self-efficacy ($p = .614$) and technology-enabled creativity self-efficacy ($p = .352$) were not.

Table 11. Technology with Creativity Self-efficacy Scales Regressed on Square Drawing Originality

<i>Scale</i>	<i>b</i>	<i>SE</i>	β	<i>t</i>	<i>p</i>
Creativity self-efficacy	.08	.03	.30	2.94	.004
Technology self-efficacy	-.02	.03	-.05	-.51	.614
Technology-enabled creativity self-efficacy	-.05	.05	-.09	-.94	.352

Note. N=114. $R=.27$, $R^2=.07$.

How well do the three self-efficacy scores predict the Real-world Problem fluency

scores? The results from the multiple regression that tested Real-world Problem fluency as the outcome variable indicated that the three sub-scales combined explained only 9% of the variance ($R^2 = .09$, $F(3, 110) = 3.80$, $p = .012$; see Table 12). The model produced a multiple correlation coefficient that was weak but statistically significant when Story Drawing originality was the outcome ($R = .31$, $p = .012$). As shown in Table 12, only creativity self-efficacy was a strong predictor of Square Drawing originality ($p = .002$), whereas technology self-efficacy ($p = .971$) and technology-enabled creativity self-efficacy ($p = .061$) were not.

Table 12. Technology with Creativity Self-efficacy Scales Regressed on Real-world Problem Fluency

<i>Scale</i>	<i>b</i>	<i>SE</i>	β	<i>t</i>	<i>p</i>
Creativity self-efficacy	.05	.02	.32	3.18	.002
Technology self-efficacy	-.001	.02	-.004	-.04	.971
Technology-enabled creativity self-efficacy	-.05	.03	-.19	-1.89	.061

Note. N=114. $R=.31$, $R^2=.09$.

How well do the three self-efficacy scores the predict Real-world Problem originality

scores? The results from the last of this series of multiple regressions that tested Real-world Problem originality as the outcome variable indicated that the three sub-scales combined explained 11% of the variance ($R^2 = .11$, $F(3, 110) = 4.34$, $p = .006$; see Table 13). The model

produced a multiple correlation coefficient that was within the weak range but was statistically significant when Real-world Problem originality was the outcome, ($R = .33$, $p = .006$). Looking at the regression coefficients in Table 13, creativity self-efficacy ($p = .001$) were strong predictors of Real-world Problem originality, whereas technology self-efficacy ($p = .743$) and technology-enabled creativity self-efficacy ($p = .073$) were not.

Table 13. Technology with Creativity Self-efficacy Scales Regressed on Real-world Problem Originality

<i>Scale</i>	<i>b</i>	<i>SE</i>	β	<i>t</i>	<i>p</i>
Creativity self-efficacy	.06	.02	.33	3.35	.001
Technology self-efficacy	.01	.02	.03	.33	.743
Technology-enabled creativity self-efficacy	-.05	.03	-.18	-1.81	.073

Note. $N=114$. $R=.33$, $R^2=.11$.

Summary of Simple Multiple Regression Tests. In sum, of the three self-efficacy scales, technology did not predict any of the online DT scores, while creativity and technology-enabled creativity predicted some of the online DT scores. Specifically, creativity self-efficacy predicted Line Meaning fluency, Square Drawing originality, Real-world Problem fluency, and Real-world Problem originality. Technology-enabled creativity self-efficacy (TCS) predicted Line Meaning fluency and originality. However, TCS was negatively correlated with these two DT scores, an effect that is worth researching further.

Stepwise Multiple Regression

Research Question 2b: What kinds of technology experience contribute to performance in online divergent thinking tests?

Forward, stepwise multiple regression modeling was conducted to 1) explore why technology-enabled creativity self-efficacy does not positively predict some of the online DT scores, and 2) look for other technology factors that may influence DT thinking. The analysis was carried out six times, each with a different online DT score as the dependent variable. The independent variables were video games (TE1), graphic design (TE2), TV watching (TE3), online video watching (TE4), video editing and sharing (TE5), picture editing and sharing (TE6), writing blogs (TE7), word processing (TE8), texting (TE9), status posting (TE10), radio/podcast listening (TE11), e-book/article reading (TE12), and emailing (TE13). These independent variables generally fell into two categories: variables TE1 to TE6 were focused on the visual/figural-based technology experience, while variables TE7 to TE13 were focused on the non-visual/verbal aspects of technology experience.

How well do different types of technology experience predict Line Meaning fluency?

The final model contained only one variable, radio/podcast listening, that explained 9% of the variance in the Line Meaning fluency scores ($R^2 = .09$, $F(1, 112) = 11.32$, $p = .001$; see Table 14). The model produced a correlation coefficient ($R = .30$, $p = .001$) that was weak but statistically significant, indicating that radio/podcast listening was a significant predictor of Line Meaning fluency.

Table 14. Technology Experience Items Regressed on Line Meaning Fluency

<i>Scale</i>	<i>b</i>	<i>SE</i>	β	<i>t</i>	<i>p</i>
radio/podcasts	.15	.05	.30	3.37	.001

Note. $N=114$. $R=.30$, $R^2=.09$.

How well do different types of technology experience predict Line Meaning originality?

A forward, stepwise multiple regression produced a final model containing three variables that explained 19% of the variance in the Line Meaning originality scores ($R^2 = .19$, $F(3, 110) = 8.47$, $p < .001$), including radio/podcast listening ($p = .021$), posting/editing pictures ($p = .001$), and status posting ($p = .001$). The model produced a multiple correlation coefficient ($R = .43$, $p < .001$) that was substantially higher than the model (shown in Table 9) that used the self-efficacy scores as predictors. The new model shown below (see Table 15) confirmed that visual technology experience (such as posting and editing pictures) was negatively correlated with Line Meaning originality. Interestingly, some non-visual technology experience (such as radio/podcast listening and status posting) were positively correlated with Line Meaning originality.

Table 15. Technology Experience Items Regressed on Line Meaning Originality

<i>Scale</i>	<i>b</i>	<i>SE</i>	β	<i>t</i>	<i>p</i>
radio/podcast listening	.10	.04	.20	2.34	.021
posting/editing pictures	-.17	.05	-.33	-3.44	.001
status posting	.16	.05	.33	3.44	.001

Note. $N=114$. $R=.43$, $R^2=.19$.

How well do different types of technology experience predict Story Drawing

originality? Forward stepwise multiple regression with Story Drawing originality as the dependent variable and the technology experience items as independent variables did not

produce a statistically significant model, suggesting that none of the technology experience types in our questionnaire was a strong predictor of Story Drawing originality.

How well do different types of technology experiences predict Square Drawing

originality? The final model contained only one variable, graphic design, which explained 4% of the variance in the Square Drawing originality scores ($R^2 = .04$, $F(1, 112) = 4.32$, $p = .040$; see Table 16). The model produced a correlation coefficient ($R = .19$, $p = .040$) that was weak but statistically significant, indicating that graphic design was a significant predictor of Square Drawing originality.

Table 16. Technology Experience Items Regressed on Square Drawing Originality

<i>Scale</i>	<i>b</i>	<i>SE</i>	β	<i>t</i>	<i>p</i>
graphic design	.28	.13	.19	2.08	.040

Note. $N=114$. $R=.19$, $R^2=.04$.

How well do different types of technology experiences predict Real-world Problem

fluency? The final model contained one variables, word processing ($p = .005$), which explained 7% of the variance in the Real-world Problem fluency scores ($R^2 = .07$, $F(1, 112) = 8.055$, $p = .005$; see Table 17). The model produced a correlation coefficient ($R = .26$, $p = .005$) that was weak but statistically significant, indicating that word processing was significant predictors of Real-world Problem fluency.

Table 17. Technology Experience Items Regressed on Real-world Problem Fluency

<i>Scale</i>	<i>b</i>	<i>SE</i>	β	<i>t</i>	<i>p</i>
word processing	.14	.05	.26	2.84	.005

Note. $N=114$. $R=.26$, $R^2=.07$.

How do different types of technology experience predict Real-world Problem

originality? The final model contained one variables, word processing ($p = .010$), which explained 6% of the variance in the Real-world Problem originality scores ($R^2 = .06$, $F(1, 112) = 6.833$, $p = .010$; see Table 18). The model produced a multiple correlation coefficient ($R = .24$, $p = .010$) that was weak but statistically significant, indicating that word processing, again, was significant predictors of Real-world Problem originality.

Table 18. Technology Experience Items Regressed on Real-world Problem Originality

<i>Scale</i>	<i>b</i>	<i>SE</i>	β	<i>t</i>	<i>p</i>
word processing	.15	.06	.24	2.61	.010

Note. $N=114$. $R=.24$, $R^2=.06$.

Summary of multiple regression. In sum, based on both the simple multiple regression and the forward regression modeling, some of the visual technology experience items actually negatively predicted the Line Meaning DT scores. In contrast, non-visual/verbal-based technology use was a strong predictor of most of the DT scores. Although the results were interesting and unexpected to some extent, these results did point out the possibility of inequivalence between paper and online versions due to the influence of prior technology experience, which brings us to the next question.

Analysis of Variance (ANOVA), Multivariate Analysis of Variance (MANOVA), and Multivariate Analysis of Covariance (MANCOVA)

Research Question 3: How do the results obtained through online divergent thinking tests differ from the results of the paper-and-pencil version?

Descriptive Statistics

Before further exploring evidence related to the research questions, the descriptive statistics related to each DT variable are provided by group and gender in Table 19.

Table 19. Descriptive Statistics by Group and Gender for DT Indices

	Basic Online				Advanced Online				Paper and Pencil			
	Female (n=45)		Male (n=12)		Female (n=36)		Male (n=21)		Female (n=28)		Male (n=22)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Line meaning fluency	1.85	0.59	2.02	0.33	1.85	0.53	1.89	0.70	1.96	0.30	2.18	0.69
Line meaning originality	1.39	0.53	1.57	0.37	1.39	0.54	1.58	0.67	1.46	0.34	1.71	0.74
Story drawing originality	2.42	1.02	2.31	0.95	2.72	1.30	2.54	1.24	3.01	1.16	2.68	1.33
Square drawing originality	2.56	1.13	2.92	1.64	2.86	1.15	2.92	1.22	3.46	0.86	2.87	1.17
Real-world problem fluency	2.20	0.52	2.34	0.57	2.37	0.64	2.47	0.87	2.39	0.45	2.50	0.68
Real-world problem originality	1.36	0.65	1.71	0.54	1.64	0.73	1.92	0.89	1.74	0.50	1.81	0.71

Two-Way Factorial Analysis of Variance (ANOVA)

Research Question 3a: With respect to the divergent thinking score, do the online groups and paper-and-pencil group differ when gender is considered?

A series of two-way factorial analyses of variance (ANOVAs) were performed to establish and explore group and gender differences with respect to each DT score. It was hypothesized that the advanced online group would achieve higher originality scores in the

Drawing test than the basic online and paper-and-pencil groups. It was also hypothesized that gender may influence scores obtained on verbal tasks, including the Line Meaning test and the Real-world Problem test. The interaction between group and gender were also explored in this analysis.

Group and gender differences in line meaning fluency scores. Results from two-way ANOVA (see Table 20) showed that neither group ($F(2, 158) = 1.70, p = .185, \eta^2 = .02$) nor gender ($F(1, 158) = 2.35, p = .128, \eta^2 = .02$) had a statistically significant effect on the Line Meaning fluency scores. No interaction effect was found between gender and group ($F(2, 158) = 0.33, p = .719, \eta^2 = .004$).

Table 20. Analysis of the Differences in Group and Gender on Line Meaning Fluency

<i>Variable</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>
group	1.05	2	0.52	1.70	.185
gender	0.72	1	0.72	2.35	.128
group * gender	0.20	2	0.10	0.33	.719

Group and gender differences in line meaning originality scores. Results from two-way ANOVA (see Table 21) showed that males scored significantly higher than females on Line Meaning originality ($F(1, 158) = 4.77, p = .030, \eta^2 = .03$). However, group had no statistically significant effect on Line Meaning originality scores ($F(2, 158) = 0.57, p = .566, \eta^2 = .01$). No interaction effect was found between gender and group ($F(2, 158) = 0.05, p = .953, \eta^2 = .001$).

Table 21. Analysis of the Differences in Group and Gender on Line Meaning Originality

<i>Variable</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>
group	0.34	2	0.17	0.57	.566
gender	1.43	1	1.43	4.77	.030
group * gender	0.03	2	0.01	0.05	.953

Group and gender differences in the story drawing originality scores. Results from two-way ANOVA (see Table 22) showed that, again, neither group ($F(2, 158) = 1.80, p = .168, \eta^2 = .02$) nor gender ($F(1, 158) = 1.09, p = .298, \eta^2 = .01$) significantly affected Story Drawing originality scores. No interaction effect was found between gender and group ($F(2, 158) = 0.10, p = .908, \eta^2 = .001$).

Table 22. Analysis of the Differences in Group and Gender on Story Drawing Originality

Variable	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>
group	5.00	2	2.50	1.80	.168
gender	1.51	1	1.51	1.09	.298
group * gender	0.27	2	0.13	0.10	.908

Group and gender differences in the square drawing originality scores. Results from two-way ANOVA (see Table 23) showed that neither group ($F(2, 158) = 1.60, p = .205, \eta^2 = .02$) nor gender ($F(1, 158) = 0.09, p = .765, \eta^2 = .001$) had a significant effect on Square Drawing originality scores. No interaction effect was found between gender and group ($F(2, 158) = 1.97, p = .143, \eta^2 = .02$).

Table 23. Analysis of the Differences in Group and Gender on Square Drawing Originality

Variable	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>
group	4.27	2	2.13	1.60	.205
gender	0.12	1	0.12	0.09	.765
group * gender	5.26	2	2.63	1.97	.143

Group and gender differences in the real-world problem fluency scores. Results from two-way ANOVA (see Table 24) showed that neither group ($F(2, 158) = 1.00, p = .370, \eta^2 = .01$) nor gender ($F(1, 158) = 1.22, p = .271, \eta^2 = .01$) had a significant effect on Real-world

Problem fluency scores. No interaction effect were found between gender and group ($F(2, 158) = 0.01, p = .987, \eta^2 < .001$).

Table 24. Analysis of the Differences in Group and Gender on Real-world Problem Fluency

Variable	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>
group	0.76	2	0.38	1.00	.370
gender	0.46	1	0.46	1.22	.271
group * gender	0.01	2	0.00	0.01	.987

Group and gender differences in the real-world problem originality scores. Results from two-way ANOVA (see Table 25) showed that males scored significantly higher than females on Real-world Problem originality ($F(1, 158) = 4.11, p = .044, \eta^2 = .03$). However, group did not have a statistically significant effect on Real-world Problem originality scores ($F(2, 158) = 1.73, p = .181, \eta^2 = .02$). No interaction effect was found between gender and group ($F(2, 158) = 0.57, p = .568, \eta^2 = .01$).

Table 25. Analysis of the Differences in Group and Gender on Real-world Problem Originality

Variable	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>
group	1.61	2	0.81	1.73	.181
gender	1.92	1	1.92	4.11	.044
group * gender	0.53	2	0.26	0.57	.568

Summary of Two-Way ANOVAs. In sum, only a gender effect was found in two of the six DT scores – the Line Meaning originality score and the Real-world Problem originality score. No group effects or interaction effects were found for the DT scores, indicating that the testing environment, whether online with different tools or in a classroom with pencil and paper, did not significantly affect the DT scores.

One-Way Multivariate Analysis of Variance (MANOVA)

Research Question 3b: Is there an overall group effect on the divergent thinking scores?

To test the overall effect of the three groups on divergent thinking (DT) scores, MANOVA was conducted with the six DT scores as dependent variables, and group as the independent variable.

Homogeneity of variance was examined at both the multivariate and univariate levels. Neither Box's M ($p = .076$) nor Levene's tests ($p = .054$ to $.515$) were significant. The multivariate effect of group difference on DT scores was also not statistically significant, with Wilk's lambda = .88, $F(12, 312) = 1.74$, $p = .057$, $\eta^2 = .06$.

One-Way Multivariate Analysis of Covariance (MANCOVA)

Research Question 3c: Is there an overall group effect on divergent thinking scores after controlling for the gender effect?

As DT scores may vary by gender, to test the overall effect of three groups on DT scores and control for the gender effect, MANCOVA was performed with the six DT scores as dependent variables, group as the independent variable, and gender as the covariate.

Homogeneity of variance was examined at both the multivariate and univariate levels. Neither Box's M ($p = .076$) nor Levene's tests ($p = .051$ to $.404$) were significant. The multivariate effect of group difference on DT scores was also not statistically significant, with Wilk's lambda = .89, $F(12, 310) = 1.62$, $p = .084$, $\eta^2 = .06$. Gender (the covariate), however, had a statistically significant effect on DT scores, with Wilk's lambda = .91, $F(6, 155) = 2.46$, $p = .027$, $\eta^2 = .09$.

DISCUSSION

To evaluate the feasibility of an online divergent thinking test, three research questions were proposed, including Research Question 1 (*How reliable are the scores that are obtained from the divergent thinking tests?*), Research Question 2 (*To what degree do technology-related skills and experience predict performance in online divergent thinking tasks?*), and Research Question 3 (*How do the results obtained from the online divergent thinking tests differ from those of the paper-and-pencil version?*). Research Question 1 addresses the reliability aspect of the test, while Research Question 2 and 3 represent an effort to provide some evidence of validity for the test. The results for research questions 1 through 3 will be discussed to help us understand the affordances provided by technology in creativity education.

Reliability of Online Divergent Thinking Test Scores

The present study used three types of tasks to assess divergent thinking abilities, including *Line Meaning*, *Drawing*, and *Real-world Problem*. At least two themes emerged as reliability evidence was being collected for each of these tasks, and both are important in order to understand some basic issues regarding the interpretation of online DT test scores.

Theme 1: Different scoring methods produced different evidence of reliability

The current study calculated two types of divergent thinking scores for each task: *Fluency* and *Originality*. While fluency scores can only be calculated using the objective method (counting the number of responses), originality scores can be calculated using various methods. The present study employed two scoring methods to calculate the originality scores for each participant: 1) subjective rating, as conducted by three external raters (taking the average rating of all three), and 2) objective ratings, as coded by the investigator (top 20%). We then analyzed the reliability of the scores to decide which method would be used for subsequent analysis. The

Drawing test did not produce objective scores of originality. Therefore, only scores from the Line Meaning test and Real-world Problem test are discussed in this section.

With regard to the subjective rating of originality, evidence of both inter-rater reliability and inter-item reliability was examined. Specifically, the Line Meaning test, which consisted of three items, produced inter-rater reliability estimates that were just inside the good range ($ICC = .61, .61, \text{ and } .62$) and an inter-item reliability estimate that fell into the poor range ($ICC = .34$). The Real-world Problem test, which consisted of two questions, produced inter-rater reliability estimates that were lower than .70 ($ICC = .60 \text{ and } .67$) and an inter-item reliability estimate that was within the fair range (Spearman-Brown Coefficient = .59). Despite the fact that the Real-world Problem test performed better than the Line Meaning test in terms of inter-item reliability estimates, the overall reliability was still poor if both inter-item and inter-rater reliability estimates were taken into consideration. Thus, the subjective rating of originality, conducted by averaging the external raters' ratings of all responses, failed to produce evidence of reliability for the Line Meaning test and the Real-world Problem test.

With regard to the objective scoring of originality, strong evidence of reliability was found. When the investigator counted the top 20% responses for each participant and used that number as the originality score, both the Line Meaning test and the Real-world Problem test yielded higher inter-item reliability coefficients ($ICC = .88$ and Spearman-Brown coefficient = .73) than with the subjective ratings. The results indicated that the estimates of internal consistency across items for these two sub-tests were adequate. Also, since no inter-rater reliability needs to be estimated for objective ratings, it had fewer error sources than the subjective ratings.

The results seem to contradict the findings of previous research that examined and compared subjective and objective scoring methods in terms of their reliability estimates. For example, Plucker, Qian and Schmalensee (2014) employed seven scoring methods to calculate originality scores for two real-world problems (Pat and Kelly problem) and found no dramatic differences between a subjective rating method and the top 20% scoring method in terms of their inter-item reliability estimates ($ICC = .70$ and $.62$, respectively). The authors also found that the correlations between raters were very high for the two problems ($r = .90$ and $.91$). Similarly, Plucker, Qian and Wang (2011) used the same methods for the Noise and Wheel problems and found no dramatic differences between the subjective rating method and the top 20% scoring method in terms of their inter-item reliability estimates ($ICC = .62$ and $.50$, respectively). The correlation between raters was high as well ($r = .89$). Silvia and colleagues (2008) used generalizability coefficients that considered both the rater facet and the task facet to estimate and compare the reliability evidence between the subjective and objective rating methods (uniqueness 1/0). They found that the objective scoring method performed more poorly than the average ratings, which stands in contrast with the findings of the present study.

Regarding these differences, several points should be noted. First, none of the studies mentioned above used exactly the same methods as were used in the present study. Plucker and colleagues used subjective ratings of the entire ideational pool, whereas the present study used averaged ratings from all the responses in the pool (which was also done in Silvia et al.'s study). However, Silvia and colleagues calculated objective scores using the 1/0 method (in which any response given by only one person received a 1; all other responses received a 0), whereas the present study used the number of top 20% responses as the objective score for each participant. Second, the recruitment and training of raters might also influence reliability. For example,

previous research has tended to utilize graduate students from the same department (e.g. educational psychology). The present study, on the other hand, employed 3 raters (graduate students) from three different departments. Despite the fact that the investigator of the present study spent 30 minutes to train the raters, this may not have been adequate compared to studies that used raters with more creativity-related knowledge. If this is true, it means that creativity-related knowledge may have an impact on the reliability of subjective scoring. Third, the type of tasks (or responses) may also affect the reliability of divergent thinking scores (Silvia et al., 2008). This may have been the case especially when the DT tests asked participants to come up with simple answers (in one or two words), as opposed to elaborated responses (such as drawing). In the present study, the number of test responses produced by the participants might have been too many for the raters (the present study produced 2,252 verbal responses for the Line Meaning test and 2,136 solutions for the Real-world Problem test), which may have resulted in fatigue and thus inconsistencies between raters and items. Similarly, Hass (2015) used an alternative use test (e.g. “tell me different uses of a brick”) and an instance test (e.g. “name things with leg”) in one of his studies, which yielded three inter-rater reliability estimates (ICCs) that ranged from .15 to .53. This last point touches upon the second theme that emerged from the current study.

Theme 2: Different types of responses produced different evidence of reliability

Generally, two types of tasks were used in the present study: tasks that asked for verbal responses and tasks that asked for drawing responses. The Line Meaning test and the Real-world Problem test belonged to the former category, while the Drawing test belonged to the latter.

A clear pattern emerged as reliability information was collected for both types of responses. Both of the two Drawing sub-tests – the Story Drawing test and the Square Drawing

test – produced inter-rater reliability estimates that were within the excellent range ($ICC = .74$ and $.83$, respectively). These estimates were higher than those obtained from the other two tests, as discussed in Theme 1 (ICC ranged from $.60$ to $.67$). Despite the fact that the inter-item reliability estimate was low for the drawing tasks (Spearman-Brown coefficient = $.55$), this was expected because the two sub-tests (*Story* and *Square*) were very different from one another compared to the sub-tests in the other DT tests.

Few studies have compared evidence of inter-rater reliability between drawing responses and verbal responses, but the results of the present study are consistent with previous studies with regard to the inter-rater reliability of drawing responses. For example, Jankowska and Karwowski (2015) designed a test of imagery ability that required drawing responses. The authors recruited 4 judges to rate all the responses in terms of vividness, originality, and transformativeness, and they found that the inter-judge reliability estimate for originality was adequate ($ICC = .89$). Similarly, Kālis, Roķe, and Krūmiņa (2014) investigated the psychometric properties of the Test for Creative Thinking–Drawing Production (TCT-DP), which was developed by Jellen and Urban (1986) in Germany. Four of the fourteen creativity criteria of the TCT-DP were unconventionality indices, which were similar to the originality index used in the present study. Using the methodology of structural equation modeling (SEM), Kālis and colleagues found adequate inter-rater reliability estimates (3 raters) for these four indices ($\rho = .80$ to $.90$).

A possible explanation for why the subjective ratings of the drawing responses yielded stronger evidence of inter-rater reliability in our study may be that the drawing responses were more interesting and were significantly fewer in number than the verbal responses. As mentioned previously, the Line Meaning test and the Real-world Problem test produced more than 4,000

responses, whereas the two drawing sub-tests produced only 110 pictures in total. The smaller number of responses, combined with the aesthetic quality of drawing, makes them easier and more pleasant to conduct subjective rating of.

Taking into consideration all of the reliability information discussed above, the investigator decided to use six DT scores for subsequent analyses. Two were fluency scores obtained from the Line Meaning test and the Real-world Problem test. The fluency scores from the Drawing test were dropped due to their invariability. The other four scores were originality scores obtained from all three sub-tests, including objective scores from the Line Meaning test and the Real-world Problem test, and two subjective scores from the Story Drawing test and the Square Drawing test.

The Relationship between Technology Use and Divergent Thinking Abilities

To answer Research Question 2 (*To what degree do technology-related skills and experience predict performance in online divergent thinking tasks?*), the present study also employed two other technology-related measurements, the *Creativity with Technology Self-efficacy Scale* (CTSS) and the *Technology Experience Survey* (TES). The CTSS consisted of three sub-scales: General Creativity Self-efficacy (GCS), General Technology Self-efficacy (GTS), and Technology-Enabled Creativity Self-efficacy (TECS). A third theme emerged when these technology-related measures were used to predict divergent thinking performance in the online test.

Theme 3: Technology Use Might Influence Online Divergent Thinking Performance

Visually-focused Technology Use and Divergent Thinking. The results showed that, while creativity self-efficacy positively predicted most of the DT scores, people's confidence

with technology use for creative activities (as measured by the TECS) negatively predicted the DT scores obtained from Line meaning tasks, including both fluency and originality scores. Further analyses using the technology experience survey yielded negative correlations between picture editing and the Line Meaning originality scores, contrasting with the hypothesis.

A possible explanation for the negative correlations may be that most people post or edit pictures for practical purposes, such as adjusting picture qualities for social media, rather than for creative purposes. If this was the case, the picture editing had then become a repetitive task rather than work that stimulated visual imagination, as measured by the Line Meaning test.

Non-visually-focused Technology Use and Divergent Thinking. The results also showed that some forms of non-visual technology use positively predicted some DT scores. This was consistent with most of the previous research investigating the influence of radio, reading, and writing on creative thinking. For example, Valkenburg and Beentjes (1997) found that radio elicited more novel responses than television in story completion tasks. Similarly, Greenfield, Farrar, and Beagles-Roos (1986) found that radio was more stimulating to imagination processes, leading to more fluency in story making, than television. Not only radio, but reading and writing also have positive influences on creativity, as the previous literature has shown (Ritchie et al., 2013; Wang, 2012).

A possible stream of exploration is whether radio listening, reading, and writing, as opposed to TV watching or video gaming, are more associated with abilities that facilitate imagination, divergent thinking, and creative problem solving. According to the reduction hypothesis proposed by Calvert and Valkenburg (2013), the generation of visual images and novel responses would be impeded by exposure to visually-focused media, which supply those images rather than stimulating their creation via verbal modalities.

All of our results regarding the relationships between types of technology experience and divergent thinking should be interpreted with caution, as they are correlational and exploratory. More research needs to be conducted to elucidate technology's influence on creative thinking. Nonetheless, the results did point out the possibility of technology influence on people's divergent thinking abilities, which may have implications for computerized DT testing with regard to the equivalence of online and paper versions.

Mode Effect and Gender Effect on Divergent Thinking Scores

Another two themes emerged while comparing DT scores between groups. Specifically, the comparison of DT scores showed no differences between the basic-online group, the advanced-online group, and the paper-and-pencil group after controlling for gender. This study also compared and found significant differences in DT scores between male and female participants.

Theme 4: No mode effect was found for DT scores or their reliability estimates

After the six DT indices were selected, the next step was to determine the differences between the online and paper versions of the test. The purpose of this section is to discuss whether there was any mode effect on people's divergent thinking performance, and the implications of this. Two kinds of information were examined: 1) differences in reliability estimates, and 2) differences in DT scores.

No mode effect on reliability estimates. The current study found no dramatic differences in either inter-item or inter-rater reliability estimates between the three groups (online-basic, online-advanced, and paper-and-pencil). Although few studies have compared the reliability

information of drawing tasks, our findings were consistent with previous research that compared reliability information for verbal responses between computer and paper versions of DT tests. For example, Lau and Cheung (2010) compared the electronic and paper-and-pencil versions of the Wallach-Kogan Creativity Tests in terms of their internal consistency (alpha), and found that the magnitudes were comparable. The tests they used only produced verbal responses. The present study has demonstrated that drawing responses can produce similar or even the same inter-rater reliability estimates across modes.

No mode effect on divergent thinking scores. Contrary to the hypothesis, we found no differences in any of the divergent thinking scores after the gender effect was controlled for. The results are consistent with the latest research on computerized DT tests (Zabramski, Gkouskos, & Lind, 2011), while contradicting findings obtained 20 years ago by Kwon (1994). Initially, our hypothesis was made on the basis of the idea that mouse input would have a negative impact on participants' drawing performance. As Zabramski and Neelakannan (2011) have claimed, feedback from mouse movements is presented on the computer screen, which is spatially separated from the user, making the entire human-computer interaction indirect. In addition, mouse movement is usually smaller than cursor movement on the screen, which would also affect drawing accuracy in the test. However, neither the present study nor the study of Zabramski and colleagues found significant differences in DT scores between different input methods, indicating that a mode effect is not present in DT testing.

An important factor that might contribute to the inconsistency between the results from now and two decades ago is computer familiarity. People who use computers often may develop motor skills that would compensate for the inaccuracy and inconvenience that comes with the use of a mouse. What is evident about the changes in computer use across time is that people

nowadays use computers and mice more often than people living 20 years ago. According to the 2014 U.S. Census, more than 80% of American households now own a computer, compared to 30% twenty years ago. College students, a major component of our sample, would have even more access to computers. Zabramski and colleagues collected their data in Sweden, a country that has the most computers per household in the world (OECD, 2015). Their sample consisted of adults aged from 20 to 38. In contrast, the sample in Kwon's (1996) study consisted of fifth and sixth graders. Therefore, both time and age would lead to differences in participants' computer familiarity, which would in turn affect their performance in computerized drawing tests.

Interestingly, Zabramski and colleagues (2013) provided a different explanation, arguing that the graphic user interface (GUI) might lead to additional cognitive load, influencing participants' drawing performance. They believed that the more elaborate GUI used in Kwon's (1996) study – as opposed to the simplified GUI used in their study – was the major factor contributing to the differences between the computerized and paper versions. This claim, however, is not supported by the findings of the present study. The present study used elaborative GUIs in both the basic-online group and advanced-online group, according to Zabramski's standards (drawing tools were visible on the screen), but no differences were found between these two online groups and the paper-and-pencil group. Neither did Zabramski and his colleagues (2013) find any differences in DT scores between the elaborate and simple GUIs.

Theme 5: Gender differences in DT scores were inconsistent

Significant gender differences were found in the Line Meaning originality scores and the Real-world Problem originality scores, with males producing significantly higher overall mean scores than females. The results added more mystery to the mixed evidence on gender

differences found in the previous literature. In fact, no consistent pattern of gender differences in creativity tests has been identified (Baer & Kaufman, 2008; Cheung & Lau, 2010), with a greater number of studies finding that girls or women score higher on verbal DT tests.

One possible explanation for the gender effect in the present study is that the participants were self-selected. Self-selected males might produce higher originality scores than females in verbal tasks. However, the results of the present study should be interpreted with caution, as there might be bias due to the fact that the females outnumbered the males by almost double. Therefore, further research needs to be done to study gender differences in creativity.

Summary of the Feasibility of an Online Divergent Thinking Test

The above analyses have yielded important implications for the use of an online divergent thinking test. All of the issues discussed above, together with the design issues of the online testing system, will be summarized and further discussed below.

Limitations and Disadvantages of the Current Online Test.

Several limitations should be noted before we discuss the potential advantages of an online creativity test. The present study represents an effort to computerize the divergent thinking test using current available technologies. One of the goals was to develop and test a system, rather than to publish a new tool. To achieve this goal, compromises needed to be made on several aspects.

First, unlike the application development done in the previous studies, none of the technologies involved in the present study were owned by the investigator. For example, the web drawing interface was designed and maintained by a third-party. The investigator only added some simple coding to adapt the tool into the testing system. One of the downsides of this

method was that the researcher was unable to have total control of the tool if any errors occurred. In fact, in the initial experimental design, the drawing interface presented to the advanced-group contained more tools, such as a straight-line burton and circle burton, to allow participants more control during the drawing process. In the actual experiment, however, these tools mysteriously disappeared, despite communication between the investigator and the tool's developer. Neither could identify the cause the problem, leading to fewer tools being shown in the drawing interface for the advanced group.

Second, an online creativity test may not be suitable for small-scale assessments. The findings in the present study have demonstrated that the online and paper versions were equivalent in almost every respect, which leads to the next question: If they are equivalent, why would we use an online test? In other words, is there value added by putting the test online or computerizing it? The answer may be no. For example, a small-scale and occasional assessment, which can usually be carried out in a classroom, does not need to be computerized. Using an online version would not only be an expensive option, but would also bring with it technological issues.

Third, the present study only touched on some of the potential of a computerized creativity test, leaving the rest to be explored by future research. For example, the development of a computerized creativity test should not just focus on putting the paper version online or on computers. With the advanced technologies available, computerized testing can do much more than replicating the paper version. This will be further discussed in the next section.

Advantages and Potential of Online Creativity Testing

With the aforementioned limitations in mind, the present study also demonstrated several advantages and the potential of an online DT test.

First, online testing can reach more people within a relatively short amount of time, at a lower cost. Paper-and-pencil testing usually requires people to be at a particular place at a certain time to take the test. If people's schedules conflict with the test time, they will need to change their schedule, or the test administrator will need to change the time or place. This issue does not apply to online testing. As long as the test takers have the link to the testing system, they can choose to do the test at the time and place that suits them. This is especially an important feature when test administrators want to collect self-selected samples in a large-scale assessment.

Second, current technologies allow test designers to develop their own creativity testing systems. Despite the pitfalls mentioned in the previous section, there are several advantages of using third-party technologies. For example, test designers do not need to possess advanced technological knowledge (such as programming or web design skill) to develop such a tool, thus reducing learning time as well as any costs associated with hiring technology assistants. In addition, since different interfaces do not significantly influence creativity test performance, test designers/administrators can experiment with different tools to determine which solutions best satisfy their purposes.

Third, technologies hold great potential in allowing people to develop automatic and adaptive creativity tests. In the present study, objective scoring of verbal responses produced high reliability estimates. This has important implications for automatic scoring because objective scoring, which is carried out by first pooling all the responses and then counting the

number of responses given by less than 20% of the sample, can be conducted via computer program if it can recognize the meaning of the responses. Given that artificial intelligence (AI) systems such as SIRI, Alexa, and Cortana, can nowadays easily recognize and respond to simple commands from people, it would not be very difficult to design a computer program that recognizes verbal responses to various types of DT tasks. If automatic scoring is feasible, then adaptive creativity testing is also possible. This of course requires more research in the future.

In summary, the findings of this study indicate that an online DT test can produce the same, reliable DT scores as the paper test does. Despite that technology use may influence people's online divergent thinking performance, it does not result in the differences between paper and online versions of the test. This form may be more useful for large-scale assessments because it makes data collection much faster and more convenient. The current study also demonstrates that online creativity testing could add more value if automatic scoring and adaptive testing are used.

Future Directions

The present study provides a foundation for continued development and evaluation of an online or computerized divergent thinking test which includes both verbal tasks and figural (drawing) tasks. The design of this study only addressed very limited aspects of the test, basically comparing online and paper versions of the test with regard to fluency and originality scores. To fully test the equivalence between online and paper versions requires further investigations into other DT indices, such as flexibility and elaboration, as well as other factors, such as completion time and drawing process. In addition, further studies would benefit from the development of automatic scoring and adaptive testing techniques to fully explore the potentials of digital technologies.

With regard to the technology influence on divergent thinking, the present study only provides correlational results that warrant more research in the future. One of the findings was the negative correlation between picture editing/posting and visual imagination/ideation. Further research could examine the purposes of picture editing/posting, the tools that are involved, and when and how relevant technologies are used in daily life, to find out specific behaviors that are associated with the possible decline of visual imagination. Another interesting finding was the positive relationships between non-visually-focused technology use and divergent thinking abilities. Further research needs to be done to find out if there are causal relationships between the two, which may provide educational implications for educators who are interesting in the cultivation of divergent thinking abilities, or more broadly, creative abilities in educational settings.

References

- Andersen, L. (2014). Visual-spatial ability: Important in STEM, ignored in gifted education. *Roeper Review*, 36, 114-121. doi:10.1080/02783193.2014.884198
- Baer, J. (1993). Why you shouldn't trust creativity tests. *Educational Leadership*, 51, 80-83.
- Baer, J. (1994). Divergent thinking is not a general trait: A multi-domain training experiment. *Creativity Research Journal*, 7, 35-46. doi:10.1080/10400419409534507
- Baer, J., & Kaufman, J. C. (2008). Gender Differences in Creativity. *The Journal of Creative Behavior*, 42, 75-105. doi:10.1002/j.2162-6057.2008.tb01289.x
- Beghetto, R. A., Kaufman, J. C., & Baxter, J. (2011). Answering the unexpected questions: Exploring the relationship between students' creative self-efficacy and teacher ratings of creativity. *Psychology of Aesthetics, Creativity, And The Arts*, 5, 342-349. doi:10.1037/a0022834
- Benedek, M., Franz, F., Heene, M., & Neubauer, A. C. (2012). Differential effects of cognitive inhibition and intelligence on creativity. *Personality and Individual Differences*, 53, 480-485. doi:10.1016/j.paid.2012.04.014
- Benedek, M., Mühlmann, C., Jauk, E., & Neubauer, A. C. (2013). Assessment of divergent thinking by means of the subjective top-scoring method: Effects of the number of top-ideas and time-on-task on reliability and validity. *Psychology of Aesthetics, Creativity, and the Arts*, 7, 341-349. doi:10.1037/a0033644

- Brown, Robert T. (1989). Creativity: What are We to Measure? In J. A. Glover, R. R. Ronning & C. R. Reynolds (Eds.), *Handbook of Creativity* (1 ed., pp. 3-32). New York: Plenum Press.
- Calvert, S. L., & Valkenburg, P. M. (2013). The Influence of Television, Video Games, and the Internet on Children's Creativity. In M. Taylor (Ed.), *The Oxford Handbook of the Development of Imagination* (pp. 438-450). New York: Oxford University Press.
doi:10.1093/oxfordhb/9780195395761.013.0028
- Chand, I., & Runco, M. A. (1993). Problem finding skills as components in the creative process. *Personality and Individual Differences*, 14, 155-162. doi:10.1016/0191-8869(93)90185-6
- Cheung, P. C., & Lau, S. (2010). Gender Differences in the Creativity of Hong Kong School Children: Comparison by Using the New Electronic Wallach–Kogan Creativity Tests. *Creativity Research Journal*, 22, 194-199. doi:10.1080/10400419.2010.481522
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6, 284-290. doi:10.1037/1040-3590.6.4.284
- Clapham, M. M. (2011). Testing/Measurement/Assessment. In M. A. Runco & S. R. Pritzker (Eds.), *Encyclopedia of creativity* (2 ed., pp. 458-464). London: Elsevier Inc.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155–159. doi:10.1037/0033-2909.112.1.155

- Dunleavy, M., Dede, C., & Mitchell, R. (2009). Affordances and limitations of immersive participatory augmented reality simulations for teaching and learning. *Journal of Science Education & Technology*, 18, 7-22. doi:10.1007/s10956-008-9119-1
- Eisinga, R., Grotenhuis, M. te, & Pelzer, B. (2012). The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? *International Journal of Public Health*, 58, 637-642. doi:10.1007/s00038-012-0416-3
- Fornell, C., & Larcker, D. F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, 18, 39-50. doi:10.2307/3151312
- Greenfield, P., Farrar, D., & Beagles-Roos, J. (1986). Is the medium the message?: An experimental comparison of the effects of radio and television on imagination. *Journal of Applied Developmental Psychology*, 7, 201-218. doi:10.1016/0193-3973(86)90029-8
- Halsey, K., Jones, M., & Lord, P. (2006). *What works in stimulating creativity amongst socially excluded young people*. Slough, Berkshire: National Foundation for Educational Research. Retrieved from <http://nfernew.dudobi.com/nfer/publications/NES01/NES01.pdf>
- Gardner, H. (1993). *Creating minds*. New York, NY: Basic Books.
- George, J. M. (2008). Creativity in organizations. In J. P. Walsh, & A. P. Brief (Eds.), *The Academy of Management Annals* (Vol. 1, pp. 439-477). New York, NY: Taylor & Francis Group/Lawrence Erlbaum Associates.
- Guilford, J. P. (1950). Creativity. *The American Psychologist*, 5, 444-454. doi:10.1037/h0063487

- Guilford, J. P. (1967). *The nature of human intelligence*. New York: McGraw-Hill
- Hass, R. W. (2015). Feasibility of online divergent thinking assessment. *Computers in Human Behavior*, 46, 85-93. doi:10.1016/j.chb.2014.12.056
- Jackson, L. A., Witt, E. A., Games, A. I., Fitzgerald, H. E., von Eye, A., & Zhao, Y. (2012). Information technology use and creativity: Findings from the Children and Technology Project. *Computers in Human Behavior*, 28, 370-376. doi:10.1016/j.chb.2011.10.006
- Jankowska, D. M., & Karwowski, M. (2015). Measuring creative imagery abilities. *Frontiers in Psychology*, 6. doi:10.3389/fpsyg.2015.01591
- Jellen, H. G., & Urban, K. K. (1986). The TCT-DP (Test for Creative Thinking-Drawing Production): An instrument that can be applied to most age and ability groups. *Creative Child & Adult Quarterly*, 11, 138-155.
- Kālis, E., Roķe, L., & Krūmiņa, I. (2014). Investigation of Psychometric Properties of the Test for Creative Thinking-Drawing Production: Evidence from Study in Latvia. *The Journal of Creative Behavior*, 50, 47-63. doi:10.1002/jocb.68
- Kaufman, J. C., Plucker, J. A., & Baer, J. (2008). *Essentials of creativity assessment*. Hoboken, N.J.: John Wiley.
- Kogan, N., & Pankove, E. (1974). Long-term predictive validity of divergent-thinking tests: Some negative evidence. *Journal of Educational Psychology*, 66, 802-810. doi:10.1037/h0021521

- Kwon, M. C. (1996). *An exploratory study of a computerized creativity test: comparing paper-pencil and computer-based versions of the Torrance tests of creative thinking* (Doctoral dissertation, Texas A&M University).
- Lau, S., & Cheung, P. C. (2010). Creativity assessment: Comparability of the electronic and paper-and-pencil versions of the Wallach-Kogan Creativity Tests. *Thinking Skills and Creativity*, 5, 101-107. doi:10.1016/j.tsc.2010.09.004
- Leeson, H. V. (2006). The Mode Effect: A Literature Review of Human and Technological Issues in Computerized Testing. *International Journal of Testing*, 6, 1-24. doi:10.1207/s15327574ijt0601_1
- Magnussen, R., Hansen, S. D., Plank, T., & Sherson, J. F. (2013). Scientific discovery games for authentic science education. In *Proceedings of the 7th ECGBL (pp. 344–351)*. Porto: Instituto Superior de Engenharia do Porto.
- Mishra, P. (2012). Rethinking technology & creativity in the 21st century: Crayons are the future. *TechTrends*, 56(5), 13-16. doi:10.1007/s11528-012-0594-0
- Mumford, M. D., & Licuanan, B. (2004). Leading for innovation: Conclusions, issues, and directions. *The Leadership Quarterly*, 15, 163-171. doi:10.1016/j.leaqua.2003.12.010
- Nusbaum, E. C., & Silvia, P. J. (2011). Are intelligence and creativity really so different? Fluid intelligence, executive processes, and strategy use in divergent thinking. *Intelligence*, 39, 36-45. doi:10.1016/j.intell.2010.11.002
- OECD. (2016). *OECD Factbook 2015-2016: Economic, Environmental and Social Statistics*, Paris, France: OECD Publishing. doi:10.1787/factbook-2015-en

Okuda, S. M., Runco, M. A., & Berger, D. E. (1991). Creativity and the finding and solving of real-world problems. *Journal of Psychoeducational Assessment*, 9, 45-53.

doi:10.1177/073428299100900104

Palaniappan, A. K. (2012). Web-based Creativity Assessment System. *International Journal of Information and Education Technology*, 2, 255-258. doi:10.7763/ijiet.2012.v2.123

Partnership for 21st Century Skills (P21). (2014). *Our mission*. Retrieved from <http://www.p21.org/about-us/our-mission>

Pásztor, A., Molnár, G., & Csapó, B. (2015). Technology-based assessment of creativity in educational context: the case of divergent thinking and its relation to mathematical achievement. *Thinking Skills and Creativity*, 18, 32-42. doi:10.1016/j.tsc.2015.05.004

Pignéguy, S. (2004). Our Didcot: An inclusive arts project devised by a learning support unit. *Support for Learning*, 19, 77-80. doi:10.1111/j.0268-2141.2004.00324.x

Plucker, J. A. (1999). Is the proof in the pudding? Reanalyses of Torrance's (1958 to present) longitudinal study data. *Creativity Research Journal*, 12, 103-114. doi:10.1207/s15326934crj1202_3

Plucker, J. A., & Esping, A. (2015). Intelligence and creativity: a complex but important relationship. *Asia Pacific Education Review*, 16, 153-159. doi:10.1007/s12564-015-9374-9

Plucker, J. A., Qian, M., & Schmalensee, S. L. (2014). Is what you see what you really get? Comparison of scoring techniques in the assessment of real-world divergent thinking. *Creativity Research Journal*, 26, 135-143. doi:10.1080/10400419.2014.901023

- Plucker, J. A., Qian, M., & Wang, S. (2011). Is originality in the eye of the beholder? Comparison of scoring techniques in the assessment of divergent thinking. *Journal of Creative Behavior*, 45, 1-22. doi:10.1002/j.2162-6057.2011.tb01081.x
- Plucker, J. A., & Renzulli, J. S. (1999). Psychometric approaches to the study of human creativity. In R. J. Sternberg (Ed.), *Handbook of creativity* (pp. 35-60). New York, NY: Cambridge University Press.
- Pretz, J. E., & Link, J. A. (2008). The Creative Task Creator: A tool for the generation of customized, Web-based creativity tasks. *Behavior research methods*, 40, 1129-1133. doi:10.3758/brm.40.4.1129
- Ritchie, S. J., Luciano, M., Hansell, N. K., Wright, M. J., & Bates, T. C. (2013). The relationship of reading ability to creativity: Positive, not negative associations. *Learning and Individual Differences*, 26, 171-176. doi:10.1016/j.lindif.2013.02.009
- Rosen, Y., & Tager, M. (2013). *Computer-based Performance Assessment of Creativity Skills: A Pilot Study*. Pearson Research Report. Retrieved from: <http://researchnetwork.pearson.com/wp-content/uploads/creativityassessmentresearchreport.pdf>
- Silvia, P. J., Martin, C., & Nusbaum, E. C. (2009). A snapshot of creativity: Evaluating a quick and simple method for assessing divergent thinking. *Thinking Skills and Creativity*, 4, 79-85. doi:10.1016/j.tsc.2009.06.005
- Silvia, P. J., Winterstein, B. P., Willse, J. T., Barona, C. M., Cram, J. T., Hess, K. I., ... & Richard, C. A. (2008). Assessing creativity with divergent thinking tasks: Exploring the

- reliability and validity of new subjective scoring methods. *Psychology of Aesthetics, Creativity, and the Arts*, 2, 68-85. doi:10.1037/1931-3896.2.2.68
- Tagiuri, R. (1960). Movement as a cue in person perception. In H. P. David & J. C. Brengelmann (Eds.), *Perspectives in personality research (pp. 175-195)*. New York: Springer.
- Tierney, P., Farmer, S. M., & Graen, G. B. (1999). An examination of leadership and employee creativity: the relevance of traits and relationships. *Personnel Psychology*, 52, 591-620. doi:10.1111/j.1744-6570.1999.tb00173.x
- Torrance, E. P. (1968). A longitudinal examination of the fourth grade slump in creativity. *Gifted Child Quarterly*, 12, 195-199. doi:10.1177/001698626801200401
- Torrance, E. P. (1974). *Torrance Tests of Creative Thinking: Norms-technical manual*. Bensenville, IL: Scholastic Testing Service.
- Torrance, E. P. (2008). *The Torrance Tests of Creative Thinking—Norms—Technical manual—Figural (streamlined) forms A and B*. Bensenville, IL: Scholastic Testing Service.
- Urban, K. K. (2005). Assessing Creativity: The Test for Creative Thinking-Drawing Production (TCT-DP). *International Education Journal*, 6, 272-280.
- Valkenburg, P. M., & Beentjes, J. W. (1997). Children's creative imagination in response to radio and television stories. *Journal of Communication*, 47, 21-38. doi:10.1111/j.1460-2466.1997.tb02704.x
- Wallach, M., & Kogan, N. (1965). *Modes of thinking in young children*. New York: Holt, Rinehart, & Winston.

- Walsh, C. S. (2007). Creativity as capital in the literacy classroom: Youth as multimodal designers. *Literacy*, 41, 79-85. doi:10.1111/j.1467-9345.2007.00461.x
- Wang, A. Y. (2012). Exploring the relationship of creative thinking to reading and writing. *Thinking Skills and Creativity*, 7, 38-47. doi:10.1016/j.tsc.2011.09.001
- Weisberg, R. W. (1993). *Creativity: Beyond the myth of genius*. New York, NY: W. H. Freeman.
- Yee, S. L. C. Y. (2015). *Performative Authoring: Nurturing Children's Creativity and Creative Self-Efficacy through Digitally-Augmented Enactment-Based Storytelling* (Doctoral dissertation, Texas A&M University).
- Zabramski, S. (2014). *Creating Digital Traces of Ideas. Evaluation of Computer Input Methods in Creative and Non-Creative Drawing. Storytelling* (Doctoral dissertation, Uppsala: Acta Universitatis Upsaliensis).
- Zabramski, S., Gkouskos, D., & Lind, M. (2011). A comparative evaluation of mouse, pen-and touch-input in computerized version of the Torrance tests of creative thinking. In *Proceedings of the 1st European Workshop on HCI Design and Evaluation* (pp. 57-61). Toulouse, France: IRIT Press.
- Zabramski, S., Ivanova, V., Yang, G., Gadima, N., Leepraphantkul, R. (2013). The Effects of GUI on Users' Creative Performance in Computerized Drawing. In *Proceedings of the International Conference on Multimedia, Interaction, Design and Innovation* (pp. 142-151). New York, NY: ACM Press.
- Zabramski, S., & Neelakannan, S. (2011). Paper equals screen: a comparison of a pen-based figural creativity test in computerized and paper form. In *Proceedings of the Second*

Conference on Creativity and Innovation in Design (pp. 47-50). New York, NY: ACM Press.

Appendix 1: Online Divergent Thinking Test (ODTT)

Instructions for Line Meaning test:

Welcome! On the next page, you are going to play a line game. You will see some lines and figures and after you have looked at each one, please tell us all of the things that the drawing makes you think of.

Here is an example:



For this figure, you could answer, "Mountain; Crack in a glass table; A squished piece of paper; Stock market trend lines; Kid's doodling; Electricity; Soundwave; Analysis chart." There are many more things you can think of, and **all of them are legitimate**.

If you are ready, please turn to the next page. Let's begin now.

Line Meaning test

List all the things that the drawing could be. You can look at it from any direction (or angle) you want.



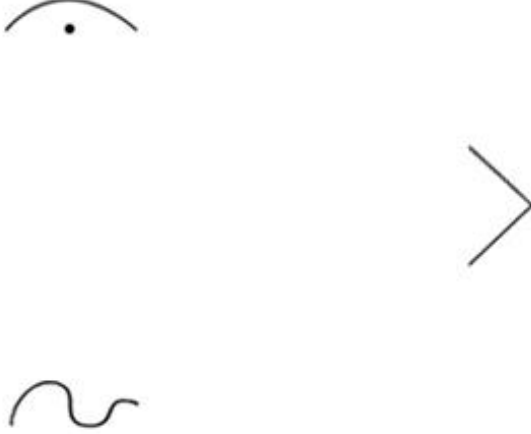




Drawing test 1

Below are the elements you've seen in the previous task. Now, how about combining these figures together to make a story? Try to **draw an interesting story** based on the figures below. And **add a title** to your story.

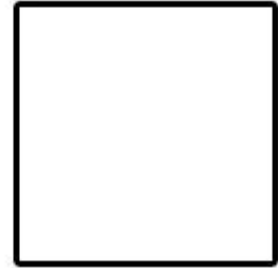
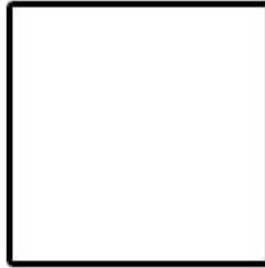
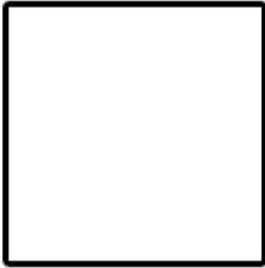
Title: _____



please draw within borders

Drawing test 2

Please add details to the squares below to make recognizable objects. Try to make these figures as **interesting and unusual** as possible. **Add names or titles** if you think the meaning of the object is not clear.



please draw within borders

Instructions for Real-world Problem test:

On the next two pages, we will describe 2 problems which may occur at school and work. Your task is to first read about the problem and then try to write down as many solutions as you can for each problem.

Here is an example:

Your favorite television show, *Game of Thrones*, was on last night. You had so much fun watching it that you forgot to do your homework. You are about to go to school this morning when you realize that your homework is due in your first class. Uh-oh . . . what are you going to do?

For this problem, you could answer, “Tell the professor that you forgot to do your homework; try to do your homework in the car or bus on the way to school; ask your roommate, boyfriend, girlfriend, or classmate to help you finish your homework; do your homework tonight and turn it in the next time the class meets; or finish your homework first than show up late for class.” There are many more answers to this problem, and **all of them are legitimate**.

Now turn to the next page, take your time, have fun.

Real-world Problem test 1

You friend Pat sits next to you in class. Pat really likes to talk to you and often bothers you while you are doing your work. Sometimes Pat distracts you and you miss an important part of the lecture, and many times you don't finish your work because Pat is bothering you. What should you do?

Think of as many ideas as you can. Try to provide ideas that no one else will think of.

This image shows a blank sheet of white paper with horizontal ruling lines. The lines are evenly spaced and extend across the width of the page. There are no margins, text, or other markings on the paper.

Real-world Problem test 2

It is the weekend, and you finally have time to take your dog for a long walk. But 30 minutes from home you remember 3-4 things you need from the store before tonight's dinner. You forgot them once before and are determined not to forget them again. But you have no pen or pencil, nor do you have a mobile phone. What could you do?

Think of as many ideas as you can. Try to provide ideas that no one else will think of.

This image shows a blank sheet of white paper with horizontal ruling lines. The lines are evenly spaced and run across the width of the page. There are no margins, text, or other markings on the paper.

Appendix 2: The Creativity with Technology Self-efficacy Scale (CTSS)

Sub-Scale	Item
General	I am good at coming up with new ideas when working on creative tasks
Creativity Self-efficacy (GCS)	I have a lot of good ideas when working on creative tasks
	I have a good imagination when working on creative tasks
	I am good at coming up with my own creative tasks
	I am good at coming up with new ways of doing creative tasks
General	I feel confident installing and uninstalling large software on my computer, tablet, or other smart devices
Technology Self-efficacy (GTS)	I feel confident learning to use new apps or software
	I feel confident troubleshooting my PCs, tablets or other smart devices when they encounter problems
	I feel confident figuring out the formats of different files
	I feel confident finding the correct programs or apps to open or process files in different formats
	I feel confident switching between different operating systems (for example, Mac, Windows, Android)
	I feel confident using cloud storage services to sync and share my files
Technology Enabled	I feel confident using drawing software or apps to create new drawings or paintings
Creativity Self-efficacy (TECS)	I feel confident making animations
	I feel confident doing web designing
	I feel confident designing 3D objects on computers or tablets
	I feel confident using video editors to do creative post-production after shooting or recording video footage
	I feel confident using (online) digital instruments to create new music (for example, online piano, online drums, online guitar)

Appendix 3: Technology Experience Survey (TES)

(with a 5-point Likert scale ranging from *never* to *always*)

Category	Item
Visually-focused Technologies	play video games
	do graphic design on computer (2D or 3D)
	watch TV
	watch online videos
	edit or share videos
	edit or post pictures
Non-visually-focused Technologies	write blogs
	use word processing software
	exchange messages (in words only) with others
	post your status (in words only) on social networking sites
	listen to (non-music) radios or podcasts online
	read E-books or articles on electronic device or computer
	exchange emails (in words only) with others