The Effects of Test Mode and Contiguity of Material on Geometry Test Scores, Cognitive Load, and Self-Efficacy

Kevin Greenberg[*]   Robert Zheng   Mike Gardner   Lauren Barth-Cohen

Department of Educational Psychology, University of Utah, Salt Lake City, Utah

* Email of corresponding author: kevin.greenberg@utahedu

Abstract

In recent years, the development and usage of computer-based tests for educational assessment has grown. The computer-based tests are typically derived from paper-based tests, with the assumption the tests being administered in different modes are equivalent. Studies examining this test mode effect have mainly focused on the test scores, but few have examined other factors important to test performance. The current study examines the test mode effect for geometry test problems, while also putting in perspective the factors of self-efficacy and cognitive load as both are significant components in performance. The results suggest test scores and cognitive load for geometry problems are similar across the test modes, however learners' self-efficacy significantly decreases when performing the geometry test problems in computer-based test mode. The findings provide insight into the test mode literature and give direction for future lines of research.

Keywords: Computer-based testing, Self-efficacy, Test mode effect, Geometry, Cognitive load

1. Introduction

Assessment, along with content and pedagogy, constitutes the three most important components in the process of teaching and learning (Popham, 2002). One of the issues surrounding assessment is its presentation format as standardized testing is transitioning from paper test mode to computer test mode (Backes & Cowan, 2019). Studies show that the computer based testing has the benefits of reducing human error in grading, quick feedback for test takers, and a smaller ecological footprint (Rausch, Seifried, & Koegler, 2016; Scalise & Gifford, 2006). Despite these benefits, researchers are concerned computer based tests (CBT) that are developed from paper based test (PBT) may change the validity and reliability, resulting in the test mode effect (Öz & Özturan, 2018). In other words, besides the vital role of psychometrics, individual factors, such as self-efficacy, motivation, and cognitive resources, can have an impact on test performance (Chua & Don, 2013). Regardless of the advancement of our knowledge in the test mode and related factors, the relationship of test mode with cognitive and affective states are still understudied. The current study investigates the test mode in mathematics by examining the cognitive and affective aspects in test performance.

1.1 Test mode effect

The test mode effect refers to differences in performance outcomes due to the mode in which the test is administered (CBT or PBT) (Prisacari & Danielson, 2017). Ideally, the mode that a test is administered in should not influence performance outcomes, however test mode research has mixed results for the comparability of test modes on performance scores (Bennett et al., 2008; Chua, 2012b, 2012a; Chua & Don, 2013; Emerson & Mackay, 2011; Johnson & Green, 2006; Nikou & Economides, 2016; Noyes & Garland, 2008; Öz & Özturan, 2018; Prisacari & Danielson, 2017; Puhan, Boughton, & Kim, 2007). Some researchers argue that the test mode does not have any effect on performance outcomes claiming CBT and PBT are equally valid and reliable (Johnson & Green, 2006; Nikou & Economides, 2016; Öz & Özturan, 2018; Prisacari & Danielson, 2017; Puhan et al., 2007). Others found CBT to be less valid and reliable compared to PBT. They attributed the differences in performance to the influence of self-efficacy and motivation (Bennett et al., 2008; Chua, 2012b, 2012a; Chua & Don, 2013; Emerson & Mackay, 2011; Noyes & Garland, 2008). Given the equivocal status in the literature, it is essential to understand the factors that may have exerted the influence on learner performance outcomes, which include self-efficacy, motivation, and the design of test.

1.2 Self-efficacy and test mode effect

Self-efficacy is a belief about one's ability in a certain domain and varies between individuals. (Bandura, 1977, 1994). Research has shown self-efficacy plays a key role in performance outcomes (Bandura, 1977, 1994; J. Lee, 2009; Multon, Brown, & Lent, 1991; van Dinther, Dochy, & Segers, 2011) and interrelates with cognitive processing in multimedia learning (Zheng, McAlack, Wilmes, Kohler-Evans, & Williamson, 2009). Recent studies demonstrate significant correlations between self-efficacy and test mode (Chua, 2012a, 2012b; Chua & Don, 2013). Chua and colleagues (2012a, 2012b, 2013) examined the test mode effect on learners' achievement performance, self-efficacy and motivation by comparing CBT and PBT in biology. Interestingly, their studies

revealed no test mode effect on performance with CBT and PBT modes as measured by achievement scores, but there were differences in test mode in terms of self-efficacy and motivation. They noticed learner's self-efficacy decreased in the PBT mode but increased in the CBT test mode. The researchers concluded that the CBT mode might have made the individuals more motivated in taking the test than did the PBT. They further speculated that the difference may be attributed to the design of the test in terms of element presentation such as temporal contiguity (Mayer, 2001).

1.3 Temporal contiguity and test mode effect
Temporal contiguity refers to presenting meaningfully related information simultaneously as opposed to presenting it separately in succession (Mayer & Moreno, 2002; Moreno & Mayer, 1999; Paas, Renkl, & Sweller, 2003; Schüler, Scheiter, Rummer, & Gerjets, 2012). An example of contiguity design would be showing the heart circulatory image, relevant description and test questions all at once whereas an example of non-contiguity design would present the above information separately. Mayer and colleagues examined the temporal contiguity effect in multimedia learning (Mayer, 2005). They found that when the information was presented simultaneously, the learners performed better on subsequent recall and transfer tests compared to the information that was presented in succession. Ginns (2006) conducted a meta-analysis of 50 studies that explored contiguity effect, finding an overall effect size of 0.85.

Since taking a test requires high level of individual concentration and attention, presenting information that causes least disruption to cognitive processing can be consequential in terms of learners' performance in assessment. Mayer and colleagues (Mayer, Steinhoff, Bower, & Mars, 1995; Moreno & Mayer, 1999) investigated contiguity effect across paper and computer modes of presentation. They found the contiguous materials were effective in both paper and computer presentations. The findings were, however, inconclusive since there was a discrepancy in the media used between the studies, as the paper mode study had static images while the computer-based study used animated material. Such discrepancy in media may allow for possible confounds. Therefore, the current study seeks to understand the relationship between the test mode (i.e., CBT vs. PBT) and the contiguity effect (i.e., contiguous vs. non-contiguous) by keeping the visual materials in identical format, that is, using static images in both conditions. Since previous research has shown the superiority of contiguous design in learning, it is expected learners in a contiguous design condition will perform better than these in a non-contiguous design condition across the test modes as measured by achievement scores.

1.4 Contiguity effect and cognitive load
It is recognized that working memory is limited which can affect learners' abilities to process information and generate relevant mental representations (Baddeley, 1992, 2002; Mayer & Moreno, 2002; Van Merriënboer & Sweller, 2005). Because of the limitation in working memory, one of the concerns is about the extent that the learner's cognitive load may influence the cognitive information processing in learning. Sweller and colleagues (Sweller & Chandler, 1991; Sweller, van Merrienboer, & Paas, 1998) identified three types of cognitive load that may potentially affect learners' performance: intrinsic, extraneous, and germane cognitive load. Intrinsic load refers to the content difficulty which is defined by the element interactivity within the content. According to Sweller and Chandler (1991), the high the level of element interactivity is, the more difficult the content becomes. Extraneous load is caused by improper design of instruction. This type of cognitive load is irrelevant to and considered detrimental to learning. Therefore, it should be eliminated. Germane load relates to the mental effort in applying working memory resources to learning (Paas et al., 2003; Tabbers, Martens, & van Merriënboer, 2004; Zheng et al., 2009).

Based on the definition of cognitive load, learners' performance in association with the design of content (e.g., contiguous vs. non-contiguous) would be affected by extraneous cognitive load. In another words, a change in extraneous cognitive load would be expected dependent on the condition the learner is in. Since previous research has shown that contiguous presentation facilitates effectively creation of mental representations for the incoming information (Mayer & Moreno, 2002), it is predicted that that when the test material is presented with a contiguous design, the learner will be able to generate mental representations with low extraneous cognitive load. That is, low extraneous cognitive load allows for extra cognitive resources to be spent in learning through an increase in germane cognitive load. In contrast, the learner will have a hard time in creating mental representation with a non-contiguous design. This is because the learner has to spend precious working memory resources in coordinating different sources of information, resulting in high extraneous cognitive load and thus reducing the learner's germane cognitive load in learning.

1.5 Contiguity effect in spatially related learning
The contiguity effect is particularly important in domains that require visual-spatial ability, such as geometry. Learning domains that have both visual and spatial elements concurrently, require visual-spatial reasoning (Lowrie & Diezmann, 2007). Consider a situation in geometry learning where the text and geometry diagram are separated

from each other. In such case the learner has to switch his/her attention back and forth between these information elements that can cause an increase in extraneous cognitive overload (Zheng & Gardner, 2019). Therefore, it is especially important in the domain of geometry to have the information elements be contiguous (e.g., presenting the geometric diagram, the text explaining the diagram, and the question in close proximity) so they can be processed simultaneously. In addition, studies have found that individual differences in spatial ability affect the performance outcomes in spatially related learning (Hannafin, Truxaw, Vermillion, & Liu, 2008; Lee, 2007; Lowrie, Ramful, Logan, & Ho, 2014; Lowrie & Diezmann, 2007). Because the current study is focused on geometry which is closely related to learners' visual-spatial abilities, the variable of spatial ability is included to understand its relation with test mode and contiguity effects in performance.

1.6 Research questions
1. Is self-efficacy influenced by test mode and continuity design in geometry test performance?
2. Do test mode and contiguity design influence learners' test performance? Is spatial ability a predicting factor in geometry test performance?
3. Does contiguity design result in lower extraneous load and higher germane load? Is the contiguity design effect consistent across the test modes as measured by intrinsic, extraneous, and germane cognitive load?

2. Method
2.1 Design
The study employed a 2x2 factorial design, where the independent variables were test mode (PBT vs CBT) and contiguity (contiguous vs. non-contiguous). The dependent variables include geometry test scores, cognitive load measures, and changes in self-efficacy. The variable of spatial ability was used as covariate in the final analysis.

2.2 Participants
Participants were recruited from a research I university in western United States. Of the 84 participants, 17 were males and 67 were females with a mean age of 23.4 years old. Sixty-six participants reported as Caucasian, 10 as Hispanic, and 8 as other. The institutional human subject approval was obtained from the Institutional Review Board (IRB).

2.3 Instrumentation
Four instruments were used in the current study. They were geometry problems, Self- and task-perception questionnaire (STPQ), Cognitive Load Questionnaire (CLQ), and Spatial Ability Subtests.
2.3.1 Geometry problems. The geometry problems were adopted from a SAT (Scholarly Aptitude Test) geometry practice problem set ("SAT Math Practice," 2020). The geometry problems included three sections: question, diagram, and multiple choices. Two versions of SAT problems were created: contiguous and noncontiguous. Based on Mayer's (2005; Mayer & Moreno, 2002) temporal contiguity design principle, the contiguous version included the three sections on a single page with a purpose to support effective information process and reduce extraneous cognitive load. The noncontiguous version was designed serving as a control to the experiment version (i.e., contiguous design). It had the three sections in succession, on three separate pages, in which participants were free to move back and forth between the pages.
2.3.2 Self- and task-perception questionnaire (STPQ). Self-efficacy was measured by the STPQ. The measure was originally developed by Lodewyk and Winne (2005) and was found to be a reliable and valid measure of self-efficacy. The study reported a medium to high reliability with Cronbach alpha ranging from 0.72 to 0.92. The instrument contains seven statements on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). The maximum possible score was 35 points and the minimum was 7 points. The self-efficacy measure was administered twice, once as a pre-measure for self-efficacy, and once as a post measure to have a change in self-efficacy score. An example question is, "I believe I will attain a high score on math problem solving."
2.3.3 Cognitive Load Questionnaire (CLQ). Cognitive load was measured using the Cognitive Load Questionnaire developed by Leppink, Paas, Van der Vleuten, Van Gog, and Van Merriënboer (2013). The questionnaire has 10 items on an 11-point Likert scale, from 0 (Not at all the case) to 10 (completely the case). The CLQ measures the 3 aspects of Cognitive Load, Intrinsic, Extraneous, and Germane load. The first three items probe into intrinsic load; an example question is, "The activity covered formulas that I perceived as very complex." Items 4, 5, and 6, pertain to extraneous load. An example of an extraneous load question is, "The instructions and/or explanations during the activity were very unclear." The remaining 4 items measure germane load, with an example question being, "The activity really enhanced my understanding of the content covered." The instrument reported a consistent reliability with Cronbach alpha of .81 for intrinsic load, .75 for extraneous load, and .82 for germane load.
2.3.4 Spatial Ability Subtest. Spatial ability was measured using the subtests in the spatial orientation section in the Kit of Factor Referenced Cognitive Tests (Ekstrom, French, Harman, & Dermen, 1976). The spatial task had

two parts. The first part was card rotation, which had two sections each with 80 mental rotation items. Each section had 10 rows with an irregular shaped card cut out that is the 'key card', followed by 8 similar cards orientated differently. The participant was to determine if the 8 following cards were similar or different than the 'key card.' The second part was the cube comparison, which consisted of 2 sections with 21 pairs of cubes that had to be determined similar or different while only being presented 3 sides of each cube. Each section of the two parts of the spatial orientation tasks were to be completed in 3 minutes. Scores could range from 0 to 202. The spatial orientation tasks were validated from a population of 11th, 12th grade, and college students, with reliability ranging from .77 to .89 for the 4 sections. The spatial ability subtest was centered as advocated by Hunter and Hamilton (2002), transforming the scores into deviation scores (mean = 0).

2.4 Procedure
All participants were informed of the nature of the study and were given informed consent before participating in the study. After consenting, participants were randomly assigned to one of the four conditions: (1) contiguous computer, in which the materials (text, image, multiple-choice options) were presented on the same page and in a computer mode, (2) noncontiguous computer, in which the materials were presented on 3 separate pages and in a computer mode, (3) contiguous paper, the materials were presented on the same page and in a paper mode, and (4) noncontiguous paper, as the materials were presented on 3 separate pages and in the paper mode. They were asked to complete the demographic information question sheet which contained questions on age, gender, interest in math, importance of math in their life and so forth. They then completed the pre self-efficacy measure and were given a basic calculator (capable of performing exponents and roots), the equation sheet provided on the SAT quantitative section, scratch paper, and a writing utensil. Participants were told they had 25-minutes to complete the 10 math problems and if they finished early they could review their work until the 25 minutes had ended. After the math problems, the subjects completed the cognitive load measure, followed by the post self-efficacy measure and the spatial ability sub-tests. The spatial ability test was given last to not overload the participants with complex cognitive tasks prior to the study. Participants received 1 research credit for completing the study.

3. Results
Statistical methods were employed to analyze the final data. An alpha level of .05 was used for all analyses. The means and standard deviations by test mode and the contiguity conditions are shown in Table 1.

Table 1. Means and standard deviations by test mode and contiguity

Mode		Contiguous	Non-Contiguous	Total
Paper	Geometry Score	4.62 (1.94)	4.33 (2.48)	4.47 (2.20)
	Intrinsic Load	17.10 (6.06)	13.29 (6.00)	15.19 (6.26)
	Extraneous Load	8.29 (6.73)	7.29 (4.48)	7.79 (5.67)
	Germane Load	12.76 (6.28)	15.76 (7.54)	14.26 (7.02)
	Change in self-efficacy	-2.67 (5.87)	0.90 (2.45)	-0.88 (4.79)
Computer	Geometry Score	5.24 (1.48)	4.40 (2.16)	4.83 (1.87)
	Intrinsic Load	14.38 (5.55)	15.65 (6.32)	15.00 (5.89)
	Extraneous Load	6.67 (3.98)	11.05 (7.49)	8.80 (6.28)
	Germane Load	13.71 (8.33)	15.34 (8.53)	14.27 (7.95)
	Change in self-efficacy	-4.43 (6.45)	-4.10 (4.32)	-4.27 (5.45)

3.1 Test mode effect on change in self-efficacy
In addressing research question 1, the analysis of variance (ANOVA) was performed. The results indicated a main effect on change in self-efficacy, $F(1, 80) = 8.68$, $p < .01$, $\eta^2 = 0.98$ in test mode (See table 1). It was interesting to observe that self-efficacy decreased in both CBT and PBT modes, with CBT mode showing a significantly larger decrease than PBT mode (See Figure 1). Contiguity was found to have no effect on change in self-efficacy. The results partially confirmed research question 1 with self-efficacy being influenced by text mode but not by contiguity design.
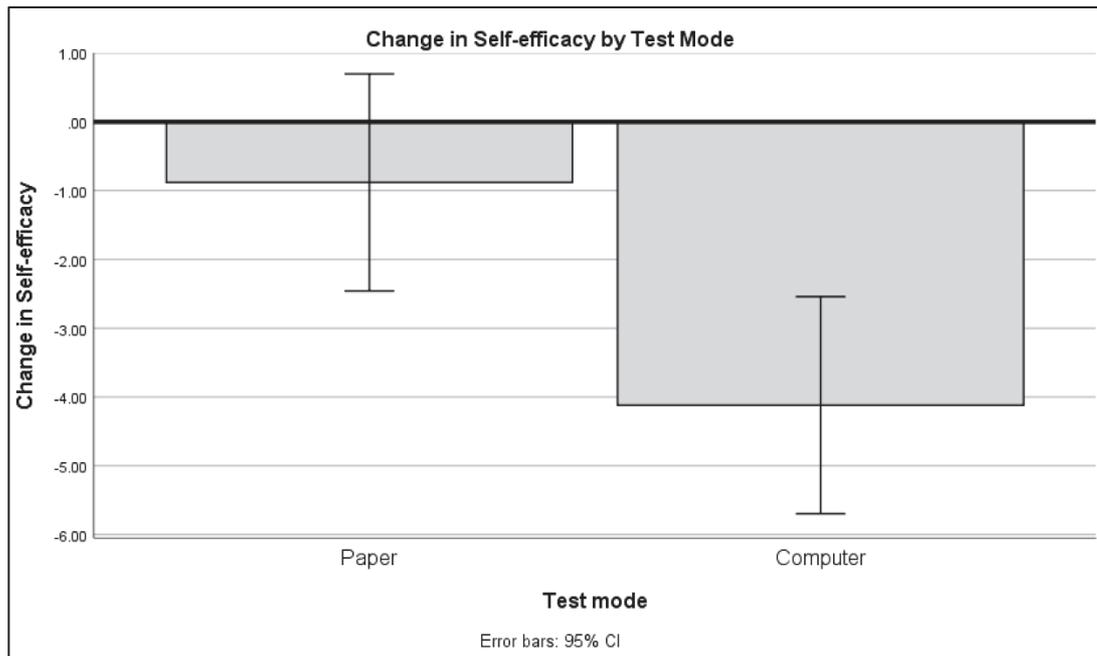
Figure 1. Change in self-efficacy by test mode

3.2 Test mode effect and contiguity effect on geometry test performance

To answer research question 2, the analysis of covariance (ANCOVA) was performed. While controlling for spatial ability ($p > .05$), the results revealed both test mode and contiguity design did not influence geometry performance (See table 1). However, spatial ability was found to be a significant covariate for geometry performance, $F(1, 78) = 10.84$, $p < .001$. A linear regression was conducted showing spatial ability accounted for thirty-four percent of variances in geometry scores ($R^2 = .34$), indicating spatial ability as a predictor for learners' performance in geometry learning (See Figure 2). The ANCOVA analysis partially confirmed research question 2 with both test mode and contiguity design failing to impact geometry performance. However, spatial ability was found to significantly account for learners' performance in geometry learning.
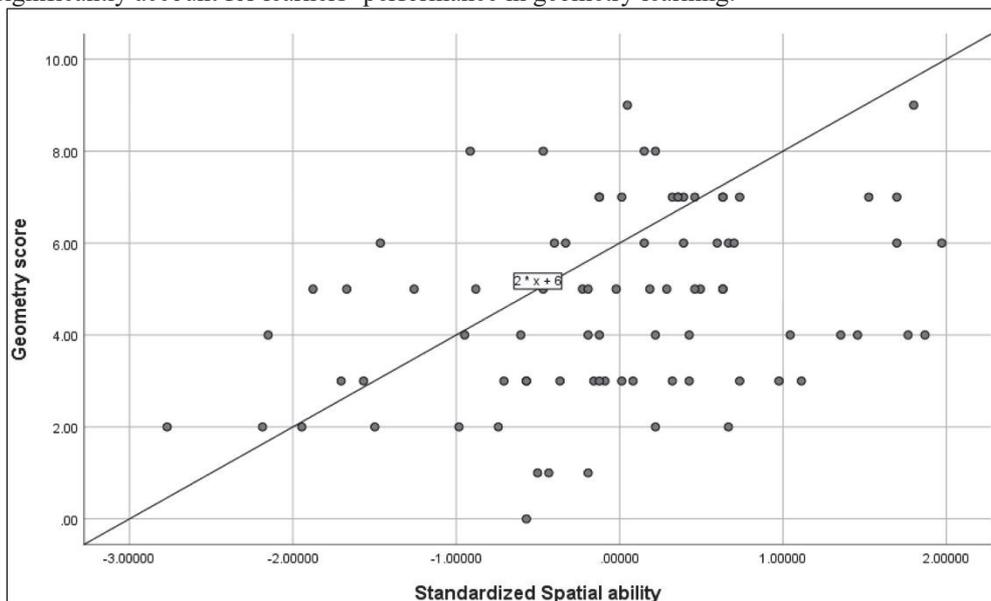

Figure 2. The regression of geometry test score by spatial ability

3.3 Effects of test mode and contiguity design on cognitive load

Research question 3 asked if contiguity design would result in lower extraneous load and higher germane load, and whether contiguity design effect would be consistent across test modes. The results of multivariate analysis of covariance (MANCOVA) revealed no significant main effects for test mode and contiguity design in terms of intrinsic, extraneous, and germane cognitive load, nor were interactions found for test mode and contiguity design

with same dependent measures (p's > .05). Therefore, the findings failed to support research question 3 as none of the cognitive load measures were significant for test mode and contiguity design.

4. Discussion

The goal of the study is to understand the relation of test mode and contiguity design with learner's self-efficacy and performance in geometry. The study has revealed several important findings that are significant to the design of assessment in geometry. The following discussions focus on the significance of findings and their implications in geometry testing.

4.1 Test mode and change in self-efficacy

The data indicates test mode significantly influenced learner's change in self-efficacy. Learners in the CBT mode had a larger decrease in their self-efficacy after completing the computer-based geometry test compared to learners taking the paper-based geometry test. The current findings do not support previous research on CBT and self-efficacy (Chua, 2012a, 2012b; Chua & Don, 2013). The discrepancy between previous findings and the findings from current study could be explained by the differences in domains. In Chua et al.'s studies, the subject domain involved biology where the learners were able to interact directly with the computer based material which may account for the difference in learners' self-efficacy between Chua et al.'s research and the current study: the current study focused on computational skills in geometry where written manipulation is typically required (81 of 84 participants used scratch paper). In the current study, the learners in the CBT mode had to transfer the geometry material from the computer to physical scratch paper when calculating the outcomes. Evidently, more effort was required on learners in the CBT mode compared to learners in the PBT mode who could simply do all the work with paper and pencil at once. This extra step of transferring work from computer to paper requires extra mental effort that can be a significant factor in influencing learner's information processing in computer-based learning and may cause decreases in self-efficacy as learners who put forth more effort may perceive themselves as having lower capabilities. In sum, the results suggest learner performance in relation to test mode and self-efficacy may be mediated by the subject domain. More research is warranted to understand the relationship between subject domain, test mode, and self-efficacy.

4.2 Test mode effect and contiguity effect on geometry performance

The findings from the study suggest the test mode and the contiguity of material did not have an effect on geometry performance. The results are consistent with the findings in previous research where computer and paper mode yielded no significant differences in performance scores (Johnson & Green, 2006; Nikou & Economides, 2016; Öz & Özturan, 2018; Prisacari & Danielson, 2017; Puhan et al., 2007).

However, the current study generated important results pertaining to learner's spatial ability and geometry test performance. It demonstrated the role of spatial ability in visual-spatially related learning like geometry. This finding aligns well with the literature on spatial ability and visual learning where learners with higher spatial ability performed better on geometry test (Hannafin, Truxaw, Vermillion, & Liu, 2008; H. Lee, 2007; Lowrie, Ramful, Logan, & Ho, 2014; Lowrie & Diezmann, 2007). Evidently, individual differences in spatial ability need to be considered in future research when studying visual-spatially related subject like geometry.

4.3 Test mode effect and contiguity effect on cognitive load

Contrary to our prediction, the test mode and the contiguity design did not have an influence on any of the three types of cognitive load. One of the possible explanations may relate to learners' prior knowledge. According to Kalyuga (2014), the design of instruction that supports novices' learning may induce higher cognitive load than on expert learners, and vice versa. Regarding the contiguity design, it was predicted that contiguous material would result in lower extraneous load and higher germane load compared to non-contiguous material. It is possible that while the contiguity design may be of a concern for low prior knowledge learners, it has less impact on expert learners since they have already had the schemas to deal with the problems, thus making the design less of an issue in learning. This may account for the non-significance of extraneous cognitive load across the test mode and contiguity design. Another potential explanation to the non-significance of three types of cognitive load is the self-reported questionnaire may be not well understood by the participants, thus affecting the outcome of load measurement. Sweller (2018) argues that learners may have trouble with self-report cognitive load questionnaires as these questionnaires require the learner to have an understanding of cognitive load types and being able to identify what makes a task difficult. More research is needed to understand the learners' approach to the cognitive load measurement, particularly in the context of contiguity design and test mode.

5. Conclusion

The current study explores the relationship between test mode and contiguity design in geometry test performance by focusing on self-efficacy, spatial ability, and cognitive load. The study reveals that learner self-efficacy

decreased in both CBT and PBT with CBT showing a larger decrease than PBT. It was found that spatial ability played a positive role in geometry problem solving in testing, suggesting the design of geometry instruction should take spatial ability into consideration. Contrary to previous research, the current study found the test mode and contiguity design had no effect on learners' cognitive load in geometry problem solving. The findings of this study are important to the field in that they contribute to the understanding of test mode and contiguity design in geometry research. It has raised the awareness of the roles of self-efficacy and spatial ability in the design of geometry learning.

Future research is needed to understand the effects of test mode and contiguity design on geometry problem solving. Improvements can be made in CBT to avoid the unnecessary switch in computation process. More studies are warranted to investigate the roles of prior knowledge and spatial ability pertaining to test mode and contiguity design in geometry testing, as well as the extent they influence learners' cognitive load and performance. In addition, more work is needed to investigate the test mode effect in other domains, such as language arts, science, social studies, etc., as differences in test mode are likely to influence learners' cognitive and performance outcomes in these areas as well. Lastly, future research must look into the effect of test mode in equitable testing practices, particularly by examining the relationship between test methods/procedures and learners' prior knowledge, cultural experience and cognitive style.

References

Backes, B., & Cowan, J. (2019). Is the pen mightier than the keyboard? The effect of online testing on measured student achievement. *Economics of Education Review*, *68*, 89–103. https://doi.org/10.1016/j.econedurev.2018.12.007

Baddeley, A. (1992). Working Memory. *Science*, *255*(5044), 556–559. https://doi.org/10.1016/S0079-7421(08)60452-1

Baddeley, A. (2002). Is Working Memory Still Working? *European Psychologist*, *7*(2), 85–97. https://doi.org/10.1027//1016-9040.7.2.85

Bandura, A. (1977). Self-efficacy : Toward a Unifying Theory of Behavioral Change. *Psychological Review*, *84*(2), 191–215.

Bandura, A. (1994). Self-Efficacy. In V. S. Ramachaudran (Ed.), *Encyclopedia of human behavior* (pp. 71–81). New York: Academic Press.

Bennett, R. E., Braswell, J., Oranje, A., Sandene, B., Kaplan, B., Yan, F., … Yan, F. (2008). Does it Matter if I Take My Mathematics Test on Computer ? A Second Empirical Study of Mode Effects in NAEP. *The Journal of Technology, Learning, and Assessment*, *6*(9), 4–38.

Chua, Y. P. (2012a). Effects of computer-based testing on test performance and testing motivation. *Computers in Human Behavior*, *28*(5), 1580–1586. https://doi.org/10.1016/j.chb.2012.03.020

Chua, Y. P. (2012b). Replacing Paper-based Testing with Computer-based Testing in Assessment: Are we Doing Wrong? *Procedia - Social and Behavioral Sciences*, *64*, 655–664. https://doi.org/10.1016/j.sbspro.2012.11.077

Chua, Y. P., & Don, Z. M. (2013). Effects of computer-based educational achievement test on test performance and test takers' motivation. *Computers in Human Behavior*, *29*(5), 1889–1895. https://doi.org/10.1016/j.chb.2013.03.008

Emerson, L., & Mackay, B. (2011). A comparison between paper-based and online learning in higher education. *British Journal of Educational Technology*, *42*(5), 727–735. https://doi.org/10.1111/j.1467-8535.2010.01081.x

Hannafin, R. D., Truxaw, M. P., Vermillion, J. R., & Liu, Y. (2008). Effects of spatial ability and instructional program on geometry achievement. *Journal of Educational Research*, *101*(3), 148–157. https://doi.org/10.3200/JOER.101.3.148-157

Johnson, M., & Green, S. (2006). On-Line Mathematics Assessment : The Impact of Mode on Performance and Question Answering Strategies. *The Journal of Technology, Learning, and Assessment*, *4*(5), 4–34.

Lee, H. (2007). Instructional Design of Web-based Simulations for Learners with Different Levels of Spatial Ability. *Instructional Science*, *35*(6), 467–479. https://doi.org/10.1007/s

Lee, J. (2009). Universals and specifics of math self-concept, math self-efficacy, and math anxiety across 41 PISA 2003 participating countries. *Learning and Individual Differences*, *19*(3), 355–365. https://doi.org/10.1016/j.lindif.2008.10.009

Lowrie, Thomas, Ramful, A., Logan, T., & Ho, S. Y. (2014). Do students solve graphic tasks with spatial demands differently in digital form? In *Proceedings of the 37th annual conference of the Mathematics Education Research Group of Australasia* (pp. 429–436). Retrieved from https://researchprofiles.canberra.edu.au/en/publications/do-students-solve-graphic-tasks-with-spatial-demands-differently-

Lowrie, Tom, & Diezmann, C. M. (2007). Solving Graphics Problems: Student Performance in Junior Grades.

Journal of Educational Research, *100*(6), 369–377.

Mayer, R. E. (2005). Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 183–200). New York.

Mayer, R. E., & Moreno, R. (2002). Aids to computer-based multimedia learning. *Learning and Instruction*, *12*(1), 107–119. https://doi.org/10.1016/S0959-4752(01)00018-4

Mayer, R. E., Steinhoff, K., Bower, G., & Mars, R. (1995). A generative theory of textbook design: Using annotated illustrations to foster meaningful learning of science text. *Educational Technology Research and Development*, *43*(1), 31–41.

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology*, *91*(2), 358–368.

Multon, K. D., Brown, S. D., & Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. *Journal of Counseling Psychology*, *38*(1), 30–38. https://doi.org/10.1037/0022-0167.38.1.30

Nikou, S. A., & Economides, A. A. (2016). The impact of paper-based, computer-based and mobile-based self-assessment on students' science motivation and achievement. *Computers in Human Behavior*, *55*, 1241–1248. https://doi.org/10.1016/j.chb.2015.09.025

Noyes, J. M., & Garland, K. J. (2008). Computer- vs . paper-based tasks: Are they equivalent ? *Ergonomics*, *51*(9), 1352–1375. https://doi.org/10.1080/00140130802170387

Öz, H., & Özturan, T. (2018). Computer-based and paper-based testing : Does the test administration mode influence the reliability and validity of achievement tests? *Journal of Language and Linguistic Studies*, *14*(1), 67–85.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive Load Theory and Instructional Design: Recent Developments. *Educational Psychologist*, *38*(1), 1–4. https://doi.org/10.1207/S15326985EP3801

Prisacari, A. A., & Danielson, J. (2017). Rethinking testing mode: Should I offer my next chemistry test on paper or computer? *Computers and Education*, *106*, 1–12. https://doi.org/10.1016/j.compedu.2016.11.008

Puhan, G., Boughton, K., & Kim, S. (2007). Examining Differences in Examinee Performance in Paper and Pencil and Computerized Testing. *The Journal of Technology, Learning, and Assessment*, *6*(3), 4–20.

SAT Math Practice. (2020). College Board. Retrieved from https://collegereadiness.collegeboard.org/pdf/sat-practice-test-1.pdf

Schüler, A., Scheiter, K., Rummer, R., & Gerjets, P. (2012). Explaining the modality effect in multimedia learning: Is it due to a lack of temporal contiguity with written text and pictures? *Learning and Instruction*, *22*(2), 92–102. https://doi.org/10.1016/j.learninstruc.2011.08.001

Tabbers, H., Martens, R., & van Merriënboer, J. (2004). Multimedia instructions and Cognitive Load Theory : *British Journal of Educational Psychology*, *74*(1), 71–81. Retrieved from www.bps.org.uk

van Dinther, M., Dochy, F., & Segers, M. (2011). Factors affecting students ' self-efficacy in higher education. *Educational Research Review*, (6), 95–108. https://doi.org/10.1016/j.edurev.2010.10.003

Van Merriënboer, J. J. G., & Sweller, J. (2005). *Cognitive load theory and complex learning: Recent developments and future directions*. *Educational Psychology Review* (Vol. 17). https://doi.org/10.1007/s10648-005-3951-0

Zheng, R., McAlack, M., Wilmes, B., Kohler-Evans, P., & Williamson, J. (2009). Effects of multimedia on cognitive load, self-efficacy, and multiple rule-based problem solving. *British Journal of Educational Technology*, *40*(5), 790–803. https://doi.org/10.1111/j.1467-8535.2008.00859.x

1. **Prepare your paper according to IISTE format**
   You can download the standard IISTE format here and prepare your paper first:
   http://www.iiste.org/sub/Final_Template_Sample.doc

2. **The authors must note that submitting an article to an IISTE journals means that**
   a. authors confirm that the articles are original;
   b. there's no conflicts of interests or any plagiarism issues;
   c. the article is not being considered elsewhere;
   d. the authors are willing to transfer the first publishing right to the publisher if the article is accepted after review.