MOJES MALAYSIAN ONLINE JOURNAL OF EDUCATIONAL SCIENCES JANUARY 2020, 8 (1)

EFFECTS OF DIFFERENT LEARNING ENVIRONMENTS ON PRE-SERVICE TEACHERS' JUSTIFICATION

Ali BOZKURT
Gaziantep University
alibzkrt@gmail.com

Yusuf KOÇ
Kocaeli University
kocyusuf@gmail.com

ABSTRACT

Justification skills should be fostered in teaching mathematics as they are highly related to effectiveness in problem solving. However, pre-service mathematics teachers do not always have the necessary skills in using mathematically convincing language and justifying their solutions. Thus, they should be introduced to a variety of instructional materials and tasks that may contribute to their justification skills. In this study, researchers investigated effect of different learning environments, including dynamic geometry environments and physical instructional materials, on 139 first year pre-service mathematics teachers' level of justifications. The participants were formed into two treatment groups: the physical instructional materials environment group and the dynamic geometry environment group. In this non-equivalent experimental group design study, the data was collected via pre- and post-tests. No statistically significant difference was found between the treatment groups. The number of mathematically convincing arguments increased at the end of the treatments. The study shows that the use of inquiry-based activities in dynamic geometry and physical instructional material environments might have helped learners improve their level of justifications in geometry.

Keywords: *Dynamic Geometry Environment, Physical Instructional Materials, Justification, Geometry Teaching*

INTRODUCTION

Justification is an essential skill that should be fostered in mathematics classrooms because it is highly associated with learners' effectiveness in problem solving (Driscoll, Wing DiMatteo, Nikula, & Egan, 2007). Students' mathematical justifications have been investigated from different perspectives, including practically-based justifications and mathematically-based justifications (Cai, 2000), proof and explanation (Hanna, 2000), conceptual and procedural explanations (Charalambous, Hill, & Ball, 2011; Kinach, 2002), and convincing mathematical explanations (Cai, 2003; Driscoll et al., 2007). The quality of students' justifications indicates the level of their geometric and algebraic thinking skills (Cai, 2000). In his investigation of Singaporean students' justifications, Cai (2003) classified student justifications into four categories: *complete and convincing arguments*, *vague or incomplete argument*, *incorrect of incomprehensible arguments*, and *no arguments*. This classification scheme indicates that the primary goal of promoting student justifications in mathematics classrooms should be to help all students provide complete and convincing arguments. When learners explain how they solve the problem and justify their solutions, their geometric thinking skills will most likely improve (Driscoll et al., 2007). Additionally, concept formation is enhanced when students are asked to explain how they solve the problem and to justify their solutions by utilizing mathematically convincing arguments in mathematics classrooms (Cai, 2003; Driscoll et al., 2007). Hence, teachers should facilitate the learning environment in a way to promote students' mathematically convincing arguments.

Previous studies on mathematical justifications are closely related to mathematical reasoning, estimation, generalizing, providing justifications, mathematical language and geometric thinking (Christou et al., 2004; Richardson, Carter, & Berenson, 2010; Sandborg, 1998). Those studies reported that a good justification is essentially important in mathematics and geometry; but it was also noted that students, pre-service teachers and even teachers cannot adequately explain and justify their solutions of mathematics problems. In a previous study on explanations (Chick, 2003), it was found out that pre-service teachers are having difficulties in using mathematically convincing language and justifying their solutions. It is recommended that in order to eliminate such difficulties students should be exposed to a variety of instructional materials and tasks that may contribute to their learning. Instructional activities and materials should be designed in a way to promote students' and pre-service teachers' justifications (Hoong & Khoh, 2003). Similar findings on justification skills and use of technology and physical manipulatives have been reported by other researchers. (Arıcı & Aslan-Tutak 2015; Carbonneau, Marley, & Selig, 2013; Pandiscio, 2002).

Effect of dynamic geometry and physical instructional materials environments

The review of the literature indicates that there are several studies that have investigated the effectiveness of physical instructional materials and dynamic geometry learning environments on student learning (Christou et al., 2004; Empson & Turner, 2006; Olkun, 2003; Arıcı & Aslan-Tutak, 2015). The review also shows that in a number of studies, students' explanations and justifications for solutions of geometry problems were the focus of interest (Hanna, 2000). Examining the effectiveness of different instructional environments on students' explanations and justifications is essential to develop effective methods to enhance justification skills; yet, the number of such studies is quite rare (McNeill, Lizotte, Krajcik, & Marx, 2006; Kelly, 2006). Investigating effectiveness of dynamic geometry software or physical instructional materials on the use of justifications is highly critical to improve our understanding of student justification skills.

Among other instructional media, dynamic geometry environments have found to be effective in enhancing students' achievement, justification skills (Erbas & Yenmez, 2011; Hadas, Hershkowitz, & Schwartz, 2000; Stols, 2012). In a recent study, Guven and Karatas (2009) found out that use of dynamic geometric software improves students' estimation, generalization, and mathematical explanations. Another piece of the literature shows that the use of manipulatives and other physical materials are helpful in helping students improve their explanations in mathematics (Hanna, 2000; Kelly, 2006; Zacharia & Costantinou, 2008). Along with these findings, in the present study, learning environments were modified to enhance first year mathematics teacher education students' explanations and justifications in geometry problem solving. In particular, the researchers investigated the effects of inquiry-based geometry instruction in dynamic geometry learning and physical instructional materials environment on participants' level of justifications.

Given the above literature support and rationale, the purpose of the present study is to investigate effect of the use of Dynamic Geometry Environment (DGE) and Physical Instructional Materials Environment (PIME) on pre-service teachers' levels of justifications. The following research questions guided the study:

1) Does the quality of the justifications of the pre-service teachers increase when they are exposed to inquiry-based geometry activities?
2) How do the level of justifications of the pre-service teachers taught in the dynamic geometry environment (DGE) and the ones taught in the physical instructional materials environment (PIME) change from the pre-test to the post-test?
3) How do the frequencies of the complete and convincing arguments category for each test item change between and within DGE and PIME groups from the pre-test to the post-test?

METHOD

In this study, a non-equivalent experimental group design was employed because researchers worked with two intact groups, which means the participants were not assigned to the treatment groups randomly. Although researchers selected the treatment groups which were as similar as possible, but we can never be sure that the groups are equal. In order to better control of the subject characteristics threat, pretest-posttest design was employed. This design provides an effective control over the subject characteristics threat (Fraenkel, Wallen, & Hyun, 2015).

Participants and Background

First-year pre-service teachers (N=139) who were enrolled in an introduction level geometry course at a public university in Turkey participated in the study for ten weeks. As seen in Table 1, while 67 participants were in the DGE group, 72 participants were in the PIME group. There were 49 females and 18 males in the DGE group, and 53 female and 19 male participants in the PIME group.

Table 1
Distribution of participants by treatment groups and gender

	DGE	PIME	Total
Female	49	53	102
Male	18	19	37
Total	67	72	139

A national mathematics teacher education curriculum is implemented throughout Turkey. All teacher education students are required to take the same courses, with the exception of some electives, at the same time. In the second semester of the first year, among other courses, they are required to take an introductory level geometry course, which is mainly geared toward understanding of fundamental geometry concepts, including two and three dimensional shapes and their properties and geometric problem solving. In the present study, researchers were the instructors of this geometry course and the participants were the students who were enrolled in the course. There was a teacher-student relation between both parties; but the course grades were given independently from their project participation. Additionally, all participants were informed about the study.

Procedures

In this non-equivalent group design, the inquiry-based geometry activities provided with the Fostering Geometric Thinking Toolkit (Driscoll et al., 2008) were implemented in both treatment groups. The main purpose of the instruction in both groups was to foster participants' geometric thinking, including their explanations and justification skills. Although the treatment groups received similar instructions, but each group used different instructional materials. So, the treatments differ by the nature of the instructional materials. In the first treatment group, the inquiry-based activities were introduced with the assistance of dynamic geometry software, Geometer's Sketchpad®. In the second treatment group, the same inquiry-based activities were introduced with the assistance of physical instructional materials. The treatments were implemented concurrently during the same academic semester and two mathematics educators taught the treatment groups. In the PIME group, the instructor met with the students in a regular classroom where physical instructional materials such as mathematics manipulatives, patty paper, and scissors were used to solve the geometry problems. In the DGE group, the instructor met with the students in a computer lab where they were working on the computer to solve the problems. The instructors met regularly and worked together to design and deliver the course activities throughout the semester. They discussed how to implement the activities and agreed on a way instruction so that their instructions were comparable. Additionally, they followed the same textbook (Driscoll et al., 2007), mathematics notes and guidebook (Driscoll et al., 2008) to control the effects of teaching.

The treatments were separately conducted in the two treatment groups for ten weeks, three hours a week. During the activities, students in both groups were asked to first individually and then collaboratively work through and solve the given geometry problems, and provide detailed explanations and justifications regarding their solutions. They were reminded that they needed to use the language of mathematics appropriately and provide convincing mathematical explanations (Driscoll et al., 2007; Driscoll et al., 2008). During the instructions, while the instructors were facilitating classroom discussions, they encouraged students to solve the problems in multiple ways, provide their justifications and use convincing mathematical justifications. The instructors interacted with small groups, encouraged within and between small group interactions, and supported their understanding of the problems. After working through the problems, students shared their solution methods and justifications with the whole class.

Treatment Group 1: Physical instructional materials environment group

The students who were in the Physical Instructional Materials Environment (PIME) group completed a set of inquiry-based geometry activities (Driscoll et al., 2008). They used a variety of physical instructional materials or tools, including regular paper, patty paper, ruler, scissors and tangram pieces to understand and solve the problems introduced in the geometry activities. The participants used the regular and patty papers to fold, draw or cut to obtain various geometric shapes asked in the problems. Tangram pieces were used to form a square, rectangle, triangle and other geometric shapes. The instructor provided all necessary materials and the activity sheet at the beginning of each session.

Treatment Group 2: Dynamic geometry environment group

The DGE participants worked through similar inquiry-based geometry activities used in the PIME group. The directions of the activities were revised for appropriate use of the Geometer's Sketchpad Program (GSP). For example, in one of the activities used in the PIME group, students were asked to use straightedge, paper and pencil to construct a line segment that is perpendicular to a given segment; whereas, in the DGE group, for constructing a perpendicular line segment, the students were asked to use the circle, segment and point functions of the GSP. In all other inquiry-based activities, similar modifications were done. In addition, in the DGE group, the GSP applets provided with the Fostering Geometric Thinking Toolkit (Driscoll et al., 2008) were used.

A sample activity

During the treatments, the participants worked through several geometry activities. An activity involving working with tangram pieces to solve geometry problems was one of them. In order to illustrate how this activity was implemented with modifications for the PIME and DGE groups is presented below.
In the PIME group, first, the participants were introduced to a tangram puzzle. They were asked to make a non-square rectangle by using all seven tangram pieces and write a convincing mathematical statement to explain that the figure has the properties of a rectangle. Then, they made a square out of all the tangram pieces and measure the area of the square with the tangram pieces. In each case, they were asked to describe how they found their answer. During the activity, first the participants worked individually and then they shared and discussed their solution with their peers and the whole class. The instructor regularly monitored student interactions and guided their discussions to facilitate their thinking. During whole group discussions, students presented their solutions and their approaches to the problem. Additionally, they shared the geometric habits of mind (Driscoll et al., 2007) that they used while working through the task. They were encouraged to justify their thoughts with convincing statements.

In the DGE group, the participants worked through the very same problems, including forming a non-square rectangle and a square via a Geometer's Sketchpad tangram puzzle available for download (http://sketchexchange.keypress.com/sketch/view/117/tangram-puzzles). The participants were already familiar with the Sketchpad, so they could easily move the tangram pieces to form the shapes.

Like in the PIME group, they made the shapes first individually and shared their solutions with others. The instructor walked around the computer stations to monitor student progress and facilitated their thinking via guiding questions. At the end, there was a whole group discussion very much like the one in the PIME group. As a result, the participants in both treatment groups worked through the same activity with different media; while in the PIME they used a regular tangram set with seven pieces, in the DGE group the participants used a virtual tangram puzzle. The instructors assumed very similar roles as they decided together.

Data collection instrument

In the present study, a Geometry Achievement Test (GAT) was administered as the pre-test to all participants to determine their level of justifications at the beginning. The same test was used as the post-test at the end of the treatment period; only the order of the questions was changed to eliminate the remembrance factor. Two mathematics educators with extensive experience in mathematics education research have compiled the test items that require students to engage in geometric reasoning, and explain and justify their solutions. There were five open-ended geometry items in the GAT. The items were adapted from various resources (Driscoll et al., 2007; Kaplan, 2011) and one item was adapted from a Japanese TIMMS Video (http://www.timssvideo.com). In the first item, the students needed to investigate the relationship between similarity, area, and height of a triangle; thus, they were asked to engage in reasoning with relationships. In the second item, after several attempts of conducting symmetrical transformation, the students were asked to determine that on a plane, the collection of infinitely many points which are equidistant from a centre form a circle. The third item was about analytic geometry. The participants were expected to generalize geometric ideas on coordinate plane. The fourth item required the participants to form a square out of four non-isosceles triangles and provide sound explanations and justifications. Finally, the fifth item demanded participants to show their understanding of area of triangles with the same base and height between two parallel lines. In all items, the participants were required to justify their solutions.

Four of the items, except item 1 which was already in Turkish, were translated from English into Turkish by two mathematics education researchers. Both researchers are competent enough in using the English language. One of them received his doctoral degree from a research university in the US and taught in the US about four years as a teacher educator; thus, he had the necessary knowledge of English and mathematics to perform the translations. The second researcher received his doctoral degree in mathematics from a large public university in Turkey. He taught mathematics in middle and high schools about eight years. He has the necessary skills and experience in translating mathematics problems from English into Turkish. The two researchers translated the items together. It was confirmed that the items were clear and understandable, and they were appropriate for Turkish middle school classrooms and university students. For construct validity, two experts in mathematics education research and a mathematics teacher reviewed all items to check whether the items had the potential to elicit student thinking and promote explanations and justifications. They conducted a group meeting to discuss the test items and reached the final form of the test prior to the pilot study.

After compiling the test items, as the pilot study, the researchers administered the GAT to 44 third-year mathematics teacher education students. For content validity, the course instructors who happened to be the researchers reviewed the test items if the course content was adequately represented by the GAT. Mainly, the items that triggered and required student justifications were included into the test. The pilot study indicated that items were clear to the students. It was also observed in the pilot study that the test took about one and a half hours to complete. The pilot study also showed that more space was needed for each question.

Data analysis

The purpose of data analysis was two-fold: to examine the level of student justifications and also to investigate the effect of inquiry mathematics instruction in dynamic geometry environment (DGE) and physical instructional materials (PIME) environments on the level of student justifications. In order to determine the level of justifications, the categorization scheme by Cai (2003) was used (Table 2).

Table 2
Descriptions of levels of justifications

Level of justification	Description
Complete and convincing argument	Correct response with complete and convincing arguments
Vague or incomplete argument	Correct or partly correct response with vague or incomplete arguments
Incorrect or incomprehensible argument	Incorrect response with computational or conceptual errors, and incomprehensible and irrelevant arguments.
No argument	No response or some incomplete visual representations; but no written arguments.

Initially, all responses were reviewed by the researchers; then, for scoring-reliability, two graduate students who were experienced in data coding and mathematics teaching worked together to code the entire data set. Two mathematics educators monitored the graduate students' data coding. The coders first individually coded twenty randomly selected student responses, 10 from the pre-test and another 10 from the post-test, 20 in total. Initially, they reviewed and coded the responses by level of justification via the coding framework provided in Table 2. Next, they recorded the level of justification for each of the five items in GAT. In this process, the researchers took complete responses as the unit of analysis. Individual responses were examined in their entirety because the whole response contained a single justification. The coders compared the coded units to determine the percentage of the units matched across each coder (Robson, 1993). They initially reached a level of agreement below 80%. In this initial round of coding, the coders and mathematics educators jointly discussed where the disagreements were and how they could come to a common understanding of each code. Then, the two coders randomly picked another set of twenty student papers and independently coded them. They reached an agreement of about 90% on their coding decisions. After having another meeting, they coded the rest of the papers (See Appendix A). In Appendix A sample coding of the second item These sample coding illustrates participants' mathematical arguments and how this qualitative data was analysed via utilizing Cai's coding scheme (2003).

Next, frequency analyses were conducted to determine the number of times each level of justification was assigned. Chi-square non-parametric tests and t-test were conducted to assess whether there were significant differences between the pre- and post-tests, and between the treatment groups. In order to obtain a holistic picture of participants' level of justification, a numerical value was assigned to each level of justification: *complete and convincing argument* (5 points), *vague or incomplete argument* (3 points), *incorrect or incomprehensible argument* (1 point) and *no argument* (0 point). If a participant argument did not provide an argument, no point was assigned to that participant. If an incorrect argument was provided, only 1 point was assigned. Vague or incomplete arguments received a point of three and complete arguments received five points. The intervals between the scoring categories are not consistent because researchers observed that the difference between a *vague or incomplete argument* and *complete and convincing argument* is more than one point.

FINDINGS

A paired-samples t-test was conducted to evaluate whether level of justifications increased at the end of the treatments.

Whole group analysis

Each participant was assigned a justification score out of 25. Although the participants were formed into two treatment groups, for this initial data analysis all participants were considered as a whole group.

The results indicated that the mean justification scores for the post-test ($\overline{X}$ = 15.46, sd = 4.10) was significantly greater than the mean justification scores for the pre-test ($\overline{X}$ = 12.10, sd = 3.43), t(139) = -10.431, p < .01). Thus, the participants' level of justification significantly improved at the end of the study.

Treatment group differences

In the present study, the participants were formed into two treatment groups, DGE and PIME. Table 3 presents the mean scores and standard deviations for the pre and post-test by these two groups.

Table 3
Mean scores and standard deviations for the pre-test and post-tests

	Pre-test			Post-test		
	N	$\overline{X}$	sd	N	$\overline{X}$	sd
PIME	72	12.00	3.44	72	15.88	.26
DGE	67	12.21	3.45	67	15.01	.90

While the mean score of the PIME group was 12.00 in the pre-test, the mean score was 12.21 for the DGE group. In the post-test, the group averages were also close (15.88 for PIME, 15.01 for DGE). Independent-samples t tests were conducted to evaluate whether the mean differences between the treatment groups were statistically significant. The t test results based on the pre-test data indicated that there was not any significant difference between the DGE and PIME groups regarding the level of justifications (t (137) = .357, p = .751). Thus, the treatment groups' mean scores were statistically equal at the beginning of the research.

Since no significant differences were observed in the pre-test data, another independent- samples t-test was conducted to determine whether any mean differences between the treatment groups exist in the post-test data. Such analysis would be enough to determine whether there was any statistically significant difference between the treatment groups regarding the participants' level of justifications. After running the independent-samples t-test, it was observed that there was not any significant difference between the treatment groups (t (137) = -1.237, p = .218). Therefore, it was found out that in the present study both teaching methods had statistically equal effect on the participants' level of justification.

Analyses by level of justifications for each item

The analysis of student responses to test items indicates that there were responses at all four levels of justification. Figure 1 shows the distribution of number of responses by the levels of justification both in the pre-test and post-test for the whole group. It should be noted that each response was coded under just one level of justification.

Figure 1 indicates that the participants provided more arguments in the post-test than in the pre-test and also frequency of *vague or incomplete* and *complete and convincing argument* categories increased at the end of the treatment. Overall, there is positive trend in regards to the quality of the arguments. For Item 1, while the most frequent category in the pre-test was *vague or incomplete argument* (102 responses), the *complete and convincing argument* became the most frequent category in the post-test (66 responses) and there were also 62 responses under the *vague or incomplete argument* category in the post-test. Regarding the *incorrect or incomprehensible argument* and *no argument* categories, there were only a few responses in the pre-test and post-test.

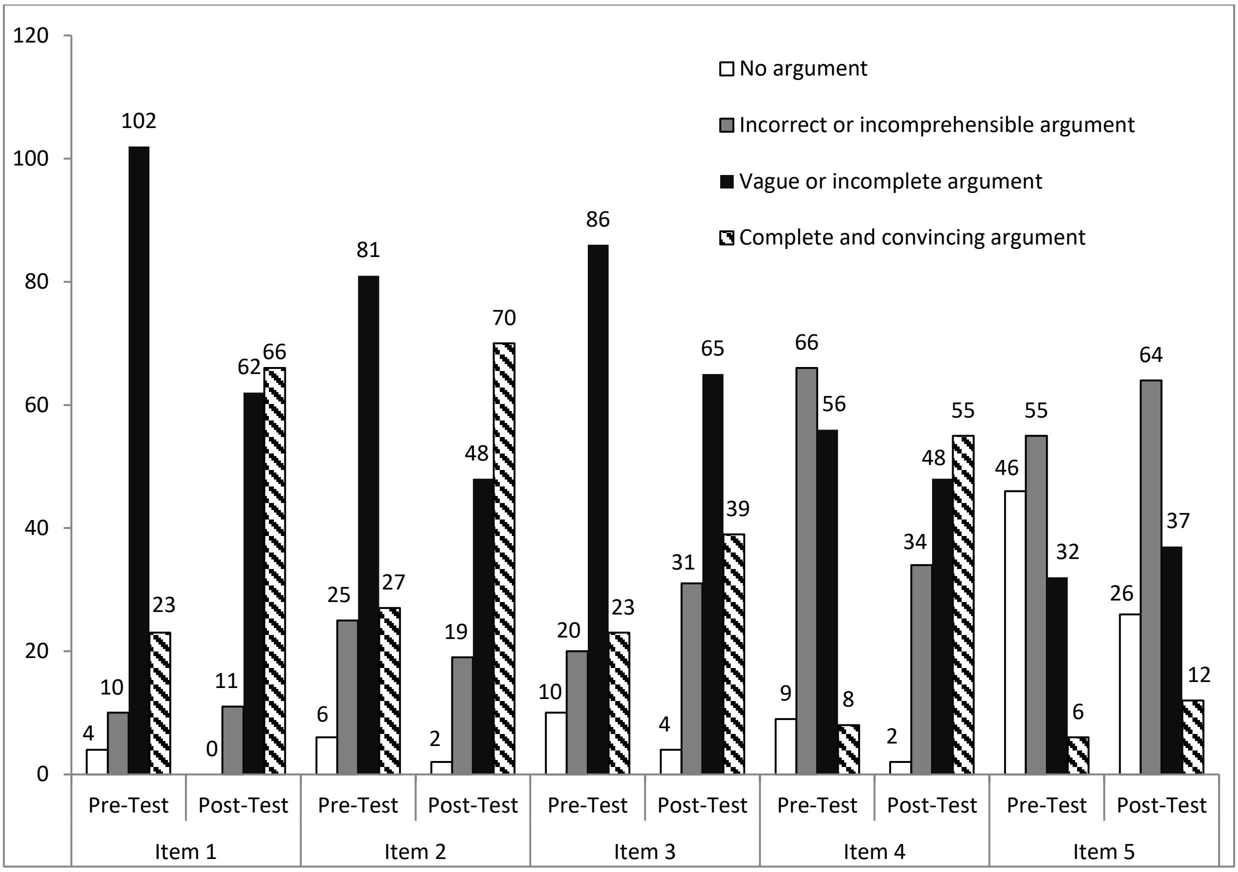

Figure 1. The distribution of responses by the levels of justification

Like in the first item, in Item 2 and Item 3 the *vague or incomplete argument* was the most frequent category in the pre-test (Item 2: 81 responses and Item 3: 86 responses); yet, in the post test, the frequency of *vague or incomplete argument* category declined to 48 in Item 2 and it declined to 65 in Item 3. Besides, in the post-test, the frequency of the *complete and convincing argument* category increased from 27 to 70 for Item 2 and from 23 to 39 for Item 3.

In Item 4, while most of the responses in the pre-test were under the *incorrect or incomprehensible argument* (66) and *vague or incomplete argument* (56) categories, in the post-tests there was a shift. Most of the responses in the post-test were coded under the *vague or incomplete argument* (48) and *complete and convincing argument* (55) categories. It should be noted that there was a considerable jump in the *complete and convincing argument* category in the post-test (from 8 to 55).

A review of Figure 1 also indicates that the level of justifications in Item 5 were considerably different from responses in the first four items. In particular, there were only a few *complete and convincing arguments* both in the pre-test and post-test, and many of the responses were coded as *no argument* or *incorrect or incomprehensible argument*. In Item 5, there was only slight improvement in the level of justifications from pre-test to post-test.

The complete and convincing arguments by treatment groups

Above findings pertain to all participants' responses to each of the items; yet, since the participants were formed into two treatment groups (PIME & DGE), the analysis also focused on the differences between the treatment groups. A major purpose of the treatments was to help all participants reach the *complete and convincing argument* category (Figure 2).

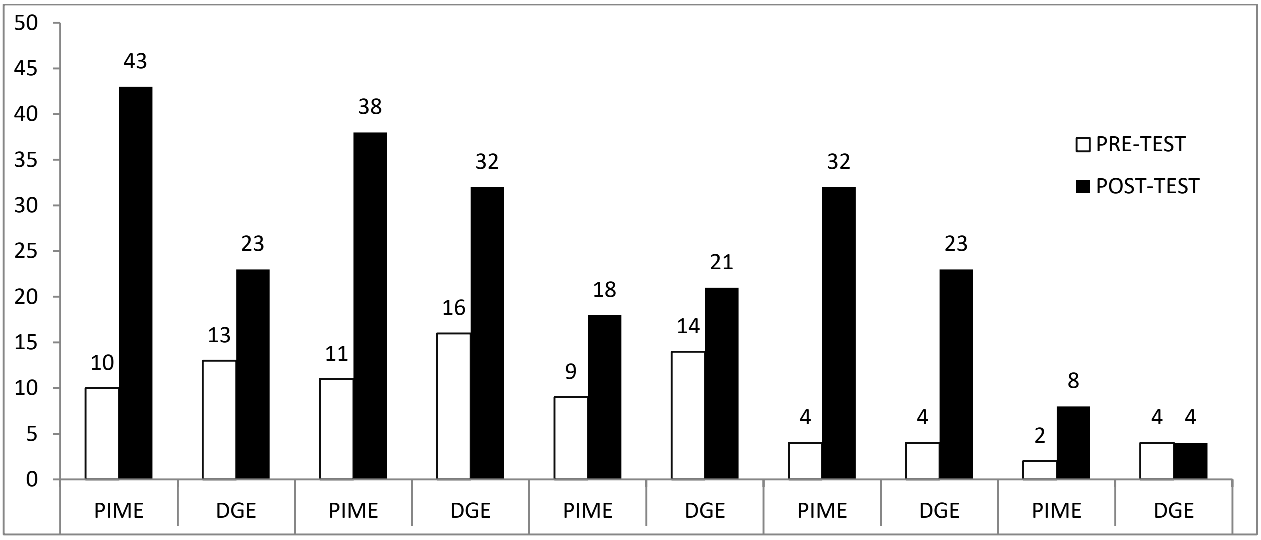

Figure 2. Complete and convincing argument category by the treatment groups

Figure 2 indicates that group frequencies of the *complete and convincing argument* category increased from pre-test to post-test. While considerable changes are seen in the first four items, in Item 5 the numbers of responses coded as *complete and convincing argument* were very low in the pre-test and post-tests. In all items, frequencies of complete and convincing argument category in the PIME group improved relatively more than the same frequencies in the DGE group.

In order to check significance of the increases from the pre-test to post-test, chi-square tests were conducted. One-sample chi-square tests were conducted to assess whether there were differences between the number of participants in the PIME and DGE groups whose responses were coded under the *complete and convincing argument* category. The analysis of the pre-test and post-test data indicates that significant results were only found in the first item of the post-test ($p= .014 < .05$). This particular finding suggests that in the post-test the number of PIME participants with the *complete and convincing arguments* were significantly more than the number of DGE participants with *complete and convincing arguments*.

Additional one-sample chi-square tests were conducted to assess whether there were significant differences between the pre-test and post-tests. In particular, the changes in the number of participants whose responses were coded under the *complete and convincing argument* category were investigated. Such analyses were done by each treatment group. In the PIME Group, significant differences were found in Item 1 ($\chi^2 = 20.547$, $p= .000 < .01$), Item 2 ($\chi^2 = 14.878$, $p= .000 < .01$), and Item 4 ($\chi^2 = 21.778$, $p= .000 < .01$). On the other hand, in the DGE Group, significant differences were found only in Item 2 ($\chi^2 = 5.33$, $p= .021 < .05$) and Item 4 ($\chi^2 = 13.370$, $p= .000 < .01$). These results suggest that in the PIME group, the number of responses coded under the *complete and convincing argument* category significantly increased in Item 1, Item 2 and Item 4; and in the DGE group, such significant differences were found in Item 2 and Item 4.

DISCUSSION

This study has investigated first year mathematics teacher education students' justifications of their solutions to geometry problems. Data analyses show that overall, the participants' level of justifications improved during the treatments. In particular, based on the pre-test and post-test comparison, it was observed that they provided more accurate, complete and mathematically sound justifications at the end of the study. Regardless of the treatment methods, participants produced sound, more reasonable, complete and convincing arguments to support their solution methods at the end of the study. Writing explanations and justifications could be a motivating factor for students and such processes might have helped them develop their thinking skills (Ma, 1999). The use of inquiry-based geometry activities might have played a significant role in enhancing the participants' justification skills because throughout the treatments in each session the participants worked on and solved non-routine geometry problems, and they justified their solutions verbally and on paper. Previous research has also indicated that the use of inquiry-based activities with dynamic geometry environments and physical materials has the potential to improve student performance in mathematics (Erbas & Yenmez, 2011; Zacharia & Costantinou, 2008).

Although the participants' level of justification improved on all five items, the improvement in the fifth item was relatively lower than the first four items. This particular finding might have been due to the nature of the fifth item, which was originally used as one of the problems of the TIMSS video study (http://www.timssvideo.com). In an analysis of the videos in the TIMSS Videotape Classroom Study, it was reported that such problems are used to foster students' mathematical thinking (Shimizu, 1999). In that item, the students were asked to understand the area of triangles with the same base and height between two parallel lines. The item demands students to relate their previous understanding of parallel lines. Thus, the item requires a comprehensive understanding of the subject for successful performance. Due to this feature, the item might have become relatively more challenging for the Turkish participants to solve and provide their justifications. In a previous research on TIMSS Video Study, Hiebert and Stigler (2000) have found out that German and US students were given considerably fewer opportunities than Japanese students to engage in deductive reasoning, doing mathematics and analyse mathematical situations. This might have been one of the reasons of Japanese students' higher TIMMS performance than students of other nations. Like German and US students, Turkish students are not also familiar with such demanding problems. For better performance, students should be given more opportunities to work through mathematical tasks demanding high level mathematical thinking (Stein & Smith, 1998). The improvement in the level of justification on all five items indicated that both treatment methods in general helped the participants improve their justification skills. In treatment groups, the participants were given open-ended geometry problems, and asked to solve and justify their solution strategies. This nature of the treatments might have positively influenced the participants' justifications. Additionally, fostering the learners' justification skills during the treatments might have helped them provide sound and reasonable justifications to support their decisions and methods in solving the problems (Charalambous, Hill, & Ball, 2011; Kinach, 2002; Sandborg, 1998).

In the present study, Dynamic Geometry Environment (DGE) and Physical Instructional Environment (PIME) were found to be appropriate settings for the participants to provide more sound and complete arguments to justify their solutions (Arıcı & Tutak, 2015; Erbas & Yenmez, 2011). During the treatments, the participants were asked to solve high level geometry problems; and, the instructors highly encouraged the participants to write down their justifications as well as solutions. Toward the end of the study, the participants were providing their justifications on their own without the instructors' prompts; hence, after a while, they might have become internally motivated to justify their solutions and such habits might have become a classroom routine. It is highly possible that the increase in the number of *complete and convincing arguments* is a direct impact of such treatment conditions. The findings of the present study suggest that when learners are expected to justify their solutions their justification skills most likely increase (Christou et al., 2004). As a matter of fact, it is well known that improvements are expected when students are given opportunities to explain their solution methods and convince others of the reasonableness of their findings (Driscoll et al., 2007).

Investigating any differences between the treatment groups indicate that in the first item there were significant mean differences in the number of *complete and convincing arguments*. In particular, the participants in the PIME improved their *complete and convincing arguments* significantly more than the participants in the DGE. In fact, we see the very same pattern in all other items; although, those differences were not statistically significant. It seems that being taught with the PIME method was more effective in producing *complete and convincing arguments*. This particular finding may be due to a number of reasons. The participants are used to the traditional paper and pencil environment; so, they might have needed more time to improve their problem-solving skills in the DGE environment. Additionally, although the DGE group received the instruction in a computer lab and they worked on the computer to solve the problems, the pre-test and post-test were given as a traditional paper-pencil test.

Thus, it might have been different if the researchers have found a way to give the assessment instruments on a computer.

In addition, the other results and findings of the present study show that the use of dynamic geometry environments and physical materials in teacher education contexts should be encouraged for enhanced justification skills (Hanna, 2000). The dynamic geometry software and physical materials allow the learner to attempt to solve the given problem in multiple ways, develop hypotheses and see the conclusions.

CONCLUSION

Based on the findings, it can be asserted that more valid conclusions can be derived from the findings if the medium of instruction and assessment were the same. Future research should utilize computer-based assessment in DGE-like treatments. The present study recommends that pre-service teachers should be given enough skills and opportunities to work with dynamic geometry software and physical instructional materials. Having first-hand experience with such tools will more likely to improve pre-service teachers' justification skills. Having said that, future studies should be conducted to investigate the effect of other instructional environments such as online learning tools on justification skills of learners. Additionally, researchers may also consider embedding such materials into geometry learning for a better understanding of student justifications. Finally, future researchers may also consider including a control group to better understand the effect of the treatment and forming another treatment group, which utilizes both dynamic geometry software and physical instructional materials.

REFERENCES

Arıcı, S., & Aslan-Tutak, F. (2015). The effect of origami-based instruction on spatial visualization, geometry achievement and geometry reasoning. *International Journal of Science and Mathematics Education*, *13*(1), 179-200. doi: 10.1007/s10763-013-9487-8.

Cai, J. (2000). Mathematical thinking involved in U.S and Chinese students' solving of process-constrained and process-open problems. *Mathematical Thinking and Learning, 2(*4), 309-340.

Cai, J. (2003). Singaporean students' mathematical thinking in problem solving and problem posing: an exploratory study, *International Journal of Mathematics Education in Science and Technology, 34*(5), 719-737.

Carbonneau, K. J., Marley, S. C., & Selig, J. P. (2013). A meta-analysis of the efficacy of teaching mathematics with concrete manipulatives. *Journal of Educational Psychology*, *105*(2), 380-400.

Charalambous, C. Y., Hill, H. C., & Ball, D. (2011). Prospective teachers' learning to provide instructional explanations: How does it look and what might it take? *Journal of Mathematics Teacher Education*, *14*, 441-463.

Chick, H. L. (2003). Pre-service teachers' explanations of two mathematical concepts. *Proceedings of the 2003 Annual Conference of the Australian Association for Research in Education* (Auckland, NZ). Retrieved from http://www.aare.edu.au/03pap/chi03413.pdf

Christou, C., Mousoulides, N., Pittalis, M., & Pitta-Pantazi, D. (2004). Proofs through exploration in dynamic geometry environments. *International Journal of Science and Mathematics Education. 2*, 339-352.

Driscoll, M., Wing DiMatteo, R., Nikula, J., & Egan, M. (2007). *Fostering geometric thinking: A guide for teachers, Grades 5-10.* Portsmouth, NH: Heinemann.

Driscoll, M., Wing DiMatteo, R., Nikula, J., Egan, M., Mark, J., & Kelemanik, G. (2008). *The fostering geometric thinking toolkit.* Portsmouth, NH: Heinemann.

Empson, B. S., & Turner E. (2006). The emergence of multiplicative thinking in children's solutions to paper folding tasks. *Journal of Mathematical Behavior 25*(1), 46-56.

Erbas, A. K., & Yenmez, A. A. (2011). The effect of inquiry-based explorations in a dynamic geometry environment on sixth grade students' achievements in polygons. *Computers & Education*, *57*(4), 2462-2475.

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2015). *How to design and evaluate research in education.* New York, NY: Mc. Graw-Hill. Inc.

Guven, B., & Karataş, İ. (2009). The effect of dynamic geometry software (Cabri) on pre-service elementary mathematics teachers' achievement about locus problems. *Ankara University Journal of Faculty of Educational Sciences, 42*(1): 1-31.

Hadas, N., Hershkowitz, R., & Schwarz B. B. (2000). The role of contradiction and uncertainty in promoting the need to prove in dynamic geometry environments. *Educational Studies in Mathematics, 44*, 127-150.

Hanna, G. (2000). Proof, explanation and exploration: An overview. *Educational Studies in Mathematics, 44,* 5-23.

Hiebert, J., & Stigler, J. W. (2000). A proposal for improving classroom teaching: Lessons from the TIMSS video study. *The Elementary School Journal*, 3-20.

Hoong, L.Y., & Khoh, L. T. S. (2003). Effects of Geometer's Sketchpad on spatial ability and achievement in transformation geometry among secondary two students in Singapore. *The Mathematics Educator, 7*(1), 32- 48.

Kaplan, E. (2011). *Ortaogretim Geometri Ders Kitabi* [High School Geometry Textbook] (2nd ed.). Ankara, Turkey: Pasa Publication.

Kelly, A. C. (2006). Using manipulatives in mathematical problem solving: A performance-based analysis. *The Montana Mathematics Enthusiast, 3*(2),184-193.

Kinach, B. M., (2002). Understanding and learning-to-explain by representing mathematics: Epistemological dilemmas facing teacher educators in the secondary mathematics "methods" course. *Journal of Mathematics Teacher Education,* 5, 153-186.

Ma, L. (1999). *Knowing and teaching elementary mathematics: Teachers' understanding of fundamental mathematics in China and the United States*. Mahwah, NJ: Lawrence Erlbaum.

McNeill, K. L., Lizotte, D. J., Krajcik, J., & Marx, R. W. (2006). Supporting students' construction of scientific explanations by fading scaffolds in instructional materials. *The Journal of the Learning Sciences*, *15*(2), 153-191.

Olkun, S. (2003). Comparing computer versus concrete manipulatives in learning 2D geometry. *Journal of Computers in Mathematics and Science Teaching, 22*(1), 43-56.

Pandiscio, E. A. (2002). Exploring the link between preservice teachers' conception of proof and the use of dynamic geometry software. *School Science and Mathematics*, *102*(5), 216-221.

Richardson, K., Carter, T., & Berenson S. (2010). Connected tasks: The building blocks of reasoning and proof. *Australian Primary Mathematics Classroom, 15*(4), 17-23.

Robson, C. (1993). *Real World Research.* Oxford: Blackwell Publishers Ltd.

Sandborg, D. (1998). Mathematical explanation and theory of Why-Questions. *The British Journal for the Philosophy of Science, 49*(4), 603-624.

Shimizu, Y. (1999). Studying sample lessons rather than one excellent lesson: A Japanese perspective on the TIMSS videotape classroom study. *ZDM*, *31*(6), 190-194.

Stein, M. K., & Smith, M. S. (1998). Mathematical tasks as a framework for reflection: From research to practice. *Mathematics Teaching İn The Middle School*, *3*(4), 268-275.

Stols, G. (2012). Does the use of technology make a difference in the geometric cognitive growth of pre-service mathematics teachers? *Australasian Journal of Educational Technology*, *28*(7), 1233-1247.

Zacharia, Z. C., & Constantinou, C. P. (2008). Comparing the influence of physical and virtual manipulatives in the context of the Physics by Inquiry curriculum: The case of undergraduate students' conceptual understanding of heat and temperature. *American Journal of Physics, 76*(4), 425-430.