Examining Item Difficulties with respect to Science Teachers' Backgrounds and their Views on Science Instruction

Soru Zorluk Derecelerinin Fen Öğretmenlerinin Özelliklerine ve Fen Öğretimi Hakkındaki Görüşlerine Göre İncelenmesi

Eren CEYLAN[1]
Ankara University

Abstract

Considering the undeniable importance of science teachers in teaching science, this study was carried out to examine the average item difficulties in TIMSS 2011 with respect to various eighth grade student groups which were formed based on their science teachers' background information and these teachers' views on science learning environments. The results indicate that the average percentages of the students who answered science items correctly are higher in the group of students whose teachers have a major degree in science than those in the group of students whose teachers have a major degree in science education. In addition, the average percentage of the students who answered the science items correctly is lower in the group of students whose teachers preferred to use instructional strategies to engage their students in science instruction in most lessons than that in the group of students whose teachers preferred to use instructional strategies to engage their students in science instruction in only some lessons.

Keywords: TIMSS 2011, science item difficulty, science teachers' backgrounds, and learning environment in science classrooms.

Öz

Fen Öğretmenlerinin fen öğretimindeki rolü yadsınamayacak kadar büyük olduğundan, bu çalışma TIMSS 2011' de bulunan sekizinci sınıf fen sorularının ortalama zorluk derecelerinin fen öğretmenlerinin özelliklerine ve fen öğretimi hakkındaki görüşlerine göre oluşturulan öğrenci gruplarına göre hesaplanmasını amaçlamıştır. Sonuçlara göre, fen alanlarından mezun öğretmenlerin öğrencilerinin bulunduğu grupta, fen sorularının fen eğitimi alanından mezun öğretmenlerinin öğrencilerinin bulunduğu gruba göre daha fazla öğrenci tarafından cevaplandığı tespit edilmiştir. Bununla beraber, fen derslerinin çoğunda, öğrencilerini fen ile ilişkilendiren etkinlikler seçen öğretmenlerin öğrencilerinin bulunduğu grupta, fen derslerinin sadece bazılarında öğrencileri fen ile ilişkilendiren etkinlikler seçen öğretmenlerin bulunduğu gruba göre fen soruları daha az öğrenci tarafından doğru cevaplanmıştır.

Anahtar Sözcükler: TIMSS 2011, fen madde güçlüğü, fen öğretmenlerinin özellikleri, fen sınıflarında öğrenme ortamı.

[1] Assist. Prof. Dr. Eren CEYLAN, Ankara University, Faculty of Educational Sciences, SSME Deparment, ernceylan@gmail.com

Introduction

Since they are the primary sources of instruction in science classes, science teachers, without a doubt, have a vital role in students' understanding of scientific knowledge and also in helping them to gain higher order thinking skills in science. Considering the standards of excellence, science teachers are defined in the literature as qualified teachers having high academic skills, teaching in the field in which they received their training, having more than a few years of experience, and participating in high-quality instruction and professional development programs (Mayer, Mullens, & Moore, 2000). One of the important attributes that contributes to teachers' qualification is defined in the literature as teachers' experience or teaching experience. However, even though some of the studies in the literature indicated a strong relationship between teachers' teaching experience and their students' academic achievements (Hanushek, 1986; Greenwald, Hedges, & Laine, 1996), some recent studies have found no statistically significant difference between novice and expert teachers' students with respect to academic achievements (Croninger et al. 2007; Rivkin et al. 2005).

The National Science Teachers Association (NSTA) and the Association for the Education of Teachers in Science (AETS) have developed standards for science teacher education. These standards are identified based on some important issues that make a science teacher very effective (Ellis, 2001). One of the standards which is about "content knowledge" refers to teachers' understanding and expression of knowledge, and their practice of contemporary science. In this respect, an effective science teacher should interrelate and interpret important concepts, ideas, and applications in their fields of licensure (NSTA, 2012; McConnell, Parker, & Jan Eberhardt, 2013). Another standard which is defined as "content pedagogy" refers to teachers' understanding of how students learn and develop scientific knowledge. The effort of science teachers to further their knowledge and understanding of content knowledge and science pedagogy is also included in this set of standards and defined as improving "professional knowledge and skills" (NSTA, 2012). Namely, once one becomes a science teacher, it is very important for him to pursue opportunities for further professional development and update his knowledge and skills of teaching science. In this respect, studies have also indicated that joining seminars, workshops, conferences, and submitting articles to the professional journals have a potential to make teachers more effective in their profession and broaden their knowledge (Yoon, Duncan, Lee, Scarloss, & Shapley, 2007). Moreover, in another standard which is related to the "learning environments" in science classrooms, an effective science teacher's ability to make plans for engaging all students in science learning by setting appropriate goals is expressed. The plans of an effective science teacher should incorporate the nature and social context of science, inquiry, and appropriate safety considerations. These plans should also emphasize the importance of scientific investigation in which students collect and interpret data using applicable science-specific technology (NSTA, 2012).

The aforementioned standards which are attributed to effective science teachers are trailed also in TIMSS (Trends in Mathematics and Science Study). TIMSS which is a project of IAE (International Association for the Evaluation of Educational Achievement) does not only assess students' science and mathematics achievements at particular grade levels, but also, with its fruitful data, gives opportunities to the researchers to investigate and compare the students' science teachers based on the aforementioned standards. TIMSS also enables researchers to link students' competency levels in science with their science teachers' background information (their education, professional development, and experience in teaching), their views about science curricula and about the instructional activities and materials (learning environments) used in the science classrooms. TIMSS have been carried out once every four year, and its latest implementation was in 2011 with the participation of 63 countries which encompassed approximately half of a million of students all around the world at fourth and eighth grade levels (Martin, Mullis, Foy, & Stanco, 2012).

In the light of the literature and the data obtained from TIMSS 2011, this study was carried out to compare eighth grade students based on their science performances in various content and cognitive levels in science with a consideration of their science teachers' different backgrounds (education, professional development, and experience in teaching) and the learning environments their science teachers' create while providing science instruction in the classrooms. The results have also been interpreted with reference to aforementioned standards that make a science teacher effective.

Methods

Item difficulty is defined as the proportion of students who answered an item correctly. Item difficulty can range 0.00 to 1.00, and the higher the value, the easier the item (Van Blerkom, 2009). In TIMSS, the performances of students in science and mathematics are reported as plausible values. In this study, average item difficulties based on the performances of certain groups of students in Turkey were calculated and used, instead of students' average science plausible values in Turkey.

Plausible values are the relative estimated scores of the students based on the average performance of all countries' students in TIMSS. Therefore, some strong or weak sides of a country with respect to some issues would not be revealed with the consideration of average performances themselves. That is why there are plenty of differential item functioning (DIF) studies in the literature which have been conducted to detect whether items on a test may be functioning differently across the participating countires (Wolf, 1998; Yıldırım & Yıldırım, 2011; Zumbo, 2007). In addition, plausible values are estimated based on the performances of countries to compare one country with another, not individual students. So, this might lead to overlooking some important issues. For example, an item related to life cycles in biology may be relatively easy for the students in Turkey but difficult for those in other countries. This item would be considered difficult with respect to all countries but it would not be so if only the performances of Turkish students were considered. Moreover, it cannot be understood from the average science plausible values whether a certain group of items are easy or difficult for Turkish students. To illustrate this, Table 1 represents the average achievement scores (i.e., plausible values) of Turkish eighth grade students in science content domains (biology, chemistry, physics, and earth science) and in cognitive domains (knowing, applying, and reasoning) (Martin, Mullis, Foy, & Stanco, 2012) and Table 2 presents the average item difficulties in science content and cognitive domains for the same students.

Table 1.
Average Achievement in Science Content and Cognitive Domains at Eighth Grade Level

	Content Domains					Cognitive Domains		
	General	Biology	Chemistry	Physics	Earth Sci.	Knowing	Applying	Reasoning
8th grade	483 (3.4)	484 (3.7)	477 (4.0)	494 (3.7)	468 (3.5)	490 (3.8)	478 (3.4)	483 (3.4)

Standard errors are given in the parenthesis

Table 2.
Average Item Difficulties in Science Content and Cognitive Domains at Eighth Grade Level

	Content Domains					Cognitive Domains		
	General	Biology	Chemistry	Physics	Earth Sci.	Knowing	Applying	Reasoning
8th grade	0.44	0.45	0.48	0.43	0.42	0.51	0.44	0.34

Standard errors are less than 0.02 for all domains

When Table 1 and Table 2 are compared, it can be realized that although the students' average score in chemistry is lower than those in physics and biology, the average item difficulty of chemistry items is higher than the average item difficulty of physics and biology which indicates more students answered items correctly in chemistry compared to physics and biology. In addition, although the average achievement score of the students in the domain reasoning is higher than their average applying scores, the average item difficulty of reasoning cognitive domain is substantially lower than the average item difficulty of applying cognitive domain.

Finally, it is also mentioned in TIMSS's reports that it is not appropriate to use plausible values as an indicator of individual students' achievement level since these values are constructed to be used in the calculation of the average achievement of the population of interest (Olson, Martin, & Mullis, 2008).

Sample

63 countries from all around the world participated in TIMSS 2011. In TIMSS 2011, two stage stratified cluster sampling was used. Schools were randomly selected with probability proportional to size, and one or more classes were selected randomly from the relevant grades in sampled schools (Joncas, 2007). Based on this design, 6928 eighth grade students from 239 schools in Turkey were sampled. This sample consisted of 3414 girls and 3514 boys. In addition, 240 eighth grade level science teachers were also selected to fill out the science teacher questionnaire in Turkey. 49% of these teachers were female and 51% of these teachers were male.

Instruments

TIMSS 2011 Science Achievement Tests for the eighth graders were used in this study. The items in science achievement tests were developed and organized based on the eighth grade level's science content or cognitive domains. So, not only the overall science performance of the students were observed, but also their performance on certain sub-domains (eg., biology in the content domain, knowing in the cognitive domain) were generated based on the items in these corresponding sub-domains. (Martin, Mullis, Foy, & Stanco, 2012).

The range of cognitive skills of the students in science is divided into three sub-domains in TIMSS. Knowing is the first cognitive domain, and it entails that students need to know certain science facts, procedures, and concepts. Applying, the second cognitive domain refers to students' ability of applying knowledge and conceptual understanding to a science problem. Reasoning which focuses of unfamiliar situations, complex contexts, and multi-step problems instead of solving routine science problems is defined as the third cognitive domain (Mullis, Martin, Ruddock, O'Sullivan, & Preuschoff, 2009). On the other hand, science content domains are organized based on four subject areas in TIMSS. The topics of biology in eighth grade level are defined as characteristics, classification, and life processes of organisms, cells and their functions, life cycles, reproduction, and heredity, diversity, adaptation, and natural selection, ecosystems, and human health. The topics of chemistry in eighth grade level are classification and composition of matter, properties of matter, and chemical change. Physics, in eighth grade level, includes the topics of physical states and changes in matter, energy transformations, heat, and temperature, light and sound, electricity and magnetism, forces and motion. Finally, the topics of earth science content domain are identified as earth's structure and physical features, earth's processes, cycles, and history, earth's resources, their use and conservation, earth in the solar system and the universe (Mullis, Martin, Ruddock, O'Sullivan, & Preuschoff, 2009).

In the science achievement tests at eighth grade level in TIMSS 2011, there were a total of 217 questions and they were presented to the students in various booklets. The numbers of the items with respect to the four science content domains - biology, chemistry, physics, and earth science were 79, 44, 55, and 39, respectively. The numbers of the items with regard to the three science cognitive domains - knowing, applying, and reasoning were 73, 92, and 52, respectively.

In addition, TIMSS 2011 science teacher questionnaire was also used in this study to delineate teachers' academic and professional backgrounds, classroom resources, instructional practices, and attitudes toward teaching. In this questionnaire, teachers were also asked to evaluate the science curriculum, instructional activities, and the materials used in science classes. The questionnaire took about 45 minutes to complete.

Analysis

IDB (International Database) Analyzer 3.0 was used to analyze the data. IDB Analyzer developed by the IEA Data Processing and Research Center (IEA-DPC) is used for combining and analyzing data from IEA's large-scale assessments such as TIMSS. The SPSS code generated by IDB

Analyzer is used to calculate the estimates of achievement and their corresponding standard errors, combining sampling and imputation variance. The main advantage of using the IDB analyzer is that it takes into account the sampling weights, thus producing results that can be generalized to the population. (IAE, 2012).

The achievement scores (plausible values) of the students in TIMSS were reported based on IRT (Item Response Theory) on a scale between 0 and 1000. In TIMSS, this scale is divided into four benchmarks (400, 475, 550, and 625) to make this scale more useful for the education community. However, in TIMSS 2011, the average difficulty indices were not reported with respect to science content and cognitive domains. The average difficulty indices were not generated for certain groups of students constituted based on their teachers' background information either. Therefore, in this study, the average item difficulties in science content and cognitive domains were generated for the eighth graders in Turkey. Also, the average item difficulties in science content and cognitive domains were calculated for certain groups of students constituted based on their science teachers' responses to the questions in TIMSS 2011 teacher questionnaire that are about background information and science learning environments. All these item difficulties were produced by using IDB Analyzer 3.0.

Results

Firstly, IDB Analyzer was used to produce average item difficulties in science content and cognitive domains for the eighth graders in Turkey (see Table 2 above). It was revealed in Table 2 that the average difficulty of all the science items in Turkey was 0.44. Similarly, the average item difficulties in the four science content domains -biology, chemistry, physics, and earth science were found to be 0.45, 0.48, 0.43, and 0.42, respectively. In addition, the average item difficulties in the three cognitive domains -knowing, applying, and reasoning were found to be 0.51, 044, and 0.34, respectively. When the average item difficulties in the content domains are compared against each other and against the general average item difficulty, it can be said that there is not a big difference between the average item difficulties of the content domains and the general average item difficulties. On the other hand, science items in the reasoning domain seem to be relatively more difficult on average than the items in the domains knowing and applying for the eighth graders in Turkey. For example, one should notice that, as compared to the standard errors, the difference 0.17 between the average item difficulties in the domains knowing and reasoning is highly significant.

The average item difficulties in science content and cognitive domains were produced for the groups of students whose teachers have different years of experience in science teaching (Table 3 and Table 4). The students were categorized based on their science teachers' years of experience in science teaching such as "20 or more years of experience", "at least 10 but less than 20 years of experience", "at least 5 but less than 10 years of experience", and "less than 5 years of experience". Table 3 and Table 4 indicates that the average percentage of students who answered science items correctly in both content and cognitive domains is higher in the group of students who have a teacher with 20 or more years of experience than the other groups of students whose teachers have less than 10 years of experience. The average item difficulties have a tendency to be higher especially in biology and chemistry content domains when the teachers have 20 or more years of experience.

Table 3.
Average Item Difficulties in Cognitive Domains based on Teachers' years of Experience

	Average item difficulties			
Teachers' years of experience	General	Knowing	Applying	Reasoning
20 Years or More	0.47	0.54	0.46	0.36
At Least 10 but Less than 20 Years	0.47	0.53	0.46	0.37
At Least 5 but Less than 10 Years	0.43	0.49	0.43	0.32
Less than 5 Years	0.42	0.49	0.41	0.31

Standard errors are less than 0.02 for all domains.

Table 4.
Average Item Difficulties in Content Domains based on Teachers' years of Experience

Teachers' years of experience	Average item difficulties			
	Biology	Chemistry	Physics	Earth Science
20 Years or More	0.47	0.51	0.44	0.43
At Least 10 but Less than 20 Years	0.46	0.50	0.45	0.44
At Least 5 but Less than 10 Years	0.43	0.46	0.41	0.42
Less than 5 Years	0.41	0.45	0.40	0.39

Standard errors are less than 0.02 for all domains.

The average item difficulties in science content and cognitive domains were produced for the groups of students whose teachers have different major degrees in science and science education. The students were categorized based on their science teachers' major degrees such as "major in science and science education", "major in science education but no major in science", and "major in science but no major in science education". Table 5 and Table 6 show that the average percentage of students who answered science items correctly is higher in the group of students whose teachers have a major degree in science but no major degree in science education than the other two groups of students whose teachers have a major in science and science education and whose teachers have a major in science education but no major in science.

Table 5.
Average Item Difficulties in Cognitive Domains based on Science Teachers' Major Degrees

Teachers' Major Degrees	Average item difficulties			
	General	Knowing	Applying	Reasoning
Major in Science and Science Education	0.44	0.51	0.44	0.34
Major in Science Education but No Major in Science	0.42	0.49	0.42	0.32
Major in Science but No Major in Science Education	0.46	0.53	0.46	0.36

Standard errors are less than 0.02 for all domains.

Table 6.
Average Item Difficulties in Content Domains based on Science Teachers' Major Degrees

Teachers' Major Degrees	Average item difficulties			
	Biology	Chemistry	Physics	Earth Sci.
Major in Science and Science Education	0.44	0.47	0.43	0.41
Major in Science Education but No Major in Science	0.43	0.46	0.40	0.41
Major in Science but No Major in Science Education	0.46	0.50	0.45	0.44

Standard errors are less than 0.02 for all domains.

The average item difficulties in science content and cognitive domains were produced for the groups of students whose teachers participated in professional development in science within the past two years. The students were categorized based on their science teachers' participation in professional development as "improving students' critical thinking or inquiry skills" and "science pedagogy and instruction". Table 7 and Table 8 reveal that the average percentages of students who answered the science items correctly vary slightly between the group of students whose teachers participated in professional development about students critical thinking and inquiry skills or science pedagogy in the past two years and the group of students whose teachers did not participate in professional development about students critical thinking and inquiry skills or science pedagogy in the past two years.

Table 7.

Average Item Difficulties in Cognitive Domains based on Participation in Professional Development

| | Improving Students' Critical Thinking or Inquiry Skills | | | | Science Pedagogy / Instruction | | |
| | Cognitive Domains | | | | Cognitive Domains | | |
	General	Knowing	Applying	Reasoning	Knowing	Applying	Reasoning
Yes	0.45	0.53	0.45	0.35	0.52	0.45	0.35
No	0.43	0.50	0.43	0.33	0.50	0.43	0.33

Standard errors are less than 0.02 for all domains.

Table 8.

Average Item Difficulties in Content Domains based on Participation in Professional Development

| | Improving Students' Critical Thinking or Inquiry Skills | | | | | Science Pedagogy / Instruction | | | |
| | Content Domains | | | | | Content Domains | | | |
	General	Biology	Chemistry	Physics	Earth Science	Biology	Chemistry	Physics	Earth Science
Yes	0.45	0.45	0.48	0.44	0.43	0.45	0.49	0.44	0.44
No	0.43	0.43	0.47	0.41	0.41	0.43	0.46	0.42	0.41

Standard Errors are less than 0.02 for all domains.

The average item difficulties in science content and cognitive domains were produced for the group of students whose teachers showed various efforts to engage the students in science instruction by using a variety of strategies. The effort for "engaging students in learning" was defined based on the science teachers' responses to the items in the questionnaire that are related to their instructional practices such as "summarizing what students should have learned from the lesson","using the questioning method to elicit reasons and explanations", "encouraging all students to improve their performance", and "praising students for good efforts". The students were categorized with respect to their science teachers' responses to these items in terms of their frequencies as "in most lessons", "in about half of the lessons", and "in some lessons". Table 9 and Table 10 indicate that the average percentage of students who answered science items correctly is lower in the group of students whose teachers prefer to use instructional practices to engage their students in science instruction in most lessons than the group of students whose teachers prefer to incorporate instructional practices to engage their students in science instruction in some lessons. It is more prominent in the average item difficulties of the reasoning cognitive domain.

Table 9.

Average Item Difficulties in Cognitive Domains based on Instruction to Engage Students in Learning

Instruction to Engage Students In Science Lessons	General	Knowing	Applying	Reasoning
Most Lessons	0.44	0.51	0.43	0.33
About Half the Lessons	0.44	0.51	0.44	0.33
Some Lessons	0.51	0.57	0.50	0.46

Standard errors are less than 0.03 for all domains.

Table 10.

Average Item Difficulties in Content Domains based on Instruction to Engage Students in Learning

Instruction to Engage Students In Science Lessons	Biology	Chemistry	Physics	Earth Sci.
Most Lessons	0.44	0.47	0.42	0.42
About Half the Lessons	0.44	0.47	0.42	0.42
Some Lessons	0.52	0.51	0.50	0.48

Standard errors are less than 0.03 for all domains.

The average item difficulties in science content and cognitive domains were produced for the groups of students whose teachers put varying levels of emphasis on science investigations in their science lessons. The science teachers were asked to respond to some questions to identify how often they ask their students to engage in such activities during science teaching as: "observe natural phenomena such as the weather or a plant growing and describe what they see"," watch the teacher demonstrate an experiment or investigation", "design or plan experiments or investigations", "conduct experiments or investigations", "use scientifi c formulas and laws to solve routine problems", "give explanations about something they are studying", and "relate what they are learning in science to their daily lives". Based on their responses to these questions about the emphasis on science investigation, the students were categorized as "about half of the lessons or more" and "less than half of the lessons". Table 11 and Table 12 reveal that the average percentages of the students who answered the science items correctly vary slightly between the group of students whose science teachers emphasized science investigation "about half of the lessons or more" and the group of students whose science teachers emphasized science investigation "less than half of the lessons".

Table 11.

Average Item Difficulties in Cognitive Domains based on Teachers Emphasis on Science Investigation

Teachers Emphasize Science Investigation	General	Knowing	Applying	Reasoning
About Half the Lessons or More	0.44	0.51	0.43	0.33
Less than Half the Lessons	0.44	0.51	0.44	0.35

Standard errors are less than 0.03 for all domains.

Table 12.

Average Item Difficulties in Content Domains based on Teachers Emphasis on Science Investigation

Teacher Emphasize Science Investigation	Biology	Chemistry	Physics	Earth Science
About Half the Lessons or More	0.44	0.48	0.43	0.42
Less than Half the Lessons	0.45	0.48	0.43	0.43

Standard errors are less than 0.03 for all domains.

Discussion and Conclusion

Since science teachers' background information and their views about the learning environments in science classrooms have a potential to affect students' performance in answering the science items correctly in TIMSS 2011, this study was carried out with the eighth grade students in Turkey by directly considering the average percentages of the students who answered the science items correctly (item difficulty) to investigate the differences among the student subgroups. These subgroups were constituted based on their science teachers' background information (their years of experience in science teaching, their major degrees, and their participation in professional development in the last two years) and their learning environments in the science classrooms (science teachers' efforts to engage them in learning and emphasizing science investigation), and they were compared based on their respective average item difficulties in science content and cognitive domains.

Firstly, item difficulties were examined with respect to the groups of students whose science teachers had various years of experience (Table 3 and Table 4). In the literature, it is stated that more experienced teachers are assigned to students of higher ability and to classes with fewer discipline problems in some schools, while lower achieving students who need more help are assigned more experienced teachers in some other schools (Harris & Sass, 2011). In Turkey, both the students and the teachers are generally assigned to the classes in school randomly. However, experienced teachers have a priority to work at schools of their own choice, and they generally prefer to be in schools that have better facilities with higher achieving students. Moreover, although the nature of this association shows some differences across several studies, job experience and job satisfaction studies indicate that teachers who have less experience are generally engaged with different issues in their jobs from more experienced ones (Maele & Houtte, 2012, Crossman & Harris, 2006). These are just some of the views appeared in the literature regarding the relevance of teaching experience to student achievement, and more detailed studies focusing on the relationship between teaching experience and students' academic achievement need to be conducted to better understand the reasons behind this situation in Turkey.

One of the striking results of study was about the teachers' major degrees (Table 5 and Table 6). The importance of the content knowledge was expressed in some studies in the literature (McConnell, Parker & Eberhardt, 2013). Goldhaber and Brewer (2000), and they indicated that teachers who have subject specific academic degrees are generally more successful than teachers teaching "out of field". Having an extensive knowledge of content and curriculum as well as pedagogical knowledge, knowledge about learners and their characteristics, and knowledge about information technology have become very crucial for science teachers in line with the requirements of 21st century (Hill & Lubienski, 2007). However, the results of this study indicate the science teachers' success who graduated from out of education faculties in Turkey. Some of the policy makers in Turkey defend the view that it should be the responsibility of science faculties to educate and train science teachers, while some support the idea that science teachers should be educated by and graduated from education faculties. In this respect, the results of this study should be carefully evaluated and some qualitative studies that focus only on this issue should be conducted.

Research also indicated that the amount of professional development for teachers, especially more than 14 hours in a semester, is an important factor that has a significant effect on student achievement (Yoon, Duncan, Lee, Scarloss, & Shapley, 2007). A meta-analysis study revealed that especially focusing on science content in the professional development affects the science achievement of students positively (Blank & de las Alas, 2009). One of the variables that were examined in this study was the teachers' participation in professional development in the past two years (Table 7 and Table 8). These results may indicate that the professional development is implemented not very effectively in Turkey. The quality of in-service training programs should be examined carefully and they should be designed with a consideration of the practical issues in science classrooms in Turkey.

Since it is very crucial to build a bridge between curriculum and instruction, student content engagement is described as the activity that brings the student and the subject matter together. It can be further defined as student's in-the-moment cognitive interaction with the instructional content (McLaughlin et. all, 2005). Indeed, the cognitive interaction between the student and the instructional content defined as engagement may take many forms from listening to the teacher to providing an explanation for the solution of a problem. Accordingly, another surprising result of the study was encountered when the average item difficulties were examined with respect to the teachers' various efforts to engage students in science instruction (Table 9 and Table 10). The importance and positive effects of engaging students in the learning process were also expressed in the literature. For example, teachers in Japan, which is the one of the successful countries in TIMSS studies, share and summarize the goals and the results of an investigation, engage students with daily life activities, and discuss the investigations in small groups (Linn, Lewis, Tsuchida, & Songer, 2000). The reasons underlying this finding of the present study should be investigated in detail with a consideration of the implementation of instructional practices in science classrooms in Turkey.

Emphasizing the inquiry activities as an effective instructional practice is very important in science education. There are many studies in the literature that investigate the effects of the activities which the teachers in this study ask their students to do on students' science achievement. It was indicated in the literature with a meta-analysis of 138 studies that there is a positive effect of employing some inquiry-based instruction on students' understanding and retention of the science content. Especially activities that led students to thinking actively and drawing conclusions from the data they gained by hands-on experiences had a positive effect on scientific understanding (Minner, Levy, & Century, 2009). On the other hand, based on the TIMSS data in Turkey, a negative relationship was found between inquiry-based activities and students' science achievement (Aypay, Erdogan, & Sozer, 2007). Similarly, in this study, it was revealed that the students whose teachers asked them to do inquiry-based activities more were not significantly different from the students whose teachers asked them to do inquiry-based activities less with respect to the average percentages of students who answered science items correctly. The reason behind this finding might be the improper implementation of such activities in science classes by teachers in Turkey.

Finally, the analysis of the TIMSS 2011 data set revealed prominent differences between the student groups based on their teachers' background information and views of learning environments in science classrooms with respect to the average percentages of the students who answered the science items correctly. For further studies, it will be very useful to analyze other countries' data sets adopting the perspective in the current study. It is also strongly recommended that similar studies be carried out based on other international studies' data sets such as PISA.

References

Aypay, A., Erdogan, M., & Sozer, M.A (2007). Variation among Schools on Classroom Practices in Science Based on TIMSS-1999 in Turkey. Journal of Research in Science Teaching, 44 (10), 1417-1435.

Blank, R. K. & de las Alas, N. (2009). Effects of teacher professional development on gains in student achievement: How meta analysis provides scientific evidence useful to education leaders. Washington, DC: The Council of Chief State School Officers.

Croninger, R., Rice, J., Rathbun, A., & Masako, N. (2007). Teacher qualifications and early learning: Effects of certification, degree, and experience on first-grade student achievement. Economics of Education Review, 26, 312–324.

Crossman, A., & Harris, P. (2006). Job satisfaction of secondary school teachers. Educational Management Administration & Leadership, 34, 29-46.

Ellis, J.D., (2001). A Dilemma in Reforming Science Teacher Education. Journal of Science Teacher Education, 12(3), 253-276.

Goldhaber, D., & Brewer, D. J. (2000). Does teacher certification matter? High school teacher certification status and student achievement. Educational Evaluation and Policy Analysis, 22(2), 129-145.

Greenwald, R., Hedges, L., & Laine, R. (1996). The effect of school resources on student achievement. Review of Educational Research, 3, 361–396.

Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. The Journal of Economic Literature, 24, 1141–1177.

Harris, D. N. & Sass, T. R. (2011). Teacher training, teacher quality and student achievement. Journal of Public Economics, 95, 798–812.

Hill, H. C., & Lubienski, S. T. (2007). Teachers' mathematics knowledge for teaching and school context: A study of California teachers. Educational Policy, 21(5), 747-768.

International Association for the Evaluation of Educational Achievement (2012). IDB Analyzer Version 3.1 User Guide. Retrieved from http://www.iea.nl/fileadmin/user_upload/IEA_Software/Installing_the_IDB_Analyzer__Version_3_1_.pdf

Joncas, M. (2007). TIMSS 2007 Technical Report: Chapter 5 TIMMSS 2007 Sample Design. International Association for the Evaluation of Educational Achievement. Boston, MA.

Linn, M.C., Lewis, C, Tsuchida, 1., & Songer, N.B. (2000). Beyond fourth-grade science: Why do U.S. and Japanese students diverge? Educational Researcher, 29, 4-14.

Maele, D.V.,., Houtte, M. V., (2012). The role of teacher and faculty trust in forming teachers' job satisfaction: Do years of experience make a difference? Teaching and Teacher Education 28, 879-889.

Martin, M.O, Mullis, I.V.S, Foy, P., & Stanco, G.M. (2012). TIMSS 2011 International Science Report. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College

Mayer, D. P., Mullens, J. E., & Moore, M. T. (2000). Monitoring school quality: An indicators report (NCES Statistical Analysis Report No. 2001-030). Washington, DC: U.S. Department of Education.

McConnell, T.J.,Parker, J.M., & Eberhardt J. (2013). Assessing Teachers' Science Content Knowledge: A Strategy for Assessing Depth of Understanding. Journal of Science Teacher Education, 24:717–743.

McLaughlin, M., McGrath, D. J., Burian- Fitzgerald, A., Lanahan, L., Scotchmer, M., Enyeart, C., & Salganik, L. (2005). Student content engagement as a construct for the measurement of effective classroom instruction and teacher knowledge. Retrieved from http:// www.air.org/files/AERA2005Student_Content_Engagement11.pdf

Minner, D. D., Levy, A. J., & Century, J. (2009). Inquiry-based science instruction—What is it and does it matter? Results from a research synthesis years 1984 to 2002. Journal of Research in Science Teaching, 47(4), 474–496.

Mullis, I.V.S, Martin, M.O, Ruddock G.J., O'Sullivan, C.Y., & Preuschoff, C. (2009). TIMSS 2011 Assessment Frameworks. TIMSS & PIRLS International Study Center Lynch School of Education, Boston College.

National Science Teacher Association (2012). 2012 NSTA Standards for Science Teacher Preparation. Retrieved from http://www.nsta.org/preservice/docs/2012NSTAPreserviceScienceStandards.pdf

Olson, J.F., Martin, M.O., & Mullis, I.V.S. (Eds.). (2008). TIMSS 2007 Technical Report. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Rivkin, S., Hanushek, E., & Kain, J. (2005). Teachers, schools, and academic achievement. Econometrica, 73, 417–458.

Van Blerkom, Malcolm L. (2009). Measurement and statistics for teachers. New York, Routledge.

Wolf, R. M. (1998). Validity issues in international assessments. *International Journal of Educational Research, 29*, 491 – 501.

Yıldırım, H. H., & Yıldırım, S. (2011). Correlates of communalities as matching variables in differential item functioning analyses. *H. U. Journal of Education*, 40, 386 – 396.

Yoon, K. S., Duncan, T., Lee, S. W. Y, Scarloss, B., & Shapley, K. L. (2007). Reviewing the evidence on how teacher professional development affects student achievement (Institute of Education Sciences Report No. REL 2007-No.033). Washington, DC: U.S. Department of Education.

Yoon, K. S., Duncan, T., Lee, S. W.-Y., Scarloss, B., & Shapley, K. (2007). Reviewing the evidence on how teacher professional development affects student achievement (Issues & Answers Report, REL 2007–No. 033). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest. Retrieved from http://ies.ed.gov/ncee/edlabs

Zumbo, B. D. (2007). Three generations of DIF analyses: considering where it has been, where it is now, and where it is going. *Language Assessment Quarterly, 4*, 223 – 233.