

Evaluation of the Professional Development Program for Secondary Math Teachers on Item Writing Related to Higher Order Thinking Skills

(Received July 12, 2019 - Accepted February 24, 2020)

Bünyamin Yurdakul¹, T. Oğuz Başokçu² and Ümran Yazıcılar³

Abstract

The fact that Turkey's achievement in large-scale international examinations is not at the expected level is associated with such factors as students, teachers, schools, curricula and assessment approaches. The literature links student-related causes mainly to their lack of higher-order thinking skills; and teacher-related causes to incompetency in writing items to measure these skills, and not participating in professional development (PD) activities aimed at measurement of the skills tested in these exams. For these reasons, the present study designed, implemented and evaluated a PD program to improve secondary math teachers' skills of writing higher-order items in compliance with Cognitive Diagnostic Models. The study was conducted with Guskey's five levels of evaluating teachers' PD in the holistic single-case design. The program was attended by 100 participants from 20 different cities of Turkey. During the study, both quantitative and qualitative data concerning the periods before, during and after the program implementations were obtained through pre and post evaluation questionnaires and worksheets. The results indicated that the PD program was satisfactory; however, it needed improvement in terms of the total time of the program, time allocated for activities and monitoring and supporting teachers' work at their institutions of service. In addition, it was revealed that the context should not be structured in writing higher-order items of mathematics, which involves multiple qualities and mathematical competence.

Key Words: Professional development, program evaluation, item writing, higher-order thinking skills, PISA, TIMMS.

Introduction

Mathematical achievement of students participating in large-scale international exams like PISA (Programme for International Student Assessment) and TIMSS (Trends in International Mathematics and Science Study) from Turkey fails to reach the expected level (Bütüner & Güler, 2017; Milli Eğitim Bakanlığı [MEB], 2016a; Mullis, Martin, Foy, & Hooper, 2016). In PISA 2015, for instance, the percentage of students at level 5 or higher (high competency level) was 10.7%, and 8.2% in all countries while it was 2.01% in Turkey. In the same year, the percentage of students

¹ Department of Educational Sciences, Division of Curriculum and Instruction, Ege University Faculty of Education, İzmir, Turkey. Email: bunyamin.yurdakul@ege.edu.tr, ORCID ID: 0000-0001-8401-7202

² Department of Educational Sciences, Division of Measurement and Evaluation in Education, Ege University Faculty of Education, İzmir, Turkey. Email: oguzbasokcu@gmail.com, ORCID ID: 0000-0002-4821-0045

³ Ministry of National Education, İzmir Provincial Directorate of National Education, İzmir, Turkey. Email: umran yazicilar@gmail.com, ORCID ID: 0000-0002-5535-9075

remaining at the low level in OECD was 23.4% whereas it was calculated as 51.3% for Turkey (MEB, 2016a; OECD, 2016). TIMSS 2015 national report addresses that of all the 8th graders who participated in the exam, only 6% had an advanced level of mathematical competency while 30% were at the low level or under (MEB, 2016b). According to Kitchen, Bethell, Fordham, Henderson, and Ruo Chen Li (2019), although an increasing number of students can reach basic levels of numeracy and literacy in PISA, only 1% of the students taking this exam in Turkey show the types of higher-order skills by age 15. This has remained unchanged over the last ten years.

Turkey's failure to attain the desired achievement in international examinations is associated with reasons related to students, teachers, schools, curricula and the assessment approaches employed (Akyüz, 2006; Karip, 2017; MEB, 2016a, 2014). Student-related factors are mainly linked to higher-order thinking skills. For example, it is reported that students participating in PISA from such countries as Finland, Korea, Canada, Australia and New Zealand are better in critical and analytical thinking than students from Turkey. Moreover, students from Turkey cannot sufficiently display higher-order skills such as abstract thinking, being critical, creativity, life-association (Aydın, Erdağ, & Taş, 2011; Aydın, Sarier, & Uysal, 2012). The challenges that teachers' face in writing items to measure higher-order thinking skills can also be considered among the reasons of this failure (Güler, 2013; İnceçam, Demir, & Demir, 2018; Şata, 2016, Thompson, 2008; Yatağan, 2014). Increasing Turkey's mathematical achievement in large-scale international examinations is considered to be closely related with the cognitive level of teacher questions used in teaching processes (Baysen, 2006; Çalık & Aksu, 2018; Sahin & Kulm, 2008) and the quality of teacher-designed tests and measurement instruments used in national examinations (Aygün, Baran-Bulut, & İpek, 2016; Güler, Özdemir, & Dikici, 2012; Karaman & Bindak, 2017; İpek & Özdemir, 2019). For example, Çalık and Aksu (2018) state that teachers and prospective teachers prefer to ask questions at lower thinking levels during teaching. Güler et al. (2012) report that questions that measure students' lower cognitive levels are mainly the focus of both the examinations designed by primary math teachers and of central placement examinations.

Assessment is a major element that enhances students' thinking and develops their skills (Alkhateeb, 2019). The related literature reports that teachers in Turkey lack adequate knowledge and skills of measurement and assessment (Çakan, 2004; Şata, 2016), as well as how to use assessment to measure higher-order thinking competencies and they, are unable to use assessment sufficiently to support learning. This may indicate their need for more quality learning opportunities on assessment. However, PD activities on assessment are limited, and participation is rather low in Turkey (Aydın, Selvitopu, & Kaya, 2018; Kitchen et al., 2019). TIMSS 2015 national report indicates that the ratio of 8th grade math teachers participating in PD activities intended for the improvement of students' critical thinking or problem-solving skills is under 26%, which

remains below the average of countries (45%) in TIMSS 2015 (MEB, 2016b). Because one of the factors affecting student achievement in examinations at international level is teacher quality (Peak et al., 1996; Vlaardingerbroek & Taylor, 2003), and teachers who receive substantial PD can boost their students (Yoon, Duncan, Lee, Scarloss, & Shapley, 2007), these results may point to teachers' needs for PD.

In the literature, PD is defined as “activities that aim to develop an individual’s skills, knowledge, expertise and other characteristics as a teacher” (OECD, 2009, p. 51); and is claimed to have positive effects on students’ learning outcomes and achievement (Loyalka, Popova, Li, & Shi, 2018). The most effective PD programs are reported to be those which are active, reflective, sustainable, and job-embedded (e.g. Darling-Hammond & McLaughlin, 1995; Darling-Hammond & Richardson, 2009; Desimone, Porter, Garet, Yoon, & Birman, 2002; Harris, 2016). PD programs may fail to improve teachers and remain ineffective on students’ learning outcomes for several reasons (Guskey, 1995; Loyalka et al., 2018; Villegas-Reimers, 2003). For instance, PD activities for teachers in Turkey are usually held in the form of face-to-face courses and/or seminars. The rate of active PD activities such as observation visits, peer and/or self-observation and coaching is low (OECD, 2019; TEDMEM, 2019). Thus, the positive effect of PD activities on teaching practices is also at a low level (TEDMEM, 2019). Although there is currently a lack of evidence concerning the effectiveness of PD programs internationally (Loyalka et al., 2018); in Turkey, this can be associated with the fact that PD activities are majorly theoretical as stated in previous studies (e.g. Arslan, 2015; Birgin, Tutak, & Türkdoğan, 2009; Özen, 2006; Sıcak & Parmaksız, 2016). Moreover, Stiggins, Griswold, and Wikelund (1989) point out the fact that teachers had been trained to teach thinking skills to some extent, but they were less often trained to assess those skills. Thomson (2008) states that math teachers have difficulty in determining higher or lower-level thinking skills and creating test items for higher-order thinking. Similarly, according to Driana and Ernawati (2019), teachers have misunderstandings about the higher-order thinking skills items, the items they write fail to reach high levels, which requires them to be trained with appropriate programs.

For all the reasons mentioned above, efficiency of a diagnostic model was tested to improve mathematical achievement of Turkey in large-scale international examinations within the scope of a project carried out with the cooperation of Ege University Faculty of Education, Izmir Provincial Directorate of National Education and TUBITAK (The Scientific and Technological Research Council of Turkey). The project aims to design measurement instruments that can measure 6th grade mathematics subject matter knowledge, together with cognitive processing skills, determine students’ level of possessing this knowledge and skills through effective statistical methods (Cognitive Diagnostic Models [CDM]), and monitor the progress of achievement by providing feedback to teachers and students about the specified learning insufficiencies and

their sources in order to increase Turkey's mathematical achievement in large-scale international examinations.

One of the critical phases of the project was test development. Within the scope of the project, a PD program was designed to set up a team of teachers who wrote items compatible with both the content and higher-order cognitive skills. The PD program was designed to upskill teachers in writing items concerning higher-order thinking skills similar to those used in large-scale international examinations like PISA and TIMSS.

Evaluation of the PD program was essential both to obtain feedback about the quality and the outcomes (Guskey, 2000; Haslam 2010; Mitchem et al., 2003) and to judge its efficiency (Fitzpatrick, Sanders, & Worthen, 2004). Evaluation of PD programs doubtlessly reveals whether the effort put is rewarded or not and provides significant information to make rational and reliable decisions about the process and activities of PD (Guskey, 2000, 2002). Therefore, each program needs to be evaluated (Fitzpatrick et al., 2004; Stufflebeam, Madaus, & Kellaghan, 2000). According to McChesney and Aldridge (2019, p. 308), "to evaluate is not merely to document that PD happened or what it was like; rather, to evaluate is to articulate the outcomes or impacts of PD". The results of the program evaluation allow for identifying the strengths and weaknesses of the PD program, making judgments about its efficiency, and providing feedback for stakeholders about the effectiveness of the program. It also helps teachers to observe their own development and the effects of PD on students' learning outcomes. In this regard, the primary purpose of the present evaluation study was to evaluate the efficiency of the PD program designed and implemented to improve secondary math teachers' skills of writing higher-order items in compliance with CDM. To this end, the following research questions were formulated:

- (1) What are the participants' reactions to the PD program?
- (2) What knowledge and skills were acquired by the participants?
- (3) What do the participants think about organization support and change?
- (4) What do the participants think about using new knowledge and skills?
- (5) What is the participants' use of new knowledge and skills?
- (6) What do the participants think about the contributions of the program to students' learning outcomes?

Methodology

Research design

The study was conducted using Guskey's five levels of evaluating teachers' PD in the evaluation of the program. Guskey (2000, 2002) suggests that the evaluation of PD programs could be performed at five levels as a) participants' reactions, b) participants'

learning, c) organization support and change, d) participants' use of new knowledge and skills and e) student learning outcomes. Each level is built upon the previous one and affects the achievement of the subsequent levels. Instead of obtaining generalizable results from the evaluation of the program, the present study aimed to analyze the program in depth, and to evaluate its efficiency. In this respect, case study research design was considered to be suitable for the evaluation model employed, and it was accordingly decided that the study would be carried out in the holistic single-case design (Yıldırım & Şimşek, 2016; Yin, 2003).

Case

The unit of analysis dealt with in the scope of this case study is the participants' skills to write items intended for higher-order thinking. The case is the PD program designed and implemented for secondary mathematics teachers.

According to Guskey (2000, 2002), an effective PD program must be planned backward. The main purpose of the PD program designed according to Guskey's approach was to improve secondary mathematics teachers' skills to write items to measure higher-order thinking skills in compliance with CDM. To this end, the following objectives were specified: To be able to 1) notice the indicators of higher-order thinking; 2) identify the levels of cognitive demand of math questions; 3) analyse the questions in terms of mathematical processes and competencies; 4) write items of different formats intended for higher-order thinking skills; 5) prepare the partial answer key for an open-ended question related to higher-order thinking skills; 6) recognize the appropriateness of a question for higher-order thinking skills; 7) determine which cognitive qualities were intended to be improved by the written program attainments. The content was chosen to be compatible with the program objectives, and was formed with a linear approach by considering the principle of prerequisite. On the first day of the educational practices, key points required for producing the prerequisite learning were delivered through question-answer, discussion, presentation materials and worksheets as part of expository teaching. On the second day, individual and group work activities were performed which were held in two sessions in three separate halls simultaneously with three different study groups guided by the project members in each hall. Worksheets were used in these activities. Assessment and testing of the program focused on determining the participants' level of achieving the objectives and its effects on the participants. To this end, evaluation questionnaires I and II and 3-point scoring instruction were used as well as Stein and Smith's (1998) Cognitive Demand Level criteria for the evaluation of the cognitive processes involved in the questions produced by the participants.

The PD program was implemented in three-hour morning and afternoon sessions over two days, totalling 12 hours. The implementation was performed by instructors who had previous experience with the topics specified in the program objectives, the

project coordinator and researchers with the necessary assistance of project scholars. A panel was held at the end of the day, in which the project coordinator and researchers participated. Products created by teachers during the process were collected to both select item writers and collect evidence concerning the PD program's level of attaining its objectives. The participants were awarded certificates at the end of the training.

Participants

100 teachers from 20 cities participated in the PD program. 36% of the participants were male and 64% female; 59% work at public schools, 28% at private schools, 9% at science-art centers, and 4% at the central units of the Ministry of National Education [General Directorate of Measurement, Assessment and Testing Services, General Directorate of Innovative and Education Technologies, R&D]. Regarding graduation, 69% of the participants had graduated from Mathematics Education programs, 31% from Mathematics programs and 59% held postgraduate degrees.

The participants had various reasons for taking part in the PD program. The most frequently emphasized reasons include maintaining PD (93.8%, $f=90$), improving their students' higher-order thinking skills (92.7%, $f=89$) and being able to write math items related to higher-order thinking skills (89.6%, $f=86$). It could be interpreted that the PD program managed to reach the appropriate target population and the participants had the necessary affective readiness prior to the implementation of the program.

The participants' knowledge and skills to write items related to higher-order thinking skills could not be measured prior to the program. This is a critical limitation in terms of evaluating the effects of the program on the question samples produced by the participants during program activities. However, the findings obtained from the evaluation questionnaire I may show how the participants produced questions before the program. Although this information does not provide strong evidence, it is still usable in evaluating the new knowledge and skills gained by the participants through the program. The participants mostly use readily available resources in their own teaching processes (73%, $f=73$). For example, the expressions "I pay attention to using foreign resources particularly." of P7, "I use questions on foreign websites." of P35 and "I try to examine and use questions asked in various examinations" of P74 indicate that the participants usually use questions from readily available resources.

The participants' most frequent purposes in item writing are to monitor students' progress (84.4%, $f=81$), to determine student achievement (82.3%, $f=79$), to prepare classroom activities or enrich their activities (81.3%, $f=78$) and to use as an assignment (74%, $f=71$). Considering all the characteristics defining the participants, it is thought that they possessed the necessary qualities to attain the program objectives prior to the PD program.

Data collection instruments

In the current study, triangulation was conducted by using multiple data collection instruments and data sources to increase the validity and reliability of the results (Yıldırım & Şimşek, 2016). In addition, Yin (2003) expresses that qualitative and quantitative data can be used in case studies. With regards to the related literature, triangulation was accepted as the main principle in the present study as well. Thus, the primary data sources consisted of the evaluation questionnaires used before and after the program and the worksheets used by the teachers during the process.

Evaluation questionnaires

The evaluation questionnaires were used before and after the program implementations. The questions in the evaluation questionnaires were designed considering the five levels proposed by Guskey (2000) for PD programs. Draft forms of the questionnaires were prepared after being improved according to the assessments, opinions and recommendations of field experts in PD, measurement and assessment, mathematics education and curriculum development.

Part I of the Evaluation Questionnaire I consisted of six closed-ended items aiming to determine the participants' 1) gender, 2) city of service, 3) professional seniority, 4) reasons for participating in the PD program, 5) the ways they obtain math questions they use or will use, and 6) their purposes in writing questions. Items 4, 5 and 6 allowed the participants to choose more than one option. The data obtained from these questions were used to describe the participants' demographic characteristics. The second part of the questionnaire consisted of 14 statements that aimed to measure the participants' reactions (4), participants' learning (2), organization support and change (3), participants' use of new knowledge and skills (2) and the reflection of the program on student learning outcomes (3) in accordance with Guskey's model. The statements were designed with responses of four degrees.

Evaluation Questionnaire II used at the end of the implementation of the PD program was designed with four degrees of responses and included 37 statements regarding the levels of participants' reactions (14), participants' learning (9), organization support and change (6), participants' use of new knowledge and skills (4) and the reflection of the program on student learning outcomes (4). A separate section was given for the participants to express other opinions and suggestions for each level of the questionnaire.

Worksheets

Transference of the knowledge and skills gained into practice was performed through individual and group activities scheduled on the second day of the PD program. Worksheets were used in these activities. The participants used the knowledge and skills they had gained in the tasks stated in worksheets.

Data analysis

For the analysis of the closed-ended items on the questionnaires frequencies (f) and percentages (%) were used while the four-degree questionnaire items were analysed using means (\bar{X}). Mean scores were calculated using the $(n-1)/n$ formula. Mean scores of the responses given to the related item were interpreted as follows: “strongly disagree” in the 1.00-1.75 range , “disagree” in the 1.76-2.50 range, “agree” in the 2.51-3.25 range and “strongly agree” in the 3.26-4.00 range. For the analysis of the opinions and suggestions section of the Evaluation Questionnaire II, participants were first tagged from 1 to 100 and similar expressions were put together. Similar expressions were grouped according to the levels proposed by Guskey (2000) for the evaluation of PD programs.

The worksheets were analyzed using qualitative descriptive analysis (Yıldırım & Şimşek, 2016). To this end, each worksheet was assessed in the first step by creating a 3-point scoring instruction designed as “0-worksheets left blank or not handed in, 1-worksheets involving defects/deficiencies, 2-proper worksheets filled in according to the instruction”. This assessment was conducted by two researchers. In the second step, the cognitive processes included in the questions produced by the participants in the worksheets were evaluated. This was performed by considering the levels of cognitive demand criteria introduced by Stein and Smith (1998). Cognitive demand level is defined as the type and level of thinking required for fulfilling a certain learning task successfully (Stein, Smith, Henningsen, & Silver, 2000). These are classified at four levels as 1) Basic cognitive level/memorization, 2) procedures without connections, 3) procedures with connections and 4) doing mathematics. The levels of basic cognitive level/memorization and procedures without connections can be considered as the lower-level cognitive demand, and procedures with connections and doing mathematics as the higher-level cognitive demand (Stein & Smith, 1998). The questions produced by the teachers in the process were defined considering the given levels, and later the questions produced for each learning task were tabulated using descriptive statistics of frequencies (f) and percentages (%).

Findings

Participants’ reactions to the PD program

Prior to the implementation of the program, the participants (n=96) took part in the PD program voluntarily ($\bar{X}=3.94$), and they stated that the program was well-designed ($\bar{X}=3.36$), the date of the program was appropriate ($\bar{X}=3.53$) and it met their expectations ($\bar{X}=3.46$).

At the end of the program implementation, the participants (n=99) think that the greeting and registration process was organized well ($\bar{X}=3.65$), program objectives were explained clearly ($\bar{X}=3.53$), the content fulfilled a critical need ($\bar{X}=3.37$) and was suitable for transference to professional life ($\bar{X}=3.40$), the process went on to

serve for the program objectives ($\bar{X}=3.36$), qualities intended to be improved were assessed using appropriate techniques and instruments ($\bar{X}=3.25$), project members helped when needed ($\bar{X}=3.72$) and the environmental conditions of the program implementation were appropriate ($\bar{X}=3.62$). Participants' opinions expressed in the open-ended item of the data collection instrument ($n=23$) at the end of the program also support the quantitative findings:

“I thank the instructors for providing us with a positive atmosphere and their contributions. I think my perspective on questions has changed...” P20

The participants explained that the PD program attained its goals, project members created a positive atmosphere, the content was instructive and useful, and that they had gained different perspectives in examining math questions related to higher-order thinking skills.

The findings obtained from the implementation show that expectations of the participants ($n=99$) of the PD program were met to a large extent ($\bar{X}=3.19$), yet there were still some deficiencies. In fact, the time allocated for worksheets, activities and various learning tasks was seen inadequate by the participants ($\bar{X}=2.97$) and the program was criticized for its total time ($\bar{X}=2.59$). This may have affected the benefit expected out of the PD. Participants' ($n=31$) opinions expressed in their responses to the open-ended item of the data collection instrument at the end of the program also support the quantitative finding that the time allocated for learning tasks and the total time of the program was insufficient. Participants thought the time was insufficient and based the evidence for this on reasons such as the fact that writing items related to higher-order thinking skills took time, many factors need to be considered when writing higher-order items, and time pressure caused them to act hastily. For them, this fact affected the efficiency of the program negatively. Insufficiency of the time spared for learning tasks occurred as an obstructive factor for the participants to transfer theoretical information into practice. The suggestion that was most frequently brought up by the participants in this respect appeared to be increasing the total time of the program as well as the time allocated for activities:

“The training we received through this program was very productive. But...more time is needed to write higher-order items.” P2

“...more time should be spared. Because there are many things to consider when writing an item.” P15

“There was a problem with time insufficiency in the activities performed. The time of the program was inadequate. Training programs of longer periods would bring about more efficient outcomes.” P17

“...The new concepts we learned were explained well. However, time was not enough to comprehend them completely. As they are new concepts to us.” P34

The fact that the participants found the total time of the program and the time allocated for activities insufficient and that they suggested increasing the time can be considered as their positive emotional reactions to the PD program. On the other hand, some participants (n=11) recommended that certain aspects of the program could be improved through the following: analysing sample questions; increasing the amount of item-writing practice; analysing more different questions; expanding the question-answer time in sample question analyses; establishing the theory-practice connection immediately rather than saving the first day for theory and the second for practice; providing more effective individual feedback on the items; and using the questions prepared by the participants as examples in the activities. Nevertheless, the high acceptance of the statement “I believe this program should be expanded” ($\bar{X}=3.82$) by the participants may indicate that their responses to the PD program are positive. In fact, some opinions expressed by the participants (n=20) in response to the open-ended item of the data collection instrument at the end of the program revealed the expectation to implement the program again and to make it a regular event and expand it, which supports the quantitative data:

“I think the training should be repeated outside İzmir and under appropriate.” P14

“We could be made more knowledgeable by providing more programs of this type.” P21

“I really want such programs to be more common.” P99

“...[PD program] could be made more comprehensive through improvements and larger masses can be reached ...” P12

The quotations above can be seen as evidence for the participants’ positive reactions to the program.

Knowledge and skills gained by the participants

Prior to the implementation of the program, the participants (n=96) stated that they could write items related to higher-order thinking skills ($\bar{X}=3.37$), but they did not already have the skills ($\bar{X}=2.66$); while they (n=99) expressed that they were able to write items related to higher-order thinking skills ($\bar{X}=3.22$) at the end of the program. The participants reported that they were able to write items related to higher-order thinking skills in different formats ($\bar{X}=3.21$) at the end of the program implementation. Based on these findings, it can be interpreted that the participants’ expectations before the program were met. Also, the participants expressed they were able to prepare the partial answer key for an open-ended question related to higher-order thinking skills ($\bar{X}=3.05$) as well. However, at the end of the program, the participants most frequently stated that they noticed the indicators of higher-order thinking ($\bar{X}=3.43$); they could recognize the appropriateness of a question for higher-order think-

ing skills ($\bar{X}=3.41$); they could determine which cognitive qualities are intended to be improved by the written program attainments ($\bar{X}=3.42$), and they were able to analyse the questions in terms of mathematical processes and competencies ($\bar{X}=3.31$). The findings could be considered to indicate that the participants made significant progress in attaining the program objectives. They also expressed that they could identify the levels of cognitive demand for math questions ($\bar{X}=3.19$). This finding obtained from the questionnaire is consistent with the achievement level attained in the activity held on the second day of the training to determine the level of cognitive demand for math questions. In this activity, the participants were given 14 sample questions involving different levels of cognitive demand for a solution. They were asked to identify the level of cognitive demand of the questions using the information they acquired. The achievement attained by the participants in this activity was 73.7%. It was determined that achievement was higher in identifying the cognitive demand levels of questions of basic cognitive level/memorization (97.9%) and procedures without connections (82.3%) in particular.

Participants' opinions about organization support and change

Before the implementation of the program, the participants' (n=96) opinions concerning organizational support and change showed that they were supported by school administrators ($\bar{X}=3.43$), and their group and other colleagues ($\bar{X}=3.31$) to participate in the PD program.

At the end of the program (n=99), there was a decrease ($\bar{X}=3.12$) in the participants' initial rate of agreement ($\bar{X}=3.32$) with the statement that they planned to share the knowledge, skills and attitudes they acquired with their groups or other colleagues in their institutions through an organization (seminar, workshop, sharing meeting, activity, analysis of the written items in terms of higher-order thinking etc.). The participants reported that their administrators would support them in using the knowledge and skills they acquired ($\bar{X}=3.34$) and that their groups and other colleagues would be in cooperation in writing items related to higher-order thinking skills ($\bar{X}=3.35$). They agreed with the statement that the knowledge and skills they acquired could create a change in the institution with a mean of 3.34. The findings indicate that the participants held positive opinions about organization support and change. On the other hand, the statement "In my future studies on measuring higher-order thinking skills, I would like to get support from the project members" was agreed by the participants with the highest mean ($\bar{X}=3.71$). This finding may suggest that they needed assistance in using the knowledge and skills they learned.

Participants' opinions about the use of new knowledge and skills

Before the program, the participants (n=96) stated they would be able to use the questions ($\bar{X}=3.52$) they would write with the new knowledge and skills they assume

to have acquired at their schools, institutions or the documents they would design etc. ($\bar{X}=3.60$). This may be an indicator that the participants were affectively ready before the program implementation. These rates remained unchanged after the implementation of the program. As a matter of fact, at the end of the program, the participants ($n=99$) think that they will be able to use the new knowledge and skills they assume to have acquired when writing questions ($\bar{X}=3.43$). These findings are consistent with those concerning the knowledge and skills that the participants believed they had acquired. The fact that they completely agreed with the statements “I can reflect what I learned onto the learning-teaching activities I will design” ($\bar{X}=3.51$), “I would like to ask questions that measure higher-order skills to my students more often.” ($\bar{X}=3.50$) and “I would like to spare more time on writing questions related to higher-order thinking skills.” ($\bar{X}=3.64$) may indicate that they are ready to use the new knowledge and skills, and the PD program had a significant affective impact on them; that is, the training was encouraging.

Participants’ use of new knowledge and skills

The second day activities as part of the training program are intended for the use of the information obtained on the first day. Although the performance exhibited on the worksheets does not provide information on the usability of the new knowledge and skills in the participants’ professional life, it may be significant in terms of showing whether the knowledge acquired on the first day was put into practice or not.

The activity intended for the program objective “to be able to prepare the partial answer key for an open-ended question related to higher-order thinking skills” was conducted in two separate parts. The first one involved writing an open-ended question related to an unstructured context (visual material/picture) and preparing its partial answer key on the given chart. In the second part of the activity, the participants were asked to prepare the partial answer key for a sample open-ended question taken from a large-scale international exam. Of the products obtained from the activities ($f=200$); 41% ($f=82$) were completed paying attention to the instruction; 7.5% ($f=15$) had defects/deficiencies. 51.5% ($f=103$) of the worksheets were either left blank or not handed in. This finding may indicate that the participants did not make the desired progress in attaining the concerning program objective. It could be because these activities were performed during the last sessions of the program.

In the two separate activities designed according to the program objective “to be able to analyse the questions in terms of mathematical processes and competencies”, the participants first noted down the attainments, cognitive level and mathematical competences of the question they were going to write. Using these specifications, they later produced open-ended questions in one activity and multiple-choice questions in another. In the task of writing open-ended questions, 73% of the participants completed the activity following the instruction. However, 7% produced work contain-

ing defects/ deficiencies in terms of determining cognitive processes or mathematical competences. 20% of the participants either left the given task blank or did not hand it in. As seen in Table 1, more than half (54.8%) of the participants' open-ended questions involving program attainments, cognitive demand levels and mathematical competencies were found to be at the level of procedures without connections.

Table 1.

Distribution of Participants' Questions Involving Multiple Qualities and Mathematical Competency by the Level of Cognitive Demand

<i>Level of Cognitive Demand</i>	Open-ended (n=80)		Multiple choice (n=85)		Total	
	f	%	f	%	f	%
1. Basic cognitive level / Memorization	3	3.2	5	5.6	8	4.4
2. Procedures without connections	51	54.8	56	62.9	107	58.8
3. Procedures with connections	33	35.5	26	29.2	59	32.4
4. Doing mathematics	6	6.5	2	2.3	8	4.4
Total	93	100	89	100	182	100

In the multiple-choice question writing task, 84% of the participants fulfilled proper work according to the instructions, while 1% produced defective work in terms of identifying cognitive processes or mathematical competencies. 15% of the participants either left the given task incomplete or did not hand it in. As presented in Table 1, 62.9% of the multiple-choice questions are at the level of procedures without connections. 42% of the open-ended questions and 31.5% of the multiple-choice questions written were at the procedures with connections and doing mathematics level. Higher-order questions involving attainments, cognitive processes or mathematical competencies were written in both question types, and that the participants were more successful at writing higher-order open-ended questions compared with writing multiple-choice questions. The reason for the failure to write an adequate number of questions related to higher-order thinking skills may be the insufficiency of the time allocated for the activities as expressed in participants' reactions.

In three different activities intended for writing questions related to higher-order thinking skills in different formats, 67% of the multiple-choice questions written by the participants concerning a limited context (structured-graphical display) stayed at the basic cognitive level/ memorization as shown in Table 2. Of the multiple-choice questions written based on an unstructured context (visual material/ picture), 66.7% are at the level of procedures without connections and 33.3% at procedures with connections. 72.1% of the multiple-choice questions written through reduction or deri-

vation from the context in a given authentic scenario (semi-structured) were at the procedures without connections level, and 26.2% of them were at the procedures with connections level.

Table 2.

Distribution of Participants' Questions Concerning the Given Context (Figure, Diagram, Visual material, Picture, Scenario...) by the Level of Cognitive Demand

Level of Cognitive Demand	Structured Graphical (n=91)		Semi-structured (reduction/derivation from context) (n=94)				Unstructured Visual material/Picture (n=88)				Total			
	Multiple choice		Open-ended		Multiple choice		True/False		Open-ended		Multiple choice		f	%
	f	%	f	%	f	%	f	%	f	%	f	%		
1	61	67.0	8	5.5	1	1.6	14	19.7	9	10.5	-	-	93	20.3
2	14	15.4	89	61.0	44	72.1	40	56.3	31	36.0	2	66.7	220	48.0
B	75	82.4	97	66.5	45	73.8	54	76.1	40	46.5	2	66.7	313	68.3
3	15	16.5	40	27.4	16	26.2	17	23.9	38	44.2	1	33.3	127	27.7
4	1	1.1	9	6.2	-	-	-	-	8	9.3	-	-	18	3.9
H	16	17.6	49	33.6	16	26.2	17	23.9	46	53.5	1	33.3	145	31.6
T	91	100	146	100	61	100	71	100	86	100	3	100	458	100

1. Basic cognitive level / Memorization, 2. Procedures without connections,

3. Procedures with connections, 4. Doing mathematics

B. Basic Level (Total), H. Higher-level (Total), T. Total

As seen in Table 2, of the open-ended questions written through reduction or derivation from the context in a given authentic scenario (semi-structured), 33.6% are at the procedures with connections (27.4%) and doing mathematics (6.2%) levels. 53.5% of the open-ended questions written based on an unstructured context (visual material/picture) are at the procedures with connections (44.2%) and doing mathematics (9.3%) levels. This may indicate that teachers are able to write more higher-order questions when they use unstructured contexts in writing open-ended questions. Moreover, it was seen that teachers could also write questions of true/false format when reduction or derivation is possible from the context in an authentic problem scenario (semi-structured). However, they did not tend to use this question type in unstructured contexts. Approximately one fourth (23.9%) of the true/false questions are at the level of procedures with connections. Regardless of the question type, 31.6% of the total questions written (f=458) are at the procedures with connections and doing mathematics levels.

Participants' Opinions Concerning the Program's Contribution to Students' Learning Outcomes

According to the participants' opinions before the implementation of the program, the PD program would affect students' mathematical achievement positively

($X_{pre}^- = 3.65$; $X_{post}^- = 3.63$), improve their thinking skills ($X_{pre}^- = 3.57$; $X_{post}^- = 3.66$) and motivate them better for the course ($X_{pre}^- = 3.51$; $X_{post}^- = 3.48$). No change was seen in the participants' opinions at the end of the program implementation. Therefore, the PD program may have positive cognitive and affective reflections on students. Some opinions of the participants ($n=11$) expressed in response to the open-ended item of the data collection instrument at the end of the program show that the PD program may be beneficial to their students as well as their professional and personal development, which supports the quantitative findings:

"I would like to work in this field, improve myself and make more effective contributions to my students' mathematical experiences by sharing what I learn with them." P1

"I believe the program was very beneficial for our professional training... I'm so glad that we learned methods of analysing math questions related to higher-order thinking skills and looked at the questions from a different viewpoint..." P57

Discussion and Conclusion

The current study aimed to evaluate the efficiency of the PD program designed and implemented to write math questions measuring higher-order thinking levels in accordance with CDM from the perspective of the participants. Findings showed that while the design, implementation and evaluation of the PD program were sufficient, it could be expanded. It is important to note that participation in the study was on a voluntary basis. According to the literature, teachers' willingness is crucial for the success of PD programs (Smith, Hofer, Gillespie, Solomon, & Rowe, 2003). Also, when they believe in the necessity and benefits of a PD program, teachers may act more willing to participate (Kazu & Kerimgil, 2008). Teachers' willingness before the program continued at the end of the program in the present study. This result may have contributed to the PD program in attaining its objectives.

Another factor for the success of PD programs is being consistent with participants' needs (Bayrakçı, 2009; Çiftçi 2008; Demirkol, 2010; Özer, 2004). In the present study, participants' main reasons for taking part in the PD program include improving students' higher-order thinking skills and being able to write items related to higher-order thinking skills. These reasons may show that the program was designed for teachers' PD needs.

Another finding is that the program objectives are appropriate for participants' needs, and the content is transferrable into professional life. This aspect of the program may have increased willingness. This conclusion supports those studies (Bümen, Ateş, Çakar, Ural, & Acar, 2012; Kazu & Kerimgil, 2008; Özer, 2004) that report an increase in willingness to participate when the participants' prior knowledge, interests

and needs are accounted for, and when they believe in the necessity of benefits of the PD program.

The participants reported that the total time of the program and the time allocated for learning tasks was insufficient for the following reasons: the time required for writing higher-order items, considering many factors when writing higher-order items and time pressure leading to acting quickly. This hindered the transference of theoretical knowledge into practice. The conclusion can be seen compatible with Desimone and others' (2002) opinions that it takes time for teachers to learn and change. The result revealed the necessity for the improvement of both the total time of the program and the time allocated for activities. The literature reports that while dealing with heavy content in insufficient periods may raise awareness of the objectives of the PD programs it could adversely affect the permanence and transference of teachers' learning into practice (Yurdakul, Uslu, Çakar, & Yıldız, 2014). Although the total time of the program and the time allocated for activities were criticized, the participants agreed with the statement concerning the need for making the program more widespread. This result shows that the program can be made continuous with some improvements.

Considering the knowledge and skills acquired, the participants made progress in the following areas: noticing the indicators of higher-order thinking; identifying the cognitive level of math questions; analysing the questions in terms of mathematical processes and competencies; writing items of different formats related to higher-order thinking skills; recognizing the appropriateness of a question for higher-order thinking skills and determining which cognitive qualities are intended to be improved by the written program attainments. In program studies, desired qualities can be gained when open, clear and attainable objectives are set and appropriate instruction is designed accordingly (Demirel, 2012; Ornstein & Hunkins, 2009). In this respect, it could be asserted that the teaching conditions of the program were adequate to facilitate the attainment of the desired objectives. However, the participants did not make sufficient progress in reaching the objective "to be able to prepare the partial answer key for an open-ended question related to higher-order thinking skills".

The participants reported being supported by their school administrators, branch groups and other colleagues in terms of organization support and change. This finding is consistent with studies that consider organization culture and administrative support important for the reflection of the targeted change onto the classroom through PD programs (Bümen et al., 2012; Opfer & Pedder, 2011; Özer, 2004). On the other hand, the participants held positive opinions about sharing the acquired knowledge, skills and attitudes in their organizations and the possible change to be created in their organizations through this sharing. However, a need arose to get help and support from the program conductors at the end of the PD program. The results imply that monitoring and providing support for participants should be an essential part of PD programs (Bayrakçı, 2009; Çiftçi, 2008; Demirkol, 2010; Özer, 2004). However, since the par-

ticipants of the present PD program were from different cities, it was not possible to monitor them without removing them from their own contexts to give feedback. For this reason, information concerning the participants' use of the new knowledge and skills and the reflection of the PD program onto students' learning outcomes could be obtained through participants' opinions. Therefore, it is considered that the PD program can be improved by adding the stages of monitoring, providing feedback and evaluating; and to this end, an interactive web portal could be designed where the participants can share the questions they write to measure higher-order thinking skills, receive effective feedback and get support when needed.

Another finding was that the participants held positive opinions about using the new knowledge and skills they acquired, which is consistent with the results of previous studies (e.g. Gültekin & Çubukçu, 2008; Özen, 2006; Tataroğlu Taşdan, & Çelik, 2014). Besides, they thought that the PD program would be reflected positively on students' learning. The results may indicate that the PD program had an important affective impact on the participants; that is, it was encouraging for them.

Another result was that the participants were able to write higher-order items involving attainments, cognitive processes or mathematical competencies in different formats and they are better at writing higher-order open-ended items compared to writing questions in other formats. This conclusion is consistent with the studies reporting that it is possible to improve teachers' skills of writing items to measure higher-order thinking skills when PD programs are organized (Aslan, 2011; Demir, Öztürk, & Dökme, 2011). Similarly, Zohar and Schwartz (2011) found that teachers used tasks requiring higher-order thinking skills more often and students' participation in metacognitive thinking was high through the course observations of 14 teachers who took part in PD activities on teaching higher-order thinking. Also, Barak and Dori (2009) developed a hybrid course for in-service teachers in a scope which bringing face-to-face classroom discussions together with online activities, interrelating teaching, learning, and assessment. The researchers found that teachers improved in posing complex questions, introducing consistent arguments and demonstrating critical thinking, and, thus, they enhanced their teaching capacity with the hybrid training they developed. In the present study, more than half the questions were produced at the higher-order in the open-ended question type in unstructured contexts. The results revealed that the context should not be structured in writing higher-order math questions involving more than one quality and mathematical competency. By nature of higher-order thinking (Lewis & Smith, 1993), producing questions in a structured or limited context may decrease the levels of cognitive demand of the questions written by participants.

The related literature reports that teachers in Turkey are insufficiently informed about large-scale international examinations, highlighting that PD activities should be organized to guide them in writing inference and application questions to increase

achievement in these exams (Güner, Sezer, & Akkuş İspir, 2013). Moreover, in 2018, Turkey published a new vision to transform its education system (MEB, 2018). According to this vision, assessment will be adapted to focus less on scores and examination results, and more on monitoring the development a student's competencies, using the information to guide their future decisions and pathways by providing in-service teacher training to improve teachers' assessment skills and restructuring national examinations to prioritise the assessment of higher-order skills like reasoning, critical thinking and interpretation (Kitchen et al., 2019). The results of the present study reveal that the evaluated PD program was designed and implemented to meet this need. On the other hand, many PD programs conducted in Turkey are not evaluated sufficiently (Bümen et al., 2012; Demirkol, 2010; Uslu, 2013). In this respect, the current PD program is critical both in terms of subject matter, and the identification of the components that need improvement. The present study is considered to possess qualities that can guide future PD programs to be designed on similar topics and their evaluation.

Acknowledgements

This study was produced from a research project (#115K531) granted by the Scientific and Technological Research Council of Turkey (TUBİTAK) and presented at the 5th International Conference on Curriculum and Instruction in Muğla, Turkey, October 26-28, 2017.

References

- Akyüz, G. (2006). Investigation of the effect of teacher and class characteristics on mathematics achievement in Turkey and European Union countries. *Elementary Education Online*, 5(2), 75-86.
- Alkhateeb, M. A. (2019). Assessing eighth-grade mathematics teachers and textbook in embodying thinking levels. *International Journal of Instruction*, 12(1), 371-388.
- Aslan, C. (2011). Effects of teaching applications for developing question asking skills on question forming skills of prospective teachers. *Education and Science*, 36(160), 236-249.
- Aslan, M. (2015). *Türkiye milli eğitim sistemindeki değişmeler ışığında hizmet içi eğitim*. (Yayımlanmamış Yüksek Lisans Tezi). İstanbul Aydın Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.
- Aydın, A., Selvitopu, A., & Kaya, M. (2018). Resources invested in education and PISA 2015 results: a comparative study. *Elementary Education Online*, 17(3), 1283-1301. doi:10.17051/ilkonline.2018.466346
- Aydın, A., Sarier, Y., & Uysal, Ş. (2012). The comparative assessment of the results of

- PISA mathematical literacy in terms of socio-economic and socio-cultural variables. *Education and Science*, 37(164), 20-30.
- Aydın, A., Erdağ, C., & Taş, N. (2011). A comparative evaluation of PISA 2003-2006 results in reading literacy skills: An example of top-five OECD countries and Turkey. *Educational Sciences: Theory and Practice*, 11(2), 665-673.
- Aygün, B., Baran-Bulut, D., & İpek, A. S. (2016). The analysis of the primary school mathematics exam questions according to the math taxonomy. *Turkish Journal of Computer and Mathematics Education*, 7(1), 62-88.
- Barak, M. & Dori, Y. J. (2009) Enhancing higher order thinking skills among inservice science teachers via embedded assessment. *Journal of Science Teacher Education*, 20(5), 459-474. doi: 10.1007/s10972-009-9141-z
- Bayrakçı, M. (2009). In-service teacher training in Japan and Turkey: A comparative analysis of institutions and practices. *Australian Journal of Teacher Education*, 34(1), 10-22. doi:10.14221/ajte.2009v34n1.2
- Baysen, E. (2006). The levels of teacher questions and student answers. *Kastamonu Education Journal*, 14(1), 21-28.
- Birgin, O., Tutak, T., & Türkdoğan, A. (2009). Primary school teachers' views about the new Turkish primary school mathematics curriculum. *e-Journal of New World Sciences Academy*, 4(2), 270-280. Retrieved from: <https://files.eric.ed.gov/fulltext/ED505000.pdf>
- Bümen, N. T., Ateş, A., Çakar, E., Ural, G. & Acar, V. (2012). Türkiye bağlamında öğretmenlerin mesleki gelişimi: sorunlar ve öneriler. *Milli Eğitim Dergisi*, 41(194), 31-50.
- Bütüner, S. Ö. & Güler, M. (2017). Gerçeklerle yüzleşme: Türkiye'nin TIMSS matematik başarısı üzerine bir çalışma. *Bayburt Eğitim Fakültesi Dergisi*, 12(23), 161-184.
- Çakan, M. (2004). Comparison of elementary and secondary school teachers in terms of their assessment practices and perceptions toward their qualification levels. *Ankara University Journal of Faculty of Educational Sciences*, 37(2), 99-114. doi:10.1501/Egifak_0000000101
- Çalık, B. & Aksu, M. (2018). A systematic review of teachers' questioning in Turkey between 2000-2018. *Elementary Education Online*, 17(3), 1548-1565. doi:10.17051/ilkonline.2018.466389
- Çiftçi, E. (2008). *Türkiye'de Milli Eğitim Bakanlığı tarafından müzik öğretmenlerine verilen hizmetiçi eğitimin incelenmesi ve müzik öğretmenlerinin hizmetiçi eğitim ihtiyaçlarının belirlenmesi*. (Yayımlanmamış Doktora Tezi). Gazi Üniversitesi, Ankara.
- Darling-Hammond, L. & McLaughlin, M. W. (1995). Policies that support professional development in an era of reform. *Phi Delta Kappan*, 76(8), 597-604.
- Darling-Hammond, L. & Richardson, N. (2009, February). Teacher learning: What

- matters? *Educational Leadership*, 66(5), 46–55.
- Demir, R., Öztürk, N., & Dökme, İ. (2011). The views of the teachers taking in-service training about alternative measurement and evaluation techniques: The sample of primary school teachers. *Procedia Social and Behavioral Sciences*, 15, 2347–2352. doi: 10.1016/j.sbspro.2011.04.105
- Demirel, Ö. (2012). *Eğitimde program geliştirme: Kuramdan uygulamaya*. Ankara: Pegem Akademi.
- Demirkol, M. (2010). İlköğretim okullarında öğretmenlere yönelik okul-temelli hizmet içi eğitim etkinliklerinin değerlendirilmesi. *Milli Eğitim Dergisi*, 188, 158-173.
- Desimone, L., Porter, A., Garet, M., Yoon, K., & Birman, B. (2002). Effects of professional development on teachers' instruction: Results from a three-year longitudinal study. *Education Evaluation and Policy Analysis*, 24(2), 81-112.
- Driana, E. & Ernawati, E. (2019). Teachers' understanding and practices in assessing higher order thinking skills at primary schools. *Journal of Teaching & Education*, 1(2), 110-118. Retrieved from: <https://journals.umkt.ac.id/index.php/acitya/article/view/233>
- Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2004). *Program evaluation: alternative approaches and practical guidelines*. Boston: Allyn and Bacon.
- Guskey, T. R. (1995). Professional development in education: In search of the optimal mix. In T. R. Guskey and M. Huberman (Eds), *Professional development in education: New paradigms and practices*. (pp. 114-131) New York: Teachers College Press.
- Guskey, T. R. (2000). *Evaluating professional development*. Thousand Oaks, CA: Corwin Press.
- Guskey, T. R. (2002). Does it make a difference? Evaluating professional development. *Educational Leadership*, 59(6), 45–51.
- Güler, H. K. (2013). Türk öğrencilerin PISA'da karşılaştıkları güçlüklerin analizi *Ulu- dağ Üniversitesi Eğitim Fakültesi Dergisi*, 26(2), 501-522.
- Güler, G., Özdemir, E., & Dikici, R. (2012). İlköğretim matematik öğretmenlerinin sınav soruları ile SBS matematik sorularının Bloom taksonomosine göre karşılaştırmalı analizi. *Erzincan Üniversitesi Eğitim Fakültesi Dergisi*, 14(1), 41-60.
- Güner, N., Sezer, R., & Akkuş İspir, O. (2013). İlköğretim ikinci kademe öğretmenlerinin TIMSS hakkındaki görüşleri. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, 33(1), 11-29.
- Gültekin, M. & Çubukcu, Z. (2008). İlköğretim öğretmenlerinin hizmet içi eğitime ilişkin görüşleri. *Sosyal Bilimler Dergisi*, (19), 185-201. Retrieved from: http://journals.manas.edu.kg/mjsr/oldarchives/Vol10_Issue19_2008/585-1576-1-PB.pdf
- Harris, J. (2016). Inservice teachers' TPACK development: Trends, models, and trajectories. In M. Herring, M. Koehler, & P. Mishra (Eds.), *Handbook of technological pedagogical content knowledge for educators* (2nd ed.) (pp. 191-205). New York,

NY: Routledge.

- Haslam, M. B. (2010). *Teacher professional development evaluation guide*. Washington, DC: National Staff Development Council.
- İnceçam, B., Demir, E., & Demir, E. (2018). Competencies of middle school teachers to prepare open-ended items used in open-ended test for in-classroom assessment. *Elementary Education Online*, 17(4), 1912-1927. doi:10.17051/ilkonline.2019.506900
- İpek, A. S. & Özdemir, E. (2019). Ortaokul matematik öğretmenlerinin yazılı sınav sorularının incelenmesi. *Akdeniz Eğitim Araştırmaları Dergisi*, 13(27), 244-262. doi:10.29329/mjer.2019.185.11
- Karaman, M. & Bindak, R. (2017). İlköğretim matematik öğretmenlerinin sınav soruları ile TEOG matematik sorularının yenilenmiş Bloom Taksonomisi'ne göre analizi. *Current Research in Education*, 3(2), 51-65.
- Karip, E. (2017). *Türkiye'nin TIMSS 2015 performansı üzerine değerlendirme ve öneriler* (TEDMEM Analiz Dizisi 5). Ankara: Türk Eğitim Derneği Yayınları.
- Kazu, İ. Y. & Kerimgil, S. (2008). Yeni atanan öğretmenlerin hizmetçi eğitime ilişkin görüşleri (Elazığ ili örneği). *Gazi Üniversitesi Endüstriyel Sanatlar Eğitim Fakültesi Dergisi*, 23, 14-30.
- Kitchen, H., Bethell, G., Fordham, E., Henderson, K. & Ruochoen Li, R. (2019). *OECD Reviews of Evaluation and Assessment in Education: Student Assessment in Turkey*, OECD Reviews of Evaluation and Assessment in Education, OECD Publishing, Paris, <https://doi.org/10.1787/5edc0abe-en>.
- Lewis, A. & Smith, D. (1993). Defining higher order thinking. *Theory into practice*, 32(3), 131-137. doi:10.1080/00405849309543588
- Loyalka, P., Popova, A., Li, G. & Shi, Z. (2018). Does teacher training actually work? Evidence from a large-scale randomized evaluation of a national teacher training program. Rural Education Action Program (REAP), Working Paper 330. Retrieved from: https://fsi-live.s3.us-west-1.amazonaws.com/s3fs-public/does_teacher_training_work.pdf
- Mitchem, K., Wells, D., & Wells, J. (2003). Using evaluation to ensure quality professional development in rural schools. *Journal of research in rural education*, 18(2), 96-103.
- MEB (2018). *2023 eğitim vizyonu*. Retrieved from: http://2023vizyonu.meb.gov.tr/doc/2023_EGITIM_VIZYONU.pdf
- MEB (2016a). *Uluslararası öğrenci değerlendirme programı PISA 2015 ulusal raporu*. Ankara: Ölçme Değerlendirme ve Sınav Hizmetleri Genel Müdürlüğü. Retrieved from: http://pisa.meb.gov.tr/?page_id=22
- MEB (2016b). *TIMSS 2015 ulusal matematik ve fen bilimleri ön raporu 4. ve 8. sınıflar*. Ankara: Ölçme, Değerlendirme ve Sınav Hizmetleri Genel Müdürlüğü. Retrieved from: https://odsgm.meb.gov.tr/meb_iys_dosyalar/2017_06/23161945_

timss_2015_on_raporu.pdf

- MEB (2014). *TIMSS 2011 Ulusal Matematik ve Fen Raporu: 8. Sınıflar*. Ankara: Yenilik ve Eğitim Teknolojileri Genel Müdürlüğü. Retrieved from: <http://odsgm.meb.gov.tr/test/analizler/docs/timss/TIMSS-2011-8-Sinif%20Raporu.pdf>
- Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2016). *TIMSS 2015 International Results in Mathematics*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: <http://timssandpirls.bc.edu/timss2015/international-results/>
- OECD (2009). Creating effective teaching and learning environments: First results from TALIS Retrieved from: <http://www.oecd.org/education/school/43023606.pdf>
- OECD (2016). *PISA 2015 results (Volume I): Excellence and equity in education*. Paris: PISA, OECD Publishing. Retrieved from: <http://dx.doi.org/10.1787/9789264266490-en>
- OECD (2019). *TALIS 2018 results (Volume I): Teachers and school leaders as lifelong learners*, TALIS, Paris: OECD Publishing. Retrieved from: <https://doi.org/10.1787/1d0bc92a-en>.
- Opfer, V. D. & Pedder, D. (2011). Conceptualizing teacher professional learning. *Review of Educational Research*, 81(3), 376-407. doi:10.3102/0034654311413609
- Ornstein, A. C. & Hunkins, F. P. (2009). *Curriculum: foundations, principles and issues*. US: Pearson.
- Özen, R. (2006). İlköğretim okulu öğretmenlerinin hizmet içi eğitim programlarının etkilerine ilişkin görüşleri: Düzce ili örneği. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 6(2), 141-160.
- Özer, B. (2004). In-service training of teachers in Turkey at the beginning of the 2000s. *Journal of In-Service Education*, 30(1), 89-100. doi:10.1080/13674580400200238
- Peak, L., et al. (1996). Pursuing excellence: a study of U.S. eighth-grade mathematics and science teaching, learning, curriculum, and achievement in international context. Initial findings from the third international mathematics and science study. Washington, DC.: Office of Educational Research and Improvement (ED). Retrieved from: <https://files.eric.ed.gov/fulltext/ED400209.pdf>
- Sahin, A. & Kulm, G. (2008). Sixth grades mathematics teachers' intentions and use of probing, guiding and factual questions. *Journal of Mathematics Teacher Education*, 11(3), 221-241. doi:10.1007/s10857-008-9071-2
- Sıcak, A. & Parmaksız, R. Ş. (2016). An evaluation of efficacy of in-service training in primary schools, *The Inonu University Journal of the Faculty of Education*, 17(1), 17-33. doi: 10.17679/iuefd.17144668
- Smith, C. Hofer, J., Gillespie, M., Solomon, M. & Rowe, K. (2003). How teachers change: A study of Professional development in adult education (Report 25). Cambridge, MA: National Center for the Study of Adult Learning and Literacy.

- Retrieved from: <https://files.eric.ed.gov/fulltext/ED508607.pdf>
- Stein, M. K., Smith, M. S., Henningsen, M. A., & Silver, E. S. (2000). *Implementing standards-based mathematics instructions: a casebook for professional development*. New York, NY: Teachers College Press.
- Stein, M. K. & Smith, M. S. (1998). Mathematical Tasks as a Framework for Reflection: From Research to Practice. *Mathematics Teaching in the Middle School*, 3(January), 268–75.
- Stiggins, R.J., Griswold, M. M., & Wikelund, K. R. (1989). Measuring thinking skills through classroom assessment. *Journal of Educational Measurement*, 26, 233–246.
- Stufflebeam, D. L., Madaus, G. F., & Kellaghan, T. (2000). *Evaluation models: viewpoints on educational and human services evaluation*. 2nd ed. Boston: Kluwer Academic Publishers.
- Şata, M. (2016). Türk eğitim sistemi'nde sınıf içi ile geniş ölçekli ölçme ve değerlendirmeye genel bir bakış. *Current Research in Education*, 2(1), 53-60.
- Tataroğlu Taşdan, B. & Çelik, A. (2014). Matematik öğretmenlerine yönelik bir mesleki gelişim programı prototipi. *NWSA-Education Sciences*, 9(3), 323-340. doi: 10.12739/nwsa.2014.9.3.1c0621
- TEDMEM (2019). *TALIS 2018 sonuçları ve Türkiye üzerine değerlendirmeler* (TEDMEM Analiz Dizisi 6). Ankara: Türk Eğitim Derneği Yayınları.
- Thompson, T. (2008). Mathematics teachers' interpretation of higher-order thinking in Bloom's taxonomy. *International Electronic Journal of Mathematics Education* 3(2), 96-109.
- Uslu, Ö. (2013). A new model proposal for evaluating teachers' professional development programs. *H. U. Journal of Education, Special Issue* (1), 359-374.
- Villegas-Reimers, E. (2003). *Teacher professional development: An international review of the literature*. Paris: UNESCO International Institute for Educational Planning.
- Vlaardingerbroek, B. & Taylor, T.G. (2003). Teacher education variables as correlates of primary science ratings in thirteen TIMSS systems. *International Journal of Educational Development*, 23, 429-438. doi:10.1016/S0738-0593(03)00014-2
- Yatağan, M. (2014) Fen ve teknoloji öğretim programının öğrenci ve öğretmen özelliklerine göre değerlendirilmesi: TIMSS 2007 ve TIMSS 2011 verileri ile bir durum analizi. (Yayımlanmamış doktora tezi). Gazi Üniversitesi, Ankara.
- Yıldırım, A. & Şimşek, H. (2016). *Sosyal bilimlerde nitel araştırma yöntemleri*. Ankara: Seçkin Yayıncılık.
- Yin, R. K. (2003). *Case study research: Design and methods*. 3rd ed. Thousand Oaks, CA: Sage.
- 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 Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest (Issues & Answers Report, REL 2007–No. 033).

- Yurdakul, B., Uslu,  ., akar, E. & Yıldız, D. G. (2014). Evaluation of the professional development program on web based content development. *Educational Sciences: Theory and Practice*, 14(4) 1427-1437. doi: 10.12738/estp.2014.4.2131
- Zohar, A. & Schwartz, N. (2005). Assessing teachers' pedagogical knowledge in the context of teaching higher-order thinking. *International Journal of Science Education*, 27(13), 1595-1620. doi: 10.1080/09500690500186592.