Education and Science tedmem

Vol 45 (2020) No 203 1-25

Investigating Elementary School Teachers' Development of Mathematical Task Implementation Quality *

Engin Ader [1]

Abstract

This mixed-method study investigates the quality of teachers' implementation of mathematical tasks and the nature of changes that occurred during an academic year with the support of a professional development program. Task implementation quality consists of three dimensions: total cognitive demand, attention to student thinking and intellectual authority. The study took place in the context of a 1-year professional development program, with the participation of four teachers in a private school. Data were gathered from class observations and interviews with teachers. Both quantitative and qualitative analyses indicated changes in the teachers' practice. The major change was in the total cognitive demand in the implementation of tasks. An analysis of changes in cognitive demand level revealed fluctuations in the total cognitive demand of mathematical tasks, indicating that changes in teaching practice are complex and often nonlinear. The findings are discussed along with potential teacher-related factors. Recommendations are made for future research on professional development programs.

Keywords

Mathematical tasks
Task implementation quality
Professional development
Cognitive demand
Intellectual authority
Student thinking

Article Info

Received: 02.25.2019
Accepted: 11.20.2019
Online Published: 04.04.2020

DOI: 10.15390/EB.2020.8544

* This article is orally presented at the 13th International Congress on Mathematical Education (ICME-13).

[1] Boğaziçi University, Faculty of Education, Department of Mathematics and Science Education, Turkey, ader@boun.edu.tr

Introduction

Approaches to mathematics teaching have undergone changes in the last three decades. The defining characteristics of the changes include a sustained emphasis on conceptual understanding, reasoning, and problem solving as opposed to conventional teaching, characterized by the presentation of facts and procedures, followed by practice using these procedures, all without a focus on underlying mathematical concepts (Stein, Correnti, Moore, Russell, & Kelly, 2017).

Enabling teachers to adapt to changes and to implement new ideas in their teaching is a major challenge that Cohen (1990) characterizes as a "paradox where teachers are the chief agents of change as well as being a major part of the problem to be corrected" (p. 326). Indeed, most studies in this area indicate that teachers are unable to adapt to curricular changes (Davis, 2003). Teachers often choose tasks that are dependent on procedural skills or memorized knowledge and even when they choose tasks with the potential to develop reasoning, problem solving skills, and conceptual understanding, they often turn such tasks into routine mathematical exercises that rely on procedural skills (Tekkumru Kısa & Stein, 2015). With changes in curricular expectations, research on teaching mathematics, particularly on teacher implementation of mathematical tasks, is on the rise (Tekkumru Kısa & Stein, 2015; Ubuz & Sarpkaya, 2014).

Professional development (PD) programs can play a crucial role in supporting teachers' learning, particularly in terms of their ability to implement classroom tasks in a way that leads to achieving the desired outcomes (Borko, 2004). Using a situative approach, Adler (2000) describes PD and teacher learning as "an increased participation in the practice of teaching" (p. 37) and becoming more knowledgeable about the profession. Borko (2004) discusses how teacher–researcher collaboration in a PD program can have an impact on teacher learning. Various studies indicate that teacher learning through such programs can be documented (Borko, 2004; Fennema et al., 1996), especially in terms of their instructional practices.

In 2005, major changes in the Turkish primary mathematics curriculum such as those described above were implemented (Ersoy, 2006). At the time of the preparation of this article, a new primary mathematics curriculum was gradually being phased in one year at a time. It is similar to the 2005 curriculum in scope and teaching/learning approaches. Mathematical tasks (often referred to as "activities" in Turkish textbooks) feature prominently in primary mathematics classrooms (Milli Eğitim Bakanlığı Talim ve Terbiye Kurulu Başkanlığı [MEB-TTKB], 2017). In mathematics education, the term "task" differentiates targeted mathematical work from all other sorts of classroom activities. Stein, Smith, Henningsen, and Silver (2000) define a mathematical task as "a segment of classroom activity devoted to the development of a mathematical idea" (p. 8). In line with this definition, a mathematical problem, a question about a mathematical idea or a collection of them can be turned into a mathematical task depending on how the teacher wishes to use it in the classroom.

Considering the emphasis on using mathematical tasks to achieve the mathematical objectives and goals listed in the curriculum, studying the use of mathematical tasks in classrooms is essential. In Turkey, the textbooks published by the Ministry of National Education (MEB) offer numerous mathematical tasks. Teachers are given the freedom to choose their tasks so long as they are in line with the objectives of the curriculum. They can use the textbook tasks as they are (or amend them), find tasks from other resources, or create their own.

The present study investigates the quality of teachers' implementation of mathematical tasks and the nature of changes that occurred over the course of an academic year with the support of a PD program. Quality of teachers' implementation of tasks is defined through three features: demanding higher level thinking from the students from the beginning to the end of the task; attending to students' thinking and building task implementation on their ideas and thinking; and establishing mathematical thinking as the ultimate criteria for evaluating credibility of mathematical ideas (Stein & Kaufman, 2010). The first feature of quality of implementation of tasks concerns the key idea of cognitive demand, which can briefly be defined as the kind and level of thinking required from students to successfully work on a task (Stein et al., 2000).

An investigation of whether a dense PD program requiring time and effort, brings about change in quality of teachers' implementation of mathematical tasks in a context of curricular changes is the major significance of the study. Potentially such a study also allows researchers to make sense of how particular aspects and mechanisms of the PD program contributes to the changes taking place. Hence, while not the focus of this study, the concluding discussion will include considerations of the PD program with an eye on changes in quality of teachers' implementation of mathematical tasks.

Implementation of Mathematical Tasks

To analyze mathematical tasks in terms of cognitive processes, Stein, Grover, and Henningsen (1996) divide mathematical tasks into two broad categories based on cognitive demand level: low-level and high-level demand tasks. Each category is divided into two sub-categories. Lower-level demand tasks are classified as either memorization tasks or procedures without connection to mathematical concepts tasks. Higher-level demand tasks are classified as either procedures with connection to mathematical concept tasks or doing mathematics tasks. Stein et al. (2000) provide a detailed explanation of these task categories in their "Task Analysis Guide."

Stein et al. (2000) also developed a Mathematical Task Framework (MTF) for delineating the implementation of mathematical tasks, composed of four phases: the task as it appears in curricular materials, the task as it is set up in the classroom, the task as it is enacted in the classroom, and student learning. In the first three phases, tasks and the way they are implemented can be examined by identifying the cognitive processes required by the tasks. The MTF can be used to evaluate teaching practice by examining the level of cognitive demand in each phase.

Stein and Kaufman (2010) maintain that, in order to establish environments where students engage in high-level thinking and reasoning, various issues need to be addressed. With this in mind, they devised a framework that presents these issues in three dimensions. First, the task itself needs to have a high level of cognitive demand, which has to be maintained at the set-up of the task and in its enactment in the classroom; that is, teachers should not allow the cognitive demand level of tasks to decline as they progress through the stages towards student learning. Teachers also need to attend to students' thinking as they work on the tasks, which is the focus of the second dimension. This dimension embodies uncovering student thinking, deciding on which student ideas are to be heard by all students and connecting these ideas in a meaningful way. The third dimension is intellectual authority in the classroom, and it is built on what the teacher and students turn to for judgments of correctness of mathematical ideas. Rather than expecting the teacher to act as the judge, mathematical reasoning and norms must be emphasized as the mathematical authority (See Appendix 1 for description of high quality implementation of tasks).

Teachers' use of tasks in classrooms has been studied in various ways. Some research has focused only on changes in cognitive demand (e.g. Charalambous, 2010). Others view teaching in a broader sense and focus on the use of tasks from this broader perspective. For example, researchers from the University of Pittsburgh developed an Instructional Quality Assessment tool for evaluating the quality of instruction using these criteria: "(1) cognitively challenging instructional tasks, (2) task implementation, or opportunities for students to engage in high-level thinking and reasoning throughout an instructional episode, (3) opportunities for students to explain their mathematical thinking and reasoning in mathematical discussions or in written responses, and (4) teachers' expectations for students' learning" (Boston, 2012, p. 79). Similarly, Hill et al. (2008) frame the quality of mathematics instruction in 6 dimensions: mathematical errors, responding to students inappropriately, connecting classroom practice to mathematics, richness of mathematics, responding to students appropriately, and mathematical language. They constructed this framework by focusing on "deficits and affordances" (p. 437). In the present study, Stein and Kaufman's (2010) quality of implementation of task framework was used so as to stay connected to teaching while implementing mathematics tasks, rather than including more generic elements of teaching practice such as classroom environment or classroom questioning.

In the related literature there are various studies on factors that influence the selection of tasks and their implementation and the maintenance of the cognitive demand of tasks. These are *teacher factors*, including teacher knowledge (e.g. Charalambous, 2010; Wilhelm, 2014), teacher skills (e.g. Tekkumru Kısa & Stein, 2015), and teacher conceptions (e.g. Wilhelm, 2014) and *contextual factors*, which include the curriculum (e.g. Stein & Kaufman, 2010), time-related issues (e.g. Henningsen & Stein, 1997), and student characteristics (e.g. Henningsen & Stein, 1997). Considering that teaching is a rich amalgam of personal and social contexts, and given the interplay between all these factors, one can imagine the challenge of effecting change in teaching practices with the aim of adapting to curriculum changes.

Teacher Professional Development

Changes in teaching practice have long been regarded as a consequence of *training* (Clarke & Hollingsworth, 2002). Only recently have such changes been accepted as a dynamic process that is a consequence of teachers' critical reflection on their own practice. Rimbey (2013) identified the key features of PD programs as establishing consistency among the learning approaches of the curriculum, targeted teacher learning and educational goals of participating schools; allowing sufficient time for learning to take place; and taking into consideration the social aspects of learning.

Numerous studies in the PD literature criticize the expectation that teachers will adopt new knowledge and change as a result of knowledge transmission in PD programs (Guskey, 2002; Huberman & Miles, 1984). PD models based on short periods of dense transmission of knowledge assume that the PD programs will bring about changes in teachers' beliefs and that these changes will be reflected in their practice. Guskey (2002), pointing to the inadequacy of this approach, suggests that PD programs should provide opportunities for teachers to think about and apply new techniques in their teaching practice, to reflect on the outcomes, and to iterate new ideas that they eventually incorporate into their practice. Guskey's model is effective in highlighting the key dilemma of the transmission of knowledge approach in PD programs and in shedding light on methodological aspects of their design.

In considering these key theoretical features in planning a PD program, it is also important to make decisions about methodology that will fulfill the requirements of the content area and fit the educational context in which the PD program will take place. Changes in the Turkish mathematics curriculum taking place since 2005, meant that teachers needed help in implementing mathematical tasks with a focus on the objectives of the curriculum. Such a program also needed to be based in schools that were willing to adopt changes in teaching approach.

Issues in the day-to-day implementation of PD programs include contextualizing discussions within teachers' practice in their own classrooms (Borko, Jacobs, Eiteljorg, & Pittman, 2008) and using video recordings (Van Es & Sherin, 2008; Tekkumru Kısa & Stein, 2015) to maximize the effectiveness of the PD program. Discussion between teachers and between teachers and facilitators has been shown to provide fruitful opportunities for reflection and learning (Anderson, Coltman, Page, & Whitebread, 2005; Perry & VandeKamp, 2000; Van Es & Sherin, 2008). Ball and Cohen (1999) indicate critical features of professional development approaches: (a) an inquiry-oriented learning environment, (b) a collective endeavor to learning, and (c) discussion based on concrete artifacts from the classroom. PD programs supporting teacher change need to incorporate these key features.

In light of the teaching needs that have arisen with the changes in the mathematics curriculum in Turkey and the trends in developing effective teacher PD programs, this study poses two research questions:

- Is there a significant difference in teachers' implementation of tasks from the beginning to the end of a 1-year PD program?
- How does task implementation change during a 1-year PD program?

While the first research question requires quantification of data from observations, the second requires content analysis of data from the same observations plus subsequent interviews with teachers. Results provided as findings for the second research question is expected to significantly contribute to making sense of the changes taking place and point towards potential avenues for further research.

Method

This is a mixed-method study comprising both qualitative and quantitative elements. It is a concurrent triangulation design in that qualitative and quantitative data were collected concurrently and the results of the analysis of both sets of data were combined to study convergence and potential differences (Creswell, 2009). The study took place in the context of a PD program, which provided rich quantifiable data for this research as well as observations and interviews conducted simultaneously for qualitative analysis.

Participants

The PD program took place in a small private primary school in Istanbul in the 2014-2015 academic year. The school did not have a strict catchment area and, due to its fees, drew students from middle to high socio-economic status (SES) families. Class sizes ranged from 13 to 18. Even though majority of students attend public schools in Turkey, the school carried common characteristics of many private schools in the country.

Four teachers participated in the study. They were enthusiastic about participating in the program and declared a particular interest in improving their implementation of mathematical tasks. All participants were female with training in elementary education, but none had a specialization in mathematics or mathematics education. Their teaching experience ranged between 7 and 36 years. Nesrin and Nil, each with more than 30 years of professional experience, were both teaching first grade. Defne, who had 8 years of experience, and Suzi, with 7, were teaching second and third grades, respectively. All names are pseudonyms.

Professional Development Program

The PD program was designed on the principle of learning in practice (Ball & Cohen, 1999) and on the principle of building change in practice through the implementation of new ideas (Guskey, 2002). These principles had two practical manifestations: the use of video recordings of the teachers' classes

(Borko et al., 2008) and the creation of a community of learners where ideas are discussed by teachers and the facilitators of the program (Van Es & Sherin, 2008).

The four teachers, the author (as the main researcher), and an assistant researcher responsible for the recordings were involved in the program. The program focused on improving the quality of teachers' implementation of mathematical tasks in their classes. We started with four workshops the week before the onset of the school year. These covered the key concepts of quality of implementation of mathematical tasks. All observations and discussions in the interviews centered on the selection and planning of tasks and on the aspects of teaching that relate to the implementation of tasks.

The PD program provided numerous opportunities for teachers to reflect on their practice, with repeated planning-teaching-reflecting cycles. Videos of either the teachers' own classes or other classes served as contextually relevant tools for reflection (Borko et al., 2008). They also allowed teachers to relive their experience, thereby supporting reflection (Van Es & Sherin, 2008). Throughout the program, teachers made various comments about noticing elements of their teaching practice that they did not pay attention to before, intentions to change aspects of their practice and the fact that thinking about their practice and discussing it with others motivated them to reconsider some of the assumptions they had about implementing mathematical tasks. These were manifestations of the opportunities provided by the program.

The duration of the program was approximately nine months (mid-September 2014 – mid-June 2015). Observations were done biweekly, and interviews with teachers were held weekly – one with the third grade teacher, one with the second grade teacher and one with the two first grade teachers. The whole group of teachers met at the end of each semester. The number of lessons and the number of tasks observed each semester are shown in Table 1.

Table 1. Number of Lessons and Tasks Observed

Teacher	Fall		Spring		Total	
	Lessons	Tasks	Lessons	Tasks	Lessons	Tasks
Defne	8	21	5	13	13	34
Suzi	6	11	6	12	12	23
Nesrin	7	21	7	15	14	36
Nil	8	18	6	13	14	31

Data Collection Tools and Instruments

The Classroom Observation Coding Instrument: Audio and video recordings were used to collect data from class observations. For the quantitative part of the study, the researcher chunked the lessons into tasks, and each task implemented in the classroom was coded for the quality of its implementation, using a revised version of Stein and Kaufman's (2010) Classroom Observation Coding Instrument (see Appendix 1), the only revision being a further subdivision of one of the codes for intellectual authority. The indicators for the codes for each variable are predetermined in the instrument, and the observers coded each task according to the indicator that best described the task implementation. The coding scheme has five variables:

- the cognitive demand of the task as it appears in published materials (values from 1 to 5),
- the cognitive demand of the task as it is set up by the teacher (values from 1 to 5),
- the cognitive demand of the task as it is enacted by the teacher and students (values from 1 to 5),
- attention to student thinking (values from 0 to 3), and
- intellectual authority (values from 0 to 3).

A total cognitive demand score − ranging from 2 to 8, was calculated according to the codes of the three phases of task implementation (first three variables on the list), with higher values indicating maintenance of high cognitive demand through the phases. Scores for attention to student thinking and intellectual authority were obtained directly by using the scoring indicators in the instrument (see Appendix 1 for a detailed account of indicators of the variables and calculation of the scores).

Before starting the formal data analysis, the author and the assistant researcher independently coded the implementation of 27 tasks observed in 12 lessons in order to check inter-rater reliability. The agreement between raters' codes for total cognitive demand, attention to student thinking, and intellectual authority were 74%, 93% and 85%, respectively, with corresponding Cohen's κ values of 0.62 (weighted κ: 0.72), 0.88 and 0.73. Percent agreement and Cohen's κ values showed good agreement between raters and were taken as evidence of reliability.

Interviews (Meetings with Teachers): In addition to class observations, semi-structured interviews were conducted with individual teachers and groups through the year. These lasted 40 minutes each. The main agenda items were:

- topics for the coming week's lessons: This involved discussion of the conceptual and procedural content for lesson plans and ideas for tasks. The researcher facilitated the interviews to maintain the focus on curriculum content and the variables in the task implementation framework. Rather than imposing ideas on the teachers, the researcher acted as "a critical peer".
- the previous lesson's observation: Videos of the lessons were shared with the teachers to be viewed on the weekend. As they watched the videos, they were to focus on the implementation of task variables, changes in cognitive demand level, and the effectiveness of the task implementation.

At the end of each semester, an interview that included all four participating teachers was held. Issues that had arisen in the implementation of tasks and the interviews were discussed with the whole group. These larger interviews created opportunities for sharing experiences and opinions.

The researcher did not use a fixed set of questions in these interviews. The discussions centered on the variables of the study — dimensions of mathematical task implementation and teachers' comments about their practice. All interviews were audio recorded.

Data Analysis

Data were taken from the recordings of lessons observed and interviews. Data from the interviews were transcribed, as were data from the lessons, which were, chunked into tasks and coded for the quantitative analysis. Nvivo was used for analysis.

For the quantitative analysis, descriptive statistical indicators of task implementation quality were obtained. For each lesson, scores were calculated by weighting the scores from the tasks in the lesson according to the amount of time each task lasted within the lesson. Hence, lessons constituted the unit of analysis since there were equal number of lessons observed per teacher yet the codes for the tasks were the underpinning elements. A profile for each teacher was obtained by cross-tabulating the scores from their lessons by the time they spent in the PD program. Variance in teacher performance accounted for by the time spent in the PD program was also calculated, using time as a proxy for the influence of the PD program. Finally, three repeated measures ANOVA tests were conducted to test the effect of the program on the quality of task implementation, with successive time slots for observations as the independent variable and the three indicators of task implementation quality as dependent variables.

Lesson observation data were scored on the three dimensions of task implementation quality. The findings of the qualitative analysis were used to triangulate the findings of the quantitative analysis, in order to document teachers' implementation of tasks and the changes that occurred during the PD program. Data from teacher interviews were open coded in order to capture the emerging themes in terms of task implementation quality and change in practice.

Peer debriefing and multiple methods to collect data (triangulation) were used to establish the trustworthiness of the researcher's interpretations in the qualitative analysis (Gay, Mills, & Airasian, 2009). In addition, a thorough conceptualization of the phenomenon under study was presented in order to document how the data analysis accurately focused on the phenomenon.

Results

Findings from the Statistical Analysis

To describe the task implementation quality, mean values were calculated for all scores for each semester (Table 2). Mean values for each teacher were calculated by weighting the codes for each task according to the time that particular task lasted relative to time on task. Weighted averages were used so as to better represent the values of task implementation at a typical moment in the lesson.

Table 2. Mean Values of the Scores for Task Implementation Quality

	Suzi		**Defne**		**Nesrin**		**Nil**	
	Fall	**Spring**	**Fall**	**Spring**	**Fall**	**Spring**	**Fall**	**Spring**
Intellectual authority (out of 3)	1,71	1,95	1,38	1,46	0,69	0,82	0,37	0,50
Attention to students' thinking (out of 3)	1,27	1,40	1,78	1,98	0,91	1,21	0,52	1,03
Total cognitive demand (2-8)	5,98	7,28	5,54	6,66	3,72	7,03	2,33	5,38

Scatter plots to illustrate the teachers' progress were constructed for the dimensions of task implementation quality: total cognitive demand (Fig. 1), attention to student thinking (Fig. 2) and intellectual authority (Fig. 3). Figure 1 depicts a tendency for scores in the second semester to gather higher up in the top right quadrant compared to the left half of the plot, indicating a relationship between total cognitive demand and time spent in the PD program. However in Figures 2 and 3, the points corresponding to separate observations of the teachers are scattered across the plot, indicating much weaker relationships between time spent in the PD program and the variables, attending to student thinking and intellectual authority.

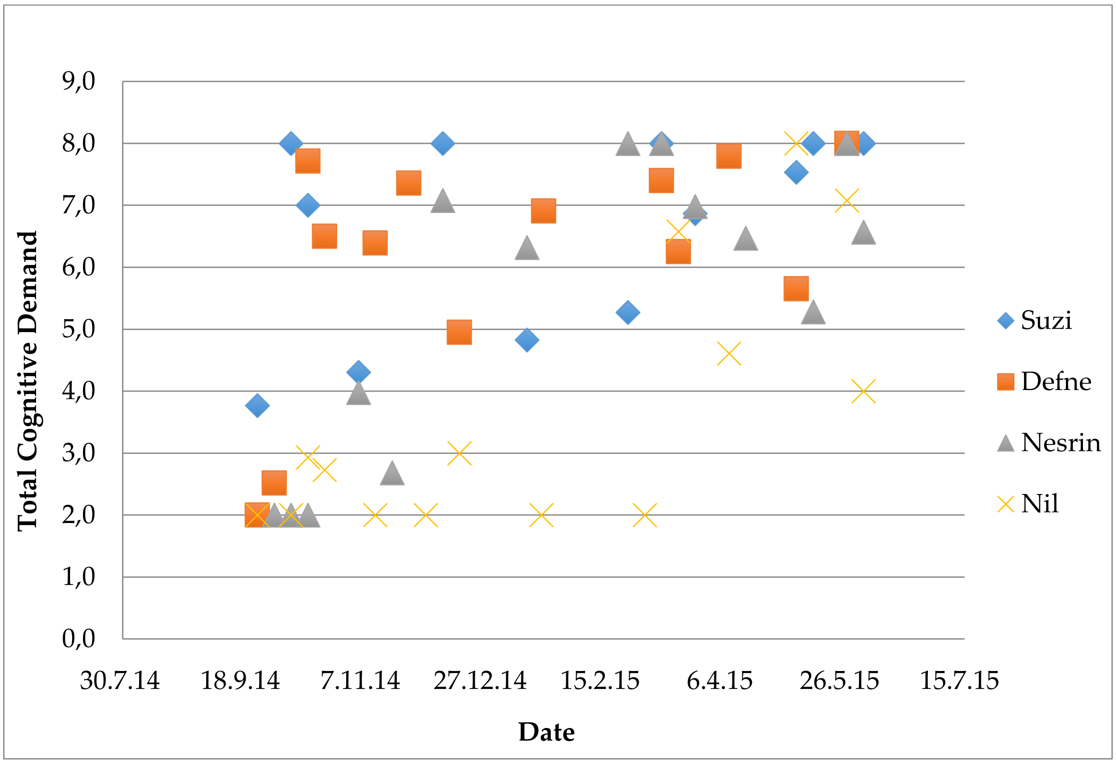

Figure 1. Total cognitive demand with respect to time

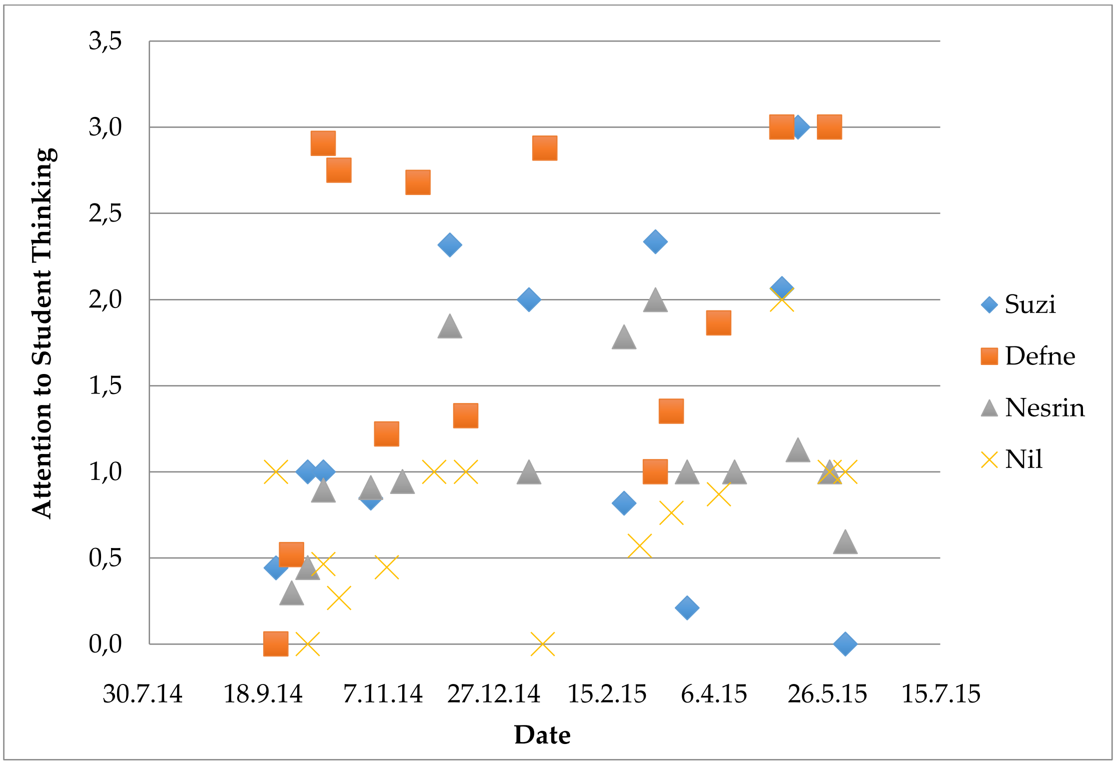

Figure 2. Attention to student thinking with respect to time

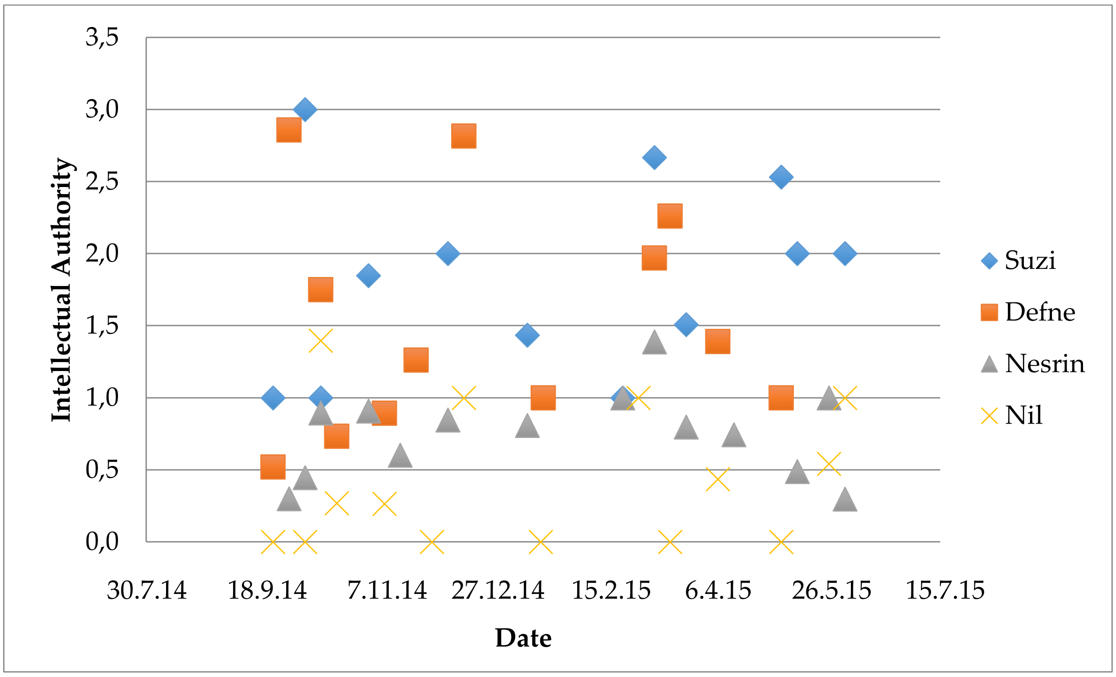

Figure 3. Intellectual authority with respect to time

To answer the first research question, correlation coefficients were calculated to determine the relationship between time spent in the PD program and codes on task implementation quality (Table 3). A significant relationship was found between the total cognitive demand and the amount of time teachers had spent in the PD program, r = .53, p (one-tailed) < .01. Attention to student thinking and intellectual authority were not significantly related to the amount of time teachers had spent in the program.

Table 3. Pearson Correlation Coefficients between Dimensions of Task Implementation Quality and Time Spent in the PD Program (whole group).

	Total Cognitive Demand	Attention to Student Thinking	Intellectual Authority
Time spent in the PD program	.53**	.23	.10

Note. N = 53; ** p < .01, one-tailed.

To determine whether teachers' implementation of mathematical tasks changed over time, a repeated measures ANOVA was conducted. The scores for task implementation throughout the year were compared, using a repeated measures ANOVA design. Variance in task implementation quality as a result of the time spent in the program was analyzed, and the between-teacher variance was excluded. A repeated measures ANOVA with combined Greenhouse-Geisser/Huynh-Feldt correction for sphericity showed that scores on the total cognitive demand of mathematics tasks were significantly affected by the time teachers spent in the PD program $F(6.72, 20.16) = 3.42$, $p < .05$, $\eta^2 = .532$. Post hoc contrasts indicated that the cognitive demand of tasks increased over time. Total cognitive demand for the last four observations (twelfth observation: M = 7.77, SD = .46; eleventh observation: M = 6.74, SD = 1.47; tenth observation: M = 6.60, SD = 1.45; ninth observation: M = 6.67, SD = .33) was significantly higher than those in the first observation (M = 2.44, SD = .88). Total cognitive demand in the last observation (M = 7.77, SD = .46) was significantly higher than that in the fourth observation (M = 4.78, SD = 1.57), the sixth observation (M = 4.78, SD = 1.37) and the ninth observation (M = 6.67, SD = .33). These findings point to an increase in total cognitive demand over time. However, a repeated measures ANOVA with Greenhouse-Geisser correction showed that scores on attention to student thinking [$F(2.16, 6.48) = 2.20$, $p > .05$] and intellectual authority [$F(1.72, 5.15) = 1.15$, $p > .05$] were not significantly affected by the amount of time teachers spent in the PD program.

To further investigate individual teacher progress throughout the PD program, the relationship between the scores on the dimensions of task implementation and time spent in the PD program were analyzed (Table 4). Similar to the results for the whole group, most of the significant links were found between total cognitive demand scores and time spent in the PD program. Three of the four teachers' total cognitive demand scores were significantly related to time: for Nesrin $r = .78$, p (one-tailed) $< .01$, for Nil $r = .74$, p (one-tailed) $< .01$ and for Defne $r = .78$, p (one-tailed) $< .05$. The only other significant link was between Nil's attention to student thinking and time in the PD program, $r = .53$, p (one-tailed) $< .01$. However, these Pearson values should be interpreted with care since the numbers of data points are below 25 and the normal approximations are getting poorer with fewer data points (Weaver & Koopman, 2014).

Table 4. Dimensions of Task Implementation Quality and Time Spent in the PD Program.

Teacher	N	Pearson correlation coefficients		
		Total Cognitive Demand	**Attention to Student Thinking**	**Intellectual Authority**
Defne	13	.52*	.37	.22
Suzi	12	.47	.22	.24
Nesrin	14	.78**	.26	.13
Nil	14	.74**	.53*	.13

Note. N = number of observed lessons; * $p < .05$, one-tailed; ** $p < .01$, one-tailed.

Analysis of Progression of Teaching with Illustrative Episodes

Teachers having prominent changes in task implementation quality: An analysis of the qualitative data on teachers' practice throughout the PD program provided evidence that corroborated the results of the statistical analysis. Similarities were observed between the practices of the two first-grade teachers in terms of changes throughout the year. At the beginning, in both Nesrin's and Nil's lessons, the total cognitive demand was relatively low (see Figure 1). The tasks they implemented in their classes were planned and maintained at a low level of cognitive demand, focusing mainly on reproducing previously learned facts and rules. These lessons were predominantly on learning numbers and counting. Both teachers cited the curriculum objectives of that time of the year targeting counting skills and knowledge of numbers as the main reasons for using low cognitive demand tasks.

The second observation of Nil's class illustrates her task implementation in the early phases of the PD program. The lesson was on the number five, representing five as a quantity, counting up to five, and finding number pairs that combine to make five. Nil asked the students to hold up fingers on one hand and then gave the instructions. The students were to answer as a whole class by indicating their answers with their fingers. During this task, Nil asked short-answer questions to focus students on pairs of numbers that combine to make five and using skills on counting up to five, e.g. hiding one finger of her hand and asking "how many fingers will look for this hidden finger?", hiding two fingers of her hand and asking "how many fingers will look for those two hidden fingers? Let's count them". Even though these ideas can be considered a preparation for addition and subtraction operations, no conceptual link was made between the task and these mathematical concepts. The task consisted of reciting the counting and subitizing knowledge to which students had already been introduced. Attention to student thinking and intellectual authority were also at a low level, since inviting student responses generated very short answers, and the teacher often confirmed students' answers without making links to reasoning.

Towards the end of the school year, both Nesrin and Nil's task implementation showed higher levels of total cognitive demand. They often encouraged the students to make connections with mathematical concepts. Nil was the only teacher who showed significant changes throughout the year in more than one dimension of task implementation: total cognitive demand and attention to students' thinking, as illustrated by an excerpt from one of her lessons during the last month of the PD program.

In a lesson on formulating problems that would use mathematical operations the students had learned, Nil showed a picture of several kinds of animals with a different number of each and asked her students to take inspiration from the picture to create word problems. She planned to spend the entire lesson on this task. The excerpt below illustrates her interaction with the students as she set up the task.

Nil:	What do we need to have in a problem? For example, if I said: There were two horses. Two more horses joined them. Would this be a problem?
Students:	(shouting in chorus) No!
Nil:	Why not?
Dilek:	Because there is no unknown number in this.
Nil:	(repeating) There is no unknown number. So, you're saying that there should be something unknown. Correct. We should be looking for something unknown. Something we can find a solution for.
Esra:	Something we can solve…but it should be a bit difficult.
Ece:	The other day we solved a difficult problem…it should be difficult like that.
Nil:	Yes…from now on, we can focus on difficult problems. OK…now, here's what I want you to do. In our problem, we can use addition, subtraction or we can use them together. And it can be a bit difficult. You can come up with problems to make us think about those things we learned.

In setting up the task, Nil reminded the students about operations and the relevant terms they had learned, and she made connections with key ideas involved in the operations. She did not suggest a particular path for the students to follow. This was a doing mathematics task which had no specific procedure for leading the learners to a product. Throughout the implementation, Nil maintained the cognitive demand by giving students time to come up with problems and by sharing some of the student-generated problems with the whole group.

Nil encouraged student input from the earliest phase of the task by asking them what they needed to pay attention to when formulating problems and gave them opportunities to share their ideas. She asked some of them to share the problems they had come up with, paying attention to whether their problems involved more than one mathematical operation. Although Nil made no explicit connection between problems and their mathematical underpinnings, her teaching moves manifested high levels of attention to student thinking that had not been apparent earlier in the PD program.

Nesrin's teaching displayed a sharp increase in total cognitive demand like Nil's task implementation, but not an increase in attention to student thinking or intellectual authority. At the end of the year, her attention to student thinking was essentially at a level where she would give students opportunities to share their responses and make explanations, but she generally did not make connections between student ideas and the purposeful selection of specific responses. She did most of the justifications herself rather than referring the students to mathematical reasoning.

A lesson on solving problems that involved addition and subtraction operations exemplifies the dynamics of Nesrin's task implementation at the end of the PD program. She gave her first graders a word problem about two children going shopping, each with the same amount of money and each making a variety of purchases. Students had to answer questions about how much money had been spent and who had spent more. Starting from the set-up of the task, Nesrin asked her students to think about what needed to be found, the approach to be adopted, and the reasoning behind their answers. She emphasized that she wanted the students to focus on the meaning underlying the operations and that she expected them to give justifications for their steps. This emphasis on the conceptual underpinning of steps demonstrated a key shift in her maintenance of cognitive demand, as opposed to implementing tasks with lower levels of cognitive demand. Nesrin created opportunities for her students to share their ideas about solution methods. The following exchange is excerpted from about the midpoint of the implementation.

Nesrin: How do I find how much money he has left? How did you do that?
Pınar: First, I added those two and found 13. Then I took away 13 from 20 and found 7.
Nesrin: This is also correct. This is another strategy. You?
Murat: I counted.
Nesrin: You counted as well as doing operations. (pointing to another student) How did you do that?
Kerem: First, I subtracted 9 from 20. Then I subtracted the 4.
Nesrin: What?
Kerem: (looking confused) Then…
Nesrin: You should explain it just the same way you did it. I think you didn't explain exactly like you did it.
Kerem: First, I added 4 and 9.
Nesrin: Why did you add them up? What did you find by adding?
Kerem: How much he spent…
Nesrin: Good. Then?
Kerem: Then I subtracted that number from 20.
Nesrin: How did you solve it? Why did you subtract instead of add?
Kerem: How much was left.
Nesrin: OK, but why did you subtract? What happened to the money that made you prefer subtraction?

In a short period of time, three different strategies were verbalized by three students. Nesrin listened to the explanation of their strategies, and she helped them express their ideas in a way the whole group could understand. She affirmed the correctness of the response. However, she did not encourage the class to make connections among these ideas or leverage conversations that would make conceptual connections between mathematical ideas. She often acted as an evaluator of student responses with respect to their mathematical validity. Throughout the lesson, there were few occurrences of reasoning as a whole group for evaluation of ideas. As a result, the quality of her implementation of this task was not high in terms of attention to student thinking and intellectual authority.

Overall, the major shift in task implementation occurred in the dimension of maintaining the cognitive demand and, in the interviews that followed, the two first-grade teachers occasionally touched on this change by referring to changes in their decisions about how they implemented tasks. Nesrin, for instance, made the following comment about one of her tasks in the last month of the PD program: "Doing the task this way meant I had to give students a lot of time to work on the task themselves, but it's necessary to expose students to high level cognitive processes. We've been talking about this throughout the year while watching our lesson videos together" (Interview, May 6, 2015). Similarly, Nil highlighted the emphasis she put on students' making sense of mathematics problems, internalizing the mathematical meanings of operations, and using these concepts to analyze problems. Even though she had elements of these in her teaching practices, she claimed that she adopted a more systematic approach through sustained efforts in the PD program (Interview, June 3, 2015).

A fluctuating progression through the program – cases of Suzi and Defne: Compared to Nesrin and Nil, Suzi and Defne had higher task implementation scores on all three dimensions from the onset of the PD program. As they progressed through the program, there were fluctuations in the quality of their task implementation. No significant changes occurred in attention to student thinking or intellectual authority.

Even though Suzi had the highest average scores on total cognitive demand in both semesters, no significant change was observed on that dimension throughout the program. From the very beginning of the program, fluctuations in the level of cognitive demand were evident in the tasks she planned. A particular issue early in the program was that she chose tasks with high cognitive demand but failed to maintain the cognitive demand level in implementing them. For instance, in her first observed lesson, she told a story and handed out cards, each of which had a multiple of 6 written on. The cards were in no particular order. She explained there was a relationship between the numbers and gave students approximately five minutes to figure out the relationship. Her goal was to get the students to notice that the numbers were multiples of 6.

Suzi set up the task in an open-ended fashion, without prescribing a particular route or procedure for the solution. She gave students time and opportunities to explore the relationship and come up with an answer. However, visiting groups as they worked on the task, she focused their discussions on the answer and directed the students towards the answer, as illustrated in the interaction below:

Emir: Miss, we found it, increasing by 6.

Suzi: Okay, you say 6 by 6? (pointing to a part of their work) But there is a mistake here. (She moved to another group. A student made an explanation that was not clearly audible in the video) OK, should we sort them in ascending order? If they can be sorted in descending order, then they can also be sorted in ascending order, right?

Even though she avoided telling the students what the answer should be, she did not encourage conceptual analysis of the pattern or suggest how it could be verified or falsified. It turned into a low cognitive demand task where students were simply reciting the multiples of 6 or ordering the numbers from the smallest to the greatest. Suzi worked through the task, asking students to explain their answers, but the answers were not sequenced nor were they discussed in a coherent way in the class. She visited the groups, evaluated their answers, either affirming them or pointing to areas where students needed to do further work. Her lowering of the cognitive demand was related to her preference for evaluating the students' work, which kept the intellectual authority at a low level.

However, there were also lessons where Suzi managed to maintain a high level of cognitive demand during the implementation of tasks. For example, in a lesson on symmetry in the spring semester, after a brief review of what students already knew about symmetry, she introduced a task where each student got a sheet of paper on which she had dripped ink, and she asked them to fold the paper. When the paper was unfolded, an amorphous symmetrical figure would appear. Students then examined the figures to see whether they were symmetrical and to explain their thoughts. During the discussion, various students described their strategies to check symmetry. Suzi also encouraged the students to think about the generalizability of the claims made in the class. Towards the end of the task, the discussion centered on the paper with the shapes shown in Figure 4 and another piece of paper. The following conversation then took place:

Suzi: Have a think about this. I'll ask the question *why*. Is there symmetry here, or not? Ali?

Ali: It's not symmetrical because when you fold it here (pointing to a vertical fold in the middle of each shape on Figure 4), you get different shapes.

Suzi: Can you explain that once again? Let me see if I've understood you correctly.

Ali: (drawing a vertical line with his finger) Here, in the middle. Look, you get different shapes.

Suzi: OK, suppose we think about *this* line (referring to the vertical fold on the paper). Consider this and think again. You are talking about one of the two shapes on this paper. What if we consider both shapes?

Ali: In that case, they are symmetrical.

Suzi: How's that?

Ali: Because when we fold it, both of them are exactly symmetrical.

Suzi: OK, thank you. Does it matter if they match when you fold the paper? Oh, now I'm a bit confused. Our conversation is all over the place. Anyone want to comment? Is this symmetrical, or not? Elif?

Elif: It's symmetrical, miss… Ali didn't look at it from the outside. (folding and unfolding the paper) When folded, all the lines match — look, Ali. This is why we can say it is symmetrical.

Suzi: You think you get the same shape when you fold it. You're right. They match. OK, Cem?

Cem: Miss, that single shape (pointing to each shape on Figure 4) is both symmetrical and not symmetrical. Because if we look at one of the shapes and divide it into two (showing a horizontal axis) two parts are not symmetrical. But if we fold the paper to match the two shapes and unfold it again, there's symmetry.

Figure 4. Student work opened up for discussion in Suzi's class

Suzi encouraged students to discuss why they thought the figures in front of them were symmetrical. She extended the idea of symmetrical shapes and linked the discussion to investigating the reflection symmetry of shapes with respect to a line. She allowed the students to think about how the idea of symmetry can be used to analyze complex shapes, thus maintaining a high level of cognitive demand. She listened carefully to students' responses and invited them to comment on each other's ideas. Refraining from immediately evaluating student ideas enabled Suzi to create a classroom environment where the accountability of student claims was judged through further mathematical reasoning by the students themselves. This was a task where high levels of total cognitive demand were accompanied by high levels of attention to student thinking and intellectual authority.

In the interview following this lesson, Suzi analysed the cognitive demand level of the task and how she tried to maintain a high level of cognitive demand upon watching video extracts from the lesson. A key point she raised during her analysis was how the sustained work during the PD program influenced some of her decisions about how to use tasks in her lessons. She emphasized starting to have more trust in the ideas discussed within the PD program as she found repeated opportunities to try them several times (Interview, March 25, 2015). This was one of the incidents where Suzi referred to the links between participation in the program and her teaching practice.

Use of materials during task implementation: An analysis of Defne's lessons also showed changes in the total cognitive demand in task implementation. Manifestations of such changes occurred in her second and third observations. She used similar materials and similar manipulatives for representing place values in both. In the second observation, the task was to compare numbers to find which one was greater; the task in the third observation focused on addition with carrying over.

In the first task, she handed out mats and base blocks and asked the students to model two 2-digit numbers with them. She also asked the students to compare the two numbers and to insert a greater-than or smaller-than sign between the numbers. Base blocks are manipulatives designed for representing tens and ones [separately] and can be useful in directing students' attention to place values. However, starting from the set-up and continuing through the implementation of the task, Defne did not refer to the tens or ones of the given numbers, nor was there any overt reference to manipulatives. Hence, no explicit connections were made with the mathematical concepts involved. The task focused mainly on producing the correct answer. This illustrates a task with low total cognitive demand resulting from low cognitive demand from the set-up until the end of the task.

Two weeks later, Defne used beans to represent ones, and cups holding ten beans represented tens for modeling numbers. She asked each student to independently model their numbers and started the task by discussing their representations. She gave the numbers 26 and 19, which students modeled at their desk. There was a whole-class discussion on key issues such as how they represented the numbers, what the two cups in 26 represented and how many ones are in a ten. This discussion was an early indicator of Defne's effort to establish and maintain conceptual links in task implementation. She then asked the students to model addition by combining all the cups and beans representing the two numbers. She asked further questions about the sum and how the materials representing the numbers need to be dealt with, such as "can there be 15 ones in the house of ones?", "what will we do with those 10 ones?"

Defne encouraged her students to make conceptual links between ten ones (i.e., ten beans) to be grouped into a ten (i.e., a cup holding ten beans) and how those two entities could be exchanged. The interaction was focused on how the manipulatives could be used for linking addition with the concept of place value. Earlier in the lesson, Defne had prepared the students for what the manipulations of the beans meant. When the students finished working with the manipulatives, she asked them to write in their notebooks the operation they had carried out with numerals. She circulated around the room to observe their work and wrote the expressions on the board (see Figure 5). She then concluded the discussion, focusing on three operations she had seen in students' notebooks. She asked which one was the most appropriate representation of addition with carrying over that they had done with the beans.

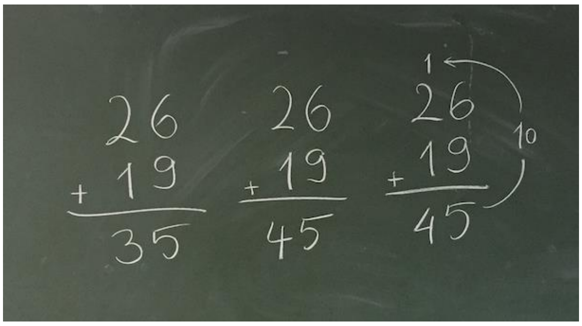

Figure 5. Three student representations of an addition operation

Defne:	When I told you to write the operation in your notebooks, I saw these expressions. I'm just writing them on the board without any comments…what I saw in your notebooks. I want you to look at these. What are the differences? Orhan, what do you see?
Orhan:	Can I say whatever I want?
Defne:	Yes. For example, what do you see on the first one, the second one and the third one? I walked around the class and saw these in your notebooks.
Orhan:	Erm…I think the third one makes more sense. The second one is correct too, but they didn't do the carry-over.
Defne:	Any other ideas? Alp?
Alp:	I agree with Orhan. In the first one, they got 35, but the answer is 45. You told us that, too.
Defne:	Why do you think they got 35?
Alp:	They might have forgotten to add the carry-over. That might be why they got 35.

Having made the conceptual links between place value and carrying over, the subsequent discussion helped Defne highlight the procedure for carrying over in addition. At the heart of her task was a simple procedure, but by making conceptual connections, she maintained the cognitive demand at a high level. Moreover, in the implementation of this task, her attention to student thinking was at a high level because she was explicitly inviting ideas for discussion and encouraging students to make a comparative analysis of these ideas.

In the interview in which this lesson was discussed, Defne watched the video extracts and commented on how asking students' comments about which of the three representations made sense helped them to clarify their understanding. Yet, she also emphasized that there needs to be repetition in the classroom about factual information so that students can remember mathematical knowledge and she claimed that such elements were missing from this lesson. Even though the quality of implementation for the lesson was high, Defne's comments showed that there would possibly be changes in her teaching practice throughout her participation in the PD program and it was not in a stable state (Interview, October 22, 2014). This was indeed manifested in the fluctuations in cognitive demand of her tasks especially in the early phases of the PD program.

A key incident – interconnections between dimensions of task implementation quality: It was in the cognitive demand dimension where significant changes occurred throughout the PD program, but there were lessons where high total cognitive demand was not accompanied by high levels of attention to student thinking and intellectual authority. For instance, the symmetry lesson in Defne's second-grade class showed how tasks that sustain student inquiries about manifestations of a mathematical idea and its applications can be impaired by paying little attention to student thinking and intellectual authority.

In the second half of Defne's lesson, she used a task where students discussed the features of shapes they had cut from a folded sheet of paper. The aim was to explain symmetry by using models. Defne first asked the students to offer explanations and followed up on an idea of one student:

Defne: Emre, come show us your shape and tell your friends what you see in your shape.
Emre: Erm…I see…
Defne: (interrupting) What do you see when you fold and then unfold it? (pointing to the two sides) What do you see here and over there?
Emre: There's a half here, another half there, another half…(trying to describe his irregular shape)
Defne: (to the class) what is the feature of Emre's shape, between its two sides when we fold and unfold it? (folding and unfolding the paper)
Dilek: They're the same.
Defne: They're the same. Emre, do you agree that they're the same?
Emre: (thinking a few seconds) Yes.
Defne: OK, but why are they the same? Everyone, have a look at your own shape. OK…Let's look at this fold line (pointing to the line on the paper). How does this line, the fold line, help us?

Defne opened up for discussion the idea of folding to see whether the two parts were congruent. She drew students' attention to the fold line by having them glue a thin string over the part they had cut out from the paper. Throughout the task, Defne asked questions to draw students' attention to key ideas relating to symmetry. She kept the total cognitive demand level high by not presenting the knowledge to the students but by asking questions to encourage them to explore and come up with the conceptual features of symmetry. She gave students the opportunity to share their ideas with the whole group, but most responses were short or simple affirmations of a previously stated idea. Since it was a whole-class discussion, Defne was unable to select individual student responses or make connections between them. There were no explanations or discussions about the thinking that student judgments were based on.

After this discussion, Defne finally gave the answer because her students were unable to reach a consensus. Acting as the mathematics authority, she provided the answer without going through the steps of reasoning. Throughout the task, Defne maintained a high level of cognitive demand by keeping the inquiry open for discussion, but attention to student thinking and intellectual authority were not at high levels.

Discussion

This study set out to determine whether changes would occur in the quality of implementation of mathematical tasks during a PD program and to investigate how such changes occurred. Both quantitative and qualitative analyses indicated changes in teachers' practice over time. The major change was in the total cognitive demand in the implementation of tasks.

Changes occurred in the total cognitive demand of tasks implemented by 3 of the 4 teachers. The more statistically significant changes were seen in the two first-grade teachers, Nesrin and Nil. At the onset of the PD program, they tended to plan low-level tasks and maintain that low level throughout the implementation; as the program progressed, however, they started implementing tasks with higher cognitive demand. Defne and Suzi, the second- and third-grade teachers, had relatively higher levels of total cognitive demand from the beginning, and they did have changes in the level of cognitive demand throughout the program. However, only the increases in Defne's quality indicators were statistically significant. Increasing total cognitive demand during task implementation was among the key aims of this PD program, recognizing that research in this area indicates a close link between student success in reasoning and problem solving and their teachers' selecting tasks with a high level of cognitive demand that is maintained throughout the implementation of the task (Stein & Lane, 1996; Stigler & Hiebert, 2004). Establishing a relationship between task implementation quality and student learning outcomes was beyond the scope of this study, but this is an area that warrants investigation (Hiebert & Grouws, 2007; Stein et al., 2017).

An analysis of changes in cognitive demand level reveals fluctuations in the total cognitive demand of mathematical tasks, indicating that changes in teaching practice are complex and often nonlinear (Van Es & Sherin, 2008). An investigation of how teacher learning occurs in such a PD program would constitute a valuable contribution to the field, possibly by analyzing changes in teachers' practice through teacher learning frameworks (M.K. Stein, personal communication, November 11, 2016).

Though not the main focus of the study, findings about implementation of tasks that indicate changes in teachers' practice and the data from corresponding interviews, point to how the PD program influenced teachers' practice. Teachers participating in this study commented in various interviews how the sustained emphasis on quality of implementation of tasks, coupled with repeated opportunities to implement tasks, watch their recordings and engage in discussions helped them establish some of their practice in the long run. PD programs that run longer than a semester, especially with in-service teachers who hold full-time positions, can be cumbersome. Still, sustained work with teachers over long periods of time is arguably the best approach to supporting their professional development, and this can most efficiently be achieved in their own professional setting (Borko, 2004; Guskey, 2002; Rimbey, 2013). The findings of the present study corroborate what has been reported in the literature in terms of changes in practice occurring through repeated cycles of observation and follow-up interviews. Further studies closely investigating the links between elements and steps of PD programs and changes in teachers' practice would shed light onto the mechanisms of this relationship.

In the analysis of the total cognitive demand dimension of task implementation, certain issues came to the forefront as to how a teacher's decisions and moves can influence the level of cognitive demand. Planning tasks whose cognitive demand is low and maintaining this low level throughout the task implementation led to failure to achieve high levels of total cognitive demand in the tasks they implemented. Additionally, shifting away from the intended use of the materials and limiting the amount of time for students to engage with the tasks without teacher intervention contributed to a low level of total cognitive demand.

Manipulatives intended to support thinking about mathematics concepts can only operate at a high level of cognitive demand if, and only if, teachers use them in a way that leads to conceptual connections (Van de Walle, Karp, & Bay-Williams, 2009). Otherwise, work with manipulatives can end up simply following established procedures or reciting memorized facts. Coaching teachers on using manipulatives is therefore an important element in developing their ability to achieve quality in task implementation. Stein and Kaufman (2010) underscore the importance of the productive use of materials as an indicator of teacher capacity for effective teaching. Another key factor in maintaining high levels of cognitive demand is the allocation of time for students to think about mathematical concepts. Taking shortcuts, teachers often point the students in a particular direction, which can lower the cognitive demand level of a task if students end up using a procedure without having made conceptual connections. Insufficient time to support conceptual connections is a dilemma for many teachers (Van de Walle et al., 2009), but they can sustain efforts to maintain tasks at a high level of cognitive demand by allocating time for students to work without close teacher steering, when they deem it as a worthwhile teaching investment.

Teachers in the present study showed little change in attention to student thinking and intellectual authority. Even though the average of the quality indicators for the second semester were higher than those for the first semester as a group, no statistically significant increase was observed over time. The only significant change over time occurred in Nil's attention to student thinking. One interpretation of these findings is that attention to student thinking and intellectual authority are dimensions of task implementation that are relatively stable and therefore more resistant to change. It can be particularly challenging to change a teacher's exercising intellectual authority in the classroom, owing to a deep-rooted *wisdom* about the role of teachers (Schoenfeld, 1994).

In a number of tasks during the PD program, implementation was at a high level of quality in terms of attention to student thinking and intellectual authority. This generally occurred in tasks where total cognitive demand was also high. There was evidence that attention to student thinking (e.g., by using students' ideas in discussion and encouraging them to exercise mathematical authority through mathematical reasoning) can help achieve high total cognitive demand. However, in some instances, a high level of total cognitive demand was accompanied by a low level of attention to student thinking and intellectual authority. Hence, the claim here is that changes in total cognitive demand do not necessarily bring about changes in these two dimensions. This is a challenge for teacher PD program designers to tackle as they plan activities to support teacher growth in all three dimensions.

The aim of this research was not to study factors that influence change in task implementation quality. Rather, it examined changes in how teachers implemented tasks. However, differences between the two first-grade teachers and the other two teachers in implementation quality and in the changes taking place suggest that curriculum content and a myriad of other factors related to individual teacher background are possible routes for further inquiry.

One key difference between the two pairs of teachers was the number of years of teaching experience. Both first-grade teachers had 30 years' experience versus under 10 years' experience for the other two teachers. The former started out with low total cognitive demand scores, but the increase they achieved was greater than that of the younger teachers. Experience has previously been reported not to have clear links to task implementation quality (e.g. Stein & Kaufman, 2010), but this might not be the case when the context is a PD program. Further research is necessary to tease out relationships between individual teacher characteristics and their ability to improve the quality of their task implementation in a PD program. While experience can be a potentially influential construct, teacher knowledge needs to be included in future research on this topic (Charalambous, 2010).

Conclusion and Suggestions

This study investigated the quality of teachers' implementation of mathematical tasks in a PD program. Increases were observed in the total cognitive demand dimension of implementation quality. The small sample size does not allow generalizing about the influence of PD programs on change in teaching practice, but the detailed contextual accounts of what took place in the classrooms can inform future research.

One potential area for future research is the interrelationship between the dimensions of implementation quality and the building of PD programs that focus on these relationships. The current study provides evidence that taking this route might be fruitful. Other factors such as mathematical content and the facilitator as intervening variables can also be investigated. Even though the role of the facilitator was not a focus in this study, other researchers highlight its potential influence on PD program outcomes (Le Fevre & Richardson, 2002).

Extending this focus to mechanisms of communication and participation in PD activities would be the next step. As findings determine how teachers behave and identify which changes occur in task implementation, research could then be conducted at multiple research sites and subsequently with multiple methods.

Acknowledgements

Research in this paper was supported by a grant from Boğaziçi University Scientific Research Projects (BAP) Fund, with project code 15D01P1 and project number 9420. The author would like to thank Vuslat Şeker for her contribution in the data collection and analysis of this study.

References

Adler, J. (2000). Social practice theory and mathematics teacher education: A conversation between theory and practice. *Nordic Mathematics Education Journal, 8*(3), 31-53.

Anderson, H., Coltman, P., Page, C., & Whitebread, D. (2005). Going it alone: The CIndLe project. *Practical Pre School, 58*, 223-242.

Ball, D., & Cohen, D. (1999). Developing practice, developing practitioners: Toward a practice-based theory of professional education. In G. Sykes & L. Darling-Hammond (Eds.), *Teaching as the learning profession* (pp. 3-32). San Francisco: Jossey-Bass.

Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher, 33*(8), 3-15. doi:10.3102/0013189X033008003

Borko, H., Jacobs, J., Eiteljorg, E., & Pittman, M. E. (2008). Video as a tool for fostering productive discussions in mathematics professional development. *Teaching and Teacher Education*, *24*(2), 417-436. doi:10.1016/j.tate.2006.11.012

Boston, M. (2012). Assessing instructional quality in mathematics. *The Elementary School Journal*, *113*(1), 76-104. doi:10.1086/666387

Charalambous, C. Y. (2010). Mathematical knowledge for teaching and task unfolding: An exploratory study. *The Elementary School Journal*, *110*(3), 247-278. doi:10.1086/648978

Clarke, D., & Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. *Teaching and Teacher Education, 18*, 947-967. doi:10.1016/S0742-051X(02)00053-7

Cohen, D. K. (1990). A revolution in one classroom: The case of Mrs. Oublier. *Educational Evaluation and Policy Analysis*, *12*(3), 311-329. doi:10.3102/01623737012003311

Creswell, J. W. (2009). *Research design: Qualitative, quantitative and mixed methods approaches* (3rd ed.). Thousand Oaks, California: Sage Publications Inc.

Davis, K. S. (2003). “Change is hard”: What science teachers are telling us about reform and teacher learning of innovative practices. *Science Education, 87*(1), 3-30. doi:10.1002/sce.10037

Ersoy, Y. (2006). İlköğretim matematik öğretim programındaki yenilikler - I: Amaç, içerik ve kazanımlar. *İlköğretim Online, 5*(1), 30-44. doi:10.17051/io.86896

Fennema, E., Carpenter, T. P., Franke, M. L., Levi, L., Jacobs, V., & Empson, S. (1996). A longitudinal study of learning to use children’s thinking in mathematics instruction. *Journal for Research in Mathematics Education*, *27*, 403-434. doi:10.2307/749875

Gay, L., Mills, G., & Airasian, P. (2009). *Educational research: Competencies for analysis and application* (9th ed.). Upper Saddle River, NJ:Merrill.

Guskey, T. R. (2002). Professional development and teacher change. Teachers and teaching: *Theory and Practice, 8*(3), 381-391. doi:10.1080/135406002100000512

Henningsen, M., & Stein, M. K. (1997). Mathematical tasks and student cognition: Classroom-based factors that support and inhibit high-level mathematical thinking and reasoning. *Journal for Research in Mathematics Education, 28*(5), 524-549. doi:10.2307/749690

Hiebert, J., & Grouws, D. A. (2007). The effects of classroom mathematics teaching on students’ learning. In F. K. Lester Jr. (Ed.), *Second handbook of research on mathematics teaching and learning* (pp. 371-404). Charlotte, NC: Information Age Publishing.

Hill, C. H., Blunk, M. L., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., Sleep, L., … & Ball, D. L. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: An exploratory study. *Cognition and Instruction, 26*(4), 430-511. doi:10.1080/07370000802177235

Huberman, A. M., & Miles, M. B. (1984). Rethinking the quest for school improvement: Some findings from the DESSI study. *Teachers College Record*, *86*(1), 34-54.

Le Fevre, D., & Richardson, V. (2002). Staff development and the facilitator. *Teaching and Teacher Education, 18*, 483-500. doi:10.1016/S0742-051X(02)00011-2

Milli Eğitim Bakanlığı Talim ve Terbiye Kurulu Başkanlığı. (2017). *Matematik dersi öğretim programı: İlkokul ve ortaokul*. Ankara: MEB.

Perry, N. E., & VandeKamp, K. J. (2000). Creating classroom contexts that support young children's development of self-regulated learning. *International Journal of Educational Research*, *33*(7), 821-843. doi:10.1016/S0883-0355(00)00052-5

Rimbey, K. A. (2013). *From the common core to the classroom: A professional development efficacy study for the common core state standards for mathematics* (Unpublished doctoral dissertation). Arizona State University, Arizona.

Schoenfeld, A. H. (1994). Reflections on doing and teaching mathematics. In A. H. Schoenfeld (Ed.), *Mathematical thinking and problem solving* (pp. 53-70). Hillsdale, NJ: Lawrence Erlbaum Associates.

Stein, M. K., & Kaufman, J. (2010). Selecting and supporting the use of mathematics curricula at scale. *American Educational Research Journal, 47*(3), 663-693. doi:10.3102/0002831209361210

Stein, M. K., & Lane, S. (1996). Instructional tasks and the development of student capacity to think and reason: An analysis of the relationship between teaching and learning in a reform mathematics project. *Educational Research and Evaluation, 2*(1), 50-80. doi:10.1080/1380361960020103

Stein, M. K., Correnti, R., Moore, D., Russell, J. L., & Kelly, K. (2017). Using theory and measurement to sharpen conceptualizations of mathematics teaching in the common core era. *AERA Open*, *3*(1), 1-20. doi:10.1177/2332858416680566

Stein, M. K., Grover, B. W., & Henningsen, M. A. (1996). Building student capacity for mathematical thinking and reasoning: An analysis of mathematical tasks used in reform classrooms. *American Educational Research Journal, 33*(2), 455-488. doi:10.3102/00028312033002455

Stein, M. K., Smith, M. S., Henningsen, M. A., & Silver, E. A. (2000). *Implementing standards-based mathematics instruction: A casebook for professional development* (1st ed.). New York, NY: Teachers College Press.

Stigler, J., & Hiebert, J. (2004). Improving mathematics teaching. *Educational Leadership, 61*(5), 12-16.

Tekkumru Kısa, M., & Stein, M. K. (2015). Learning to see teaching in new ways: A foundation for maintaining cognitive demand. *American Educational Research Journal, 52*(1), 105-136. doi:10.3102/0002831214549452

Ubuz, B., & Sarpkaya, G. (2014). The investigation of algebraic tasks in sixth grades in terms of cognitive demands: Mathematics textbooks and classroom implementations. *Elementary Education Online, 13*(2), 594-606.

Van de Walle, J. A., Karp, K. S., & Bay-Williams, J. M. (2009). *Elementary and middle school mathematics: Teaching developmentally*. New York, NY: Pearson Education.

Van Es, E. A., & Sherin, M. G. (2008). Mathematics teachers' "learning to notice" in the context of a video club. *Teaching and Teacher Education, 24*(2), 244-276. doi:10.1016/j.tate.2006.11.005

Weaver, B., & Koopman, R. (2014). An SPSS macro to compute confidence intervals for Pearson's correlation. *The Quantitative Methods for Psychology, 10*(1), 29-39. doi:10.20982/tqmp.10.1.p029

Wilhelm, A. G. (2014). Mathematics teachers 'enactment of cognitively demanding tasks: Investigating links to teachers' knowledge and conceptions. *Journal for Research in Mathematics Education*, *45*(5), 636-674. doi:10.5951/jresematheduc.45.5.0636

Appendix 1. Codebook for Using the Classroom Observation Coding Instrument for Tasks

(adapted from Classroom Observation Coding Instrument, Stein & Kaufman 2010)

<u>Cognitive Demand (preliminary coding)</u>

Decide on the scores by thinking about the majority of the students the majority of the time.

a. Instructional task in the lesson: ______________________

Length: ____________

b. Cognitive demand of the task as it appeared in written/resource materials (circle one):

1 No mathematical activity

2 Memorization

3 Use of Procedures Without Connections to Meaning, Concepts or Understanding

4 Use of Procedures With Connections to Meaning, Concepts or Understanding

5 Doing Mathematics

c. Cognitive demand of the task as it was set up by the teacher (circle one):

1 No mathematical activity

2 Memorization

3 Use of Procedures Without Connections to Meaning, Concepts or Understanding

4 Use of Procedures With Connections to Meaning, Concepts or Understanding

5 Doing Mathematics

d. Cognitive demand of the task as it was enacted by students and teacher (circle one):

1 No mathematical activity

2 Memorization

3 Use of Procedures Without Connections

4 Use of Procedures With Connections

5 Doing Mathematics

Decision rules for coding the cognitive demand of tasks:

- *Refer to the cognitive demand descriptions in TAG when making decisions.*
- *If set up and enactment are interwoven, code set up same as "Enactment".*
- *If there is no available written task and the set up is interwoven, code written task and set up same as "Enactment".*
- *Procedure or concept needs to be mathematically correct or accurate for the teaching to be labelled as memorization, procedures without connections, procedures with connections or doing mathematics.*

<u>Total Cognitive Demand</u>

Cognitive demand score for materials to setup + cognitive demand score for setup to enactment (possible scores from 2 to 8)

Maintenance of cognitive demand, materials to setup:

Based on coding of each observed task using the following scale:

1 point—The teacher maintained a low level of cognitive demand from one phase to the next.

2 points—The teacher transformed a task from a high level of cognitive demand to a low level of cognitive demand.

3 points—The teacher maintained a high level of cognitive demand between two phases but transformed the task from DM to PWC or from PWC to DM. Although the teacher still maintained a high level of cognitive demand, the nature of that cognitive demand essentially shifted in a way that was not consistent with the materials or the teachers' setup. Thus, a teacher received fewer points than

if he or she had maintained the same type of high-level cognitive demand from one phase to another. The rare occurrence of raising a task from a low level of cognitive demand to a high level of cognitive demand would also be credited in this category.

4 points—The teacher maintained the same high level of cognitive demand from one phase to another without transforming the task into another type of high-level demand or to a lower level of cognitive demand.

Maintenance of cognitive demand, setup to enactment:

Based on coding of each observed task; coded with the same point system as cognitive demand, materials to setup (above)

Teacher Work to Uncover Student Thinking (circle one):

0 The teacher did no work to uncover student thinking; he or she did most of the talking in the lesson and/or asked questions with short or one-word answers.

1 The teacher did some work to uncover student thinking by asking some open-ended questions, by asking for some explanations, by arranging for public sharing of student responses, and/or by listening respectfully.

2 In addition to #1 above, the teacher purposefully selected certain students to share their work during whole-class discussion because she wanted the whole class to hear about the mathematical approach the student took. However, the teacher did not sequence or connect students' responses in a mathematically meaningful way (i.e., to move the class toward the mathematical goal of the lesson).

3 In addition to #1 and #2 above, the teacher sequenced or connected students' responses in a mathematically meaningful way to make student thinking productive for the class as a whole (i.e., to move the class toward the mathematical goal of the lesson).

Decision rules for coding the teacher work to uncover student thinking:

- *The purposeful selection criteria should have judged according to what was observed in the lesson (e.g. what the teacher tells or does), not on what the teacher expressed elsewhere about the students.*
- *One difference between 1 and 2 is that in 1 teacher allows limited opportunities for uncovering student thinking, while in 2 teacher provides a balance between teacher talk and student talk in uncovering thinking.*

Intellectual Authority (circle one):

0 Judgments about correctness were derived from the text or the teacher, with no appeal to mathematical reasoning.

1 Judgments about correctness were mostly derived from the text or the teacher. Nevertheless, some appeals to mathematical reasoning were made.

2 Judgments about correctness were primarily (most of the time) derived from mathematical reasoning and discussion during the class. It was primarily (most of the time) the teacher engaging in reasoning and modeling for the students.

3 Judgments about correctness were primarily (most of the time) derived from mathematical reasoning and discussion during the class. It was primarily (most of the time) the students engaging in reasoning through the teachers' support.

Decision rules for coding intellectual authority:

- *If teachers use both mathematical reasoning and text or the teacher as the judgement criteria for correctness, their teaching should be coded as 1. This is what most of the teachers would be coded as.*