Eurasia Journal of Mathematics, Science & Technology Education, 2016, **12**(3), 387-409

Examining the Pattern of Middle Grade Mathematics Teachers' Performance: A Concurrent Embedded Mixed Methods Study

Sunyoung Han
Sungkyunkwan University, REPUBLIC OF KOREA
Sevket Ceyhun Cetin & Shirley M. Matteson
Texas Tech University, USA

•Received 27 April 2015• Accepted 28 July 2015•Published online 14 November 2015

The purpose of this study was to reveal the pattern in a structure of teachers' instructional performance. Specifically, seven teachers' performances (Standards and Objectives, Presenting Instructional Content, Lesson Structure and Pacing, Activities and Materials, Academic Feedback, Grouping Students, and Teacher Content Knowledge) were investigated. An embedded mixed methods approach analyzed 426 reports concerning 175 classroom-capturing videos. The findings showed a structured pattern and the relationships among the teachers' performances. Also, multiple group analysis was employed to examine the differences between beginning and experienced teachers' performance patterns. The beginning and experienced teachers presented different paths regarding the four performances: Activities and Materials, Grouping Students, Standards and Objectives, and Lesson Structure and Pacing. The findings of the current study have implications for teacher professional development.

Keywords: teaching performance, TAP rubric, video-based training, professional development

INTRODUCTION

Classrooms are complicated environments where teachers exhibit diverse behaviors relevant to pedagogical components such as curriculum, instructional strategies, resources, and learning activities. Teachers' behaviors in the classroom are influenced by many external and internal factors. External factors refer to the classroom environment, including physical elements such as whiteboards, computers, and projectors. Internal factors influencing teachers' classroom performances are educational ideology, curriculum, and instructional strategies. The teacher's behavior during a lesson influences other elements. For example, teachers utilizing high quality grouping strategies mitigate the loss of instructional time. In

Correspondence: Sunyoung Han,
Sungkyunkwan University, 25-2 Sungkyunkwan-ro, Jongno-gu, Seoul, Republic of Korea.
E-mail: sy.han@skku.edu
doi: 10.12973/eurasia.2016.1206a

ISSN: 1305-8223

other words, systematic patterns should be evident from the interactions of teachers' performance factors.

Investigating and understanding patterns of teaching performance is important because those patterns provide critical information for educators and policy makers in designing the vision of teacher education and practical professional development (PD) (Newman, Ridenour, Newman, & Demarco, 2003). This is why teachers' internal factors, should receive more attention when designing PD. However, most researchers have studied only isolated segments of individuals' teaching practice, rather than examining the systematic structure of teachers' repeated or multiple performances in classrooms (Alonzo, 2002; Carreker, Joshi, & Boulware-Gooden, 2010; Garet, Porter, Desimone, Birman, & Yoon, 2001; Hattie & Timperley, 2007; Hill et al., 2008; Jang, Reeve, & Deci, 2010; Stahl, 1994; Thompson, 2009). Heller, Daehler and Shinohara (2003) noted, "Any of the studies is meaningful in itself, but the sum, like a mosaic, presents a broader picture and more convincing evidence than separate pieces" (p. 36). Therefore, the present concurrent embedded mixed method study was designed to examine the pattern of teaching performance in mathematics classrooms. Specifically, the study examined the patterns involving seven areas: Standards and Objectives (S&O), Presenting Instructional Content (PIC), Lesson Structure and Pacing (LS&P), Activities and Materials (A&M), Academic Feedback (AF), Grouping Students (GS), and Teacher Content Knowledge (TCK). These areas comprise part of the Teacher Advancement Program (TAP) rubric developed by the National Institute for Excellence in Teaching (NIET, 2011).

State of the literature

- Teachers' behaviors in the classroom are influenced by many external and internal factors.
- Only isolated segments of individuals' teaching practice have been studied rather than examining the systematic structure of teachers' repeated or multiple performances in classrooms.
- Experience and beginning teachers showed differences in the dimensions of content knowledge, pedagogical knowledge, and pedagogical content knowledge.

Contribution of this paper to the literature

- Teachers' performance components during a lesson had direct and indirect relationships with each other in constructing a structural model.
- Curriculum sequence reflecting the structure of teachers' performance needs to be considered when designing professional development.
- The professional development needs of beginning and experienced teachers are not the same and require uniquely different approaches.

TEACHER'S PRACTICE (PERFORMANCE/ BEHAVIORS)

Teachers' practice and its quality has been a core topic in teacher education field. Schacter and Thum's (2004) study resulted in a new tool for measuring teacher quality and capacity to improve student achievement. The measurement tool called TAP rubric has been distributed across the U.S. and utilized for teacher's professional development. As the TAP rubric was employed to evaluate the participating teachers' performances in mathematics classrooms in the current study, it was appropriate to use the framework from the foundational studies of TAP rubric (e.g., Jerald & van Hook, 2011; Schacter & Thum, 2004) in analyzing the data. In the process of developing the measure, Schacter and Thum (2004) elaborated the factors influencing student achievement within three major components: teaching quality, teacher productivity, and classroom composition. The teaching quality was measured using 12 teaching performance standards (Schacter & Thum, 2004). Although, Schacter and Thum (2004) defined the factors evaluating effecting teaching practice, the relationship among teaching performance standards was not investigated. The current study adopted several of the teaching practice factors from Schacter and Thum's (2004) study, but was more focused on the structure of teaching performance and the embedded relationships in it. The following presents a review of literature on each teaching performance standard involved in the analyses of this study.

Content knowledge

Researchers have found that teachers' content knowledge can positively impact their knowledge of students. In a study by Heller et al. (2003), elementary teachers who attended a PD focusing on subject content were more likely to pay attention to student thinking. Teachers with more sufficient content knowledge tended to be more aware of students' understanding of content and were more likely to involve students in inquiry based learning (Alonzo, 2002; Sanders, Borko, & Lockard, 1993). Teachers with stronger content knowledge also posed more questions to encourage students' inquiries (Alonzo, 2002; Sanders et al., 1993). Teachers with weaker content knowledge tended to employ more direct instruction for students' procedural understanding rather than present activities that built conceptual understanding (Sanders et al., 1993). The teacher's knowledge of content not only indirectly impacts the teaching strategies used, but also directly influences students' knowledge. Gess-Newsome and Lederman (1995) reported that teachers tended to teach content for students in the way they earned their own knowledge.

Presenting instructional content

Presenting content in a manner that positively influences students' learning has remained as one of the major challenges for educators. No single research study has helped educators determine the features of the best method(s) for PIC because of the effects of many factors. For instance, Good and Brophy (1996) specified that effective mathematics teaching included modeling provided by a teacher, product type of questions, and having a smooth transition from the teacher modeling. The teacher held a key role in PIC in reaching well-defined learning objectives. Teacher mathematical knowledge is a critical factor on PIC (Hill et al., 2008) and education researchers have focused on how to increase and empower teachers' mathematics knowledge. Shulman (1986) defined the concept of mathematical knowledge and contended that teachers' mathematical knowledge influenced PIC. Shulman (1986) tried to answer the following question: What knowledge do teachers need to communicate effectively? In other words, what are the "most useful forms of representation...The most powerful analogies, illustrations, examples, explanations, and demonstrations—in a word, (knowing) the ways of formulating the subject that make it comprehensible to others" (Shulman, 1986, p. 9). According to Shulman's (1986) framework, teacher mathematical knowledge consisted of content knowledge, pedagogical content knowledge (PCK), and curricular knowledge that allowed teachers to utilize effective skills and proper instructional materials for each topic.

Activities and materials

Compared to the traditional classrooms that normally used a didactic instructional approach, current mathematics classrooms are employing more diverse activities and materials. In addition, the positive impacts of activities and materials on student academic achievement and attitude have been examined (Remillard, Herbel-Eisenmann, & Lloyd, 2011; Reys, Reys, Lapan, Holliday, & Wasman, 2003). However, it was also insisted that simply employing activities and materials for instruction does not guarantee its benefit to students (Tarr et al., 2008). Carbonneau, Marley, and Selig (2013) conducted a meta-analysis of the literature on the use of manipulatives for the purpose of assessing how utilizing concrete manipulatives impacted teaching mathematics in comparison to abstract symbolic instruction. The findings of Carbonneau et al. (2013) suggested that the effectiveness of manipulatives was moderated by several instructional characteristics, such as the perceptual richness of a manipulative, level of instructional guidance, and students' development status. Results from the problem-solving, aggregated, and retention data in the Carbonneau et al. (2013) study provided evidence that the support of high levels of guidance, along with the use of manipulatives, had a greater effect on student mathematic achievement, in contrast to low instructional guidance. Other researchers have also reported that teachers' high-structured instruction, including stating explicit expectations and giving clear directions about the students' learning activity, reinforced students' engagement in learning activities by mainly enriching students' awareness of control over learning outcomes and perceived competence (Jang et al., 2010). Jang et al. (2010) suggested teachers initiate learning activities to deliver expectations and instructions with great precision, as well as provide feedback in order to increase perceptions of personal control and competence.

In addition to the effect of teachers' presenting content and standards prior to classroom activities, TCK has direct influence on one's ability to identify and design appropriate activities (Carreker et al., 2010). For example, teachers who held the highest level of knowledge concerning phonemes, syllables, and morphemes, were reported to be more capable of detecting learner spelling errors and identifying the most suitable instructional activities to meet students' needs (Carreker et al., 2010). Garet et al. (2001) confirmed the findings of Carreker et al. (2010). The findings of Garet et al. (2001) indicated that the more skills and knowledge of teachers were enhanced in regard of determining curriculum standards and extending knowledge of mathematics, substantially positive changes in teaching practices were found. Changes in teaching practices included the cognitive challenge of activities in mathematics classroom and the manner in which technology (e.g., calculator or computer) was combined into instruction (Garet et al., 2001).

Even though it was examined that teacher knowledge, curriculum materials and activities were critical factors deciding the quality of instruction, there were very few studies investigating how these factors interactively influence to each other (e.g., Brown, 2009; Remillard, 2005). Therefore, it was necessary to explore how the instructional activities and materials might be affected by other instructional factors through the construct model representing the relationships among teaching performances.

Standards and objectives

Instruction that is implemented based on standards is called standards-based instruction (Green, 2007). In the best practice of teaching, the standards and objectives representing what teachers are teaching and what students should be learning are communicated in the beginning of and mentioned in the remainder of the lesson. Additionally, lesson objectives should be aligned to the standards, even though a specific standard may cover more than one lesson objective (Delandshere & Arens, 2001). The standard and objective of a lesson should be aligned to the instructional strategy and evaluation in standards-based instruction. Moreover, the teacher's assessment and evaluation plan must directly relate back to the standards. Additionally, teachers need to be clear about their expectations on student's knowledge and skills based on the standards (Green, 2007).

Standards-based instruction has tended to encourage the use of student-centered teaching strategies (Thompson, 2009). Thompson (2009) investigated the effect of standards-based instruction on mathematics and science achievement of 10,000 sixth through ninth grade students. Standards-based instruction comprised classroom activities that encourage students' participation rather than utilizing lecture-based instruction. Examples of student-centered standards-based instructional practices included incorporating scenarios from daily life events into activities in class, utilizing inquiry-based projects such as making models out of recycled materials, using manipulatives or hands-on materials (e.g. base ten blocks and algebra tiles), applying concepts of science to environmental issues, and employing technologies such as calculators and computers in classrooms. Results of Thompson's study provided rigorous support for standards-based instruction activities such as the use of hands-on and technology, inquiry-based learning, and cooperative learning-projects-based activities in math and science classrooms, as key contributors to student academic performance.

Grouping students

Grouping students effectively has emerged as an underutilized strategy of best practice. Most instructional approaches have concentrated on small groups of students (Schumacker & Lomax, 2010). For example, cooperative, project based, and differentiated learning has been implemented with small groups of students within a class. In a study examining students in small groups, an improvement in their academic achievement was found in contrast to those in non-grouping learning environment (Burris, Hubert, & Levin, 2006; Iyer, 2013). However, in some cases, grouping strategies showed a non-statistically significant or negative impact on student academic achievement (Chen, Lam, & Chan, 2008). These contradictory finding have implied that grouping strategies might influence student attitude or academic achievement differently, depending on the teacher's ability to implement a grouping strategy.

Grouping students has been regarded as a critical factor influencing the teaching practice. The relationships between the teaching strategy of GS and the other instructional factors have been studied (Hallam & Ireson, 2005; Stahl, 1994). For example, according to Stahl (1994), objectives for student learning outcomes, unambiguous task-completion instruction, and AF were essential elements of successful cooperative learning within class. Teachers' clear expression in regard of the targeted knowledge and abilities, which students were expected to obtain, helped maximize the effect of within-class grouping strategy. A teacher's concise and clear explanation on how students were to accomplish the group task was important in small group instruction (Hallam & Ireson, 2005). In addition, to ensure better small group instruction, appropriate AF was important to students (Hallam & Ireson, 2005; Stahl, 1994).

Lesson structure and pacing

Lesson planning and the actual implementation of the designed lesson are two crucial elements of effective lesson structure and brisk pacing. Student learning is dependent upon the teachers' preparation, knowledge and skills, and reflection for lesson (Simpson, 2006). Darling-Hammond et al. (2005) stated, "Decisions made explicitly and implicitly during the *planning and interactive phases of teaching* [emphasis added] influence what students learn and are influenced by the teacher's intentions for and vision of student learning" (p. 183). However, having a well-designed lesson plan does not guarantee that the implementation of the lesson will be flawless and result in high student engagement. Creating a well-designed lesson is different with implementing it in the actual classroom. For better implementation of a good lesson plan, teachers need to understand "how their goals for instruction should be related to the assignments and assessments they devise, the activities they plan, the materials they select, the feedback they give, and the ways in which they interact with students" (Darling-Hammond et al., 2005, p.183).

In addition to the structure of a lesson, another important factor is the pacing of the lesson. According to the existing research, there is a strong correlation between academic learning time, “defined as the amount of time a student spends engaged in an academic task that s/he can perform with high success” (Fisher & Berliner, 1985, p. 8), and student achievement (Andrews, 2011; Ottmar, Decker, Cameron, Curby, & Rimm-Kaufman, 2014). Academic learning time is deeply dependent on the A&M employed in a lesson. That is, an inappropriate level of A&M for students’ prior knowledge and background results in a decrease in both student attention and allocated instruction time (Wehby, Symons, Canale, & Go, 1995). For example, A&M above students’ current skills make them frustrated and disengaged (National Research Council, 1999). In addition, having too easy of a level of A&M can promote boredom, make students disengaged with the tasks (Andrews, 2011), and does not challenge students to think (Oliver & Reschly, 2007). In both of these extreme situations, it was noted that inappropriate A&M decrease the actual meaningful academic learning time.

In addition to the difficulty with student attention, getting started on lessons, making discrete transitions among activities, and inappropriate pacing caused loss of instructional time (Gettinger & Seibert, 2002). Jang et al. (2010) stated, “When teachers provide high structure by communicating clear expectations and framing students’ learning activity with explicit directions and guidance, these instructional acts support students’ engagement by keeping students on task, managing their behavior, and avoiding chaos during transitions” (p. 588).

Academic feedback

Academic feedback has been identified as one of the most potent tools that can be used to increase learning of students (Hattie, 2009). According to Hattie and Timperley (2007), the way it is given can also determine the efficiency of feedback. Sadler (1989) claimed that feedback needed to have the following three conditions to make it effective: “(a) possess a concept of the standard (or goal, or reference level) being aimed for, (b) compare the actual (or current) level of performance with the standard, and (c) engage in appropriate action that leads to some closure of the gap” (p. 121).

On the other hand, van den Bergh, Ros, and Beijaard (2013) demonstrated that, despite knowing the features discussed before (Hattie & Timperley, 2007; Sadler, 1989) and expected to be seen in classrooms, effective feedback is rarely given in actual practice. In a case study by van den Bergh et al. (2013), 32 middle school teachers’ exchanges with their students were examined and the type of feedback these teachers provided while interacting with students was investigated. According to the results of the study, about half of 1,485 teacher-student interactions contained feedback, but only 5% of the feedback was related to the learning goals of the lessons.

VIDEO-BASED PROFESSIONAL DEVELOPMENT

The use of videos as a learning tool for in-service educators has increased because of “its unique capability to capture the richness and complexity of elusive classroom practice” (Zhang, Lundeberg, Koehler, & Eberhardt, 2011, p. 454). One of the benefits of capturing videos of classroom episodes has been that the segments may be watched individually or collaboratively as a form of PD. Video viewing has allowed for the examination of the practices of the teacher, the behaviors of students, and the interactions of both groups. Video has also become a permanent record of practice, which has been beneficial as recalling what has occurred during a particular teaching episode has not always been accurate.

Video-based learning as PD has not always been regarded as an effective tool. For example, Borko, Jacobs, Eiteljorg, and Pittman (2008), in their two-year study of a mathematics PD program, found that video-focused group discussions lacked depth.

Self-critiques or comments made to peers lacked specificity and there was a general reticence in sharing something that could be perceived as negative or critical. Kleinknecht and Schneider (2013), in a study with 10 eighth grade mathematics teachers, noted differences in the behaviors between groups of teachers watching videos of others and those watching their own videos. The teachers watching their own videos were generally less critical and less able to identify alternatives to situations appearing in the videos.

Zhang et al. (2011) examined three types of video-based learning commonly used in PD: published videos, peer videos, and personal videos. The 26 participants were K-12 science teachers engaged in problem-based learning. Zhang et al. (2011) noted that the science teachers regarded viewing their own videos as being most useful. Once past the unease of watching themselves on video, 19 of the 22 teachers watched his/her video two or more times. "Video helped teachers analyze their teaching from different angles, such as the teacher, the students, content, and their interaction, and enabled teachers to see themselves as the students saw them" (Zhang et al., 2011, p. 458).

The viewing of published videos has been found by some teachers to be the least helpful (Zhang et al., 2011). As noted by van Es (2012), "if the video comes from non-participating teachers' classrooms, it can seem too distant from their practice, making it difficult to delve into the particulars of practice represented in the clips or to take away important principles" (p. 184).

Video clubs have been promoted as a productive PD tool (Borko, Koellner, Jacobs, & Seago, 2011; Sherin & van Es, 2009). Members of the video club can focus on specific concerns brought to light by watching and analyzing videos from their classrooms. When PD includes video analysis, highly adaptive approaches should be emphasized (Borko et al., 2011). In other words the PD provided should be driven by what is seen or heard on the videos. However, usually PD has focused on highly specified approaches, such as a specific strategy, which may or may not apply to a participant's teaching situation. In those instances, participants only minimally engage in the PD activities. Teachers should to be exposed to activities that focus them on specific aspects of the video and lead them into deeper discussions on content and pedagogical issues. Borko et al. (2011) stated,

> to be an effective tool for teacher leaning in PD, video representations of teaching must be selected to address specific learning goals (e.g., enhancing teachers' specialized content knowledge, improving teachers' ability to analyze students' incorrect solution strategies) and incorporated into activities designed to scaffold teachers' progress towards those goals. (p. 180)

Video-based PD has the potential to be a powerfully reflective tool. Zhang et al. (2011) contended "teachers can take fuller advantage of the power of their video when they have control over their video and can observe their video in its entirety multiple times, editing and selecting clips that are most informative for other teachers."(p. 492). Thus video-based learning can serve as both an individual and collaborative tool for reflective practice.

BEGINNING AND EXPERIENCED TEACHERS

Researchers have studied how beginning and experienced teachers showed different teaching performance (Borko & Livingston, 1989; Cardona, 2008; Cortina, Miller, McKenzie, & Epstein, 2015; Leinhardt, 1989), and there have been many studies comparing experienced and beginning teachers in general. However, there were very few studies contrasting experienced and beginning teachers in terms of teaching performance in classroom. Therefore, the current study has the purpose to reveal the pattern in teachers' performance in classrooms and to use it in designing teacher PD, as the comparison between beginning and experienced teachers' performance might contribute to developing more effective teacher PD.

Experience and beginning teachers showed differences in the dimensions of content knowledge, pedagogical knowledge, and pedagogical content knowledge (Cortina et al., 2015). Among these dimensions, the differences in pedagogical knowledge and pedagogical content knowledge between experience and beginning teachers have been illuminated (Berliner, 2001; Cortina et al., 2015). As expert teachers had more experience, they were more likely to show performances based on better pedagogical knowledge than novice teachers (Berliner, 2001; Blömeke, Felbrich, Müller, Kaiser, & Lehmann, 2008). Leinhardt (1989) compared novice and expert mathematics teachers in terms of lesson agendas, lesson segments, and model of explanation. The expert teacher in this study was identified by reviewing the students' achievement growth scores during a 5-year period. Also, the novice teachers were student teachers. They were involved in student teaching in the last semester of their certification program. In terms of the preparation in lesson agendas, experts showed much more detailed and richer plans and demonstrated more instructional actions than novice teachers. In addition, novice teachers spent more minutes for transition during the lesson and showed bigger variability in the time for guided practice than experts (Blömeke et al., 2008). Overall, experienced teachers had better skills to monitor lessons and to control the pace than beginning teachers (Cortina et al., 2015).

In terms of pedagogical content knowledge, experience and beginning teachers also revealed differences (Cortina et al., 2015). Novice teachers were also shown to have fewer cases of explanations during lessons than experts (Leinhardt, 1989). Experts were more likely to use representations and examples that were more familiar or already known to students, whereas novice teachers used new models or representations to teach new mathematics concepts. That is, experts had better knowledge of their students and were willing to employ explanations based on students' prior knowledge, which is consistent with the finding of the study by Cardona (2008). Cardona (2008) contrasted novice teachers with experts in teaching the concept of chance in mathematics classrooms. According to the findings of this study, novice teachers were more likely to underestimate students' difficulties and their explanations were not related to students' thinking, whereas experts tended to connect the teaching to students' thinking and played a role as a facilitator in lesson. In addition, novice teachers' explanations more often included errors in terms of mathematical content, which influenced students and hindered them from learning the mathematical concepts. In the sense of PIC, novice teachers were less effective and could not make students to understand why the answers were right or wrong. Finally, experts observed in the study by Leinhardt (1989) kept referring to the goals and objectives in every explanation, whereas novice teachers were not able to connect the goals to the explanation.

HYPOTHESIS DEVELOPMENT AND RESEARCH QUESTIONS

The present study examined the causal relationships among the teachers' performances based on hypotheses developed from the literature review. The specific hypotheses of the current research are:

H1. Teacher mathematical knowledge is necessary for the effective presentation in visualizing and transferring instructional content.

H2. Teachers' content knowledge influences how to design activities and how to organize materials.

H3. The presentation skill for standards and objectives decides the instructional strategies of activities and materials.

H4. Teachers' ability to present learning standards and objectives influences the skill of grouping students.
H5. The way that teachers present instructional contents affects the employed activities, and teachers' performance in presenting instructional content is also influenced by how well the employed activities and materials are implemented.
H6. Academic learning time is deeply dependent on the activities and materials employed in a lesson.
H7. Teachers' ability to present instructional content influences the skill of grouping students.
H8. Informative feedback requires the clear statement of standards and objectives.
H9. Informative feedback requires appropriate presentation of instructional content such as modeling by the teacher to demonstrate his or her performance expectations.
H10. Sufficient academic feedback allows the grouping work to be more effective.

The hypothesized relationship among teachers' performances was represented in Figure 1. Based on these 10 hypotheses, the concurrent embedded mixed methods study addressed the following research questions:

1. What is the structural model representing a pattern of relationships among teaching performances?
    a. Does the structural model confirm the hypotheses?
    b. How does the qualitative analysis provide enhanced understanding of the structural model?
2. What, if any, differences are there between beginning and experienced teachers on the paths of the structural model?

METHODS

This study was designed as a concurrent embedded mixed methods study (Creswell, 2007) in order to explore the patterns of mathematics teachers' performance practices. Secondary data collected in a federally funded research project using the TAP rubric (NIET, 2009) was used in both quantitative and qualitative phases of the study. Although the both quantitative and qualitative data

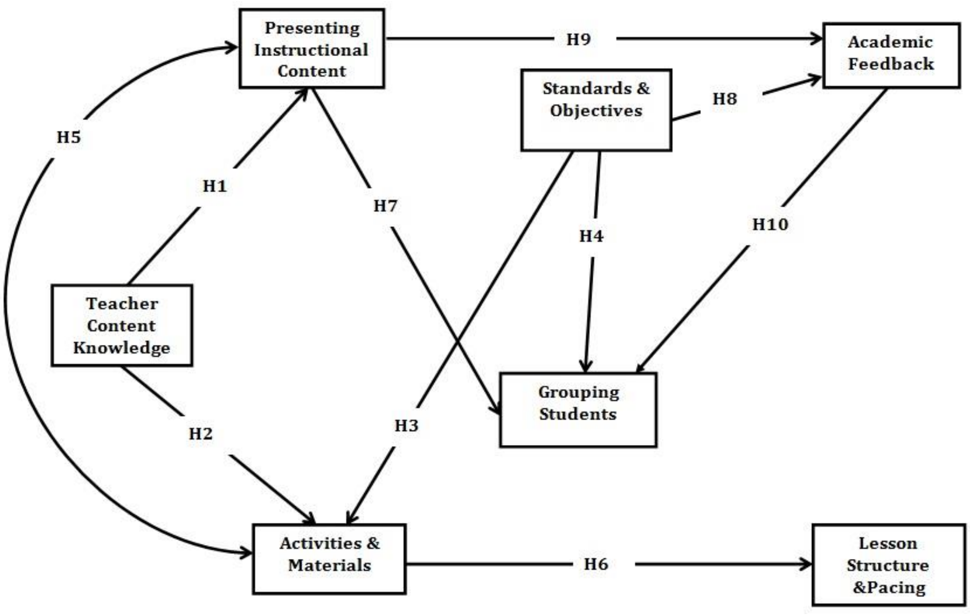

Figure 1. Initial model reflecting 10 hypotheses

were collected at the same time, the qualitative data is embedded into the quantitative data and so the priority of the mixed method study was given to the quantitative phase.

The rationale for using mixed methods in this study is that employing qualitative research assisted in the interpretation of the quantitative findings. Utilizing quantitative research in the first phase was valuable in testing the ten hypotheses, which were developed based upon the related literature on teaching practice. The observation scores collected from teachers' classroom videos were used in this phase. In the process of quantitative analysis, some relationships among TAP indicators were found to be significant, but those were not hypothesized in the quantitative phase. Therefore, the qualitative research was utilized as the second phase of the study by using written comments of raters as the qualitative data in order to explain the relationships among the TAP indicators. In other words, in the second phase of the study, the raters' comments were analyzed to enhance the understating of the quantitative results of the first phase. Using the concurrent embedded mixed method study provided a complete understanding of the patterns of teaching practices by analyzing quantitative data first and supporting the findings with qualitative research findings (Punch, 1998).

Context of the study

This study was conducted as part of a federally funded research project to investigate the patterns of mathematics teachers' performance practices. The foundation of the project was that teachers articulate their competency in the domains of content and pedagogy when they reflect on their own practice in light of student learning. It has been noticed that "reflective practice is a process that helps teachers think about what happened during a classroom lesson, why it happened, and what could be done next time to make it happen more successfully" (Hiebert, Morris, Berk, & Jansen, 2007, p. 50). Therefore, the purpose of this project was to boost the mastery of grades six through nine mathematics teachers regarding to pedagogy and content knowledge (subject-area) by providing Competency-based feedback through multiple rubric-scored observations.

Data collection procedures

Participants

Participation was voluntary in the federally funded research project. During the 2012-2013 school year, a total of 35 grades six through nine mathematics teachers from an urban school district collaborated with state university researchers, both located in the southwestern U. S. Twenty-nine of the teachers taught at one of five middle school campuses, and the remaining six teachers taught at one of the two high school campuses in this particular school district. 60% of the teachers were primarily beginning educators, who had zero-five years of experience as shown in Table 1. The multiple rubric-scored observations of teachers participated in the project were used as the data source of this present study. In other words, the data source consisted of existing data; therefore, the sampling of the present study was convenience sampling (Creswell, 2007).

The participating teachers agreed to video one lesson of their choice for each of the first five 6-weeks grading periods. In this particular research project, each 6-weeks grading period was defined as a cycle; thus each teacher videoed five lessons in total of five cycles through the school year. The videoed lessons ranged about 45 minutes to 80 minutes, depending upon each school schedule. In each cycle, videoed lessons were viewed and scored by the teachers, master coach and university personnel using TAP rubric. Mathematics coaches and university personnel were

Table 1. Demographics and responsibilities of raters in research project

Raters	Demographics	Responsibilities
35 Teachers	Teaching experience of the teachers were, 0-5 years: 21 teachers 6-10 years: 5 teachers 11-20 years: 5 teachers 20 and more: 4 teachers	Each teacher in the research project was charged with, • videoing a lesson per cycle • viewing and scoring his/her own video lesson
7 Mathematics Coaches (MCs)	5 middle school MCs 2 high school MCs *All the coaches, who had been mathematics teachers prior to their selection for that position, were recruited by the particular school district.	Each MC was charged with, • leading professional communities • disaggregating data • scoring videoed lessons of teachers • providing support to the teachers.
5 University Personnel (UP)	2 doctoral students in mathematics education were assigned to one high school and one middle school, which was the feeding school of the assigned high school. 3 master students were assigned to the remaining three middle schools. *All the UP had teaching experience prior to their selection for that position.	Each UP was charged with, • viewing and scoring assigned teachers' video lessons • providing feedback to the campus coach of his/her assigned school(s).

other two groups of raters that were assigned to specific school campuses. The demographics and responsibilities of the MCs and UP have been reported in Table 1.

In the research project, university personnel was charged with sharing their reports with the campus mathematics coach, who then merged the information with their TAP rubric scores and comments. The mathematics coach determined refinement and reinforcement areas among the seven indicators and relayed that information to the teacher. The coach then gave constructive feedback to the teacher with the purpose of increasing his/her pedagogy and content knowledge competency.

Data sources

The measurement instrument used in this study was adopted from rubric (NIET, 2009), which has been examined in terms of its reliability and validity (Jerald & van Hook, 2011). The reliability and validity concern regarding rubric has been addressed by requiring all evaluators to undergo training and pass a certification test (Jerald & van Hook, 2011). The rubric consists of three dimensions (i.e., Instruction, Designing and Planning Instruction, and The Learning Environment). In this present study, we used the dimension of Instruction with seven indicators. The selection of the seven indicators was based on discussions among experts in TAP rubric, regional educational administrators, and mathematics educators. In the research project, university personnel and all the mathematics coaches participated in three days of NIET training on the TAP rubric and were certified as an observer and rater after passing the certification exam. The scores and comments of the raters were the data sources for the quantitative and qualitative data as explained below in details.

Quantitative data

The rubric used a rating scale of 1-5. The score of "1" meant "Unsatisfactory (performance)," "3" meant "Proficient," and "5" was "Exemplary. For each video, each rater submitted separate numerical scores for each of the seven indicators. The amount of data collected included 426 reports that were written for the 175-videoed lessons of the 35 mathematics teachers.

Qualitative data

In addition to the numerical scores, the three different rater groups (teacher, master coach, and university personnel) provided detailed comments based upon what teachers and students said or did in the videoed lessons as justifications of their scores. When reviewing the qualitative data, we noted that the written comments of mathematics teachers and mathematics coaches were not as detailed as the comments of university personnel. Therefore, the researchers, who hold expertise in qualitative research, decided to analyze only the reports written by university personnel for the qualitative phase of the mixed method study to guarantee the validity of qualitative findings. Among the total 133 written reports of UP, 48 reports (36%) were selected as the data of qualitative phase by using diverse techniques of case-selection, which requires researchers to group cases first, then select at least two cases from each group by using statistical or non-statistical approach. One of the non-statistical approaches is choosing extreme cases from each group, which was defined as "well advised" approach (Gerring, 2007, p. 98). The purpose was making the large qualitative data sets more manageable. According to Gerring (2007), while selecting cases from large qualitative data sets, two objectives, representativeness and variation, need to be achieved. To achieve these two objectives in this study, the researchers grouped all reports of UP into school campuses and reporting periods. While grouping into school campuses, the six high school teachers from two high schools were combined into one group. Then, the reports of highest and lowest performing teachers of each cycle were selected based on the scores across the seven indicators for each school. In total, eight teacher reports (4 high and 4 low performing teacher reports) were selected from each one of the five middle school campuses and the high school group. As a result, the total of 48 university personnel reports consisted of the data set of the qualitative phase.

Data analysis

Quantitative

This study used structural equation modeling (SEM) using Mplus 7.0 (Muthén & Muthén, 2012). To estimate the coefficients, a structural model was hypothesized based on the theoretical framework and represented using a path diagram. Also, to evaluate the measurement model, we employed four fit indices: Chi-square (χ^2) values and degrees of freedom (df), Root Mean Square Error of Approximation (RMSEA), Comparative Fit Index (CFI), and Standardized Root Mean Square Residual (SRMR). Multiple indices were utilized to resolve the complexity in assessing the fit of a model (Schumacker & Lomax, 2010). In addition, group difference analysis was utilized to compare the models between experienced and novice teachers.

Qualitative

The analysis was completed by one of the UP, who holds expertise in qualitative research. The qualitative data analysis was conducted in support for the quantitative research findings by exploring the relationships among indicators that were found significant, but were not hypothesized in the quantitative phase of the concurrent embedded mixed method study. Therefore, constant-comparison method (Glaser & Strauss, 2012) was employed to examine the recurrent words and phrases, known as open coding (Corbin & Strauss, 2008) across the indicators in each selected written report. Next, the recurrent words and phrases across indicators were utilized to determine the major categories, which were then divided, merged, and developed for further analysis, known as axial coding (Corbin & Strauss, 2008).

In qualitative studies, the rigor of the study is addressed with the trustworthiness of the data analysis (Lincoln & Guba, 1985). Credibility for the study has been demonstrated through peer debriefing and the prolonged engagement of the researchers with the data. The researchers have analyzed the data across multiple cycles, discussed the codes with each other, recoded data as needed, and debriefed the excerpts used in the qualitative findings. To address dependability, the comments from different reports, which were written by different university personnel, have been used across multiple cycles. Confirmability has been addressed through the audit trial coding system in which the raters, university personnel (UP), were coded as 'UP 1' to 'UP 5'. The initials of TAP indicators were coded, such as GS (Grouping students), and then the page number of the comments or excerpts in the data set was reported (i.e., UP2, A&M, p. 17). Also, transferability has been demonstrated through the description of the data, the purposeful selection of the 48 reports (24 reports each for high and low performing teachers) from the 380 reports and the usage of excerpts taken from university personnel reports.

Multiple group path analyses

To answer the research question, "Does group membership depending on teachers' teaching experience moderate the relations specified in the model?" multiple group path analyses was employed (Kline, 1998). Three multiple group path analyses were conducted. The first multiple group path analysis was to test if there was an overall difference between beginning and experienced teachers. Next, a series of 16 separate multiple group path analyses was conducted to identify those paths on which beginning and experienced teachers differed significantly from one another. Finally, a third multiple group path analysis, which incorporated the results of these 16 separate analyses, was conducted. The third analysis allowed confirmation that beginning and experienced teachers presented statistically significantly differences on some specified paths from one another, whereas the first analysis only informed how much the assumed baseline model was aligned to each beginning and experienced teacher group. The teachers were grouped by the criteria of teaching experience. That is, teachers who taught six years or more were assigned to experienced teachers and teachers who had less than six years were assigned to beginning teachers.

RESULTS

Structural model: Quantitative analyses

The correlation coefficients between variables were reported in Table 2. To reveal the structural model representing relationships among teaching performances, path models with observed variables were tested. The initial model based on 10 hypotheses showed poor fit (χ^2=351.339, *df*=9, non-significant; RMSEA=0.299; CFI=0.710; SRMR=0.181). To adjust the model, some paths were added and removed according to the results from correlation coefficients and fit indices. The four model fit indices of the final model determined that this model provided a more than adequate fit to the data (χ^2=10.393, *df*=5, non-significant; RMSEA=0.050; CFI=0.996; SRMR=0.019) (see Figure 2).

The structural coefficients depicted to what extent each indicator affected the other indicators (see Table 3). Consistent with H1 and H2, TCK impacted to teaching

Table 2. Correlation coefficients between variable of teacher performance

Variables	A	B	C	D	E	F	G
(A) TCK	1.000						
(B) PIC	0.334**	1.000					
(C) A&M	0.355**	0.613**	1.000				
(D) LS&P	0.315**	0.591**	0.551**	1.000			
(E) S&O	0.264**	0.634**	0.588**	0.556**	1.000		
(F) GS	0.240**	0.497**	0.560**	0.468**	0.515**	1.000	
(G) AF	0.342**	0.589**	0.488**	0.519**	0.540**	0.450**	1.000

*Note. **p<0.01*

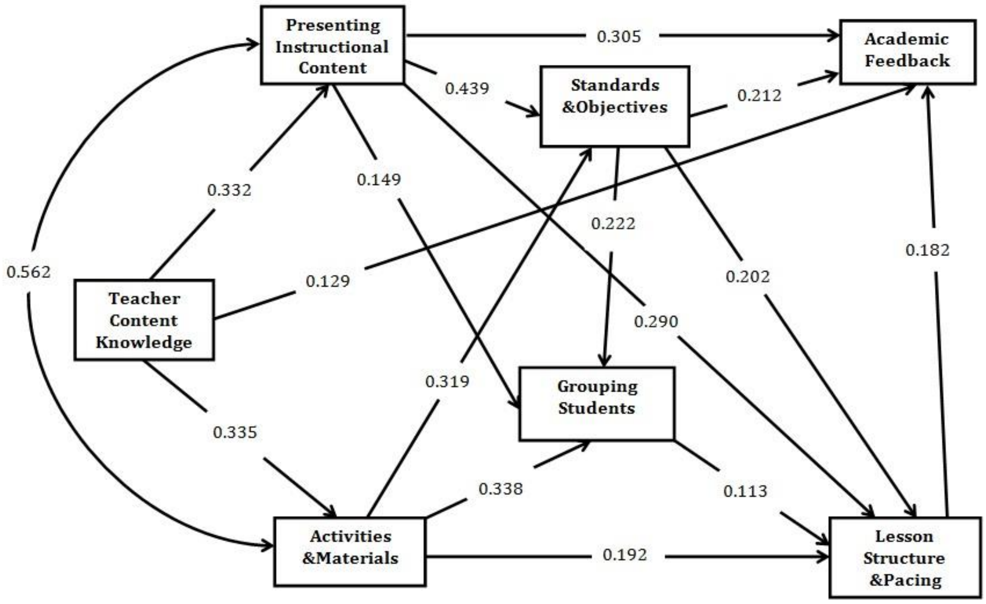

Figure 2. Structural equation model of patterns of mathematics teachers' performances

Table 3. Estimates for the paths in the Structural Model

Path	Estimates	Standardized estimates	Standard error	*p*-value	Hypothesis
TCK→PIC	0.415	0.332	0.045	<0.001	H1
TCK→A&M	0.460	0.355	0.046	<0.001	H2
PIC→S&O	0.459	0.439	0.042	<0.001	–
A&M→S&O	0.322	0.319	0.044	<0.001	⇔H3
PIC→GS	0.169	0.149	0.053	0.005	H7
A&M→GS	0.370	0.338	0.050	<0.001	–
S&O→GS	0.241	0.222	0.052	<0.001	H4
PIC→LS&P	0.273	0.290	0.050	<0.001	–
A&M→LS&P	0.174	0.192	0.051	<0.001	H6
S&O→LS&P	0.182	0.202	0.050	<0.001	–
GS→LS&P	0.094	0.113	0.046	0.014	–
TCK→AF	0.171	0.129	0.042	0.002	–
PIC→AF	0.324	0.305	0.051	<0.001	H9
S&O→AF	0.216	0.212	0.049	<0.001	H8
LS&P→AF	0.205	0.182	0.048	<0.001	–
PIC ⇔ A&M	0.306	0.562	0.034	<0.001	H5

performances such as PIC and A&M. The impacts of TCK on PIC (β=0.332) and A&M (β=0.335) were statistically significant. The path from TCK to AF (β=0.129) was also statistically significant in the final model, which was not hypothesized based on literature. In addition, GS was found to significantly relate to PIC (β=0.149), A&M (β=0.338), and S&O (β=0.222). The impacts of PIC and S&O on GS were supported by H7 and H4, respectively. The relationship between A&M and S&O showed a different direction comparing to H3. As expected in H5, teacher's PIC and A&M skills influenced interactively (β=0.562). PIC and SO had statistically significant impacts on AF (β=0.305, β=0.212, respectively), which were supported by H9 and H8, respectively. Finally, as aligned to H6, A&M was found to significantly impact on LS&P (β=0.192).

Moreover, the statistically significant relationships between PIC and S&O (β=0.439), TCK and AF (β=0.129), S&O and LS&P (β=0.202), A&M and GS (β=0.338), PIC and LS&P (β=0.290), GS and LS&P (β=0.113) and LS&P and AF (β=0.182) were found, which were not hypothesized.

Additional evidence: Qualitative analyses

The subset of 48 purposefully chosen reports formed the foundation of the more detailed analysis to support the quantitative results of the first phase. The frequency of the relationships that occurred in multiple reports were counted and reported in Table 4. The top three relationships (A&M-GS, TCK-AF, and PIC-S&O) were observed in more than half of the 48 reports. Therefore, the most representative excerpts have been utilized to explain these three relationships. The researchers selected the excerpts after several peer-debriefing sessions.

Activities & materials → grouping students

The relationship between A&M and GS is the highest reported relationship among the seven non-hypothesized relationships. According to detailed qualitative analysis, in 32 of the total 48 reports, the raters reported that the activities used while teaching mathematics provided opportunities for student-to-student interaction so that the activities entailed GS. For example, in one of the reports, a rater noted that the provided activity forced students to work individually and with their group mates in order to complete the task. The rater explained:

> The group work activity was challenging for students. They [Students] had to work with partners and utilize prior learning to complete the assignment. Also, they had to verbally explain their reasoning behind the answer to this problem… Students had a great deal of time for student-to-student interaction… They had to participate during the group work, and they had to fill out their worksheet for each problem that was presented. (UP2, A&M, p. 17)

The opportunity for students to work in small groups was also reflected in the GS indicator of the same report. The activity provided students with a role in their group and forced them to contribute to the group work. In light of that, the activity

Table 4. Frequency of the seven relationships in the qualitative data

Relationships between indicators	Frequency of the relationships (N=48)	Percentage (N=48)
A&M→GS	32	67%
TCK→AF	25	52%
PIC→S&O	25	52%
GS→LS&P	13	27%
A&M→LS&P	10	21%
LS&P→AF	8	17%

held each member of the group accountable to complete the group work.

In contrast, the rest of the reports (16 out of 48) noted that the individually based adopted or prepared activities provided for very little interaction among students; therefore, no group work was evident in these reports. For example, another rater stated in the AM indicator of a report, "There were three activities: the warm-up, the paper activity, and the closure activity… These activities did not provide a student to student interaction." (UP1, A&M, p. 63) Also, without the student-to-student interaction, the rater revealed its negative impact on GS in the GS indicator of the same report: "From beginning to the end of the lesson, the class was like a whole group. Students usually worked on the activities individually." (UP1, GS, p. 69)

In summary, the qualitative data analysis revealed students had opportunities to work in groups based on the nature of the activities that were adopted or prepared by the mathematics teachers.

Teacher content knowledge → academic feedback

Through the qualitative data analysis, evidence showing the effects of TCK on AF was found in 25 out of 48 reports. Analysis of these 25 reports indicated that teachers who had strong content knowledge provided AF with high quality. Use of appropriate academic language and having sufficient content knowledge seemed to be crucial components of academically focused feedback. Furthermore, when clarifying the confusion of students on the procedural and conceptual knowledge, TCK seemed noteworthy when providing AF in the right place at right time and was commonly provided through a dialogue with students, instead of just transferring knowledge. For example, in one of the lessons, it was noted that students were confused with isolating an unknown (x) on one side of an equation to solve the given equation (x+2 = 5). The dialogue between the teacher and students were:

> Teacher: You know by looking at that I want the x by itself isolated, so if I said you need to use either an addition or subtraction to maintain equality what would you do?
> Student: Addition.
> Teacher: Addition?
> Students [more than one student answered]: No
> Teacher: Ok. I like that you said that, so what would I add to get x by itself? What ideas do you have? What could I add to both sides?
> Student: You could add what's on the left to the right.
> Teacher: Ok, what is this right here? [x+2]
> Student: 2
> Teacher: 2? Couldn't I add something to make zero pairs and make them go away? [The teacher explained both methods using addition of negative numbers and using subtraction of positive numbers to isolate x on the right side.] (UP3, AF, p. 34)

The conversation between the teacher and students was also placed in the TCK indicator. The rater used the same evidence to show how accurately the teacher explained how to isolate the variable. It was also noted that "The teacher used the following terminology correctly and accurately: reciprocal, inverse operation, equation, expression, multiplication, division, addition, subtraction, and zero pairs. All notations on the board were accurate as the teacher solved equations." (UP3, TCK, p. 35)

Presenting instructional content → standards & objectives

The relationship between PIC and SO was evident in 25 out of 48 reports. In half of these reports, the raters noted that having logical sequence throughout a lesson provided opportunities to revisit and review students' prior learning and use the learning as the foundation for newly taught concept(s). For instance, the objective of a lesson was, "I can write, solve, and justify a linear system," and a rater reported the relationship between sequence and prior learning in the following excerpt:

> The sequencing and segmenting of the lesson was logical. The lesson began with a bell ringer that was used to remind students' prior knowledge about slope-intercept form, continued with a worksheet that was used to teach how to write, solve and justify a solution of a linear system, finished with a closure… (UP1, PIC, p. 7)

Another crucial element of PIC emphasized in the majority of the reports was modeling that was provided by teachers to demonstrate their performance expectation(s) from students. The effects of modeling on explaining to students what they are expected to do were reported in the following excerpt:

> The teacher provided various models of how to solve one-step equations throughout the lesson to demonstrate performance expectations.
> Teacher: x+2 = 5…. [T]here is an equal sign on your paper so you are going to put x+2 one side and 5 on the other, and I know what you are thinking…tell me what you put in..?
> Student: Green bar.
> Teacher: Green bar plus..?
> Students: 2 tiles.
> Teacher: And then..?
> Student: Five square blocks, no negatives.
> Teacher: No negatives, so you know by looking at that I want the x by itself isolated, so if I said you need to use either an addition or subtraction to maintain equality what would you do? (UP3, PIC, p. 38)

In contrast, lack of modeling reported in 8 reports caused the expectations to be unclear and made students confused about what they were supposed to do in a given task. The negative effects of absence of modeling were reported for the expectations of individual and group works as well. For instance, a rater provided the following quote as evidence of the absence of modeling and what it caused. "Teacher did not model any instruction or what needed to be done throughout the lesson of what was expected at the end of lesson. Students seemed lost." (UP4, PIC, p. 91)

Multiple group path analyses

Based on the findings through the quantitative and qualitative analyses using observed scores on teaching performance, the best fitting overall baseline model was designed (Figure 2). The first multiple group path analysis was conducted to test whether the overall baseline model would differ when this model was applied to beginning and master teachers separately. The model fit indices for the first multiple group path analysis indicated a good model fit (χ^2=18.455, *df*=10, *p*=0.048; RMSEA=0.064; CFI=0.992; SRMR=0.027). For the beginning teacher group, the path from GS to LS&P was not statistically significant, whereas for the experienced teacher group the paths from GS to LS&P, S&O to LS&P, LS&P to AF, and PIC to GS were not statistically significant.

To examine the paths on which beginning and experienced teacher group showed significant difference in terms of factor loading, 16 constrained path analyses were employed. In each path analysis, one path was constrained. The result of each constrained model was compared to that with all paths freed. The results of 16 chi-square tests indicated that beginning and experienced teachers differed significantly from one another on the paths leading from AM to GS and S&O to LS&P (p<0.05)

Table 5. Results of Chi-Square tests

Path	Chi-Square	df	p
All Freed	18.455	10	
TCK->PIC	19.414	11	0.327
A&M<->PIC	19.293	11	0.360
TCK->A&M	19.087	11	0.427
PIC->S&O	18.478	11	0.879
A&M->S&O	18.814	11	0.549
PIC->GS	19.031	11	0.448
A&M->GS	26.642	11	0.004*
S&O->GS	19.14	11	0.408
PIC->LS&P	19.083	11	0.428
AM->LS&P	18.959	11	0.478
SO->LS&P	23.799	11	0.021*
GS->LS&P	18.67	11	0.643
TCK->AF	18.455	11	1
PIC->AF	18.476	11	0.885
S&O->AF	19.313	11	0.354
LS&P->AF	18.567	11	0.738

Note. * $p<.05$.

(see Table 5). The final multiple-group path model specifying these two constrained paths was compared to the model having all 16 paths constrained. The result of this chi-square difference test was statistically significant (p=0.01).

DISCUSSION

Without question, the teacher is a critical factor in student learning. Concerning best practices of teaching, factors influencing teachers' performance during a lesson have been studied (Alonzo, 2002; Carreker et al., 2010; Garet et al., 2001; Hattie & Timperley, 2007; Hill et al., 2008; Jang et al., 2010; Stahl, 1994; Thompson, 2009). However, few studies have investigated the internal relationships among the teachers' behavioral factors. Thus, this study is important in terms that it contributes to the scholarly significance on understanding the pattern of mathematics teachers' teaching practice. This is the first study investigating internal factors influencing teachers' performance during lessons. We assumed that there should be causal relationships among the teachers' behaviors through the review of literature. The findings regarding the pattern in a structure of teachers' performance and the relationships among teachers' behaviors within the structure will provide theoretical and practical implication for effective teacher PD.

This study demonstrates the relationships among teachers' performance components during a lesson. In terms of the relationship among the seven indicators, the results from the path analyses generally support the theoretical literature and numerous prior studies (Alonzo, 2002; Carreker et al., 2010; Garet et al., 2001; Hattie &Timperley, 2007; Hill et al., 2008; Jang et al., 2010; Stahl, 1994; Thompson, 2009). The previous studies on teachers' practice were more focused on direct effect between two indicators. For example, in the study by Sanders et al. (1993), the relationship between TCK and teaching activity was mainly discussed. On the other hand, in the present study the relationship among the seven indicators was investigated in one model simultaneously. Therefore, it was possible to see the sequential indirect effect among the components of teachers' practices. For instance, the effect from TCK to SO might be explained through the paths from TCK to PIC and from PIC to SO (or from TCK to AM and from AM to SO), even though it was not found in the model of this study. This is the benefit from the comprehensive model including more teachers' performance components as consisted by Heller et al. (2003).

By contrast, the present study showed some distinctions from previous research (Garet et al., 2001; van den Bergh et al., 2013) or uncovered relationships that had not been found in the previous studies. Garet et al. (2001) demonstrated that teachers having better teaching content knowledge tended to design a lesson, including well-aligned sub-objectives, which were connected to students' prior knowledge. This indicates that there is a relationship between TCK and S&O in rubric. However, there was not a direct path from TCK to S&O in the model developed. Also, the effect from S&O to AF was inconsistent with the finding of van den Bergh et al.'s (2013) study. That is, van den Bergh et al. (2013) explained that not every AF provided was related to the learning goal of the lesson, whereas the finding from the current study found that teachers' skills on S&O influence their skills on AF. In addition, some findings were not hypothesized from the literature review. However, these findings were supported through the analysis of the comments that supported the rubric ratings. The qualitative analysis of comments supported the findings of strong relationships for A&M to GS, TCK to AF, and PIC to S&O, although we did find evidence of other possible relationships.

The employed mixed method approach increases the reliability and validity of the findings of this study. The findings from quantitative and qualitative data and analyses are complementary to each other. As the main topic of this study is teachers' practice during a lesson, and there are diverse factors that influence their behaviors, we could not only rely on quantitative analyses. In order to make sense of and contextualize the findings, we also had to analyze the qualitative data to fully understand a generalized structure of teaching performance by quantitative analyses.

An important implication of the study is that curriculum sequencing needs to be considered for designing PD. In the sense that the structural model in this study is developed based on teachers' actual performances, the sequence among the performances might need to be considered in designing teacher PD. For example, we suggest additional training PIC and A&M as the first focus because PIC and A&M were connected with other following teaching performances. As consistent with Brown (2009) and Remillard et al. (2011), employing A&M was not independent, but relevant to other teaching performances. Teachers' TCK affected how to utilize A&M in mathematics lessons (Carreker et al., 2010), and the employed A&M influenced other factors such as S&O, GS, and LS&P. That is, training A&M at the first step might help provide teachers with the overview of instructional factors. Causal relationships can be used as resource to decide the subsequence of components for the effective teacher PD. According to Reigeluth (1999), the sequence of the curriculum for teacher PD makes a difference, thus further and future investigations are warranted. Experimental research comparing effects of PDs having different sequences might be possible.

The findings from the multiple group path analysis imply that PD needs to be prepared for each beginning and experienced teacher group. The difference between beginning and experienced teachers on the path from A&M to GS might be indicative that experienced teachers are more likely to have detailed plans on instructional actions and activities than beginning teachers (Leinhardt, 1989). That is, whenever experienced teachers plan A&M of a lesson, they already consider the appropriate grouping strategy. Also, experienced teachers better understand students' difficulties and their explanations are more related to students' background (Cardona, 2008). Thus, many other factors may have contributed to the increase of the experienced teachers' loading coefficient on the path from A&M to GS.

There was also a difference between beginning and experienced teachers on the path from S&O to LS&P. As opposed to the path from A&M to GS, the beginning teachers' path showed a higher path factor-loading coefficient on the path from S&O to LS&P. This finding indicates that beginning teachers show bigger variability in terms of LS&P and are affected by other factors, especially from the practice regarding S&O in this study, which was consistent with the finding in Leinhardt (1989). One of the descriptors in the S&O indicator is "Sub-objectives are mostly aligned to the lessons' major objective" (NIET, 2011, p. 14), which is strongly related to the LS&P indicator. In other words, when teachers present aligned sub-objectives during a class, they are more likely to have a coherent and well-paced instruction. However, experienced teachers might not display the sub-objectives during a class, but have them in their mind, thus their instructional structure is not affected from the mentioned S&O. This suggests that training on sub-objectives and how to state them clearly may help beginning teachers to remember the objectives and create a more coherent structure throughout the lesson.

Additionally, the statistically insignificant path loading from S&O to LS&P underlines that experienced teachers are less likely to be affected in terms of LS&P by other factors. As consistent with literature, experienced teachers are able to monitor class progress and have better flexibility in implementing a designed lesson plan (Blömeke et al., 2008; Cortina et al., 2015). Moreover, experienced teachers have better pedagogical content knowledge than novice teachers (Berliner, 2001; Blömeke et al., 2008) and thus they might tend to have more effective teaching strategies for each specific mathematics content, which will prevent them from wasting time during the class.

The video-based PD allowed the collection of rich quantitative and qualitative data. Video can be effective and reflective tool for capturing the teachers' classroom practices (Zhang et al., 2011). The videos enabled the analysis of teacher's performances and the relationship of behaviors during the performance. From experience we know that it is difficult to observe live lessons and make detailed notes that provide evidence of teacher behaviors and support the rubric score. However, video-captured lessons permit raters to watch the videos and look at the teachers' performance repeatedly if necessary. This practice also supports the reliability of ratings across multiple raters and strengthens the findings of this study. On the other hand, we also found that comments made by teachers for self-evaluation lacked depth and detail, which was pointed out by Borko et al. (2008) and Kleinknecht and Schneider (2013). For this reason, we excluded the teachers' self-comments in analyzing qualitative analyses, whereas their self-reported numerical scores were included in the quantitative analyses.

It is also important to address the limitations of the present study and to provide suggestions for further study. The impact of video-based PD on teachers' performance was not an interest of the study, therefore the improvement of teachers' practice was not considered in the analyses. However, self-review of video was indicated as the most useful video-based PD approach (Zhang et al., 2011). That is, teachers were more comfortable to open their classes rather than the classes were observed by others and felt less anxiety with video-capturing their classes. Therefore, we suggest a further study examining the impact of video-based PD on teachers' performance when teachers are made aware of the relationships between specific rubric indicators and look for specific examples in their own videos.

ACKNOWLEDGEMENT

The data of this study is based upon the work supported by the U.S. Department of Education Investing in Innovation (i3) under grant number u411c110102. Any opinions, findings, and conclusions or recommendations expressed in this document are those of the authors and do not necessarily reflect the views of the U.S. Department of Education. We want to acknowledge the contribution by all the members of the RE: FLECT project team. We also want to thank the teachers who volunteered to participate in this project.

REFERENCES

Alonzo, A. C. (2002). *Evaluation of a model for supporting the development of elementary school teachers' science content knowledge*. Retrieved from http://files.eric.ed.gov/fulltext/ED465613.pdf

Andrews, M. C. (2011). *Meaningful engagement in educational activity and purposes for learning*. Stanford University.

Berliner, D. C. (2001). Learning about and learning from expert teachers. *International Journal of Educational Research, 35*, 463–482.

Blömeke, S., Felbrich, A., Müller, C., Kaiser, G., & Lehmann, R. (2008). Effectiveness of teacher education. *ZDM Mathematics Education, 40*, 719–734.

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*, 417-436. doi:10.1016/j.tate.2006.11.012

Borko, H., Koellner, K., Jacobs, J., & Seago, N. (2011). Using video representations of teaching in practice-based professional development programs. *ZDM Mathematics Education, 43,* 175-187.

Borko, H., & Livingston, C. (1989). Cognition and improvisation: Differences in mathematics instruction by expert and novice teachers. *American Educational Research Journal, 26*, 473-498

Brown, M. W. (2009) The teacher-tool relationship: Theorizing the design and use of curriculum materials. In J. T. Remillard, B. A. Herbel-Eisenmann & G. M. Lloyd (Eds), *Mathematics teachers at work: Connecting curriculum materials and classroom instruction* (pp. 17–36). New York, NY: Routledge.

Burris, C. C., Heubert, J. P., & Levin, H. M. (2006). Accelerating mathematics achievement using heterogeneous grouping. *American Educational Research Journal, 43*(1), 105-136.

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

Cardona, O. L. Z. (2008). *Teachers' Understanding of Students' Conceptions about Chance: An Expert-novice Contrast* (Doctoral dissertation, University of Georgia).

Carreker, S., Joshi, R. M., & Boulware-Gooden, R. (2010). Spelling-related teacher knowledge: The impact of professional development on identifying appropriate instructional activities. *Learning Disability Quarterly*, *33*(3), 148-158.

Chen, R. W., Lam, S. F., & Chan, J. C. (2008). When high achievers and low achievers work in the same group: The roles of group heterogeneity and processes in project based learning. *British Journal of Educational Psychology, 78*, 205–221.

Cortina, K. S., Miller, K. F., McKenzie, R., & Epstein, A. (2015). Where low and high inference data converge: Validation of CLASS assessment of mathematics instruction using mobile eye tracking with expert and novice teachers. *International Journal of Science and Mathematics Education, 13*(2), 389-403.

Creswell, J. W. (2007). *Qualitative inquiry and research design: Choosing among five approaches* (2nd ed.). Thousand Oaks, CA: Sage.

Csikszentmihalyi, M., Rathunde, K., & Whalen, S. (1993). *Talented teenagers: The roots of success and failure*. New York: Cambridge University Press.

Darling-Hammond, L., Banks, J., Zumwalt, K., Gomez, L, Sherin, M. G., Griesdorn, J., & Finn, L. (2005). Educational goals and purposes: Developing a curricular vision for teaching. In L. Darling-Hammond & J. Bransford (Eds.), *Preparing teachers for a changing world: What teachers should learn and be able to do* (pp. 169-200). San Francisco: Jossey Bass.

Delandshere, G., & Arens, S. A. (2001). Representations of teaching and standards-based reform: are we closing the debate about teacher education? *Teaching and Teacher Education*, *17*, 547-566.

Fisher, C. W., & Berliner, D. C. (Eds.) (1985). *Perspectives on instructional time.* New York: Longman.
Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal, 38*, 915-945.
Gerring, J. (2007). *Case Study Research: Principles and Practices.* New York: Cambridge University Press.
Gess-Newsome, J., & Lederman, N. G. (1995). Biology teachers' perceptions of subject matter structure and its relationship to classroom practice. *Journal of Research in Science Teaching, 32*(3), 301–325.
Gettinger, M., & Seibert, J. K. (2002). Best practices in increasing academic learning time. *Best Practices in School Psychology IV*, *1*, 773-787.
Good, T., & Brophy, J. (1996). *Looking in classrooms* (7th ed.). New York: Harper and Row.
Green, J. (2007). Standards from start to finish. *Leadership*, *37*(1), 14.
Glaser, B. G., & Strauss, A. L. (2012). *The discovery of grounded theory: Strategies for qualitative research* (7th ed.). NJ: Aldine Transaction.
Halla, S., & Ireson, J. (2005). Secondary school teachers' pedagogic practices when teaching mixed and structured ability classes. *Research Papers in Education, 20*(1), 3-24.
Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112.
Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London and New York: Routledge.
Heller, J. I., Daehler, K., & Shinohara, M. (2003). Connecting all the pieces: Using an evaluation mosaic to answer an impossible question. *Journal of Staff Development, 24*, 36–41.
Hiebert, J., Morris, A. K., Berk, D., & Jansen, A. (2007). Preparing teachers to learn from teaching. *Journal of Teacher Education, 58*(1). 47-61. doi: 10.1177/0022487106295726.
Hill, H. C., 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*, 430-511.
Iyer, R. B. (2013). Relation between cooperative learning and student achievement. *International Journal of Education and Information Studies, 3*(1), 21-25.
Jang, H., Reeve, J., & Deci, E. L. (2010). Engaging students in learning activities: It is not autonomy support or structure but autonomy support and structure. *Journal of Educational Psychology*, *102*(3), 588-600.doi:10.1037/a0019682
Jerald, C. D., & van Hook, K. (2011). *More than measurement: The system's lessons learned from designing better teacher evaluation systems*. Retrieved from http://files.eric.ed.gov/fulltext/ED533382.pdf
Kline, R. B. (1998). *Principles and practices of structural equation modeling*. New York: Guilford Press.
Kleinknecht, M., & Schneider, J. (2013). What do teachers think and feel when analyzing videos of themselves and other teachers teaching? *Teaching and Teacher Education,33*, 13-23. doi.org/10.1016/j.tate.2013.02.002
Leinhardt, G. (1989). Math lessons: A contrast of novice and expert competence. *Journal for Research in Mathematics Education, 20*(1), 52-75.
Lincoln, Y., & Guba, E. G. (1985). *Naturalistic inquiry.* Beverly Hills, CA: Sage.
Muthén, L. K., & Muthén, B. O. (2012). *Mplus: The comprehensive modeling program for applied researchers: User's guide*. CA: Muthén & Muthén.
National Institute for Excellence in Teaching. (2009). *Research summary: Updated September 2009*. Retrieved from http://www.tapsystem.org/publications/tap_research_summary_0909.pdf
National Institute for Excellence in Teaching. (2011). *Evaluation system handbook.* Santa Monica, CA:Author.
Newman, I., Ridenour, C. S., Newman, C., & DeMarco, G. M. P. Jr. (2003). A typology of research purposes and its relationship to mixed methods. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 167-188). Thousand Oaks, California: Sage.
Oliver, R. M., & Reschly, D. J. (2007). *Effective classroom management: Teacher preparation and professional development.* Washington, DC: National Comprehensive Center for Teacher Quality.

Ottmar, E. R., Decker, L. E., Cameron, C. E., Curby, T. W., & Rimm-Kaufman, S. E. (2014). Classroom instructional quality, exposure to mathematics instruction and mathematics achievement in fifth grade. *Learning Environments Research, 17*(2), 243-262.

Reigeluth, C. M. (1999). The elaboration theory: Guidance for scope and sequence decisions. In C. M. Reigluth (Ed.), *Instructional design theories and models: A new paradigm of instructional theory* (Vol. 2) (pp. 425-453). Mahwah, NJ: Lawrence Erlbaum Associates.

Remillard, J. T. (2005) Examining key concepts in research on teachers' use of mathematics curricula. *Review of Educational Research, 75*(2), 211–246.

Remillard, J. T., Herbel-Eisenmann, B. A., & Lloyd, G. M. (Eds.). (2011). *Mathematics teachers at work: Connecting curriculum materials and classroom instruction*. New York, NY: Routledge.

Reys, R., Reys, B., Lapan, R., Holliday, G., & Wasman, D. (2003). Assessing the impact of *standards*-based middle grades mathematics curriculum materials on student achievement. *Journal for Research in Mathematics Education, 34*(1), 74–95.

Sadler, R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, *18*, 119-144.

Sanders, L. R., Borko, H., & Lockard, J. D. (1993). Secondary science teachers' knowledge base when teaching science courses in and out of their area of certification. *Journal of Research in Science Teaching*, *30*, 723-736.

Schacter, J., & Thum, Y. M. (2004). Paying for high-and low-quality teaching. *Economics of Education Review*, *23*, 411-430.

Schumacker, E. S., & Lomax, R. G. (2010). *A beginner's guide to structural equation modeling* (3rd ed.). New York: Routledge.

Sherin, M. G., & van Es, E. A. (2009). Effects of video club participation on teachers' professional vision. *Journal of Teacher Education, 60,* 20-37.doi: 10.1177/0022487108328155

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational researcher*, *15*(2), 4-14.

Stahl, R. J. (1994). The essential elements of cooperative learning in the classroom. *ERIC Digest*. Retrieved February, 2014 from http://files.eric.ed.gov/fulltext/ED370881.pdf

Thompson, C. J. (2009). Preparation, practice, and performance: An empirical examination of the impact of standards-based instruction on secondary students' math and science achievement. *Research in Education*, *81*, 53-62.

van den Bergh, L., Ros, A., & Beijaard, D. (2013). Teacher feedback during active learning: Current practices in primary schools. *British Journal of Educational Psychology*, *83*(2), 341-362.

Van Es, E. A. (2012). Examining the development of a teacher learning community: The case of a video club. *Teaching and Teacher Education, 28,* 182-192. doi:10.1016/j.tate.2011.09.005

Wehby, J. H., Symons, F. J., Canale, J. A., & Go, F. J. (1998). Teaching practices in classrooms for students with emotional and behavioral disorders: Discrepancies between recommendations and observations. *Behavioral Disorders, 24*(1), 51-56.

Zhang, F., Lundeberg, J., Koehler, M. J., & Eberhardt, J. (2011). Understanding affordances and challenges of three types of videos in professional development. *Teaching and Teacher Education, 27,*454-462. doi:10.1016/j.tate.2010.09.01

◈◇◈