

## Validation of The Mathematical Knowledge For Teaching Statistics Instrument<sup>1</sup>

*Oktay MERCİMEK<sup>2</sup>, Ayhan Kürşat ERBAŞ<sup>3</sup>*

### Abstract

The purpose of this study is to assess preservice teachers' mathematical knowledge for teaching statistics (MKT-S) and identify the relationship between the components of this knowledge. For this purpose, MKT-S instrument consisting of two dimensions, 'content knowledge' (CK) and 'pedagogical content knowledge' (PCK) was developed, and applied to 659 preservice middle school mathematics teachers (PTs). Confirmatory factor analysis showed that CK and PCK are two different dimensions of mathematical knowledge of teaching statistics. Third year and fourth year preservice teachers' factor scores were significantly different and fourth year preservice teachers' factor scores were slightly higher. It was found that CK factor scores were highly correlated with PCK factor scores. The reliability levels were 0.65 for CK factor scores and 0.76 for PCK factor scores. MKT-S instrument developed in this study has several implications for teacher education. MKT-S instrument can be used to evaluate efficiency of PTs' mathematical knowledge for teaching statistics. Instrument can be adapted for in-service teachers.

**Keywords:** Teacher knowledge, content knowledge, pedagogical content knowledge, teaching statistics, Item Response Theory, Confirmatory Factor Analysis.

### İstatistik Öğretimine Yönelik Matematiksel Bilgi Ölçeğinin Geçerliğinin Sağlanması

#### Öz

Bu çalışmanın amacı matematik öğretmen adaylarının istatistik öğretimine yönelik matematiksel bilgilerinin (MKT-S) ölçülmesi ve bu bilginin yapıtaşları arasındaki ilişkinin belirlenmesidir. Bu amaçla, 'alan bilgisi' (CK) ve 'pedagojik alan bilgisi' (PCK) olmak üzere iki bölümden oluşan MKT-S ölçeği geliştirilmiş ve 659 ilköğretim matematik öğretmeni adayı (ÖA) üzerinde uygulanmıştır. Doğrulayıcı faktör analizi, CK ve PCK faktörlerinin, istatistik öğretimine yönelik matematiksel bilginin iki ayrı boyutu olduğunu göstermiştir. Üçüncü ve dördüncü sınıf öğretmen adaylarının faktör puanları arasında istatistiksel olarak anlamlı fark bulunmuştur ve dördüncü sınıf öğretmen adaylarının faktör puanlarının az da olsa daha yüksek olduğu görülmüştür. CK puanlarının PCK puanlarıyla yüksek derecede ilişkisinin olduğu saptanmıştır. Ayrıca CK faktör puanlarının güvenilirliği 0,65 düzeyinde iken PCK faktör puanlarının güvenilirliği 0,76 düzeyindedir. Bu çalışmada geliştirilen MKT-S ölçeğinin sonuçlarının öğretmen eğitimi için çeşitli çıkarımları bulunmaktadır. Geliştirilen ölçek ÖA'larının istatistik öğretimine yönelik matematiksel bilgilerinin etkinliğini değerlendirmede kullanılabilir ya da çalışan öğretmenlere yönelik adapte edilebilir.

**Anahtar Kelimeler:** Öğretmen Bilgisi, Pedagojik Alan Bilgisi, Alan Bilgisi, istatistik öğretimi, Madde Tepki Kuramı, Doğrulayıcı faktör analizi.

## INTRODUCTION

Researchers and teacher educators need to clarify what knowledge teachers require to teach effectively. Lee Shulman was one of the pioneers that started theorizing teacher knowledge (Shulman, 1986; Shulman, 1987). Shulman attempted to describe the knowledge that is required to teach a specific subject and he described the teaching knowledge as having three categories: Subject matter content knowledge, pedagogical content knowledge (PCK) and curriculum knowledge. His subject matter content knowledge refers to facts, concepts,

<sup>1</sup> This article is based on doctoral dissertation "Assessment Of Preservice Mathematics Teachers' Knowledge For Teaching Statistics" that supervised by Prof. Dr. Ayhan Kürşat ERBAŞ at Middle East Technical University.

<sup>2</sup> Assistant Professor Dr., Kastamonu University, Faculty of Education, omercimek@kastamonu.edu.tr

<sup>3</sup> Professor Dr., Middle East Technical University, Faculty of Education, erbas@metu.edu.tr

and theorems in a domain. It also includes why a fact or concept is true, and how knowledge is generated in the domain. He describes the PCK as

... goes beyond knowledge of subject matter per se to the dimension of subject matter knowledge for teaching

...the most useful forms of representations of ideas, the most powerful analogies, illustrations, examples and demonstrations, ...an understanding of what makes the learning of specific topics easy or difficult: the conceptions and preconceptions that students of different ages and backgrounds bring with them to the learning (Shulman, 1986, p.9).

Shulman's third category, curriculum knowledge, includes understanding why a topic is included in curriculum, why we teach it in a certain level, and other alternative curriculum materials to teach.

An, Kulm and Wu (2004) defined teacher knowledge as the knowledge of effective teaching. According to them, knowledge of effective teaching consisted of three sub-dimensions; namely knowledge of content, knowledge of curriculum, and knowledge of teaching. They also placed 'teaching' (similar to PCK in Shulman's definition) as a core component of knowledge of effective teaching. The teaching dimension also had five sub-scales: knowing students' thinking, building on students' math ideas, promoting students' thinking, addressing students' misconception and engaging students in mathematics learning (see Figure 1).

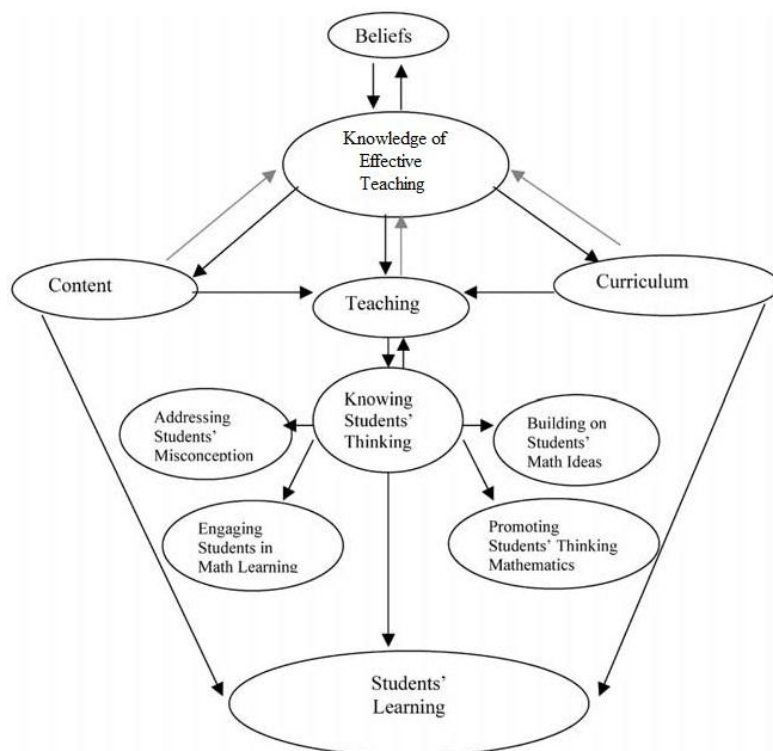


Figure 1. The Network of Knowledge of Effective Teaching (An, Kulm & Wu, 2004)

Another attempt to distinguish between different components of teacher knowledge, namely PCK and content knowledge (subject matter content knowledge in Shulman's definition), came from German researchers (Krauss, Brunner, Kunter, Baumert, Blum, Neubrand, & Jordan, 2008). They conducted a study on 198 secondary mathematics teachers to explore the relationship between the PCK and content knowledge (CK). Their study also compared teachers with respect to their teacher training program which qualifies them whether to teach in Gymnasium (GY), an academic track, or non-Gymnasium, e.g., Realschule, Sekundarschule.

Their PCK test consisted of three subscales:

- i. Task, knowledge of mathematical tasks
- ii. Student, knowledge of student misconceptions and difficulties
- iii. Instruction, knowledge of mathematics specific instructional strategies.

The study mainly resulted that GY and NGY teachers differed in their both PCK and CK level. Moreover, they found that cognitive connectedness, latent correlation between CK and PCK, is dependent on the level of mathematical expertise. Even though loadings for indicators were not significantly different, the latent correlation between PCK and CK was 0.61 in the NYG group and 0.96 for the GY group. Very strong relationship between PCK and CK in the GY group raised the question whether PCK and CK is separable constructs for these highly knowledgeable teachers. Another result was that PCK and CK form one body of connected knowledge that almost indistinguishable in the group of GY teachers. However, for the NYG group PCK and CK categories were separate constructs. Their results may imply that it is very difficult to construct CK or PCK items for highly knowledgeable teachers. For example, a highly knowledgeable teacher may offer more than one approaches to handle a mathematical task using his/her deeply connected content knowledge without thinking pedagogical aspects of the task.

An additional attempt for conceptualizing the mathematics teacher knowledge comes from Hill, Ball & Schilling (2008) as a product of their progress on measuring mathematical knowledge for teaching (Ball, 2002; Hill, Schilling & Ball, 2004; Hill, Rowan, Ball, 2005). In their 2008 article, Hill, Ball, & Schilling (2008) defined teacher knowledge (mathematical knowledge of teaching) having two major dimensions: Subject matter knowledge and pedagogical content knowledge. Their subject matter knowledge included 'common content knowledge', knowledge at the mathematical horizon' and 'specialized content knowledge'. Their pedagogical content knowledge also included three subdimensions: 'Knowledge of content and students (KCS)', 'knowledge of content and teaching (KCT)', and knowledge of curriculum'.

Hill et al.'s (2008) domain map for mathematical knowledge for teaching implies that subject matter knowledge and pedagogical content knowledge are separate constructs, and more importantly 'Knowledge of Content and Teaching' and 'Knowledge of Content and Students' can be independently observable from each other.

Even though teacher knowledge studies generally focused on inservice teachers (Incikabi, 2013; Incikabi, & Sancar-Tokmak, 2013), recent attempt for theorizing the teacher knowledge of preservice teachers came as international comparative studies, MT21 (Schmidt et al., 2007) and TEDS-M (Tatto et al., 2008). These studies influenced mainly by theory of Shulman (1986, 1987) and adapted the work of Fan and Cheong (2002). These research projects hypothesized the mathematical knowledge for teaching having two dimensions: CK and PCK. They also hypothesized that PCK has at least three components: curricular knowledge; knowledge of planning for mathematics teaching and learning (pre-active); and enacted mathematics knowledge for teaching and learning (interactive).

In their late article (Hill et al., 2008), they tried to clarify a component of PCK: Knowledge of Content and Students (KCS). They found that KCS is a multidimensional construct; however, cause of multidimensionality was not the specification of the domain. They explain that "different amounts on mathematical reasoning, knowledge of students, and perhaps even on a special kind of reasoning about students' mathematical thinking" (p. 395) caused multidimensionality. Even though they tried to construct KCS items that teachers would use knowledge of students, their follow up interviews showed that about forty percent of teachers used mathematical reasoning and twenty percent of teachers used test-taking skills to find correct answer in a multiple-choice KCS item

Even though there are many studies that are trying to define teacher knowledge, there are limited number of studies that work on components of teacher knowledge. Studies that are aiming to differentiate the components of teacher knowledge generally cover all mathematics topics because constructing items for each objective for each topic is very difficult. Therefore the purpose of this study is creating a teacher knowledge test especially for statistics topics and investigate the relationship between the components of teacher knowledge for preservice elementary mathematics teachers.

## METHODOLOGY

This study adopted the theoretical framework of Knowledge for Teaching Mathematics instrument of TEDS-M [(Tatto et al., 2008) see also MT21 (Schmidt et al., 2007)] for measuring mathematical knowledge of teaching statistics of preservice teachers. Four main reasons led to choose this framework:

- This framework defines components clearly in a fashion of expected objectives
- These objectives can be set as a teacher education standards, and can be measurable with paper and pencil tests
- Framework is specially developed and appropriate for preservice mathematics teachers' knowledge structure.
- Assessment type is suitable for large-scale assessment.

This framework was influenced mainly by theory of Shulman (1986, 1987), and adapted the work of Fan and Cheong (2002). In these research projects, the mathematical knowledge for teaching were hypothesized to have two dimensions: CK and PCK.

CK dimension of TEDS-M framework (Tatto et al., 2008) has three main cognitive domains that are parallel to TIMSS 2007 framework (Mullis et al., 2005). These main cognitive domains are knowing, applying and reasoning. Knowing domain includes recall, recognize, compute, measure and classify/order; applying domain includes select, represent, model, implement and solve routine problems; and reasoning analyze, generalize, synthesize/integrate, justify and solve non-routine problems.

All PCK objectives in MT21 framework were too broad to cover in this study. Thus, objectives were examined, and objectives that are appropriate to assess statistics knowledge were included in the study. For example, 'choosing assessment formats' objective was not selected for this study because this objective was related to general mathematics education. Some objectives were also too broad and could be defined as a combination of other objectives. For example 'planning mathematical lessons' can be measurable as combination of 'planning or selecting appropriate activities', 'selecting possible pathways and seeing connections within the curriculum', 'Identifying different approaches for solving mathematical problems' and so on. Then PCK objectives that will be included in the test were limited to six main objectives and defined to guide the item development process. Other technical details can be found on Mercimek (2013).

### *Population and Sample*

Total number of preservice elementary mathematics teachers, who are in third year or fourth year in the program, were about 5248 (OSYM, 2008 and 2009). In final administration, MKT-S was applied to 659 preservice middle school mathematics teachers (approximately 13 % of population) from eight public universities. Sample was not selected randomly and two factors taken into account during sampling: (1) university capacity for preservice middle school mathematics teachers and (2) convenience to travel between universities. Table 1 presents the distribution of preservice teachers to universities. The final implementation sample consisted from 421 (65.7%) third year and 220 (34.3) fourth year preservice teachers.

Furthermore, 435 (68.5%) of the participants were female and 200 (31.5%) of them were male preservice teachers.

Table 1. Distribution of Sample

Institution	Region	N
University A	Northern Anatolia	46*
University B	Northern Anatolia	56*
University C	Northern Anatolia	126*
University D	North-west Anatolia	64*
University E	Middle Anatolia	40
University F	Middle Anatolia	79
University G	Eastern Anatolia	148
University H	Western Anatolia	100
TOTAL		659

\* Different preservice teachers participated in Final implementation

Final instrument was applied to preservice teachers from eight institutions across Turkey by researcher in fifty minutes sessions. Final instrument was especially applied at the end of spring semester to assure that each participant received required knowledge to solve items.

## RESULTS

To validate the factor structure of the MKT-S instrument, it was needed to test that proposed two-factor model (Model II) better fits to data than one-factor model (Model I). To achieve this goal, results were acquired using MLR estimator (Muthen & Muthen, 1998-2012). One-factor solution (Model I) contained 68 parameters while two factor solution (Model II) contained 69 parameters. A chi-squared difference test conducted for assessing the fit of these two models. Results are summarized in Table 2. Chi-Squared difference test results showed that two-factor model (Model II) significantly ( $\Delta\chi^2(1)=7.95$ ,  $p< 0.01$ ) better fitted to data than one-factor model (Model I). Standardized factor loading are shown in Figure 2.

Table 2. Fit Indices for Model I and Model II for MLR Estimator.

Model	Log Likelihood	Scaling Correction Factor	Number of Parameters	BIC <sub>adj.</sub> *	$\chi^2$ Difference	df ( $\Delta\chi^2$ )	p-value ( $\Delta\chi^2$ )
Model I	-8449.10	1.0279	68	17123.68			
Model II	-8441.15	1.0274	69	17111.09	7.95	1	0.0048

\* Adjusted Bayesian Information Criterion

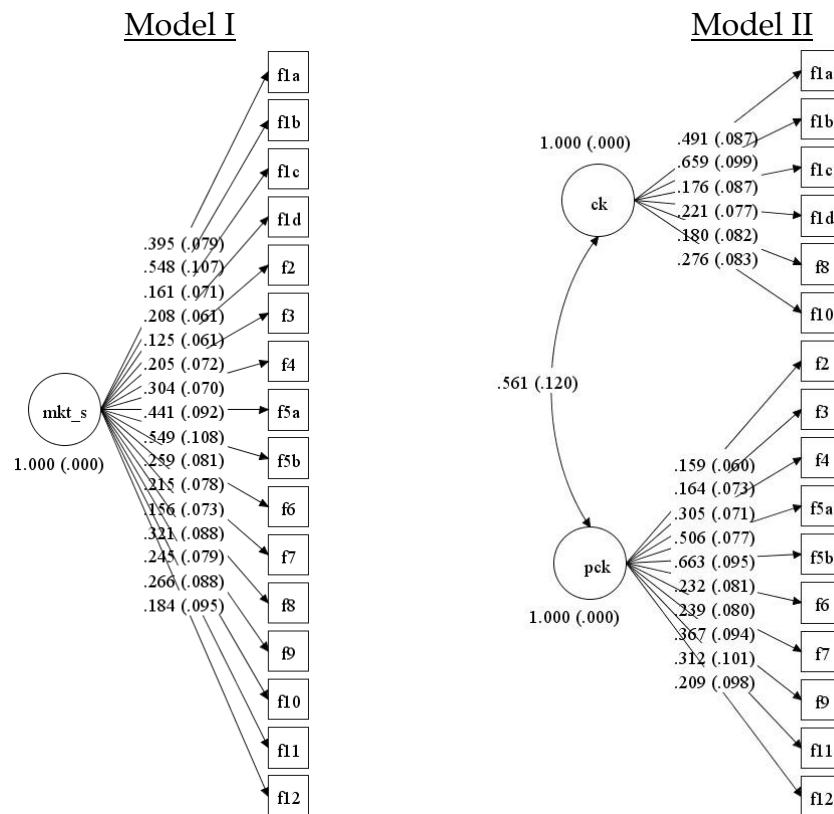


Figure 2. Standardized Loadings for Model I and Model II Using MLR Estimator.

Since MLR estimator is only useful for model comparison, fit of these two models also tested using WLSMV estimator and results are summarized in Table 3.

Table 3. Fit Indices for Model I and Model II for WLSMV Estimator.

Model	$\chi^2$	df	p-value	$\chi^2/df$	RMSAE	CFI	TLI	WRMR
Model I	166.05	104	0.0001	1.596	0.030	0.781	0.748	0.961
Model II	151.33	103	0.0014	1.469	0.027	0.830	0.802	0.915

Even though Model II seemed to fit to data better than Model I, Chi-squared differences cannot be computed directly from WLSMV output because Chi-square differences for WLSMV is not distributed as chi-square. However, there is a DIFFTEST option in Mplus, which utilizes chi-square testing for nested models. When these two models tested using DIFFTEST command, Model II showed significantly better fit than Model I,  $\Delta\chi^2(1)=11.549$ ,  $p < 0.001$ .

Results of both estimators significantly favored Model II. These results led to conclude that items contained in MKT-S instrument do not uniformly measure a single construct (teacher knowledge). Instead, pedagogical content knowledge (PCK) and content knowledge (CK) are two different constructs that both of them had their own characteristics.

After confirming two-factor structure of the instrument, modification indices reported by Mplus were examined. Recommended modifications were correlating item F.1D with F.1C, and correlating F5B with F5A. Recommended correlations clearly made sense because both indices were related to items which share same stem even though they seek different information. First modification conducted was correlating item F1D with F1C resulting in Model IIA and second modification conducted was correlating item F1D with F1C resulting in Model IIB. Confirmatory factor analysis results, using WLSMV estimator, are summarized in Table 4.

Table 4. Fit Indices for Model IIA and Model IIB for WLSMV Estimator.

Model	$\chi^2$	df	p-value	$\chi^2/df$	RMSAE	CFI	TLI	WRMR
Model IIA (F1D with F1C)	137.945	102	0.0103	1.352	0.023	0.873	0.851	0.870
Model IIB (F5B with F5A added to Model IIA)	124.999	101	0.0530*	1.237	0.019	0.915	0.900	0.821

\*Model significantly fitted to data at 0.05 level

Even though it seemed each modification improved model fit, DIFFTEST command of Mplus had to be applied to test chi-square differences. The test results summarized in Table 5.

Table 5. Results for Comparing Chi-square Values

Compared Models	$\Delta\chi^2$	df	p-value
Model II and Model IIA	14.189*	1	0.0002
Model IIA and Model IIB	10.708*	1	0.0011

\*Differences tested using DIFFTEST option of Mplus

DIFFTEST results showed first modification significantly improved fit of Model II , so Model IIA fitted to data better than Model II ( $\Delta\chi^2(1)=14.189$ ,  $p < 0.001$ ), and second modification significantly increased fit of Model IIA so Model IIB fitted to data better than Model IIA ( $\Delta\chi^2(1)=10.708$ ,  $p < 0.01$ ).

Final model, Model IIB, significantly fitted to data,  $\chi^2(101)=124.999$ ,  $p > 0.05$  (Barret,2007). Most of the other fit indices also showed good fit of model. For example,  $\chi^2/df$  was 1.137 and it was lower than most conservative cut-off value of 2. RMSAE was 0.019, and it was lower than 0.05. WRMR was 0.821 and it was lower than 1.0. On the other hand, CFI and TLI indices showed poor fit.

Since the  $\chi^2$  statistic was not significant, it was concluded that proposed model (Model II) is currently best model that fitted to data and standardized loadings of Model IIB are shown at Figure 3

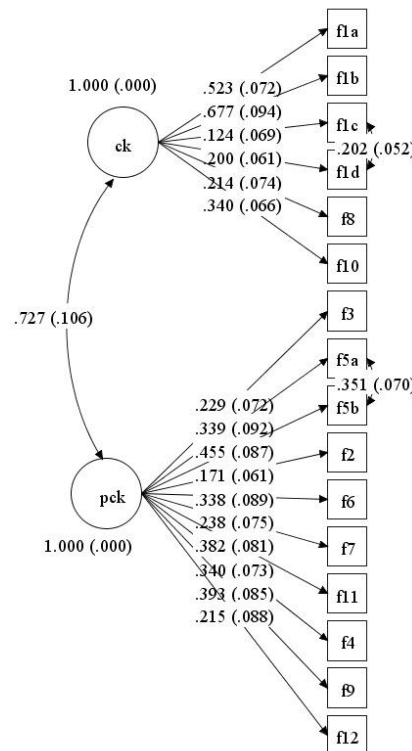


Figure 3. Standardized Factor Loadings for Model IIB.

Next step in validation process was comparing results with previously conducted TEDS-M study. Since this study and TEDS-M study used similar framework, it was possible to compare results by following the method that explained by Blömeke, Houang and Suhl (2011). To achieve this goal, researcher also constrained the factor loadings to be same within each factor. Table 6 shows the comparison of this study with TEDS-M study.

Table 6. Comparison of Results with TEDS-M Study.

Model	Factor Loading for CK items	Factor loadings for PCK items	$R^2$	
			CK	PCK
TEDS-M	0.34 (0.00)***	0.30 (0.01)***	0.12 (0.00)	0.09 (0.00)
Current Study	0.306 (0.028)***	0.299 (0.023)***	0.094 (0.017)	0.089 (0.014)

\*\*\* $p < 0.001$ . Parenthesis represent standard errors.

A comparison of the results showed that this study had similar finding for both CK and PCK factors. Only clear differences were observed for CK factor were average loading was 0.34 and  $R^2$  was 0.12 for TEDS-M study while average loading was 0.306 and  $R^2$  was 0.094 for this study. Since TEDS-M study covered more items than this study, it was concluded that differences were small and arbitrary. Even though TEDS-M study covers a broad range of topics and this study covers only some of statistics topics, results of MKT-S instrument was consistent with MKT-S study.

After validating factor structure of the MKT-S instrument, factor scores for CK and PCK were calculated using Maximum Likelihood (MLR) estimator of Mplus (Muthen & Muthen, 2006). Then tests were conducted for factor scores against predetermined variables.

First, the correlation between factor scores and preservice teachers' Introduction to Statistics and Probability-I (ISP-I) grades were checked. Correlation between CK score and ISP-I grade was 0.305 ( $p < 0.001$ ), and correlation between Mathematics PCK score and ISP-I grade was 0.273 ( $p < 0.001$ ). Small but significant positive correlation was found between factor scores and preservice teachers' ISP-I grades. Since the ISP-I course covers much broader content than this study, it was concluded that scores obtained from MKT-S was instrument consistent with preservice teachers' ISP-I course grades.

As a second step, factor score differences were tested for third and fourth year preservice middle school mathematics teachers. After checking assumptions, MANOVA test was conducted. MANOVA test results revealed that differences existed among third year and fourth year preservice teachers,  $F(2, 638) = 5.076$ ,  $p = 0.007$ , partial eta squared = 0.016. Then independent samples-t-tests were conducted for follow up analysis. Results showed that fourth year preservice teachers CK factor score was significantly ( $p < 0.01$ ) higher than third year preservice teachers, and fourth year preservice teachers PCK factor score was significantly ( $p < 0.05$ ) higher than third year preservice teachers. Even though differences were significant, effect size for CK factor was small and effect size for PCK factor was barely small.

Additionally IATA software was used to assess the reliability of scores obtained from the MKT-S instrument because IATA displays test information function along with the IRT scores, and provides a holistic reliability coefficient which is defined as the proportion of variability in observed scores that can be explained by variation in true scores (Cartwright, 2013).

For CK scores, reliability coefficient provided by IATA, which is based on average  $SEM^2$  of IRT scores, was 0.65. For PCK scores, reliability coefficient provided by IATA, which is based on average  $SEM^2$  of IRT scores, was 0.76.

It was found that correlation between CK scores and PCK scores was very high ( $r = 0.78$ ,  $p < 0.001$ ). This high correlation coefficient implied that content knowledge and pedagogical content knowledge dimensions were closely related to each other for preservice teachers.



Therefore, a high content knowledge score was generally corresponding to a high pedagogical content knowledge score.

## DISCUSSION AND CONCLUSION

It was found that MKT-S instrument has two dimensions, and content knowledge and pedagogical content knowledge are two different knowledge forms of teaching knowledge. The structure of MKT-S instrument was also compared with other researchers' results (Blömeke, Houang, & Suhl, 2011), and found that results of MKT-S instrument was in line with these researchers.

To provide concurrent validity evidences, the relationship between preservice teachers' 'Introduction to Statistics and Probability (ISP-I)' grades and MKT-S scores (CK and PCK scores) was examined. It was found that there was a small relationship between ISP-I grades and CK scores ( $r=305$ ,  $p<0.001$ ), and a small relationship between ISP-I grades and PCK scores ( $r=273$ ,  $p<0.001$ ). This result can be explained by the nature of the items in the MKT-S instrument. Even though items were part of the ISP-I course content; items were required preservice teachers to think on more abstract level.

Even though fourth year preservice teachers significantly got better CK and PCK scores than third year preservice teachers, effect sizes for differences were very small. This result suggested that preservice teachers gain a small amount of information for teaching statistics topics during fourth year in the program.

It was found that the reliability of the CK scores was 0.65. Even though reliability is lower than industry standard of 0.7, the low number of items that consists CK scores could explain this situation, and larger number of items may result in a more reliable CK instrument. Reliability was also higher for high CK scores and this could be due to absence of items that has medium difficulty. Since the one purpose of the instrument was to assess the adequacy of content knowledge of the preservice teachers, the items generally aimed this purpose, and items were challenging to seek deep information of preservice teachers.

Reliability of PCK scores was 0.76, and it was little higher than industry standard of 0.70. This level of reliability can be considered enough for the purposes of this study. In this study, the general pedagogical levels of preservice middle school mathematics teachers regarding to statistics topics were tried to be pictured, and aim was not defining cut-off values that important decisions (such as hiring for a job or passing from a course) will be made upon these values. Low level of the reliability is also can be explained by content of the PCK items. Even though both average related PCK items and graphics related PCK items aim to measure pedagogical knowledge for statistics topics, confirmatory factor analysis results show that mean loading for average related items were higher than mean loadings of graphics related items. Even though PCK scores aimed to picture the general pedagogical levels of preservice middle school mathematics teachers, it makes sense that an average related PCK items will seek different information from a graphics related PCK item.

Low reliability levels of MKT-S instrument can also be connected to content coverage. As Shulman (1987) stated "Pedagogical content knowledge ... presents the blending of content and pedagogy into an understanding of how particular topics, problems, or issues are organized, presented, and adapted to diverse interest and abilities of learners, and presented for instruction" (p. 8). Therefore, it may be asserted that content is a key factor for pedagogical content knowledge (or teacher knowledge general), and teachers' knowledge is differently organized for different concepts of statistics curriculum. However, because of the insufficient number of items, it was not possible test whether knowledge structures differs for concepts of statistics, and it was assumed that both content knowledge and pedagogical content knowledge of the preservice teachers would be parallel for both central tendency and graph related topics.

It was found that PCK scores were highly correlated with CK scores ( $r=0.78$ ,  $p<0.001$ ). The result is similar to finding reported by other researchers. For example, Krauss et al. (2009) found similar latent correlation ( $r=0.79$ ) between content knowledge and pedagogical content knowledge, and correlation was even higher ( $r=0.96$ ) for teachers who possess high level of CK and PCK. Even though it was explicitly tested and found that two-factor structure better fitted the data, a high correlation among factors of an instrument brings the question that whether factors could be collapsed to construct a single factor. It was found that this result was pretty much in line with the nature of pedagogical knowledge because it is a trivial fact that teaching a mathematics topic properly for any person requires an understanding about topic but knowing mathematical content does not always result in good teaching (Borko et al., 1992). Therefore, it was an expected result for this study.

Even though classical test theory can be used score preservice teachers factor scores of MKT-S instrument, this study used IRT scoring of the factor scores, which took account of both difficulty of an item compared to other items and difficulty of each score level of item. However, it should be also noted that reliability levels also estimated using IRT because other methods are not possible when instrument administered balanced incomplete booklet design. Therefore, researchers may implement complete MKT-S instrument in order to analyze psychometric properties of the instrument under classical test theory.

## REFERENCES

- An, S., Kulm, G., & Wu, Z. (2004). The pedagogical content knowledge of middle school mathematics teachers in China and the U.S. *Journal of Mathematics Teacher Education*, 7(2), 145-172.
- Ball, D. L. (2002). Knowing mathematics for teaching: Relations between research and practice. *Mathematics and Education Reform Newsletter*, 14(3), 1-5.
- Barrett, P. (2007). Structural equation modelling: Adjudging model fit. *Personality and Individual Differences*, 42, 815-824.
- Blömeke, S., Houang, R. T., & Suhl, U. (2011). TEDS-M: Diagnosing teacher knowledge by applying multidimensional item response theory and multiple-group models. *IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments*, 4, 109-129. Retrieved May 10, 2013, from <http://www.ierinstitute.org/dissemination-area.html>
- Borko, H., Eisenhart, M., Brown, C. A., Underhill, R. G., Jones, D., & Agard, P. C. (1992). Learning to teach hard mathematics: Do novice teachers and their instructors give up too easily?. *Journal for Research in Mathematics Education*, 194-222.
- Cartwright, F., (2013). IATA Manual. Retrieved June 25, 2013, from [www.polymetrika.org](http://www.polymetrika.org)
- Fan, L., & Cheong, N. P. (2002). Investigating the sources of Singaporean mathematics teachers' pedagogical knowledge. In D. Edge & B. H. Yeap (Eds.), *Mathematics education for a knowledge-based era : Vol. 2* (pp. 224-231). Singapore: AME.
- Hill, H. C., Ball, D. L., & Schilling, S. G. (2008). Unpacking pedagogical content knowledge: Conceptualizing and measuring teachers' topic-specific knowledge of students. *Journal for Research in Mathematics Education*, 39(4), 372-400.
- Hill, H.C., Rowan, B., & Ball, D.L. (2005) Effects of teachers' mathematical knowledge for teaching on student achievement. *American Educational Research Journal*, 42, 371-406.
- Hill, H.C., Schilling, S.G., & Ball, D. L. (2004). Developing measures of teachers' mathematics knowledge for teaching. *Elementary School Journal*, 105, 11-30.
- İncikabı, L. (2013). TPAB-temelli "Teknolojiyle Öğretim" dersi: Matematik öğretimi için bilgisayar oyunu tasarlama. In T. Yanpar-Yelken, H. Sancar-Tokmak, S. Özgelen, & L. İncikabı, (Eds.), *Fen ve matematik eğitiminde teknolojik, pedagojik alan bilgisi temelli öğretim tasarımları*, (pp. 221-238). Ankara: Anı Yayıncılık.
- İncikabı, L., & Sancar-Tokmak, H. (2013). Integrating technology into mathematics teaching: A TPACK (Technological, Pedagogical, Content knowledge) - based course design for college students. In J. Keengewe (Ed.), *Research Perspectives and Best Practices in Educational Technology Integration* (pp. 288-303). Hersley, PA: IGI Global.
- Krauss, S., Brunner, M., Kunter, M., Baumert, J., Blum, W., Neubrand, M., & Jordan, A. (2008). Pedagogical content knowledge and content knowledge of secondary mathematics teachers. *Journal of Educational Psychology*, 100(3), 716-725.
- Mercimek, O. (2013). Assessment of Preservice Mathematics Teachers' Knowledge For Teaching Statistics (Unpublished doctoral thesis). Middle East Technical University, Turkey
- Mullis, I. V. S., Martin, M. O., Ruddock, G. J., O'Sullivan C. Y., Arora, A. & Erberber, E.. (2005). *TIMSS 2007 assessment frameworks*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College
- Muthen, L. K., & Muthen, B. O. (2006). IRT in Mplus. Retrieved March 25, 2013, from <http://www.statmodel.com/download/MplusIRT1.pdf>
- Muthén, L.K. and Muthén, B.O. (1998-2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén
- OSYM (2008). Table 4: University Student Capacities for Bachelor Degree Level. Retrieved September 23, 2010, from [http://dokuman.osym.gov.tr/pdfdokuman/arsiv/2008/2008\\_OSYS\\_TERCIH\\_KILAVUZU/6\\_tablo4.pdf](http://dokuman.osym.gov.tr/pdfdokuman/arsiv/2008/2008_OSYS_TERCIH_KILAVUZU/6_tablo4.pdf)

- OSYM (2009). Table 4: University Student Capacities for Bachelor Degree Level. Retrieved September 23, 2010, from [http://dokuman.osym.gov.tr/pdfdokuman/arsiv/2009/2009\\_OSYS\\_TERCIH\\_KILAVUZU/tablo4.pdf](http://dokuman.osym.gov.tr/pdfdokuman/arsiv/2009/2009_OSYS_TERCIH_KILAVUZU/tablo4.pdf)
- Schmidt, W., Tatto, M.T., Bankov, K., Blomeke, S., Cedillo, T., Cogan, L., Han, S. I., Houang, R., Hsieh, F. J., Paine, L., Santillan, M., & Schwille, J. (2007). *The preparation gap: Teacher education for middle school mathematics in six countries* (MT21 report). East Lansing, MI: Michigan State University. Retrieved February 25, 2010, from <http://usteds.msu.edu/MT21Report.pdf>
- Shulman, L. S. (1986). Those Who Understand: Knowledge Growth in Teaching. *Educational Researcher*, 15(2), 4-14
- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard educational review*, 57(1), 1-23.
- Tatto, M. T., Schwille, J., Senk, S., Ingvarson, L., Peck, R., & Rowley, G. (2008). *Teacher Education and Development Study in Mathematics (TEDS-M): Conceptual framework*. East Lansing, MI: Teacher Education and Development International Study Center, College of Education, Michigan State University.