PAMUKKALE JOURNAL OF SPORT SCIENCES
ISSN: 1309-0356
dergipark.org.tr/en/pub/psbd

Pamukkale J Sport Sci, 12(2), 29-45, 2021

Research Article

The Validity and Reliability of General Measurement and Evaluation Efficiency Scale for Physical Education Teachers

Cevdet CENGİZ[1*] Nurdan GÜMÜŞBAŞ[2]

[1*]*Faculty of Sport Sciences, Department of Teaching Physical Education, Çanakkale Onsekiz Mart University, Turkey*
[2]*Physical Education Teacher, Akcansa Fine Arts High School, Çanakkale, Turkey*

ABSTRACT

Keywords
Evaluation,
In-service teachers,
Measurement,
Physical education

Article History
Received 6 April 2021
Revised 23 August 2021
Accepted 25 August 2021
Available online 30 August 2021

* **Corresponding Author**:
Cevdet CENGİZ
E-mail Address:
cevdetcengiz@gmail.com

Measurement and evaluation are critical elements in curriculum programs and for physical education (PE) teachers. Therefore, the purpose of this study is to establish the validity and reliability of the Measurement and Evaluation of Common Competency Perception (MECCP) Scale for in-service PE teachers (n=878). For construct validity, Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) (n=440) were conducted with two different samples. According to the EFA results (n=438), the three-factor MECCP Scale contained 23 items. After CFA (n=440), the three-factor structure was 23 items using Schumacker and Lomax LISREL analyses. Findings indicated good fit indices with $\chi 2/sd=4.83$, RMSEA=0.09, CFI=0.96, IFI=0.96, and NFI=0.95. In addition, Cronbach alpha scores were calculated as 0.50-0.77. In conclusion, the MECCP Scale was found to be valid and reliable for Turkish in-service PE teachers and is therefore highly recommended to evaluate in-service PE teacher's perceptions of the measurement and evaluation of common competency and also to establish new experimental designs to support the development of general efficiency levels of measurement and evaluation by PE teachers.

Citation: Cengiz, C., & Gümüşbaş, N. (2021). The Validity and Reliability of General Measurement and Evaluation Efficiency Scale for Physical Education Teachers. *Pamukkale Journal of Sport Sciences*, 12(2), 29-45.

INTRODUCTION

Education is the most effective tool for shaping the future of a country and a critical factor in the development of an individual, as well as the development of society (Brennan & Teichler, 2008; Yeşilyaprak, 2006). In this context, measurement means tracking the qualifications of individuals and expressing the results with digits or symbols under specific rules; while evaluation describes the interpretation of the obtained measurement results and completion of the decision-making process utilizing specific criteria (Bell, 2001; Çelik, 2005; Turgut & Baykul, 2010).

Measurement and evaluation are used to determine if a specific education program is as successful as desired and whether the students have gained the intended knowledge, qualifications, and attitudes. Similarly, measurement and evaluation ensure the determination and solving of problems at each level of the education process by continuous monitoring (Ministry of National Education (MoNE), 2009; 2013; 2018). In physical education (PE), measurement and evaluation are defined as the determination of students' level of success with different resources and the collection of information about student performance, which includes the quantization stages of this information (Harrison et al., 2001).

When constructivism emerged from behaviorism in educational psychology, curriculum theorists and developers began to develop curricula based on constructivist learning theories (Von Glaserfeld, 1995). This approach mainly focuses on the understanding and knowledge learned through new experiences and in a social context (Zuckerman, 2003). Different forms of cooperative learning have proved motivating and effective (Perrenet & Terwel, 1997). The constructivist approach has three implications in the learning environment; (1) learning is an active process, (2) the learner has prior knowledge, and (3) the learner takes responsibility for their learning (MoNE, 2009; 2013; Yager, 1991). Therefore, the purpose of measurement and evaluation is to improve the quality of student learning by providing specific needs and measuring the characteristics of teachers and students. Assessment must be integrated with instruction in the classroom environment, which indicates that the meaning of assessment tasks will depend on the environment. Assessment data and use become part of the daily reality of the classroom and the role of learning. Many researchers have successfully implemented PE curricula in different periods and contexts (Azzarito & Ennis, 2003; Dyson, 2012; Rovegno & Dolly, 2006). Constructivist perspectives on learning focus on students being actively engaged in constructing knowledge and understanding. This puts the teacher in a critical position; in that, the teacher should encourage students to explore their surroundings,

discover knowledge, solve problems, and pursue critical thinking (Grennon-Brooks & Brooks, 1993).

As in all other school subjects, measurement and evaluation are crucial for PE teachers and their lessons, which support theoretical education with practical physical motion. In this context, it is a professional necessity for PE teachers to apply measurement and evaluation as required to observe and evaluate students' success, and to be equipped with the essential knowledge and qualifications for measuring and evaluating (Uçar, 2001). In addition, teaching quality (Darling-Hammond, 2014), teacher effectiveness, and student learning (Darling-Hammond & Snyder, 2000; Goldhaber, 2015; Popham, 2013) should continuously be improved as much as possible.

Measurement and evaluation performance criteria have been determined as necessary for the general efficiency of the teaching profession (Avşar, 2012; Nartgün, 2008; Smith, 2005). Various studies have addressed the measurement and evaluation efficiency of teachers. The results have shown that teachers often consider themselves to be insufficient in these areas, that they need in-service training, and that they utilize those measurement techniques with which they feel most comfortable, such as written examinations and observing student performance and student interest in the lesson (Bıçak & Çakan, 2004; Gelbal & Kelecioğlu, 2007). Specifically, PE teachers have so far experienced a shortage of time during the application of alternative assessment and evaluation methods (Özkoparan & İnan, 2018) and possess insufficient information and equipment in terms of the implementation of both traditional and alternative measurement and evaluation tools (Özgül & Kangalgil, 2018). In some studies, it is seen that teachers have problems, particularly in the application of such techniques, and are influenced by generally accepted practices (Çalık, 2007; Gömleksiz & Bulut, 2007). Zhang and Burry-Stock (2003) show that the perception of measurement and evaluation efficiency levels among primary and secondary school teachers is that they need improving, with the hypothesis that achieving higher education levels would correlate with feeling more qualified. A practical, comprehensive assessment is recommended by Salimin et al. (2015); that are, assessing the cognitive, affective, and psychomotor domains of learning. Daniel and King (1998) indicate that primary and secondary school teachers do not have extensive knowledge about measurement and evaluation and lack even a basic understanding of statistics.

Teachers may consider themselves sufficient in terms of the desired qualification skills in the field of measurement and evaluation in general, but they show differences based on their years of experience, subjects studied, the number of students in their classes (Özbaşı &

Çıkrıkçı-Demirtaşlı, 2013) as well as a general absence of a valid and reliable way of measuring student achievement (Rink, 2013). In Turkey, studies have shown that PE teachers do not find themselves sufficiently versed in utilizing information to measure and evaluate the performance of their students efficiently, have sufficient time to apply alternative assessment tools (Özkoparan & İnan, 2018), and they do not consider themselves skilled in using the measurement and evaluation tools available in the literature (Özgül & Kangalgil, 2018; Şirin et al., 2009; Yaykın, 2015). Baş and Beyhan (2016) reported that teachers had a low level of self-efficacy in educational measurement and evaluation, both in knowledge and skill sub-dimensions. Historically, teachers have found themselves insufficient and faced many problems in the field of measurement and assessment (Ulutaş & Erman, 2011). Therefore, it is crucial to assess PE teachers' measurement and evaluation levels and to improve their efficacy levels through experimental designs. The present study provides a new scale to measure what constructs need to be developed in terms of measurement and evaluation.

In Turkey, the standards and performance indicators related to the general efficiency of teachers have been determined by the Ministry of National Education and the Council of Higher Education (YÖK, 2008). Three key areas have been established; Professional Knowledge, Professional Skills, and Attitudes and Values. In the professional skills category, Assessment and evaluation are categorized. Specifically, this aims that the teacher "uses the methods, techniques, and tools of assessment and evaluation that fit the purpose." In a study on the general efficiency of teachers, the perception of available measurement and evaluation efficiency is divided into various sub-dimensions and performance indicators (MoNE, 2002). The Measurement and Evaluation of Common Competency Perception (MECCP) Scale, developed by Nartgün (2008), have been proven to be effective in measuring the general efficiency perception of preservice teachers. This scale comprises three sub-dimensions, and the desired coefficients were established for the internal consistency and reliability of all items. Similar findings were evident for preservice PE teachers (Arslan, İlker & Demirhan, 2013).

Research has shown that PE teachers do not feel confident in measuring and evaluating; they prefer more practical applications and oral examinations in measuring student outcomes, and while they utilize measurement and evaluation applications in compliance with the defined standards, their performance in this regard is not at the desired level (Avşar, 2012; Şirin et al., 2009; Şirinkan & Erciş, 2009; Şirinkan & Gündoğdu, 2011; Tagele & Bedilu, 2015; Yaykın, 2015). The research objective of the present study is to test the validation and reliability of the MECCP Scale for in-service physical education teachers in Turkey.

METHODS

This research is a cross-sectional study based on the survey model (Büyüköztürk et al., 2012), which is one of the quantitative research methods.

Study Groups

The study population was comprised of in-service PE teachers (n=878). PE teachers who agreed to participate in the research and lived in the seven different geographic regions of Turkey were contacted. Next, ethical permission was obtained for the questionnaire to be used from the Ethics Committee for Clinical Research of Çanakkale Onsekiz Mart University. The PE teachers gave consent to participate in personal meetings or via the Internet. The purposive sampling method was used (Büyükoztürk et al., 2012).

Inclusion criteria for participation was that the teachers be currently employed in service at public or private schools and had graduated from physical education teaching and sports departments. Participants were mainly from different cities in one of the seven geographical regions in Turkey (Adana=53, Ankara=102, Aydın=30, Bursa=37, Çanakkale=83, Çorum=74, Diyarbakır=20, Edirne=6, Elazığ=24, Gaziantep=64, Mersin=27, İstanbul=159, İzmir=65, Konya=35, Trabzon=50, Van=49).

Data Collection Tools

Data were collected with questionnaire forms. The data collection tool, the MECCP Scale, was developed by Nartgün (2008) for preservice teachers and originally consisted of 24 items. The questionnaire was validated for Turkish preservice students, and the scale is a five-point Likert-type scale, where “5” represents "sufficient", and “1” illustrates "insufficient." There are three sub-dimensions of the scale: basic concepts (6 items), measurement techniques (9 items), and statistical analysis and reporting (9 items). Higher points from the full scale and each sub-dimension show that the preservice teachers find themselves sufficient for general efficiency in measurement and evaluation. In comparison, lower points are obtained by those who feel insufficient (Çelik & Arslan, 2012). The correlation values of the MECCP Scale for PE teachers were determined and were found to be high (Cronbach's α=0.94).

Data Collection Procedure

The questionnaires (n=1107) were administered in all seven geographic regions; cities were selected based on their population. The purposive sampling method was followed. Forms were excluded if they were incompletely filled out or not returned. Those filled out in cities with insufficient responses were also excluded, yielding the final total of acceptable questionnaires to be (n=878). Consent to participate was given by the PE teachers in personal

meetings or via the Internet. Official permission from schools to carry out the survey was sent, and the questionnaires were filled out online. In addition, some PE teachers were found to apply the questionnaire in person.

From the results of the test-retest questionnaire, Cronbach alpha values for each sub-dimension were over 0.70 (basic concepts: 0.88, measuring techniques: 0.89, statistical analysis and reporting: 0.92).

Data Analysis

To adapt this questionnaire for the PE teachers, Exploratory Factor Analysis (EFA) (n=438) and Confirmatory Factor Analysis (CFA) (n=440) were implemented with LISREL (Schumacker & Lomax, 2010) having an alpha level ($p<0.05$).

RESULTS

Exploratory Factor Analysis (EFA)

In relation to the EFA (SPSS), the stated variance was 62.49%. The original scale, consisting of three sub-dimensions and 24 items, was therefore reduced in accordance with the EFA results to three sub-dimensions and 23 items (in the sub-dimension of measuring techniques, item 14 was deleted as being less than <0.50). Examining the sub-dimensions in which these 23 items are found, the sub-dimension of basic concepts contains six items (nos. 1, 2, 3, 4, 5, 6), that of measurement techniques includes eight items (nos. 7, 8, 9, 10, 11, 12, 13, 15), and that of statistical analysis and reporting contains nine items (nos. 16, 17, 18, 19, 20, 21, 22, 23, 24) (see Table 1).

The data collected in this study were utilized after being processed with the assumptions from EFA. The questionnaire factor weights that are part of the three-dimensional structure of the scale are summarized in Table 1.

Confirmatory Factor Analysis (CFA)

With regard to the CFA, LISREL analysis (Schumacker & Lomax, 2010) showed that the three sub-dimensions consisted of 23 items in total (basic concepts = 6 items, measurement techniques = 8 items, statistical analysis and reporting = 9 items) (Figure 1).

Reviewing the fit indices, all values are seen to be good: $\chi 2/sd.=4.83$, RMSEA=0.092, CFI=0.96, IFI=0.96, and GFI=0.94. It is stated in the literature that the admissible values of these parameters are RMSEA<0.06-0.08, CFI>0.95-0.90, GFI>0.95-0.90 and IFI>0.95-0.90 (Byrne, 2001; Hu & Bentler, 1999) (see Table 2). Thus, the results of the present study are consistent with the ranges presented in the literature (Kline, 1994). In conclusion, good values strengthened the CFA results.

In addition to the fit indices of the CFA, it is also important that the results of the t-test are significant for all items ($p<0.05$), as listed in Figure 2. Modifications between items "1 and 2", "20 and 21", and "21 and 22" were made due to better model fit indices.

The findings showed that there are three sub-dimensions in the questionnaire (Basic concepts, Measurement techniques, Statistical analysis, and reporting) with 23 items were acceptable. The survey asked the physical education teachers to rate all the 23 items on a five-point scale.

Table 1. Item loads resulting from Exploratory Factor Analysis (EFA) (n=438)

Subscales in the original scale	Item loads		
Item 1	0.64		
Item 2	0.70		
Item 3	0.76		
Item 4	0.63		
Item 5	0.73		
Item 6	0.63		
Item 7		0.74	
Item 8		0.70	
Item 9		0.69	
Item 10		0.77	
Item 11		0.68	
Item 12		0.67	
Item 13		0.50	
Item 14		0.53	
Item 15			0.73
Item 16			0.60
Item 17			0.64
Item 18			0.72
Item 19			0.76
Item 20			0.76
Item 21			0.71
Item 22			0.75
Item 23			0.75

(After deleting item number 14, total variance explained = 62,49 %)

Table 2. Measurements of investigation model

Compliance Measures	Good Compliance	Acceptable Compliance	Measurement Model
χ^2/sd.	$0 \leq \chi^2/sd \leq 2$	$2 < \chi^2/sd \leq 5$	4.83
GFI	$0.95 \leq GFI \leq 1$	$0.90 \leq GFI < 0.95$	0.94
CFI	$0.97 \leq CFI \leq 1$	$0.95 \leq CFI < 0.97$	0.96
IFI	$0.97 \leq IFI \leq 1$	$0.90 \leq IFI < 0.97$	0.96
RMSEA	$0 \leq RMSEA \leq 0.05$	$0.05 < RMSEA \leq 0.08$	0.92

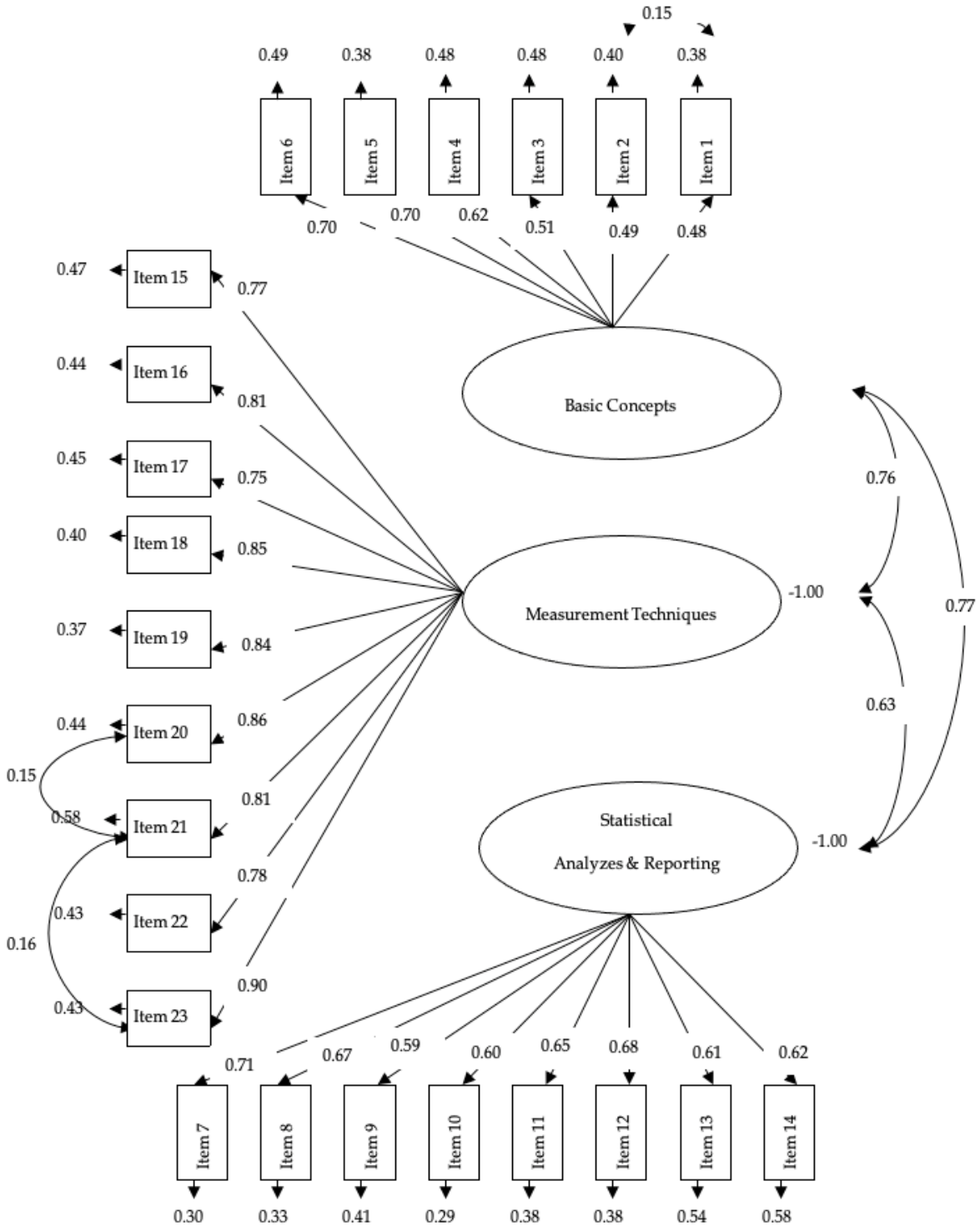

Figure 1. The results of Confirmatory Factor Analysis (CFA) (n=440)

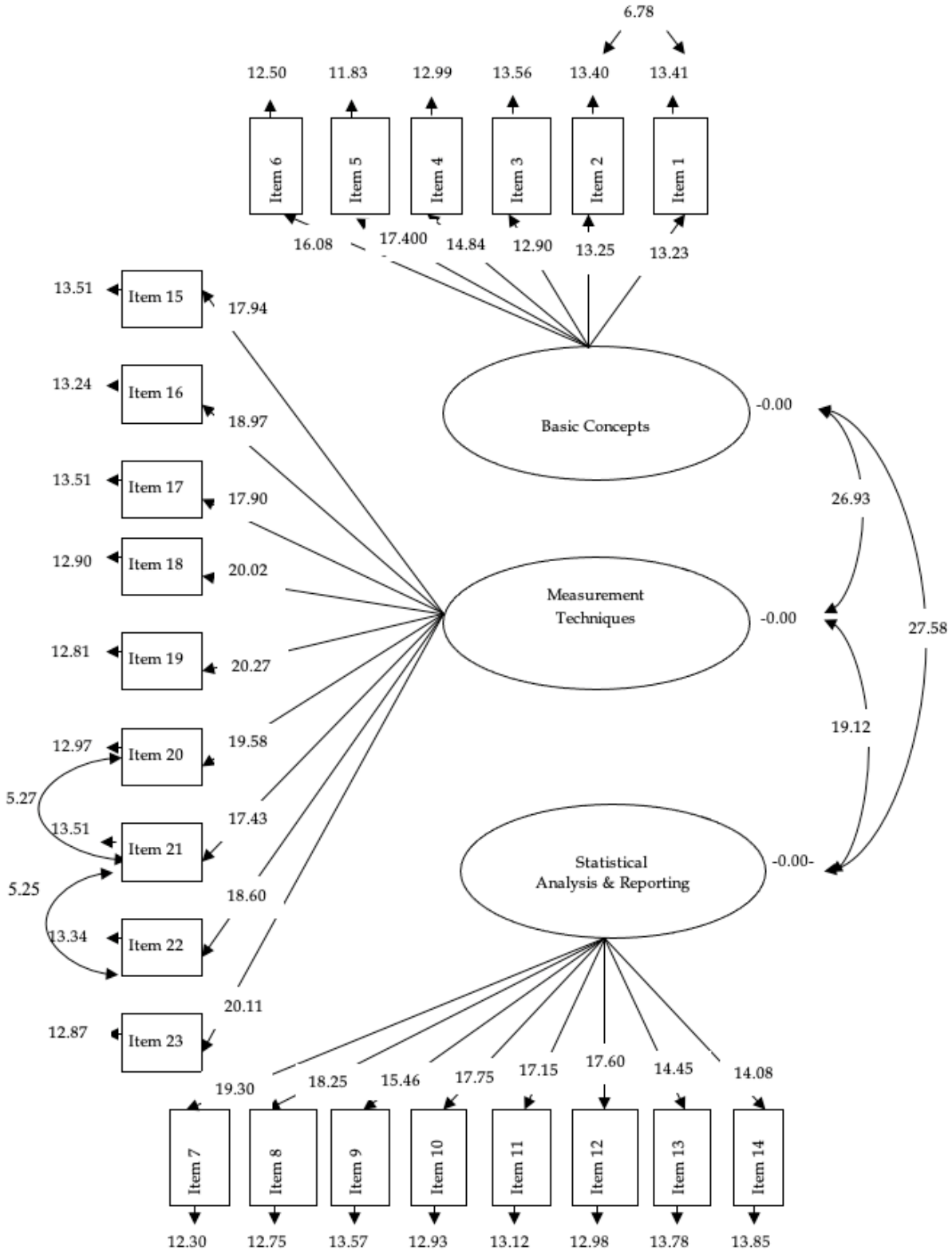

Figure 2. T-Test results of CFA (n=440)

DISCUSSION

With respect to the obtained findings, it was determined that the MECCP Scale for PE teachers in service is utilizable with its three sub-dimensions and 23 items. High scores obtained from the scale and all sub-dimensions show that in-service PE teachers perceive their general skills in measuring and evaluation efficiency to be sufficient, while on the contrary, low scores can be interpreted as a sign that they find their skills insufficient. Based on the analysis results, it was found that one item present in the original scale of Nartgün (2008) was not being utilized by PE teachers in service. This item was therefore removed, and the relevant performance indicators were determined within the three sub-dimensions. Studies in the literature regarding the measuring and evaluation efficiency of PE teachers in service, in terms of efficiency levels in the educational sciences, were said to be limited in number and problematic (Lacy & Hastad, 2006; Wise et al., 1991).

In the study of Nartgün (2008) regarding preservice teachers, the original MECCP Scale contained six items in the first sub-dimension, nine items in the second sub-dimension, and nine items in the third sub-dimension. On the other hand, Karaca (2004) had developed a 4-dimensional scale for assessing the perceptions of measuring and evaluation efficiency of preservice teachers. Different from preservice teachers' perceptions of measurement and evaluations, teacher efficiencies as a whole are shown in one dimension in general, while in studies regarding available efficiencies of the teaching profession, measurement, and evaluation efficiencies are divided into sub-dimensions, and performance indicators are determined. In the present study, the utilized performance indicators (basic concepts, measurement techniques, statistical analysis and reporting) for the evaluation of PE lessons by in-service PE teachers in terms of perceptions of measurement and evaluation efficiency were divided into three sub-dimensions and applied on a scale. The adapted scale would help researchers and stakeholders to understand PE teachers' basic concepts, measurement techniques, statistical analysis, and reporting knowledge related to measurement and evaluation. In addition, both the quality of students' learning and being able to measure the knowledge of PE teachers would be possible through the scale.

Another study on the validity and reliability of the self-efficacy of PE teachers was conducted by Ünlü et al. (2008). It stated that the reliability coefficient of their scale was high for all sub-dimensions, with six sub-dimensions related to general teacher efficacy. In particular, the reliability coefficient for ten items in the monitoring and evaluation of educational development dimensions was high. In the present study, however, the scale

consists of three sub-dimensions and the reliability coefficient is high for all sub-dimensions of the scale. Items are more specific to general skills in measurement and evaluation, and therefore the scale has the potential to determine and develop those constructs.

The data collection tool that was utilized during this research was limited to the cities and in-service PE teachers that participated. One item in measurement techniques (item 14) was removed, concerning portfolios. The reason may be that other items were written similarly in their sentence structure. This shows that PE teachers' basic knowledge about measurement and evaluation is insufficient (Daniel and King, 1998). Different cities with a random sampling procedure are recommended for future research. The three sub-dimensional adapted version for in-service PE teachers in this study was found to be a valid and reliable measuring tool. Further research to determine the level of in-service teachers' perceptions of general efficiency in measurement and evaluation is strongly suggested.

CONCLUSIONS

It is concluded that the scale can be utilized to determine the necessary efficiency perception levels and to assess perceptions of general efficiency in measurement and evaluation among in-service PE teachers. It is assumed that further national and international studies on the range of measurement and evaluation efficiency perceptions of PE teachers would have a significant impact on the relevant persons and institutions; as such studies remain limited in number to date. It is recommended to establish new experimental designs to support the development of the general efficiency levels of measurement and evaluation of PE teachers, to review the educational processes of PE teachers in their courses regarding the obtained perceptions of their levels of efficiency, and to include other topics concerning measurement and evaluation in teachers' ongoing vocational training and tutorials (Richards & Templin, 2011). It is also important to compare the level of PE teachers in different countries that are implementing the constructivist theory in their curriculum, and their general efficiency related to measurement and evaluation levels.

Acknowledgments

The study has been presented partially in the 8th International Congress of Research in Education at Manisa, Turkey. The authors would also like to thank the Physical Education Teachers for their valuable comments and their participation.

Authors' contributions

The authors confirm sole responsibility for the following: study conception and design, data collection, analysis and interpretation of results, and manuscript preparation.

Declaration of conflict interest

The authors declare there is no conflict of interest.

REFERENCES

Arslan, Y., Erturan-İlker, G., & Demirhan, G. (2013). Geliştirilen bir ölçme ve değerlendirme gelişim programının beden eğitimi öğretmen adaylarının ölçme ve değerlendirmeye ilişkin algılarına etkisi. *Educational Sciences: Theory & Practice*, 13(2), 1107-1124.

Avşar, Z. (2012). *Creating initial physical education teacher education standards for Turkey.* (Unpublished master's thesis). Eskisehir, Turkey: Department of Educational Sciences Curriculum and Instruction Program, Graduate School of Educational Sciences, Anadolu University. Retrieved from http://tez2.yok.gov.tr/

Azzarito, L., & Ennis, C.D. (2003). A sense of connection: Toward social constructivist physical education. *Sport, Education and Society*, 8, 179-198. https://doi.org/10.1080/13573320309255

Baş, G., & Beyhan, Ö. (2016). Öğretmenlerin eğitimde ölçme ve değerlendirmeye yönelik özyeterlik algılarının bazı değişkenler açısından incelenmesi. *Journal of Measurement and Evaluation in Education and Psychology*, 7(1), 18-32. https://doi.org/10.21031/epod.34199

Bell, S.S. (2001). *Beginner's guide to uncertainty of measurement*, Issue: 2, United Kingdom: National Physical Laboratory.

Bıçak, B., & Çakan, M. (2004). *Views of the high school teachers about assessment and evaluation practices in classroom.* Milli Eğitim Bakanlığı, Middle School Restructuring Symposium Proceedings, Ankara.

Boyd, W. (1966). *The History of Western Education*, 8th ed., revised by E. J. King (London: Adam & Charles).

Brennan, J., & Teichler, U. (2008). The future of higher education and of higher education research. *Higher Education*, 56, 259–264. https://doi.org/10.1007/s10734-008-9124-6.

Büyüköztürk, Ş. (2002). Factor analysis: basic concepts and using to development scale. Ankara University, School of Education, Department of Educational Sciences, *Educational Administration in Theory and Practice*, 8(32), 470-483. Retrieved from https://dergipark.org.tr/en/pub/kuey/issue/10365/126871.

Büyüköztürk, Ş., Çakmak, E.K., Akgün, O.E., Karadeniz, S., & Demirel, F. (2012). *Scientific Research Methods.* Ankara: Pegem Akademi.

Byrne, B.M. (2001). *Multivariate applications book series. Structural equation modeling with AMOS: Basic concepts, applications, and programming.* Mahwah, NJ, U.S.: Lawrence Erlbaum Associates Publishers.

Çalık, S. (2007). *A study on the perceptions of classroom teachers about measurement and evaluation process of new primary school curriculum.* In Proceedings of the 16. National Educational

Sciences Congress (pp. 252-253). Tokat, Turkey: Gaziosmanpaşa University, Faculty of Education.

Çelik, D. (2005). *What should be the Assessment in schools?* MEB Publications, Ankara.

Çelik, Z., & Arslan, Y. (2012). Determination of prospective physical education teachers' measurement and evaluation common competency perceptions. *Selçuk University Journal of Physical Education and Sport Science*, 14(2), 223-232.

Daniel, L.G., & King, D.A. (1998). A Knowledge and use of testing and measurement literacy of elementary and secondary teachers. *Journal of Educational Research*, 91(6), 331-344. https://doi.org/10.1080/00220679809597563

Darling-Hammond, L. (2014). One piece of the whole: Teacher evaluation as part of a comprehensive system for teaching and learning. *American Educator*, 38(1), 4-13. Retrieved from http://search.proquest.com/docview/1651832231?accountid=10920.

Darling-Hammond, L., & Snyder, J. (2000). Authentic Assessment of teaching in context. *Teaching and Teacher Education*, 16, 523-545. https://doi.org/10.1016/S0742-051X(00)00015-9.

Dyson, B. (2002). The implementation of cooperative learning in an elementary physical education program. *Journal of Teaching in Physical Education*, 22, 69-85. https://doi.org/10.1123/jtpe.22.1.69.

Gelbal, S., & Kelecioğlu, H. (2007). Teachers' proficiency perceptions of about the measurement and evaluation techniques and the problems they confront. *Hacettepe University Journal of Education*, 33, 135-145. https://dergipark.org.tr/en/pub/hunefd/issue/7805/102347.

Goldhaber, D. (2015). *Teacher effectiveness research and the evolution of U.S. teacher policy* (The Productivity for Results Series No. 5). George W. Bush Institute, Education Reform Initiative. Retrieved from http://gwbcenter.imgix.net/Resources/gwbi-teacher-effectiveness-research.pdf.

Gömleksiz, M.N., & Bulut, I. (2007). An assessment of the implementation of new science and technology curriculum. *Hacettepe University Journal of Education*, 32(32), 76-88. https://dergipark.org.tr/en/pub/hunefd/issue/7804/102313.

Grennon-Brooks, J., & Brooks, M.G. (1993). *In search of understanding: The case for constructivist classrooms.* Alexandria, VA: Association for Supervision and Curriculum Development.

Harrison, J.M., Blakemore, C.L., & Buck, M.M. (2001). *Instructional strategies for secondary school physical education.* New York: McGraw-Hill. Humanities, Social Sciences & World Languages

Hu, L.T., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. https://doi.org/10.1080/10705519909540118

Karaca, E. (2004). Development of a likert type competence perception scale towards measurement and evaluation competencies of teacher candidates. *Dumlupınar University Journal of Institute of Social Sciences*, 10, 15-29.

Kline, P. (1994). *An Easy Guide to Factor Analysis:* New York: Routledge.

Lacy, A.C., & Hastad, D.N. (2006). *Measurement & evaluation in physical education and exercise science.* San Francisco: Pearson: Benjamin Cummings.

Ministry of National Education (MoNE) (2002). *Teacher Competencies.* MEB: EARGED, Milli Eğitim Basımevi, Ankara.

Ministry of National Education (MoNE) (2009). *Physical Education Learning Teaching Program and Guide (9-12. sınıflar) Öğretim Programı ve Kılavuz.* Directorate of State Books, Ankara.

Ministry of National Education (MoNE) (2013). *Physical Education Learning Teaching Program and Guide (5-8. sınıflar).* Directorate of State Books, Ankara.

Ministry of National Education (MoNE) (2018). *Secondary School Physical Education and Sport Learning Curriculum.* Ankara.

Nartgün, Z. (2008). Measurement and evaluation common competency perception scale for prospective teachers: a validity and reliability study. *Journal of Abant Izzet Baysal University Faculty of Education*, 8(2), 85-94. Retrieved from https://dergipark.org.tr/en/pub/aibuefd/issue/1495/18083.

Özbaşı, D., & Çıkrıkçı-Demirtaşlı, N. (2013). Primary school teachers' perceptions of their competencies regarding measurement and evaluation in terms of some variables. *Ankara University, Journal of Faculty of Educational Sciences*, 46(2), 25-46. https://doi.org/10.1501/Egifak_0000001293.

Özgül, F., & Kangalgil, M. (2018). Beden eğitimi öğretmenlerinin ölçme ve değerlendirme araçları ile ilgili görüşlerinin incelenmesi, *SPORMETRE*, 16(2), 60-71. https://doi.org/10.1501/Sporm_0000000355.

Özkoparan, O., & İnan, M. (2018). Research on the teachers' views about the use of alternative assessment and evaluation methods for physical education and sports courses. *Journal of Physical Education and Sports Studies*, 10(1), 23-38. doi:10.30655/besad.2018.0

Perrenet, J.C., & Terwel, J. (1997). Learning together in multicultural groups: a curriculum innovation. *Curriculum and Teaching*, 12(1), 31-44. https://doi.org/10.7459/ct/12.1.04.

Popham, W.J. (2013). *Evaluating America's teachers: Mission possible?* Thousand Oaks, CA: Corwin.

Richards, K.A., & Templin, T. (2011). The influence of a state mandated induction assistance program on the socialization of a beginning physical education teacher. *Journal of Teaching in Physical Education*, 20, 340-357. https://doi.org/10.1123/jtpe.30.4.340.

Rink, J.E. (2013). Measuring teacher effectiveness in physical education. *Research Quarterly for Exercise and Sport*, 84(4), 407-418. https://doi.org/10.1080/02701367.2013.844018

Rovegno, I. & Dolly, J.P. (2006). *Constructivist perspectives on learning.* In Handbook of physical education (Eds) D. Kirk, D. Macdonald and M. O'Sullivan (pp. 242–261). London: Sage.

Salimin, N., Jani, J., Shahril, M.I., & Elumalai, G. (2015). Validity and reliability of comprehensive assessment instruments for handball and badminton games in physical education. *Asian Social Science*, 11(23), 12-21. Retrieved from http://search.proquest.com/docview/1721369991?accountid=10920.

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

Şirin, E.F., Çağlayan, H.S., & İnce, A. (2009). The perceptions of physical education teachers about competency levels of measurement-evaluation subject in the new primary school teaching program. *Gazi Journal of Physical Education and Sport Science*, 19(1), 25-40. Retrieved from https://dergipark.org.tr/tr/pub/gbesbd/issue/28044/304857.

Şirinkan, A., & Erciş, S. (2009). Research for teaching methods and measuring and evaluation criteria applied in physical training and sports classes in primary education schools. *Niğde University Journal of Physical Education and Sport Sciences*, 3(3), 184-189.

Şirinkan, A., & Gündoğdu, K. (2011). The perceptions of teachers in relation to physical education curriculum and instructional plans. *Elementary Education Online*, 10(1), 144-159.

Smith, K. (2005). New methods and perspectives on teacher evaluation. In D. Beijaard et al. (Eds.), *Teacher professional development in changing conditions*, (pp. 95-114). Netherlands: Springer.

Tagele, A., & Bedilu, L. (2015). Teachers competence in the educational Assessment of students: the case of secondary school teachers in the amhara national regional state. *The Ethiopian Journal of Education*, 35, 163-191.

Turgut, M.F., & Baykul, Y. (2010). *Eğitimde Ölçme ve Değerlendirme*. Ankara: Pegem Akademi.

Uçar, A. (2001). *The evaluation of primary education teachers performance in Istanbul.* (Unpublished master's thesis), Marmara University, Institute of Educational Sciences. Retrieved from http://tez2.yok.gov.tr/

Ulutaş, S., & Erman, M. (2011). Cumhuriyetin kuruluşundan günümüze Türkiye'de uygulanmış olan ilkokul, ortaokul ve ilköğretim okulları öğretim programlarında ölçme ve değerlendirme. *Journal of Measurement and Evaluation in Education and Psychology*, 2(1), 148-154. Retrieved from https://dergipark.org.tr/tr/pub/epod/issue/5806/77237

Ünlü, H., Sünbül, A.M., & Aydos, L. (2008). The study of physical education teachers' self-efficacy scale of validity and reliability. *Ahi Evran University Kırşehir Education Faculty Journal*, 9(2), 23-33. Retrieved from https://dergipark.org.tr/tr/pub/kefad/issue/59525/856024.

Von Glaserfeld, E. (1995). *Radical constructivism: A way of knowing and learning.* Bristol, PA: The Falmer Press.

Wise, S.L., Lukin, L.E., & Roos, L.L. (1991). Teacher beliefs about training in testing and measurement. *Journal of Teacher Education*, 42(1), 37–42. https://doi.org/10.1177/002248719104200106.

Yager, R.E. (1991). The constructivist learning model: Towards real reform in science education. *Science Teacher*, 58(6), 52-57.

Yaykın, N. (2015). *Examining Turkish in-service physical education teachers' levels of measurement and evaluation.* (Unpublished master's thesis). Çanakkale, Turkey: Health Sciences Institute, Çanakkale Onsekiz Mart University. Retrieved from http://tez2.yok.gov.tr/

Yeşilyaprak, B. (2006). *Guidance services in education.* Ankara: Nobel Yayınları.

YÖK (2008). *Teacher competencies.* Retrieved November 11, 2008 from yok.gov.tr./egitim/ogretmen/akr2/ek3/ek31.html.

Zhang, Z., & Burry-Stock, J.A. (2003). Classroom assessment practices and teachers' self-perceived assessment skills. *Applied Measurement in Education*, 16(4), 323-342. 10.1207/S15324818AME1604_4.

Zuckerman, G. (2003). The learning activity in the first years of schooling: The developmental path toward reflection. In A. Zoulin, B. Gindis, V.S. Agevey and S.M. Miller (Eds),

Vygotsky's educational theory in cultural context (pp. 177-199). Cambridge, UK: Cambridge University Press.

Appendices

Beden Eğitimi Öğretmenleri için Ölçme ve Değerlendirme Genel Yeterlik Algısı Ölçeği (In Turkish)

Bu ölçek görev yapmakta olan beden eğitimi öğretmenlerin ölçme ve değerlendirme genel yeterlik algılarını tespit etmek amacıyla hazırlanmıştır. Üç boyuttan oluşan ölçekte toplam 23 genel yeterlik ifadesi bulunmaktadır. Her bir ifade kapsamıyla birlikte verilmiştir.

ÇokYeterliyim (5) **Yeterliyim (4)** **Orta Düzeyde Yeterliyim (3)** **Yetersizim (2)** **Çok Yetersizim (1)**

	İfadeler Kapsam	**Genel Yeterlikler**				
		[5]	**[4]**	**[3]**	**[2]**	**[1]**
1	Doğrudan, dolaylı ve türetilmiş ölçme yapabilirim.	5	4	3	2	1
2	Tanılayıcı, biçimlendirici ve değer biçmeye dönük değerlendirme yapabilirim.	5	4	3	2	1
3	Nitel/nicel, sürekli/süreksiz, bağımlı/bağımsız/kontrol değişkenleri belirleyebilirim.	5	4	3	2	1
4	Adlandırma/sınıflama, sıralama, eşit aralıklı, oranlı ölçekler hazırlayabilirim.	5	4	3	2	1
5	Görünüş, kapsam, yapı, benzer ölçekler, yordama geçerliği yapabilirim.	5	4	3	2	1
6	Ölçme hatası, güvenirlik, güvenirlik türleri: test-tekrar test, paralel formlar, eşit yarılar, iç tutarlılık, puanlayıcılar arası uyum yapabilirim.	5	4	3	2	1
7	Kazanımın yapısına uygun çoktan seçmeli soru tipini belirleme, yazma, uygulama ve puanlama yapabilirim.	5	4	3	2	1
8	Kazanımın yapısına uygun kısa cevaplı soru yazma, uygulama ve puanlama yapabilirim.	5	4	3	2	1
9	Kazanımın yapısına uygun eşleştirme tipi soru yazma, uygulama ve puanlama yapabilirim.	5	4	3	2	1
10	Kazanımın yapısına uygun Doğru/yanlış tipi soru yazma, uygulama ve puanlama yapabilirim.	5	4	3	2	1
11	Kazanımın yapısına uygun açık uçlu soru yazma, uygulama ve puanlama yapabilirim.	5	4	3	2	1
12	Kazanımın yapısına uygun performans görevi belirleme, uygulama ve puanlama yapabilirim.	5	4	3	2	1
13	Öğrenci ürün dosyasına dayalı ölçme sürecini uygulayabilirim.	5	4	3	2	1
14	Duyuşsal ve psikomotor nitelikleri ölçebilirim.	5	4	3	2	1
15	Madde güçlük indeksi ve madde ayırıcılık gücü hesaplayıp yorumlayabilirim.	5	4	3	2	1
16	Frekans dağılımları ve grafiksel gösterimler hazırlayabilirim.	5	4	3	2	1
17	Ortalama, ortanca, mod vb. hesaplayıp yorumlayabilirim.	5	4	3	2	1
18	Ranj, standart sapma, varyans vb. hesaplayıp yorumlayabilirim.	5	4	3	2	1
19	Normal dağılım, çarpıklık, basıklık vb. hesaplayıp yorumlayabilirim.	5	4	3	2	1
20	Verilerin yapısına uygun korelasyon tekniğinin belirlenmesi, hesaplayıp yorumlayabilirim.	5	4	3	2	1
21	T-testi, F testi vb. istatistikleri hesaplayıp yorumlayabilirim.	5	4	3	2	1
22	Mutlak değerlendirme, bağıl değerlendirme vb. yapabilirim.	5	4	3	2	1
23	Madde güçlük indeksi ve madde ayırıcılık gücü hesaplayıp yorumlayabilirim.	5	4	3	2	1