

# The Impact of a Developed Measurement and Evaluation Development Program on Pre-service Physical Education Teachers' Perceptions Related to Measurement and Evaluation\*

Yunus ARSLAN<sup>a</sup>  
Nevşehir University

Gökçe ERTURAN İLKER<sup>b</sup>  
Pamukkale University

Gıyasettin DEMİRHAN<sup>c</sup>  
Hacettepe University

## Abstract

The purpose of this study was to investigate the impact of the Measurement and Evaluation Development Program on pre-service physical education teachers' general perceptions and competency perceptions related to alternative assessment in physical education, and their competency perceptions related to educational measurement and evaluation. The subjects were 26 pre-service physical education teachers ( $\bar{x}_{Age} = 23.26 \pm 1.21$ ). In this research, a pre-test/post-test control group experimental design was used. The experimental group participated in an 8-week Measurement and Evaluation Development Program and a 3-week field experience. The results indicated that the Measurement and Evaluation Development Program was effective in changing pre-service physical education teachers' general perceptions ( $p < .05$ ) and competency perceptions ( $p < .05$ ) related to alternative assessment in physical education, and their competency perceptions ( $p < .05$ ) related to educational measurement and evaluation, positively.

## Key Words

Physical Education, Measurement and Evaluation Development Program, Alternative Assessment, Perception, Competency Perception, Preservice Physical Education Teacher.

The purpose of physical education (PE) does not include only the short term sport-specific gains, it is also for the development of physical fitness for an active lifestyle (Sallis & McKenzie, 1991; Shepard, 2000). In view of information that physical education also aims to *create desired changes in individuals' behaviors* (Ertürk, 1998, p. 12), *physical,*

*psychomotor, cognitive, emotional and social gains defined in physical education curriculums should be measured and evaluated* (Milli Eğitim Bakanlığı [MEB], 2008, p. 216). Alternative assessment tools (self-assessment, peer assessment, rubric, portfolio, check list etc.) were also suggested alongside the traditional forms of assessment in the new physical

\* This research is a part of first author's Doctoral Thesis.

- a** Yunus ARSLAN, Ph.D., is currently an assistant professor at Department of Physical Education and Sport Education. His research interests include measurement and evaluation in physical education and sport, teacher education, methodology [quantitative, qualitative ve mixed paradigm] and advanced statistic applied SPSS and LISREL. Correspondence: Nevşehir University, Faculty of Education, Department of Physical Education and Sport Education, 50300, Nevşehir, Turkey. E-mail: yarслан@nevsehir.edu.tr Phone: +90 384 228 1000/2339.
- b** Gökçe ERTURAN İLKER, Ph.D., Contact: Pamukkale University, School of Sport Sciences and Technology, Denizli, Turkey, E-mail: gokce.erturan@gmail.com.
- c** Gıyasettin DEMİRHAN, Ph.D. Contact: Hacettepe University, School of Sport Sciences and Technology, Ankara, Turkey, E-mail: demirhang@gmail.com.

education curriculums which came into force in the 2006-2007 academic year for elementary and secondary schools and in 2009-2010 academic year for high schools (MEB, 2007, 2009). But as stressed by Lacy and Hastad (2006, p. 106), one of the biggest challenges confronting today's PE teacher is that of developing meaningful assessment tools that evaluate students' progress on achieving stated performance objectives. This situation falls directly under the responsibility of physical education teacher education (PETE) programs.

On the other hand, related studies (Arslan & Çelik, 2010; Avşar, 2009; Kneer, 1986; Veal, 1988; Yılmaz & Gündüz, 2008) showed that in-service PE teachers and pre-service PE teachers (PPETs) perceived that measurement and evaluation classes in PETE programs were insufficient. Similar findings (Birgin & Tutak, 2006; Brookhart, 1993; Campbell & Evans, 2000; Gök & Şahin, 2009; Özseveç, Çepni, & Demircioğlu, 2004; Sağlam-Arslan, Devocioğlu-Kaymakçı, & Arslan, 2009; Wise, Lukin, & Roos, 1991) are prominent in the other research areas. Study results (Çakan, 2004; Daniel & King, 1998; Gök & Şahin; Sağlam-Arslan et al., 2009; Şirin, Çağlayan, & İnce, 2009) that inadequate education causes insufficient perceptions related to measurement and evaluation and, alternative assessment, specifically.

In several studies (DeLuca & Klinger, 2010; Lund & Veal, 2008; Otero, 2006; Struyven, Dochy, & Janssens, 2008; Wood, 1996) it is emphasized that pre-service teachers should be given the opportunity for more planning and implementing of assessment; however there is no study examining the impact of an implemented developed program. This situation needs to be examined, as it is an intriguing matter. Currently, little is known about PPETs' perceptions related to measurement and evaluation, whether those perceptions are changed by their special development program experiences. Karp and Woods (2008), highlighted that such information can be useful to PETE programs charged with facilitating PPETs to develop as highly qualified beginning teachers who conduct meaningful assessments with their students. From this point of view, it was aimed to investigate the impact of the Measurement and Evaluation Development Program (MEDP) on PPETs' general perceptions and competency perceptions related to alternative assessment in physical education, and their competency perceptions related to educational measurement and evaluation.

## Purpose

The purpose of this study was to investigate the impact of the Measurement and Evaluation Development Program on pre-service physical education teachers' general perceptions and competency perceptions related to alternative assessment in physical education, and their competency perceptions related to educational measurement and evaluation.

## Method

### Research Design

In this research, a pre-test/post-test control group experimental design was used.

### Subjects

The subjects were 26 PPETs ( $\bar{x}_{Age} = 23.26 \pm 1.21$ ) from a Turkish university. 7 subjects were females ( $\bar{x}_{Age} = 23.14 \pm 1.77$ ) and 19 were males ( $\bar{x}_{Age} = 23.31 \pm 1.00$ ). The PPETs were randomly placed into either an experimental group (3 Females, 10 males) or a control group (4 Females, 9 Males). The study was conducted at the same time as the PPETs' teaching practice course (internship). The PPETs took a measurement and evaluation course in their second year (fourth semester) and there is no practical section for this course (Theoretical: 3, Practical: 0).

### Instruments

**Physical Education Alternative Assessment Perception Scale (PEAAPS):** As a result of an examination of resources (Büyükoztürk, 2007, 2009; Tavşancıl, 2006; Tezbaşaran, 1996) PEAAPS was developed by the first researcher. The PPETs' perception related to alternative assessment in physical education was assessed using 21 items. PPETs rated each item on a 7-point scale, ranging from 1 (*not at all true for me*) to 7 (*very true for me*). Exploratory factor analysis revealed that the PEAAPS has a structure where the first factor explains 37.24%. Cronbach's alpha value was .85 for the PEAAPS.

**Physical Education Measurement and Evaluation Competency Perception Scale (PEMECPS):** The PEMECPS originally developed by Şirin et al. (2009), was used to measure the PPETs' competency perception related to measurement and evaluation in physical education. The scale has 31 items and three subscales (*selecting appropriate measurement tools, applying measurement tools and evaluation the data*). The PPETs rated each item on a 5-point scale, ranging from 1 (*Strongly Disagree*) to 5 (*Strongly*

Agree). Exploratory factor analysis revealed that the PEAAPS has a structure with three factors where the first factor explains 35.75%. Confirmatory factor analyses determined that the PEMECPS displays ( $RMSEA = .06$ ,  $CFI = .96$ ,  $GFI = .91$ ,  $AGFI = .90$ ,  $NFI = .93$ ) acceptable values (Frias & Dixon, 2005; Schermelleh-Engel, Moosbrugger, & Müller, 2003; Schumacker & Lomax, 2004; Sümer, 2000) for the PPETs. Cronbach's alpha values for the PEMECPS, selecting the appropriate measurement tools, applying measurement tools and evaluating the data were: .93, .88, .87, and .87, respectively.

**Measurement and Evaluation Common Competency Perception Scale for Prospective Teachers (MDCCPS):** The MDCCPS originally developed by Nartgün (2008) was used to measure the PPETs' common competency perception related to measurement and evaluation in general education. The scale has 24 items and three subscales (*basic concepts, measurement techniques and statistical analysis and reporting*). The PPETs rated each item on a 5-point scale, ranging from 1 (*I am very incompetent*) to 5 (*I am very competent*). Confirmatory factor analyses (Crocker & Algina, 1986; Jöreskog & Sörbom, 1993; Tabachnick & Fidell, 2007) determined that the MDCCPS displays ( $RMSEA = .07$ ,  $CFI = .95$ ,  $GFI = .92$ ,  $AGFI = .90$ ,  $NFI = .92$ ) acceptable values (Frias & Dixon, 2005; Schermelleh-Engel et al., 2003; Schumacker & Lomax, 2004; Sümer, 2000) for PPETs. Cronbach's alpha values for the MDCCPS, basic concepts, measurement techniques and statistical analysis and reporting were .90, .72, .85, and .86, respectively.

**Measurement and Evaluation Concept Test (MECT):** Before the main study, MECT was used to determine experimental and control groups' equality. The MECT, developed by Arik (2006), has 13 items with two sections. In the first section, there is concept information with two options, and in the second there is concept description with four options. Misconception emerges after giving right answers for the first section and wrong answers for the second section. Test-retest reliability of MECT for PPETs was  $R = .86$ .

**Focus Groups:** In the current study, the focus group aimed to support the quantitative data with the qualitative data. Using the maximum variation sampling, the focus group were conducted with 5 PPETs from the experimental group to get more elicited information about the intervention.

## Procedure

Permission was sought from the Governor of Denizli, the Directorship of School of Sport Sciences and Technology and an ethical committee to conduct this

study. The Physical Education Alternative Assessment Perception Scale (PEAAPS), Physical Education Measurement and Evaluation Competency Perception Scale (PEMECPS), Measurement and Evaluation Common Competency Perception Scale for Prospective Teachers (MDCCPS) and Measurement and Evaluation Concept Test (MECT) were applied as pre-tests to the PPETs in experimental and control groups. It was found that there were no significant differences between groups according to the MECT, PEMECPS and MDCCPS pre-test scores ( $p > .05$ ). The experimental group then participated in an 8-week MEDP and a 3-week field experience. The MEDP was generated from the related researches in physical education (Baumgartner, Jackson, Mahar, & Rowe, 2003; Harrison, Blakemore, & Buck, 2001; Lacy & Hastad, 2006) and general education (Atilgan, Kan, & Doğan, 2007; Turgut & Baykul, 2010; Yıldırım, 1999). During the MEDP, the PPETs participated in seminars which included educational measurement and evaluation and alternative assessment in physical education; and conducted three sample classes using alternative assessment tools in the field experience schools. The PEAAPS, PEMECPS and MDCCPS were applied as post-tests to experimental and control groups at the end of the MEDP, and a focus group conducted.

## Data Analysis

To address the main study purposes, results are presented as means and standard deviations to summarize the data. Mann-Whitney U test was applied to gain scores which gathered from PEAAPS, PEMECPS and MDCCPS. Obtained qualitative data was analyzed using descriptive analysis.

## Results

Statistical results indicated that the MEDP was effective in changing the PPETs' general perceptions ( $p < .05$  see, Table 1) and competency perceptions ( $p < .05$ ; see, Table 2) related to alternative assessment in physical education, and competency perceptions ( $p < .05$ ; see, Table 3) related to educational measurement and evaluation, positively.

**Table 1.**  
Means and Standard Deviations of the PEAAPS

Variable	Group	Pre-test		Post-test		Gain scores
		M	Sd	M	Sd	
PEAAPS	Experimental (n=13)	4.82	0.61	6.33	0.44	1.51*
	Control (n=13)	4.49	0.33	4.86	0.61	-0.37

\*Significantly different from gain score of control group ( $p < .05$ )

The present quantitative study results were also supported by qualitative data. Two examples of expressions made by PPETs after intervention are following:

“...If I consider my status before participating in this study, I think I am better at measurement and evaluation in physical education and general education. I can apply alternative assessments tools better henceforward...”

“...We are more knowledgeable now. In the future, we are going to be able to better...”

**Discussion**

Previous study findings (Arslan & Çelik, 2010; Campbell & Evans 2000; Çakan, 2004; DeLuca & Klinger, 2010; Gelbal & Kelecioğlu, 2007; Gök & Şahin, 2009; Karaca, 2003; Kilmen, Kösterelioğlu, & Kösterelioğlu, 2007) showed that in-service

and pre-service teachers perceived themselves as incompetent in the area of measurement and evaluation. It can be said that in-service and pre-service teachers should be given the opportunity for more planning and implementation of assessments. It is teacher education programs’ responsibility to equip their students, and in-service teachers should also participate in the professional development programs related to measurement and evaluation. Rink, Jones, Kirby Mitchell, and Doutis (2007), stated that the applicability of education reforms hinge upon taking the responsibility of change by teachers and teachers’ support of reform.

In the present study, the MEDP was an attempt to close the gap between the theoretical and the practical content knowledge of measurement and evaluation, and alternative assessment in physical education. The current study results are in agreement with those of DeLuca and Klinger (2010), Karp and Woods

**Table 2.**  
*Means and Standard Deviations of the PEMECPS*

Variables	Group	Pre-test		Post-test		Gain scores
		M	Sd	M	Sd	
SAMT	Experimental (n= 13)	31.76	8.20	51.69	2.49	19.93*
AMT		17.92	4.53	29.15	2.44	11.23*
ED		28.61	6.66	44.53	3.01	15.92*
PEMECPS		78.30	17.97	125.38	6.13	47.08*
SAMT	Control (n= 13)	29.76	7.06	31.69	5.43	1.93
AMT		16.46	3.28	17.53	4.27	1.07
ED		24.76	5.84	25.69	8.35	0.93
PEMECPS		71.00	14.18	74.92	15.37	3.42

**SAMT:** Selecting appropriate measurement tools, **AMT:** Applying measurement tools, **ED:** Evaluation the data; \*Significantly different from gain score of control group ( $p < .05$ ).

**Table 3.**  
*Means and Standard Deviations of the MDCCPS*

Variables	Group	Pret-test		Post-test		Gain Scores
		M	Sd	M	Sd	
BC	Deney (n= 13)	13.61	3.12	23.38	3.09	9.77*
MT		21.30	9.20	38.38	3.06	17.08*
SAR		16.46	6.07	35.69	3.22	19.23*
MDCCPS		51.38	14.03	97.46	6.95	46.08*
BC	Kontrol (n= 13)	13.07	3.63	13.46	4.71	0.39
MT		21.69	5.43	19.38	6.34	-2.31
SAR		20.15	4.46	20.92	8.12	0.77
MDCCPS		54.92	11.56	53.76	17.21	-1.16

**BC:** Basic concepts, **MT:** Measurement techniques, **SAR:** Statistical analysis and reporting; \*Significantly different from gain score of control group ( $p < .05$ ).

(2008), Lund and Veal (2008), Sağlam-Arslan, Avcı, and İyibil (2008), James, Griffin, and France (2005), Mintah (2003) and Struyven et al. (2008). These studies all suggested that there should be opportunities to plan and implement assessment tools for pre-service teachers in teacher education programs. Previous study findings of Aktağ and Walter (2005), Brinkerhoff (2006), Köseoğlu, Yılmaz, Gerçek, and Soran (2007) stressed that there is strong relationship between competency perception and experience, so they are also in line with our results.

Our findings call for the need to revise the measurement and evaluation courses which should also contain theoretical and practical content knowledge of alternative assessment in PETE programs, and also in the other teacher education programs.

## References/Kaynakça

Aktağ, I. ve Walter, J. (2005). Öğretmen adaylarının mesleki yeterlilik duygusu. *Spor metre, Beden Eğitimi ve Spor Bilimleri Dergisi*, 4, 127-131.

Arık, S. R. (2006). *İlköğretim öğretmenlerinin ölçme ve değerlendirme alanındaki kavram yanlışlarının belirlenmesi*. Yayınlanmamış yüksek lisans tezi, Ankara Üniversitesi, Eğitim Bilimleri Enstitüsü, Eğitim Bilimleri Anabilim Dalı, Ölçme ve Değerlendirme Programı, Ankara.

Arslan, Y. ve Çelik, Z. (2010, Kasım). *Aday beden eğitimi öğretmenlerinin ölçme ve değerlendirme genel yeterlik algılarının belirlenmesi*. 11. Uluslararası Spor Bilimleri Kongresi'nde sunulan sözel bildiri, Antalya.

Atılğan, H., Kan, A. ve Doğan, N. (2007). *Eğitimde ölçme ve değerlendirme*. Ankara: Anı Yayıncılık.

Avşar, Z. (2009). İlköğretim öğrencilerinin beden eğitimi dersi değerlendirme süreci ile ilgili görüşlerinin belirlenmesi. *Spor Bilimleri Dergisi*, 20 (3), 81-89.

Baumgartner, T. A., Jackson, A. S., Mahar, M. T., & Rowe, D. A. (2003). *Measurement for evaluation in physical education & exercise science*. New York: Mc Graw Hill.

Birgin, O. ve Tutak, T. (2006, Eylül). Yeni ilköğretim matematik öğretimi programına ilişkin öğretmen görüşleri. *I. Ulusal Matematik Eğitimi Öğrenci Sempozyumu Bildiri Özetleri Kitabı* içinde (s. 22). İzmir: Dokuz Eylül Üniversitesi.

Brinkerhoff, J. (2006). Effects of a long duration, professional development academy on technology skills computer self-efficacy and technology integration beliefs and practices. *International Society for Technology in Education*, 39 (1), 22-43.

Brookhart, S. M. (1993). Teachers' grading practices: Meaning and values. *Journal of Educational Measurement*, 30(2), 123-142.

Büyükoztürk, Ş. (2007). *DeneySEL desenler: Ön test-son test kontrol grubu desen ve veri analizi* (2. bs). Ankara: Pegem Yayınları.

Büyükoztürk, Ş. (2009). *Sosyal bilimler için veri analizi el kitabı* (10. bs). Ankara: Pegem Yayınları.

Campbell, C., & Evans, J. A. (2000). Investigation of preservice teachers' classroom assessment practices during student teaching. *The Journal of Educational Research*, 93 (6), 350-355.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart and Winston.

Çakan, M. (2004). Öğretmenlerin ölçme-değerlendirme uygulamaları ve yeterlik düzeyleri: İlk ve ortaöğretim. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 37 (2), 99-114.

Daniel, L.G., & King, D. (1998). A knowledge and use of testing and measurement literacy of elementary and secondary teachers. *Journal of Educational Research*, 91 (6), 331-344.

DeLuca, C., & Klinger, D. A. (2010). Assessment literacy development: identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice*, 17 (4), 419-438.

Ertürk, S. (1998). *Eğitimde program geliştirme*. Ankara: Meteksan.

Frias, C. M., & Dixon, R. A. (2005) Confirmatory factor structure and measurement invariance of the memory compensation questionnaire. *Psychological Assessment*, 17 (2), 168-178.

Gelbal, S. ve Kelecioğlu, H. (2007). Öğretmenlerin ölçme ve değerlendirme yöntemleri hakkındaki yeterlik algıları ve karşılaştıkları sorunlar. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 33, 135-145.

Gök, B. ve Şahin, A. E. (2009). İlköğretim 4. ve 5. sınıf öğretmenlerinin değerlendirme araçlarını çoklu kullanımı ve yeterlik düzeyleri. *Eğitim ve Bilim*, 34 (153), 127-143.

Harrison, J. M., Blakemore, C. L., Buck, M. M. (2001). *Instructional strategies for secondary school physical education*. New York: McGraw Hill.

James, A. R., Griffin, L., & France, T. (2005). Perceptions of assessment in elementary physical education: A case study. *Physical Educator*, 62 (2), 1-12.

Jöreskog, K., & Sörbom, D. (1993). *Lisrel 8: Structural equation modeling with the SIMPLIS command language*. USA: Scientific Software International.

Karaca, E. (2003). Öğretmen adaylarının ölçme ve değerlendirme yeterliklerine ilişkin Likert tipi bir yeterlik algısı ölçeğinin geliştirilmesi. *Dumlupınar Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 9, 179-198.

Karp, G. G., & Woods, M. L (2008). Preservice teachers' perceptions about assessment and its implementation. *Journal of Teaching in Physical Education*, 27, 327-346.

Kılmen, S., Kösterelioğlu, M. A. ve Kösterelioğlu, İ. (2007). Öğretmen adaylarının ölçme ve değerlendirme araç ve yaklaşımlarına ilişkin yeterlik algıları. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 7 (1), 129-140.

Kneer, M. (1986). A description of physical education instructional theory/practice gap in selected secondary schools. *Journal of Teaching in Physical Education*, 5, 91-106.

Köseoğlu, N., Yılmaz, M., Gerçek, C. ve Soran, H. (2007). Bilgisayar kursunun bilgisayara yönelik başarı, tutum ve öz-yeterlik inançları üzerine etkisi. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 33, 203-209.

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

Lund, J. L., & Veal, M. L. (2008). Measuring pupil learning-how do student teachers assess within instructional models? *Journal of Teaching in Physical Education*, 27, 487-511

Millî Eğitim Bakanlığı (MEB). (2007). *Beden eğitimi dersi (1-8. Sınıflar) öğretim program ve kılavuzu*. Ankara: Devlet Kitapları Müdürlüğü.

- Milli Eğitim Bakanlığı (MEB). (2008). *İlköğretim beden eğitimi (1-8. Sınıflar). Öğretmen kılavuz kitabı*. Ankara: Devlet Kitapları Müdürlüğü.
- Milli Eğitim Bakanlığı (MEB). (2009). *Ortaöğretim beden eğitimi dersi öğretim programı (9-12. Sınıflar)*. Ankara: Devlet Kitapları Müdürlüğü.
- Mintah, J. K. (2003). Authentic assessment in physical education: Prevalence of use and perceived impact on students' self-concept, motivation, and skill achievement. *Measurement in Physical Education and Exercise Science*, 7 (3), 161-174.
- Nartgün, Z. (2008). Öğretmen adayları için ölçme ve değerlendirme genel yeterlik algısı ölççeği: Geçerlik ve güvenilirlik çalışması. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 8 (2), 85-94.
- Otero, V. K. (2006). Moving beyond the "get it or don't" conception of formative assessment. *Journal of Teacher Education*, 57, 247-255.
- Özsevgeç, T., Çepni, S. ve Demircioğlu, G. (2004, Ekim). *Fen bilgisi öğretmenlerinin ölçme-değerlendirme okur-yazarlık düzeyleri*. VI. Ulusal Fen Bilimleri ve Matematik Eğitimi Kongresi'nde sunulan bildiri, Marmara Üniversitesi, İstanbul.
- Rink, J., Jones, L., Kirby, K., Mitchell, M., & Doutis, P. (2007). Teacher perceptions of a physical education statewide assessment program. *Research Quarterly for Exercise and Sport*, 8 (3), 204-215.
- Sağlam-Arslan, A., Avcı, N. ve İyibil, G. (2008). Fizik öğretmen adaylarının alternatif ölçme ve değerlendirme yöntemlerini algılama düzeyleri. *Dicle Üniversitesi Ziya Gökalp Eğitim Fakültesi Dergisi*, 11, 115-128.
- Sağlam-Arslan, A., Devocioğlu-Kaymakçı, Y. ve Arslan, S. (2009). Alternatif ölçme-değerlendirme etkinliklerinde karşılaşılan problemler: Fen ve Teknoloji öğretmenleri örneği. *Ondokuz Mayıs Üniversitesi Eğitim Fakültesi Dergisi*, 28, 1-12.
- Sallis, J. F., & McKenzie, T. L. (1991). Physical education's role in public health. *Research Quarterly for Exercise and Sport*, 62 (2), 124-137.
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, 8, 23-74.
- Schumacker, R. E., & Lomax, R.G. (2004). *A beginner's guide to structural equation modeling* (2nd ed.). New Jersey: Lawrence Erlbaum Associates, Publishers.
- Shepard, R. C. (2000). The legacy of physical education: Influences on adult lifestyles. *Pediatric Exercise Science*, 12 (1), 34-50.
- Struyven, K., Dochy, F., & Janssens, S. (2008). The effects of hands-on experience on students' preferences for assessment methods. *Journal of Teacher Education*, 59 (1), 69-88.
- Sümer, N. (2000). Yapısal eşitlik modelleri: Temel kavramlar ve örnek uygulamalar. *Türk Psikoloji Yazıları*, 3 (6), 74-79.
- Şirin, E. F., Çağlayan, H. S. ve İnce, A. (2009). Beden eğitimi öğretmenlerinin yeni ilköğretim programındaki ölçme-değerlendirme konusundaki yeterlilik düzeylerine ilişkin algıları. *Gazi Beden Eğitimi ve Spor Bilimleri Dergisi*, XIV (1), 25-40.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston: Allyn and Bacon, A Pearson Education Company.
- Tavşancıl, E. (2006). *Tutumların ölçülmesi ve SPSS ile veri analizi* (3. bs). Ankara: Nobel Yayın Dağıtım.
- Tezbaşaran, A. A. (1996). *Likert tipi ölçek geliştirme kılavuzu*. Ankara: Türk Psikologlar Derneği Yayınları.
- Turgut, M. F. ve Baykul, Y. (2010). *Eğitimde ölçme ve değerlendirme*. Ankara: Pegem Akademi.
- Veal, M. L. (1988). Pupil assessment perceptions and practices of secondary teachers. *Journal of Teaching in Physical Education*, 7, 327-342.
- 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.
- Wood, T. M. (1996). Evaluation and testing: The road less traveled. In S. Silverman, & C. Ennis (Ed.), *Student learning in physical education: Applying research to enhance instruction* (pp. 199-219). Champaign, IL: Human Kinetics.
- Yıldırım, C. (1999). *Eğitimde ölçme ve değerlendirme*. Ankara: ÖSYM Yayınları.
- Yılmaz, G. ve Gündüz, N. (2008). Ankara merkez ilköğretim okullarında görevli beden eğitimi ve spor öğretmenlerinin ölçme ve değerlendirme tekniklerinin uygulanışına ilişkin görüşleri. *SPORMETRE Beden Eğitimi ve Spor Bilimleri Dergisi*, VI (3), 103-111.