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; Shep-ard, 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 education curriculums.

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 (x̄age = 23.26±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
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; Özsevgeç, Çepni, & Demircioglu, 2004; Sağlam-Arslan, Devecioglu-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 (x̄_age = 23.26±1.21) from a Turkish university. 7 subjects were females (x̄_age = 23.14±1.77) and 19 were males (x̄_age = 23.31±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 (Büyüköztürk, 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 evaluating 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 Arık (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-re test 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 was conducted with 5 PPETs from the experimental group to get more elicit 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 (Atılgan, 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>
<thead>
<tr>
<th>Variable</th>
<th>Group</th>
<th>M</th>
<th>Sd</th>
<th>M</th>
<th>Sd</th>
<th>Gain scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>PEAAPS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Experimental (n=13)</td>
<td>4.82</td>
<td>0.61</td>
<td>6.33</td>
<td>0.44</td>
<td>1.51*</td>
</tr>
<tr>
<td></td>
<td>Control (n=13)</td>
<td>4.49</td>
<td>0.33</td>
<td>4.86</td>
<td>0.61</td>
<td>-0.37</td>
</tr>
</tbody>
</table>

*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 & Kelecioglu, 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 M | Sd    | Post-test M | Sd    |
| SAMT       | Experimental (n=13) | 31.76 | 8.20 | 51.69 | 2.49 |
| AMT        | Experimental (n=13) | 17.92 | 4.53 | 29.15 | 2.44 |
| ED         | Control (n=13) | 28.61 | 6.66 | 44.53 | 3.01 |
| PEMECPS    | Control (n=13) | 78.30 | 17.97 | 125.38 | 6.13 |
| SAMT       | Control (n=13) | 29.76 | 7.06 | 31.69 | 5.43 |
| AMT        | Control (n=13) | 16.46 | 3.28 | 17.53 | 4.27 |
| ED         | Control (n=13) | 24.76 | 5.84 | 25.69 | 8.35 |
| PEMECPS    | Control (n=13) | 71.00 | 14.18 | 74.92 | 15.37 |

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 M | Sd    | Post-test M | Sd    |
| BC        | Deney (n=13) | 13.61 | 3.12 | 23.38 | 3.09 |
| MT        | Deney (n=13) | 21.30 | 9.20 | 38.38 | 3.06 |
| SAR       | Deney (n=13) | 16.46 | 6.07 | 35.69 | 3.22 |
| MDCCPS    | Deney (n=13) | 51.38 | 14.03 | 97.46 | 6.95 |
| BC        | Kontrol (n=13) | 13.07 | 3.63 | 13.46 | 4.71 |
| MT        | Kontrol (n=13) | 21.69 | 5.43 | 19.38 | 6.34 |
| SAR       | Kontrol (n=13) | 20.15 | 4.46 | 20.92 | 8.12 |
| MDCCPS    | Kontrol (n=13) | 54.92 | 11.56 | 53.76 | 17.21 |

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


