academicJournals

Vol. 6(8) pp. 60-69, November 2015
DOI: 10.5897/JPESM2015.0238
Article Number: CFA134756145
ISSN 1996-0794

http://www.academicjournals.org/JPESM

Journal of Physical Education and Sport Management

Full Length Research Paper

Validity and reliability studies for scale of evaluating physical education teachers based on student ratings (SEPETBSR)

Numan Bahadir Kayisoglu

Karabuk University, Hasan Dogan School of Physical Education and Sports, Karabuk, Turkey.

Received 31 August 2015; Accepted 13 October, 2015

The purpose of the present study is to develop a scale based on student ratings for evaluating physical education teachers to be used in secondary schools. The work-group of the research consists of 892 students who participated voluntarily and were selected randomly among 9th, 10th, 11th, and 12th grade students who study at secondary schools in Karabuk province centre and its Safranbolu district in 2013-2014 academic year. Statistical analyses were conducted on the obtained data, necessary arrangements were conducted and Scale of Evaluating Physical Education Teachers based on Students Ratings (SEPETBSR) was found to consist of 3 factors and 48 items. KMO (Kaiser-Meyer-Olkin Measure of Sampling Adequacy) value of the scale was found as .97, and Bartlett Test of Sphericity value was calculated as 24238.74; and these values were significant at .01 level. Cronbach alpha internal consistency coefficients were calculated as .96 for the first factor *General Performance*; .91 for the second factor *Course Performance*, .71 for the third factor *Negative Attitude*; and .95 for the whole scale. Consequently, *Scale of Evaluating Physical Education Teachers based on Students Ratings (SEPETBSR)* was found to be a valid and reliable scale that can be used by researchers who study at the subject field.

Key words: Evaluating physical education teacher, students' evaluation, scale development.

INTRODUCTION

Studies in the field of education show that teachers play an important role in the achievement of students, and when teachers believe the importance of that role in students' learning, students' success increases at an alarming rate (Bandura, 1994; Tschannen et al., 1998).

With the developments and changes we are going through, job definition and responsibilities of teachers have been changing and expanding. Teachers are no more the provider of knowledge or the presenter of a skill model, but they have a role of easing both learning environment and learning processes. With this new understanding, teachers who are in interaction with students constantly can guide learning and its ways, implement the curriculum, lead teaching, and

E-mail: bahadirkayisoglu@karabuk.edu.tr. Tel: (+90) 505 481 2896 Fax: (+90).

evaluate both teaching and students (Beydogan, 2002; Bircan, 2003; Kavcar, 2003; Sonmez, 2003; Oddens, 2004).

Teachers are one of the most important components of an education system. Teachers have various roles including being an analyst, curriculum developer, material developer, mentor, team member, researcher, and professional (Richards and Lockhart, 1996).

Although there is no consensus on how to assess teacher quality, scholars do agree that the improvement of teaching is the most important step that we can take toward improving the quality of education of our students and learning outcomes (Perlman and McCann, 1998). Why is measuring *teaching effectiveness* so important? Because the evidence produced is used for major decisions about our future in academe (Berk, 2005). According to Erdogan (1991) measuring effectiveness is performance. Performance evaluation is the process in which a manager evaluates the performance of the employees in accordance with the pre-determined standards through comparison and measurement (Palmer, 1993), and it is the method used to determine to what extent employees attain the objectives set for them (Luecke, 2010).

In this context, the assumption that the ones who can evaluate an organization best, are the ones who are related to the human resources of that organization, constitutes the 360 degrees evaluation perception.

Managers need data from various resources in order to understand what is going on in an organization, to lead development and to increase the performance of the employees. Organizations that value development, and need the opinions from many resources in the decisions related to employees developed the upwards feedback idea as “360 degrees feedback”. In this approach, performance data are collected from employees, employees’ superiors, subordinates, and internal and external customers (Dessler, 1991).

Making many evaluations due to use of 360 degrees performance evaluation method may result in the mistake that senior managers who do not have the opportunity to observe especially teachers’ performance avoid making good or bad evaluations and give average scores or they evaluate many evaluatees through comparisons (Can et al., 2009). Additionally, an effective evaluation system can enable one to make distinctions between successful and unsuccessful teachers, and awarding success. Performance management can contribute to the formation of an education management understanding in which individual and organizational interests are considered together (Rebore, 2001).

According to Kocak (2006), taking students’ opinions in teachers’ performance evaluation is significant. Because students are the ones who know the teachers best in terms of teaching, and are affected from this. Besides, the numerosity of students compared to inspectors or managers’ evaluation will provide more reliable, objective and significant results. Moreover, the teachers’ preference for their own performance evaluation is listed according to order of priority as students, common branch teachers, other teachers, school managers, and parents.

Scale development studies in educational studies are important for the development of education and the performance evaluation of educators (Unlu, 2008). Review of the related literature did not present any scales for evaluation physical education teachers’ performance based on student ratings, which increases the importance of the present study. In accordance with this importance, the purpose of the present research is developing a scale for evaluating physical education teachers’ performance working at secondary schools based on student ratings.

METHODOLOGY

Scale of Evaluating Physical Education Teachers based on Students Ratings (SEPETBSR) was developed in accordance with validity study and reliability study steps as suggested by experts in the field (Tavsancil, 2002; Kagitcibasi, 2005; Sakalli, 2001; Karasar, 1998; Balci, 2005). In the present study, content validity and construct validity of the scale were tested.

Development of the scale

Before drafting *Scale of Evaluating Physical Education Teachers based on Students Ratings (SEPETBSR),* a comprehensive literature review was conducted. Then, “What do you think the important measures for the evaluation of a physical teacher education are?” question was asked to 65 randomly selected secondary school students, and asked for written answers to this open-ended question. In addition, through literature review a 70-item question repository that can be used in physical teacher education evaluation based on student ratings was created. 5 academic members of the subject field and a physical education teacher conducted required analyses and evaluations; some items were excluded, some more items were included, and consequently 55 items were included in the scale. After experts examined the scale draft, Turkish Language teachers corrected the spelling and expression mistakes. After the corrections, the scale was ready for pre-implementation. For scale draft, 5-level Likert type scaling was adopted. This type of scaling was developed by Likert, and it is considered as the most practical scaling type (Cetin, 2006; Tezbasaran, 1997). The grading is as *“Totally Agree, Agree, Agree to Some Extent, Disagree, Totally Disagree”*. The items were ordered within the scale, answering scale was created, and scale instructions were written, and finally the scale was ready for implementation.

Implementation of the scale and data analysis

Data of the present research were collected from students who study at secondary school in Karabuk Province centre and its Safranbolu District. After official approval was taken from Karabuk Provincial Directorate For National Education, 55-item scale was implemented on randomly selected 482 males and 410 females; a total of 892 voluntary students who study at secondary schools in Karabuk Province centre and its Safranbolu District. Data obtained in the present research were analysed with SPSS 15.0 (Statistical Package for Social Sciences) and LISREL 8.51 (Linear Structural Relation Statistics Package Program).

Table 1. KMO and Bartlett test results.

Kaiser-Meyer-Olkin Measure of Sample Adequacy		.97
Bartlett Sphericity Test	X^2	24238.74
	Sd	1128
	P	.000

FINDINGS

Exploratory factor analysis validity study

Construct validity of the scale was tested with factor analysis. With this purpose, data obtained in pre-implementation were tested for conformity to factor analysis. Table 1 shows Kaiser-Meyer-Olkin (KMO) and Bartlett test results conducted to test conformity of the data for factor analysis.

As shown in Table 1, calculated KMO conformity measure is 0.97. Leech et al. (2005); Sencan (2005) and Tavsancil (2001) stated that critic value was 0.50, and factor analysis could not be conducted under this value. Compared to the critical value, Kaiser-Meyer-Olkin value of the scale is between ".90-1.00", which is very high range (Buyukozturk e al., 2010). Bartlett Sphericity Test score for the same data was calculated as 24238,74 and significant at .01 level (X^2_{1128}=24238.74). These values show that, data obtained from pre-implementation are convenient for factor analysis. The quantity of the sample is adequate for factor analysis (Table 2).

Factor analysis results conducted via principal components analysis are presented. As shown in Table 2, there are 3 factors with eigenvalues higher than 1.5. These three factors explain the 44.90% of the total variance. Considering the initial eigenvalues, 3 factors were obtained as the eigenvalue for the first factor (17.84), eigenvalue for the second factor (2.17), and eigenvalue for the third factor (1.54) were higher than 1.5. It was concluded that 44.90% explained variance by 3 factors was at an acceptable level. As for before and after rotation values, the changes in variance explained by the factors were as; there was a decrease in the first factor, and there were increases in the second and third factors. The variance explained by the first factor makes up the 25.31% of total variance, the variance explained by the second factor makes up the 15.30% of total variance, and the variance explained by the third factor makes up the 4.09% of the total variance.

Figure 1 presents 3 major break points, and other break points are lower than 1.5 eigenvalue, and so they are not taken into consideration. This graphic shows that there are only three break points with eigenvalue over 1.5, and therefore the scale consists of 3 factors. Table 3 shows the load values of the factors in the pre-implementation form.

Table 3 shows that items 1-2-24-25-26-27-28-29-30-31-32-33-34-35-36-37-38-39-42-43-44-45-46-47-48-50-51-52-53-54-55 have the highest factor load value in the first factor; items 6-7-8-9-10-11-12-13-14-15-16-17-18-19 have the highest factor load value in the second factor; and items 23-40-41 and 49 have the highest factor load value in the third factor. Factor loads of the items in the first factor range between .41 and .70; factor loads of the items in the second factor range between .44 and .68; and factor loads of the items in the third factor range between .41 and .66. According to these load values, the scale consists of three factors and all the items have enough load values to be included in the scale. Factor analysis showed that items 3-4-5-20-21-22-28 had factor loads lower than .40 which was set as the limit value; so these items were excluded from the scale. According to Tabachnick and Fidell (2001), if the load value of an item is .40, the critical value, that item is "mediocre" (Cited in: Buyukozturk et al., 2010). In order to increase the explained variance of a factor, the limit value was determined as .40 factor load.

Table 2. Factor eigenvalues and explained variances.

Factor	Initial Eigenvalue			Total after Rotation		
	Total	Vary %	Cum %	Total	Vary %	Cum %
1	17.84	37.16	37.16	12.15	25.31	25.31
2	2.17	4.53	41.69	7.44	15.50	40.81
3	1.54	3.21	44.90	1.96	4.09	44.90

Item analyses based on item scale correlation are presented in Table 3. According to these findings, correlations values range between r=.31 (i23) and r=.75(i31); and are significant at .05 level. Item scale correlations of 48 items in the ultimate scale form are acceptable; which means all 48 items in the scale are qualified enough to be included in the scale.

After validity studies, 48 items with factor load values over .40 and item correlation values over .30 were included in the scale under 3 factors in the ultimate form of the scale. Thirty items in the first factor were titled *General Performance*, 14 items in the second factor were titled *Course Performance*, and 4 items in the third factor were titled *Negative Attitude*.

Validity study

As a proof of construct validity, the scales used in the present research were tested for conformity Turkish

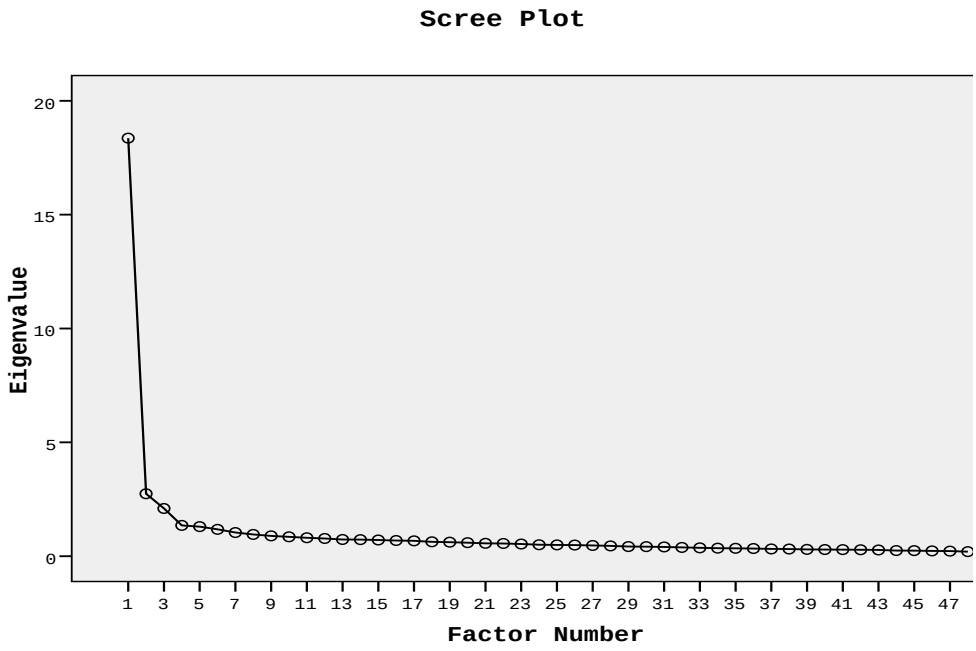

Figure 1. Scree-Plot graphic.

Table 3. Factor loads and total correlations of items.

İtem	1. Dim.	2. Dim.	3. Dim.	r	İtem	1.Dim.	2 . Dim.	3. Dim.	r
i1	.43			.54	i50	.55			.66
i2	.41			.52	i51	.58			.63
i24	.57			.62	i52	.64			.71
i25	.63			.71	i53	.61			.58
i26	.65			.69	i54	.53			.58
i27	.62			.71	i55	.49			.47
i29	.51			.61	i6		.44		.47
i30	.62			.68	i7		.52		.50
i31	.67			.75	i8		.48		.58
i32	.52			.57	i9		.68		.70
i33	.58			.61	i10		.60		.67
i34	.59			.66	i11		.60		.65
i35	.70			.73	i12		.46		.53
i36	.66			.71	i13		.63		.65
i37	.66			.73	i14		.66		.70
i38	.68			.74	i15		.67		.63
i39	.59			.64	i16		.69		.67
i42	.63			.62	i17		.55		.63
i43	.64			.66	i18		.55		.59
i44	.71			.75	i19		.55		.63
i45	.64			.66	i23			.41	.31
i46	.62			.68	i40			.66	.60
i47	.63			.70	i41			.66	.58
i48	.61			.60	i49			.60	.50

culture. Confirmatory factor analysis (CFA) was used to test whether the construct of the 3-factor, 48-item physical education teacher evaluation scale was confirmed. Confirmatory factor analysis (CFA) aims to determine to what extent a factorial model that consists of factors (latent variables) formed by many observable variables conforms to real data. The model to be examined may define a construct determined by using the data of an experiential study or a construct built based on a theory (Sumer, 2000). Many fit indexes are used to evaluate the validity of a model in CFA. Most frequently used ones among these are (Cole, 1987; Sumer, 2000); Chi-Square Goodness (χ^2), Root Mean Square Error of Approximation (RMSEA), Comparative Fit Index (CFI), Non-Normed Fit Index (NNFI), Normed Fit Index (NFI), and Goodness of Fit Index (GFI). The observed values in a scale model in X2/d<3; 0<RMSEA<.05; .97≤NNFI≤1; .97≤CFI≤1; .95≤GFI≤1 and .95≤NFI≤1 range indicate perfect fit, values in 4<X2/d<5; ,05<RMSEA≤.08; SRMR≤,08; .95≤NNFI≤.97; 0.95≤CFI≤0.97; .90≤GFI≤.95 and .90≤NFI≤.95 range indicate acceptable fit (Kline, 2005; Sumer, 2000).

CFA was conducted in order to evaluate whether the 3-factor and 48-item construct of the scale was confirmed. In the first CFA, it was examined whether there were any items with statistically not significant t values. According to this examination, there were no items with not significant t values.

Fit indexes were found as χ^2=3143.79, χ^2/sd= 2.92, CFI=.98, NNFI=.97 and NFI=.96, RMSEA=.080, SRMR=.057. When the data related to the factorial construct of the scale were investigated it was concluded that fit indexes of the scale were at sufficient level. Fit index values showed that there was a perfect fit and error values RMSEA and SRMR values showed acceptable fit. Fit statistics calculated with CFA showed that 3-factor construct of the scale were fit with the collected data in general.

Figure 2 shows that ultimate form of the scale formed of 48 items and 3 factors. Table 4 shows the regression values and t values related to the items. It shows that calculated regression coefficients and t values are significant and the model is confirmed. In the first factor, I37 is the most important item with R^2=0.70; and i55 is the least important factor with R^2=.27. In the second factor, I14 is the most important factor with R^2=.68, and I7 is the least important item with R^2=.32. In the third factor, I41 is the most important factor with R^2=.72, and I23 is the least important factor with R^2=.13. In general, I41 is the most important item of the scale with R^2=.72 value, and I23 is the least important item of the scale with R^2=.13.

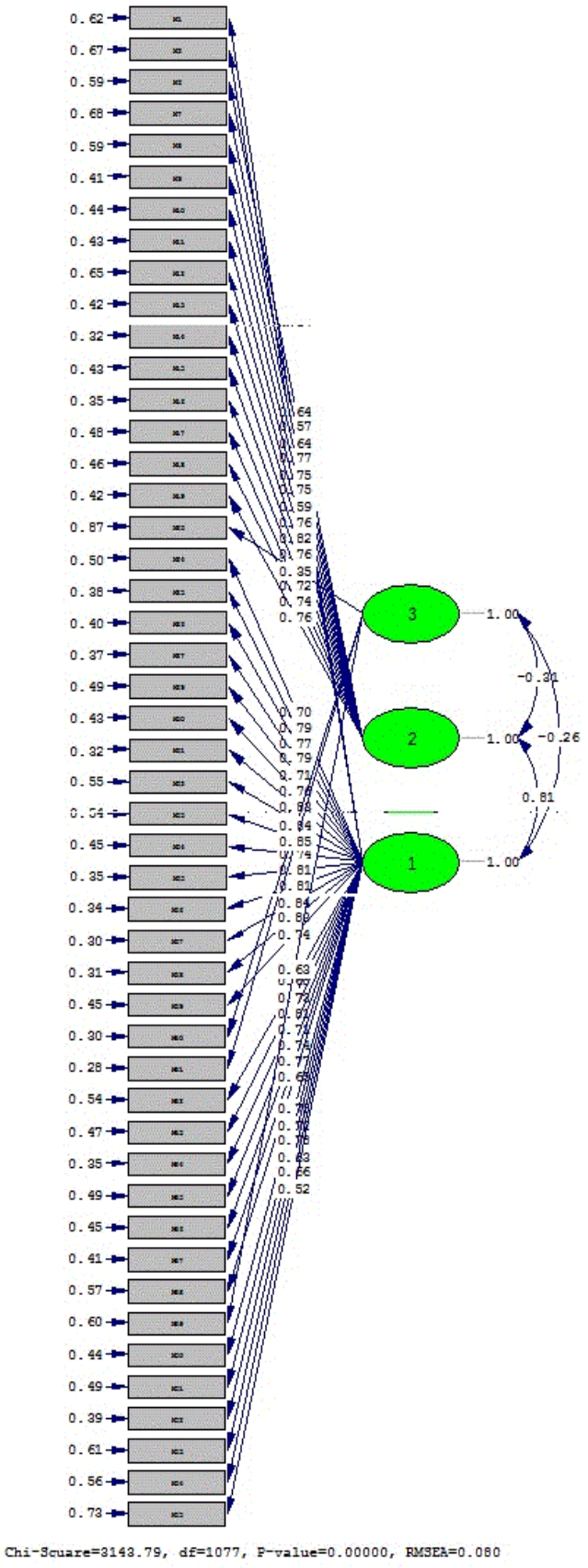

Figure 2. Path diagram related to the scale.

Reliability study

Table 5 shows the Cronbach alpha internal consistency

Table 4. CFA Regression and T Values.

First factor			Second factor			Third factor		
Items	**R^2**	**t**	**Items**	**R^2**	**T**	**Items**	**R^2**	**t**
I1	.38	20.25	I6	.41	20.96	I23	.13	10.07
I2	.33	18.63	I7	.32	18.08	I40	.70	27.17
I24	.50	23.98	I8	.41	21.08	I41	.72	27.59
I25	.62	28.17	I9	.59	26.93	I49	40	19.38
I26	.60	27.39	I10	.56	25.67			
I27	.63	28.44	I11	.57	26.06			
I29	.51	24.35	I12	.35	19.11			
I30	.57	26.45	I13	.58	26.58			
I31	.68	30.14	I14	.68	29.74			
I32	.45	22.54	I15	.57	26.23			
I33	.46	22.71	I16	.65	28.83			
I34	.55	25.71	I17	.52	24.40			
I35	.65	28.98	I18	.54	25.33			
I36	.66	29.33	I19	.58	26.58			
I37	.70	30.69						
I38	.69	30.36						
I39	.55	25.68						
I42	.46	22.85						
I43	.53	25.07						
I44	.65	29.08						
I45	.59	24.35						
I46	.55	25.60						
I47	.59	26.95						
I48	.43	21.80						
I50	.56	25.98						
I51	.51	24.51						
I52	.61	27.63						
I53	.39	20.66						
I54	.44	22.20						
I55	.27	16.49						

Table 5. Alpha reliability coefficients related to dimensions.

	1st Dimension	**2nd Dimension**	**3rd Dimension**	**Scale**
Number of Items	30	14	4	48
Cronbach α	.96	.91	.71	.95

coefficients of the items determined for each dimension and the test in general for the reliability of the scale. Because Cronbach alpha coefficients were calculated taking all items into consideration, it was the best coefficient that reflected the general reliability of the test in general among all calculation types (Ozdamar, 2004).

Table 5 shows that Cronbach alpha internal consistency coefficients are .96 for the 1st dimension, .91 for the 2nd dimension, .71 for the 3rd dimension, and .95 for the whole scale. These indicate acceptable internal consistency for all dimensions of the scale. Reliability coefficients above show that items have high-level reliability internally. Tezbasaran (1997: 47) stated that acceptable reliability coefficient for a likert type scale is close to 1. According to these findings reliability of all measurement tools used in the present research are at high levels.

DISCUSSION

Student ratings have been used for many years to

evaluate the performance of teachers in their classes (Stronge, 1997) and students are good sources of information about their instructors as knowing their situation well and observing them closely (Peterson, 1987). Review of studies on students' perceptions of teachers' effectiveness (Tuan et al. 2000; Hills et al. 2005; Shadreck and Issac 2012) has stated that students' expectance from teachers are strong content knowledge, effective pedagogical skills, and social competence.

Students' ratings are mostly used among university level (Marsh, 2007a) however, they can also be used in primary school level (Fauth et al., 2014) and secondary school level (De Jong and Westerhof, 2001; Kunter et al., 2008). From this point of view, the scale for secondary school students developed in the present study is considered as an applicable scale in terms of the level of students.

There are not many researches on evaluating physical education teachers in terms of attitude, behaviour, and performance. This kind of scale will contribute to increase the quality of education, to analyse the performance of physical education teachers, to review the educational processes and to plan the in-service training activities. Student ratings have dominated as the primary measure of teaching effectiveness for the past 30 years (Seldin, 1999). However, over the past decade there has been a trend toward augmenting those ratings with other data sources of teaching performance. Such sources can serve to broaden and deepen the evidence base used to evaluate courses and assess the quality of teaching (Arreola, 2000; Braskamp and Ory, 1994; Knapper and Cranton, 2001). Similarly, Penny (2003) stated that most researchers believe that the results of student ratings provide evaluators valid, reliable and valuable data concerning the quality and effectiveness of teaching (Penny, 2003) Concordantly, evaluating the performance of physical education teachers using various data resources such as peer ratings, self-evaluation, videos, alumni ratings, administrator ratings, teaching scholarship and learning outcome measurements (Berk, 2005) besides student ratings is important in terms of the validity and the reliability of the findings. It should be acknowledged that several researchers examining the reliability of student ratings have attempted to investigate the stability of student ratings across time, courses and instructors (Carle, 2009; Marsh, 2007b).

Conclusion

The present study tested the validity and reliability of *Scale of Evaluating Physical Education Teachers based on Students Ratings (SEPETBSR).* The ultimate form of the 48-item scale is presented in Appendix-1. According to factor analysis, the scale consists of three factors and the first factor, General Performance involves 30 items, second factor Course Performance involves 14 items, and the third factor Negative Attitude involves 4 items. In the 5-level likert type scaling type, the items in General Performance and Course Performance dimensions will be scores as *(1)Totally Disagree*, *(2) Disagree*, *(3) Undecided*, *(4) Agree*, and *(5) Totally Agree* and items in Negative Attitude dimension will be scores as *(5)Totally Disagree, (4) Disagree, (3) Undecided, (2) Agree, and (1) Totally Agree*. Accordingly, the lowest score to be got from the scale is 35, and the highest is 175. Findings of the present study show that *SEPETBSR* is a valid and reliable measurement tool. This scale can contribute to the experts studying in the subject field as a multiple evaluation tool to evaluate physical education teachers.

Finally the use of student evaluations of teaching performance has been an important but controversial tool in the improvement of teaching quality. It is important to use this type of scale that considers the students' opinions while constituting the 360 degrees evaluation in order to increase sufficiency of education although it is regarded as a contradictive method.

RECOMMENDATIONS

The scale developed for secondary school students can be adapted to students at other levels of education, and can be used as a multiple data resource for evaluating teachers' performance based on students' ratings, and therefore it can contribute to the subject field. Data obtained from this scale can be compared with various data resources such as peer ratings, self-evaluation, videos, alumni ratings, administrators' ratings in further researches. Moreover, it is suggested that a reliability and validity study can be implemented in order to find standard norms in Turkey as obtained data are limited with working group in Karabuk.

Conflict of Interests

The authors have not declared any conflict of interests.

REFERENCES

Arreola RA (2000). Developing a comprehensive faculty evaluation system: A handbook for college faculty and administrators on designing and operating a comprehensive faculty evaluation system (2nd ed.). Bolton, MA: Anker.

Balci A (2005). Sosyal Bilimlerde Arastirma Yontem Teknik ve İlkeler. (Besinci Baski). Ankara: Pegem A. Yayincilik.

Bandura A (1994). Self-efficacy. In V. S. Ramachaudran (Ed.), Encyclopedia of human behavior (Vol. 4, pp. 71-81). New York: Academic Press. (Reprinted in H. Friedman [Ed.], Encyclopedia of mental health. San Diego: Academic Press, 1998).

Berk RA (2005). Survey of 12 strategies to measure teaching effectiveness. Int. J. Teaching Learning Higher Educ. 17(1): 48-62.

Beydogan O (2002). Ogretim Stratejilerindeki Degismeler ve Ogretmenlerin Degisen Rolleri. Cagdas Egitim, 27 (287): 34-39.

Bircan İ (2003). Egitimde Yeni Yonelimler Gelismis Ulkelerde Sinif

Ogretmeni Yetistirme Uygulamalari. Egitimde Yansimalar: VII. Cagdas Egitim Sistemlerinde Ogretmen Yetistirme Ulusal Sempozyumu Bildirileri Kitabi, Sivas: Cumhuriyet Universitesi, 44-47.
Braskamp LA, Ory JC (1994). Assessing faculty work. San Francisco: Jossey-Bass.
Can H, Kavuncubasi S, Yildirim S. (2009). Kamu ve ozel kesimde insane kaynaklari yonetimi. Siyasal Kitabevi.
Carle AC (2009). "Evaluating College Students' Evaluations of a Professor's Teaching Effectiveness across Time and Instruction Mode (Online vs. Face-to-face) Using a Multilevel Growth Modeling Approach." Comp. Educ. 53: 429–435.
Cole DA (1987). Utility of confirmatory factor analysis in test validation research. J. Consult. Clin. Psychol. 55: 584-594.
Cetin S (2006). Ogretmenlik Meslegi Tutum Olceginin Gelistirilmesi, Gazi Universitesi Endustriyel Sanatlar Egitim Fakultesi Dergisi, 18, 28-37.
Dessler G (1991). Personnel/human resource management. Prentice Hall.
De Jong R, Westerhof KJ (2001). The quality of student ratings of teacher behaviour. Learn Environments Res. 4(1): 51-85.
Erdogan İ (1991). İsletmelerde Personel Secimi ve Basari Degerleme Teknikleri. İ.U.İsletme Fak.Yayin,No 28, İstanbul.
Fauth B, Decristan J, Rieser S, Klieme E, Buttner G (2014). Student ratings of teaching quality in primary school: Dimensions and prediction of student outcomes. Learning and Instruction, 29: 1-9.
Hills HC, Rowan B, Ball DL (2005) Effects of teachers' mathematical knowledge for teaching on student ability. Am. Educ. Res. J. 42(2):371–406.
Kagitcibasi C (2005). Yeni İnsan ve İnsanlar. (Onuncu Basim). İstanbul: Evrim Yayinevi.
Karasar N (1998). Bilimsel Arastirma Yontemi. (Sekizinci Basim). Ankara: Nobel Yayin Dagitim.
Kavcar C (2003). Alan Ogretmeni Yetistirme. Egitimde Yansimalar: VII Cagdas Egitim Sistemlerinde Ogretmen Yetistirme Ulusal Sempozyumu Kitabi (pp.81-89), Sivas: Cumhuriyet Universitesi.
Kline RB (2005), Principles and Practice of Structural Equation Modeling (2nd Edition ed.). New York: The Guilford Press.
Knappe, C, Cranton P (2001). Fresh approaches to the evaluation of teaching (New Directions for Teaching and Learning, No. 88). (Eds.) San Francisco: Jossey-Bass.
Kocak R (2006). The Validity and Reliability of the Teachers' Performance Evaluation Scale. Educ. Sci. Theory Prac. 6(3): 799-808.
Kunter M, Tsai YM, Klusmann U, Brunner M, Krauss S, Baumert J (2008). Students' and mathematics teachers' perceptions of teacher enthusiasm and instruction. *Learn. Instr. 18*(5): 468-482.
Leech NL, Barrett KC, Morgan GA (2005). SPSS for intermediate statistics: Use and interpretation. Psychology Press.
Luecke R (2010). (Cev: Asli Ozer). Performans Yonetimi. İstanbul: Turkiye İs Bankasi Kultur Yayinlari. 2. Baski.
Marsh HW (2007a). Students' evaluations of university teaching: a multidimensional perspective. In R. P. Perry, & J. C. Smart (Eds.), The scholarship of teaching and learning in higher education (pp. 319e384). New York: Springer.
Marsh HW (2007b). "Do University Teachers Become More Effective with Experience? A Multilevel Growth Model of Students' Evaluations of Teaching over 13 Years." J. Educ. Psychol. 99: 775–790.
Oddens DAM (2004). Kisilik Nitelikleri Acisindan Hollanda'da Mesleki Egitim İcin Ogretmen Egitimi Egilimleri. Mesleki ve Teknik Egitimde Ogretmen Egitimi Uluslararasi Konferansi, Ankara, 37-44.
Ozdamar K (2004). Paket Programlar İle İstatistiksel Veri Analizi I, Eskisehir, Kaan Kitapevi
Palmer J (1993). (Cev: Dogan Sahiner). Performans Degerlendirmeleri. Kisisel Gelisim ve Yonetim Dizisi 9. İstanbul: Rota Yayin Yapim Tanitim. 1.Baski.
Penny AR (2003) Changing the agenda for research into students' views about university teaching: four shortcomings of SRT research, Teaching Higher Educ. 8(3): 399–411.
Perlman B, McCann LI (1998). Students' pet peeves about teaching. Teaching Psychol. 25(3): 201-203.
Peterson KD (1987). Teacher evaluation with multiple and variable lines of evidence. Am. Educ. Res. J. 24, 311 – 317.
Rebore RW (2001). Human resources administration in education: A management approach. Allyn & Bacon, A Pearson Education Company, 75 Arlington Street, Boston, MA 02116.
Richards JC, Lockhart C (1996). Reflective teaching in second language classrooms. New York: Cambridge University Press.
Sakalli N (2001). Sosyal Etkiler. Ankara: İmge Kitabevi.
Seldin P (1999). Changing practices in evaluating teaching: A practical guide to improved faculty performance and promotion/tenure decisions (Vol. 10). Jossey-Bass.
Shadreck M, Issac M (2012) Science teacher quality and effectiveness: Gweru urban junior secondary school students' points of view. Asian Soc. Sci. 8(8):160–165.
Sonmez V (2003). Egitimin Tarihsel Temelleri. V, Sonmez (Ed.), Ogretmenlik Meslegine Giris. Ankara: Ani Yayincilik.
Stronge J (Ed.) (1997). Evaluating Teaching. A guide to current teaching and best practice (Thousand Oaks, CA, Corwin Press).
Sumer N (2000). Yapisal esitlik modelleri: Temel kavramlar ve örnek uygulamalar. Turk Psikoloji Yazilari, 3(6): 49-74.
Tavsancil E (2002). Tutumlarin Olculmesi ve SPSS ile Veri Analizi. Ankara: Nobel Yayin Dagitim.
Tezbasaran A (1997). Likert Tipi Tutum Gelistirme Kilavuzu. (İkici Baski). Ankara: Turk Psikologlar Dernegi Yayinlari.
TschannenMM, Hoy AW, Hoy WK (1998). Teacher efficacy: Its meaning and measure. Rev. Educ. Res. 68(2): 202-248.
Tuan HL, Chang HP, Wang KH, Treagust DF (2000). The development of an instrument for assessing students' perceptions of teachers' knowledge. Int. J. Sci. Educ. 22(4):385–398.
Unlu H, Sunbul AM, Aydos L (2008). Beden Egitimi Ogretmenleri Yeterlilik Olcegi Gecerlilik ve Guvenirlik Calismasi. Ahi Evran Universitesi Kirsehir Egitim Fakultesi Dergisi (KEFAD), 9(2): 23-33.

Appendix 1. The scale of evaluating physical education teachers based on students ratings (SEPETBSR).

ÖĞRENCİ GÖRÜŞLERİNE DAYALI BEDEN EĞİTİMİ ÖĞRETMENİ DEĞERLENDİRME ÖLÇEĞİ **A- Genel Performans Boyutu**	**Tamamen Katılıyorum**	**Katılıyorum**	**Orta Düzeyde Katılıyorum**	**Katılmıyorum**	**Hiç Katılmıyorum**
1.Öğretmenimiz derse zamanında başlar.					
2. Öğretmenimiz ders saatinde yanımızda olur.					
3. Öğretmenimiz belli bir spora yatkınlığı olan öğrencilere kendilerini geliştirebilmeleri için fırsatlar yaratır.					
4. Öğretmenimiz bütün öğrencilerin derse aktif olarak katılımını sağlar.					
5. Öğretmenimiz eğitsel oyunlarla dersi daha eğlenceli hale getirir.					
6. Öğretmenimiz her öğrenciye yeterli tekrar olanağı sağlamaya çalışır.					
7. Öğretmenimiz derslerde mevcut fiziksel olanakları etkili bir şekilde kullanır.					
8. Öğretmenimiz öğrenme güçlüğü çeken arkadaşlarımızla özel olarak ilgilenir.					
9. Öğretmenimiz bizim dikkatimizi çeker ve bizi öğrenmeye güdüler.					
10. Öğretmenimiz okul içi spor faaliyetleri düzenler.					
11. Öğretmenimiz okul içi spor faaliyetlerini düzenlerken organizasyonun her aşamasında öğrencilere görev verir.					
12. Öğretmenimiz egzersiz ve sporda örnek yaşantı tarzına sahiptir.					
13. Öğretmenimiz etrafına pozitif enerji verir.					
14. Öğretmenimiz vücut dilini etkili kullanır.					
15. Öğretmenimiz her zaman iletişime açıktır.					
16. Öğretmenimiz etkileyici bir iletişime sahiptir.					
17. Öğretmenimiz jest ve mimikleri ile iletişimi kuvvetlendirir.					
18. Öğretmenimiz ile okul dışı sorunlarımızı rahatlıkla paylaşabiliriz.					
19.Öğretmenimiz diğer öğretmenlerle işbirliği içinde çalışır.					
20. Öğretmenimiz kendimizi ifade edebilmemize yönelik demokratik bir ortam oluşturur.					
21. Öğretmenimiz ders sonunda dersin değerlendirmesini yapar.					
22. Öğretmenimiz derste performans ve proje değerlendirmelerini kullanır.					
23. Öğretmenimiz değerlendirme sürecine bizim de katılmamıza olanak verir.					
24. Öğretmenimiz gözlem, anket, görüşme gibi değerlendirme tekniklerinden yararlanır.					
25. Öğretmenimiz değerlendirme yaparken çaba ve davranışlarımızı göz önüne alır.					
26. Öğretmenimiz öğrencilere eşit mesafededir.					
27. Öğretmenimiz okulun iyileştirilmesinde ve geliştirilmesinde çevre olanaklarını etkin olarak kullanabilir.					

ÖĞRENCİ GÖRÜŞLERİNE DAYALI BEDEN EĞİTİMİ ÖĞRETMENİ DEĞERLENDİRME ÖLÇEĞİ	Tamamen Katılıyorum	Katılıyorum	Orta Düzeyde Katılıyorum	Katılmıyorum	Hiç Katılmıyorum
28. Öğretmenimiz öğrencinin gelişimi ile ilgili olarak ailelerle sürekli bilgi alışverişinde bulunmak üzere yazılı/sözlü iletişim kurar.					
29. Öğretmenimiz ders dışı faaliyetlere (okullar arası müsabakalar) katılmaya isteklidir.					
30.Öğretmenimiz öğretim sürecinde teknolojiyi kullanır. (Bilgisayar, projeksiyon, internet vb.)					
B- Ders Performansı Boyutu					
31. Öğretmenimiz derse başlamadan önce yoklama yapar.					
32. Öğretmenimiz fiziksel aktivite başlamadan önce ısınma yaptırır.					
33.Öğretmenimiz derse başlamadan önce gerekli güvenlik tedbirlerini alır.					
34. Öğretmenimiz neyi öğrendiğimiz konusunda bizi bilgilendirir.					
35. Öğretmenimiz niçin öğrendiğimiz konusunda bizi bilgilendirir.					
36. Öğretmenimiz derse başlarken gerekli malzemeyi hazır bulundurur.					
37. Öğretmenimiz bir beceriyi öğretirken model kullanır.					
38. Öğretmenimiz bir beceriyi öğretirken kendisi gösterir.					
39. Öğretmenimiz öğrenme aşamasında bize yardım eder.					
40. Öğretmenimiz dersi işlerken hatalarımızı söyler.					
41. Öğretmenimiz dersi işlerken hatalarımızı düzeltir.					
42. Öğretmenimiz dersi işlerken konuyla ilgili teorik bilgileri aktarır.					
43. Öğretmenimiz malzemelerin düzenli kullanılmasına önem verir.					
44.Öğretmenimiz malzemelerin doğru kullanılması konusunda bizi bilgilendirir.					
C- Olumsuz Davranış Boyutu					
45. Öğretmenimiz derslerde sürekli aynı aktiviteleri yaptırır.					
46.Öğretmenimiz zaman zaman fiziksel şiddete başvurur.					
47.Öğretmenimiz zaman zaman sözlü şiddete başvurur.					
48. Öğretmenimiz notu silah olarak kullanır.					