

Construct validity and internal reliability of physiotherapy teacher evaluation scale

Thangamani Ramalingam A¹, Dibyendunarayan D. Bid², S.N. Senthilkumar³

¹Lecturer; BPT, MSc (Psy), PGDRM

²Senior Lecturer; MPT (Ortho), PGDSPT; (Ph.D.)

³Orthopedic Senior Physiotherapist; MPT

^{1,2}Sarvajanik College of Physiotherapy, Rampura, Surat-395003, Gujarat, INDIA

³S.J Nursing Home, Nagercoil, Tamilnadu, INDIA

***Corresponding author**
Thangamani Ramalingam A

Email:
atramalingam@gmail.com

Article History
Received: 08.10.2017
Accepted: 14.10.2017
Published: 10.11.2017



Abstract: Evaluation of teaching is not a science and there is already a considerable body of knowledge discussing these relations about teaching evaluation. There is no simple system for evaluating the quality of faculty teaching in physiotherapy education field. The main objectives of this study was to develop a scale which consists of items seeking an opinion from students on both the clinical and academic areas of the teaching of a physiotherapy teacher and to establish the content construct validity and internal reliability of the components of the scale. Physiotherapy teacher evaluation scale (PhyTES), a self-reported Likert's type scale which consists of 22 items to be attempted by the students to evaluate a physiotherapy teacher. The participants were from Sarvajanik College of physiotherapy, a premier institute affiliated to Veer Narmad South Gujarat University. A principal component analysis was run to establish construct validity of the items in the scale and an item analysis was done to check the reliability of the scale components and its Cronbach's alpha. It was gratifying to note that, the reduction of the items from 22 in the original instrument to 15 in the final version resulted in a significant improvement in reliability, validity or consistency of the scale And the component Factors or dimensions which were extracted from the analysis were named as Academic interest, personal attributes, concern for teaching, time management, and student concern.

Keywords: construct validity, internal consistency, physiotherapy teacher evaluation

INTRODUCTION

The research evidence shows that student motivation and attitudes towards learning are closely linked to student-teacher relationships. Most teachers are good at creating beneficial relationships with their students. Their ability to create effective learning environments that favors student achievement depends on the relationship they build up with their students [1, 2]. Teachers spend more time interacting and working directly with students are perceived as supportive and effective teachers. Effective teachers are keen to encourage student participation in decision-making, allow humor in their classroom and playful too [3]. Useful teacher-student interactions are essential in linking academic success with personal achievement. A teacher must guide the student in aligning his personal goals with his academic goals. Guided students show stronger self-confidence and greater personal and academic success than those without this teacher interactions [4]. Students are very likely to develop stronger relations with teachers who are friendly and supportive and will show more interest in courses taught by these teachers [5].

Student learning could be maximized if students are inspired and motivated by the teacher, and then taught [6]. However, clinical teaching is believed to be different from the traditional classroom teaching because it requires key attributes including one-to-one evaluation and small-group management skills. Involving students in a humanistic and rigorous approach to practice, and making them a professional were associated with a positive perception of effective teachers [7, 8].

Evaluation of teaching is not a science; there is still much to learn. But there is already a considerable body of knowledge about teaching evaluation [9, 10]. There is no simple system for evaluating the quality of faculty teaching in physiotherapy field of education. However, by thinking carefully about the purposes of evaluation and the methods of evaluation that suit those purposes, one can frame a reliable evaluation system [11, 12]. The academic community is waiting to add that knowledge as we will not be able to recognize and reward teaching without a better system for evaluating a teacher [13, 14].

In the above context, considering the need for evaluating the teachers of physiotherapy the aim of the study was to create a tool for students for evaluation of a physiotherapy teacher in the terms of physiotherapy training and learning. The main objectives of the study were to:

- Develop a scale which consists of items seeking an opinion from students on both the clinical and academic area of teaching of a physiotherapy teacher.
- To establish the content construct validity and internal reliability of the components of the scale.

METHODOLOGY

Development of the instrument

Physiotherapy teacher evaluation scale (PhyTES), a self-reported Likert's type scale which consists of 22 items to be attempted by the students to evaluate a physiotherapy teacher. Initially, there were more than 30 items created based on the existing theories and reviewing available teacher evaluation instruments and there was a consideration of stakeholders like students, teachers and other authorities involved in education to finalize the 22 items which were felt important for the scale construction. An example item is "The teacher has an optimistic & enthusiastic view towards academics." Responses were on a Likert-type scale, ranging from 1 = "Strongly Disagree", 2 = "Disagree", 3 = "Uncertain", 4 = "Agree", 5 = "Strongly agree".

The 22 item scale used for data collection is shown below:

- The teacher has adequate cognitive abilities
- He / She is very polite with the students.
- He / She never hesitate to compliment the students appropriately.
- The teacher has an optimistic and enthusiastic view towards academics.
- The teacher always explains the theory and practical aspect of the subject with examples which is taken by him/her.
- The teacher encourages the students to ask questions, regarding the subject.
- I feel that the teacher is having good leadership qualities.
- I like the personality of my teacher.
- The teacher is punctual and well-disciplined inside the classroom.
- The teacher listens patiently and understands clearly to what the students say.
- The teacher plans the academic schedule in the beginning of the year.
- The teacher is having good communication skills and decision-making abilities during the clinic session.
- The teacher refers many books and gives more details regarding the subject which is really useful for exams.
- The interaction between the teacher and students is good.
- The teacher never lies and accepts his / her mistakes and tries to correct it.
- The teacher has got good fluency and word power in the medium of language.
- The teacher has more interest in upgrading the profession through participating in case presentation, seminars, and other academic maneuvers.
- The teacher's lecture is not at all boring in nature.
- My teacher is careful in not hurting the self-respect and feeling of the students.
- I believe that my teacher uses innovative methods for teaching, not replacing the knowledge by technology.
- Mostly the teacher covers the syllabus without fail.
- I am proud to get this person as my teacher.

Participants

The participants were from Sarvajani College of Physiotherapy, a premier institute affiliated to Veer Narmad South Gujarat University. The students of under graduation to post-graduation were included in the study. Students were asked to volunteer to participate in the questionnaire after being explained its purpose by the research instructors arranged to collect data on behalf of the researcher as the researcher himself the person who was going to be evaluated by the students who were taught by the researcher as a teacher for the study. In total data was collected from 124

participants who were present and given consent on the day of execution of the study. Students were assured of anonymity and that the information obtained would be used by the researcher for the purpose of an article only.

Sample size

A convenience sampling method was used in this observational study. As the aim of the study was to prove the construct validity of the tool, the sample size should be decided based on the requirement for factor analysis. Factor analysis is a technique that requires a large sample size. Factor analysis is based on the correlation matrix of the variables involved, and correlations usually need a large sample size before they stabilize [15] [16]. An advice regarding sample size is: 50 cases is very poor, 100 is poor, 200 is fair, 300 is good, 500 is very good, and 1000 or more is excellent. If factor loadings were kept above 0.5, a sample size of 120 is satisfactory [17, 18].

n	50	100	200	300	600	1000
Loading	0.722	0.512	0.384	0.298	0.210	0.162

Procedure

Participants completed a modified-PhyTES at the end of their academic year schedule. The instrument was administered by the group instructors, along with a measure of academic self-efficacy, personality, and emotional intelligence, as part of a larger study. A standard protocol for the administering was used and the instructions were given before the data collection. The questionnaire was administered in a lecture venue. Participants were informed of the purpose, benefits, and risks of the study, as well as their right to withdraw at any stage. Questionnaires were handed out and collected by the researcher, but there was no interaction between the students and the instructors after the procedure had been explained.

Statistical analysis

The collected data were cleaned and processed for analysis. A principal component analysis was run to establish construct validity of the items in the scale. The acceptable level of communalities and factor loadings for items would be 0.5 and Eigenvalue more than one would be considered for component factors. An item analysis was done to check the reliability of the scale components and its Cronbach's alpha. Data were analyzed using IBM SPSS Statistics for Windows, Version 20.0. Armonk, NY: IBM Corp. Statistical significance was set at $p < 0.05$ (Two-tailed).

For the data reduction the following norms were considered: Principal component analysis, Varimax rotation, Communalities > 0.5 , Factor loading > 0.5 (as the study sample size is more than 120), Sample size 124, KMO/MSA > 0.45 , Anti-image correlation matrix > 0.45 , Correlation matrix $> 30\%$ and Eigen value > 1 .

RESULTS

Factor analysis with 22 items

Initially, the factorability of the 22 items of modified-PhyTES was examined. Several well-recognized criteria for the factorability of a correlation were used. Firstly, all the 22 items correlated at least 0.3 with at least one other item and determinant was 0.001. Moreover, inspection of correlation matrix revealed that more than 30% of correlations are significant at the 0.01 level, suggesting reasonable factorability. Secondly, the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was 0.798, above the recommended value of 0.6, and Bartlett's test of sphericity was significant [$\chi^2(231) = 868.04, p < 0.05$]. The diagonals of the anti-image correlation matrix were all over 0.5, supporting the inclusion of each item in the factor analysis (Table-1). Finally, the communalities were all above 4, further confirming that each item shared some common variance with other items. Given these overall indicators, factor analysis was conducted with all 22 items.

Principal components analysis was used because the primary purpose was to identify and compute composite coping scores for the factors underlying the modified-PhyTES scale. The initial analysis considering factors more than one Eigenvalue produced a six-factor solution with 59.08% total variance. The initial Eigenvalues showed that the first factor explained 28.42% of the variance, the second factor 7.22% of the variance, and a third factor 7.01% of the variance. The fourth, fifth and sixth factors had eigenvalues of just over one, each factor explaining 6.211%, 5.503%, and 4.732% without rotation loadings. Three, four and five-factor solutions were examined, using a varimax rotation of the factor loading matrix and they explained 42.64%, 48.84%, and 54.35% of the variance. A six-factor solution was preferred because of the 'leveling off' of Eigenvalues on the scree-plot after six factors, and the insufficient number of primary loadings and difficulty of interpreting the fourth factor and subsequent factors. There was little difference between the varimax and oblimin solutions, thus both solutions were examined in the subsequent analyses before deciding on varimax rotation for the final solution (Table-2 & Figure-1).

Factor analysis with 18-items

During several steps, a total of seven items were eliminated because they did not contribute to a simple factor structure and failed to meet a minimum criterion of having a primary factor loading of 0.5 or above and no cross-loading of 0.3 or above. The items “I am proud to get this person as my teacher.” “The teacher has got good fluency and word power in the medium of language.” “The teacher always gives examples & explains the theory & practical aspect of the subject which is taken by him/her” and “She/he never hesitate to compliment the students appropriately” did not load above 0.5 on any factor. They are item numbers 3,5,16 and 22 and removed from analysis for the next step, reducing the number of items for analysis to 18. The KMO measure of sampling adequacy was 0.783, above the recommended value of 0.6, and Bartlett’s test of sphericity was significant [$\chi^2 (153) = 677.248, p < 0.05$]. The analysis considering factors more than one eigenvalue produced a five-factor solution with 59.67% total variance. The initial Eigenvalues showed that the first factor explained 29.53% of the variance, the second factor 8.73% of the variance, and a third factor 7.83% of the variance. The fourth and fifth factors had Eigenvalues of just over one, each factor explaining 7.142 & 6.46% (Table-3).

Factor analysis with 15-items

The items “The teacher encourages the students to ask questions, regarding the subject”, “The teacher listens patiently to what the students say & understand clearly” and “The interaction between the teacher & students is good” had factor loadings and communalities less than 0.5. They were item numbers 6, 10 and 14 from the 22-items original scale. After removing them a principle-components factor analysis of the remaining 15 items, using varimax and oblimin rotation was conducted, with the five factors explaining 66.01% of the variance. The KMO measure of sampling adequacy was 0.774, above the recommended value of 0.6, and Bartlett’s test of sphericity was significant [$\chi^2 (105) = 559.77, p < 0.05$]. The initial eigenvalues showed that the first factor explained 30.93% of the variance, the second factor 10.329% of the variance, and a third factor 8.801 % of the variance. The fourth and fifth factors had eigenvalues of just over one, each factor explaining 8.47% & 7.48% without rotation. Other rotations also provided the best-defined factor structure like varimax without much difference. And the component Factors which were extracted from the analysis were named as *Academic interest, personal attributes, concern for teaching, time management and student concern* (Table-4 & -5 and Figure-2).

Item analysis

The internal consistency was tested by Cronbach’s coefficient for each of the factor components. The observed coefficients ranged from 0.569 to 0.732, indicating a moderate reliability. By convention, a lenient cut-off of 0.60 is common in exploratory research; alpha should be at least 0.70 or higher to retain an item in an “adequate” scale. Many researchers require a cut-off =0.80 for a good scale [19] [20]. Moreover split-half reliability was also carried out for scale reliability which showed Cronbach’s alpha 0.751 for first eight items and 0.725 for second seven items. And the Guttman split-half coefficient was 0.769. The parallel and strict parallel assumptions the reliability of the 15 item scale was 0.836 & 0.812.

Table-1: Anti-image correlation matrix for 22-items scale

Items	1	2	3	4	5	6	7	8	9	10	11
item1	.855										
item2	-.023	.754									
item3	.039	-.183	.673								
item4	-.168	-.030	-.304	.747							
item5	-.157	-.078	.109	-.074	.809						
item6	-.027	.178	-.064	-.032	-.116	.860					
item7	-.149	-.203	-.035	.234	-.052	-.120	.807				
item8	-.169	-.229	.061	.079	-.142	-.209	-.059	.770			
item9	-.083	-.226	.206	-.164	.016	-.114	-.004	.087	.781		
item10	.100	.084	-.132	.091	-.086	-.151	.019	-.063	-.096	.867	
item11	.208	-.053	.081	-.395	-.195	-.013	-.161	.047	-.084	.060	.770
Items	12	13	14	15	16	17	18	19	20	21	22
item12	.779										
item13	.033	.763									
item14	-.248	-.177	.788								
item15	-.144	-.255	.027	.791							
item16	-.232	.039	.125	.105	.772						
item17	-.236	-.256	.086	-.036	-.124	.827					
item18	.240	-.146	-.176	-.219	-.235	.141	.830				
item19	.020	.083	-.065	-.192	-.018	-.132	-.147	.815			
item20	-.206	-.026	.112	.126	.088	-.014	-.234	-.338	.615		
item21	.021	-.383	.046	.210	-.140	.077	-.050	.167	-.075	.786	
item22	.025	-.185	-.157	.017	-.159	-.138	.035	-.057	-.084	-.023	.914

Table-2: Total variance for 22-items scale explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6.252	28.417	28.417	6.252	28.417	28.417	2.597	11.803	11.803
2	1.587	7.215	35.633	1.587	7.215	35.633	2.378	10.808	22.611
3	1.542	7.009	42.641	1.542	7.009	42.641	2.164	9.837	32.448
4	1.366	6.211	48.852	1.366	6.211	48.852	2.036	9.252	41.701
5	1.211	5.503	54.355	1.211	5.503	54.355	1.987	9.030	50.730
6	1.041	4.732	59.088	1.041	4.732	59.088	1.839	8.357	59.088
7	0.978	4.444	63.531						
8	0.943	4.286	67.818						
9	0.842	3.826	71.643						
10	0.801	3.643	75.286						

Table-3: Total variance explained for 18-items scale

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.315	29.528	29.528	5.315	29.528	29.528	2.579	14.328	14.328
2	1.572	8.734	38.262	1.572	8.734	38.262	2.462	13.677	28.005
3	1.409	7.825	46.087	1.409	7.825	46.087	2.310	12.833	40.838
4	1.286	7.142	53.229	1.286	7.142	53.229	1.741	9.669	50.507
5	1.162	6.458	59.686	1.162	6.458	59.686	1.652	9.179	59.686
6	0.982	5.457	65.144						
7	0.822	4.565	69.709						
8	0.777	4.315	74.024						
9	0.741	4.115	78.140						
10	0.677	3.760	81.900						
11	0.600	3.332	85.232						
12	0.554	3.078	88.310						
13	0.474	2.634	90.944						
14	0.397	2.203	93.147						
15	0.365	2.030	95.177						
16	0.343	1.905	97.082						
17	0.296	1.642	98.723						
18	0.230	1.277	100.000						

Table 4: Factor loadings and communalities based on a principle components analysis with varimax rotation for 15-items version physiotherapy teacher evaluation scale (N = 124)

Items	Factor1 ACADEMIC INTEREST	Factor2 PERSONAL ATTRIBUTES	Factor3 CONCERN FOR TEACHING	Factor4 TIME MANAGEMENT	Factor5 STUDENT CONCERN	Communalities
The teacher has adequate abilities (memory, intelligence and creativity).		0.553				0.532
He / She is very polite with the students.		0.638				0.682
The teacher has an optimistic & enthusiastic view towards academics.	0.633					0.536
I feel that the teacher is having good leadership qualities.		0.644				0.662
I like the personality of my teacher.		0.763				0.613
The teacher is punctual & well-disciplined inside the classroom.				0.674		0.628
The teacher plans the academic schedule in the beginning of the year.	0.743					0.684
The teacher is having good communication skills and decision-making abilities during the clinic session.	0.743					0.652
The teacher refers many books and gives more details regarding the subject which is really useful for exams.			0.673			0.743
The teacher never lies and accepts his / her mistakes and tries to correct it.			0.695			0.722

Table-4 continued on the next page.

Items	Factor1 ACADEMIC INTEREST	Factor2 PERSONAL ATTRIBUTES	Factor3 CONCERN FOR TEACHING	Factor4 TIME MANAGEMENT	Factor5 STUDENT CONCERN	Communalities
The teacher has more interest in upgrading the profession through participating in case	0.680					0.562
The teacher's lecture is not at all boring in nature.			0.687			0.683
My teacher is careful in not hurting the self- respect and feeling of the students.					0.708	0.713
I believe that my teacher uses innovative methods for teaching, not replacing the knowledge by technology.					0.838	0.741
Mostly the teacher covers the syllabus without fail.				0.804		0.748

Table-5: Total variance explained for 15-items final version

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.640	30.933	30.933	4.640	30.933	30.933	2.390	15.932	15.932
2	1.549	10.329	41.261	1.549	10.329	41.261	2.138	14.254	30.186
3	1.320	8.801	50.062	1.320	8.801	50.062	1.931	12.874	43.060
4	1.271	8.471	58.532	1.271	8.471	58.532	1.733	11.553	54.613
5	1.122	7.479	66.011	1.122	7.479	66.011	1.710	11.398	66.011
6	0.834	5.563	71.574						
7	0.691	4.605	76.179						
8	0.675	4.499	80.677						
9	0.632	4.213	84.890						
10	0.517	3.450	88.340						
11	0.440	2.935	91.276						
12	0.381	2.542	93.818						
13	0.350	2.333	96.151						
14	0.298	1.985	98.135						
15	0.280	1.865	100.000						

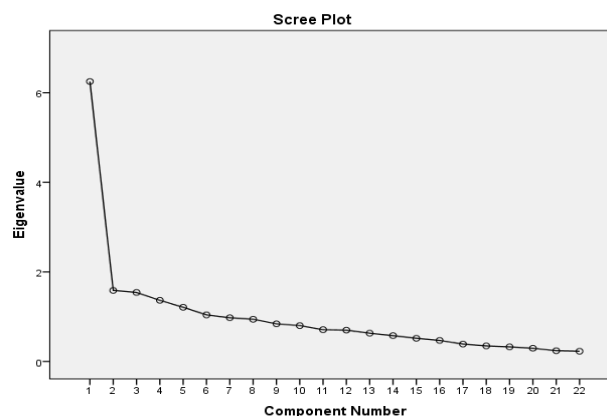


Fig-1: Scree plot for 22-items scale

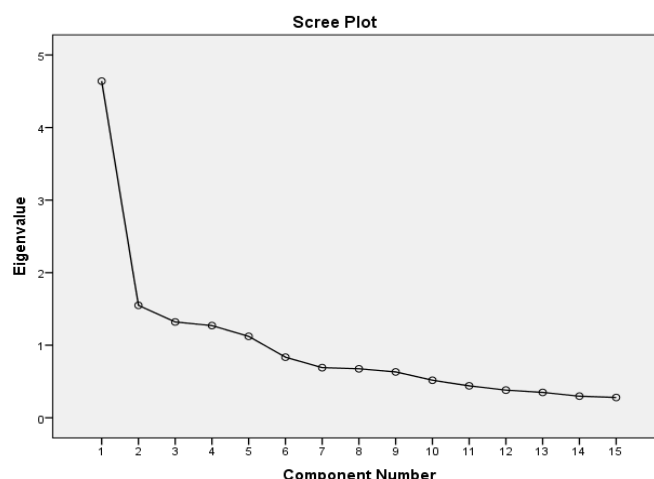


Fig-2: Scree plot for 15-items scale

DISCUSSION

Based on specific educational principles, the modified-PhyTES was developed and all teacher's evaluations are based on the hypothesis that students are the best experts to assess their teachers [19]. The exclusion of seven items was done by the logical and pragmatic approach which preserved the key components in the questionnaire be retained. Furthermore, the remaining 15-items which covered major aspects of teaching evaluation were more simply and clearly phrased for the physiotherapy students. Hence, it was gratifying to note that, the reduction of the items from 22 in the original instrument to 15 in the final version resulted in a significant improvement in reliability, validity or consistency of the scale. Researchers and practitioners agree that teaching is complex and consisting of multiple dimensions; and tools of evaluation of teachers should reflect this multidimensionality. Multidimensionality is important because it provides a more sophisticated and realistic assessment of the various aspects of teaching [21, 22]. McKeachie agreed that student ratings are valid measures of teaching, but that contextual variable may influence the level of ratings and argued that the problem lies in who is using student ratings [23]. Thus more attention should be directed toward methods of ensuring more careful and thoughtful use. Similarly, Greenwald concluded that much disagreement exists about the validity and usefulness of evaluation done by student of teachers. The considerable concern revolves around construct validity, the degree to which student evaluation of teaching measure what they purport to measure [24]. Studies have found generally high correlations between students' and faculty members' evaluations of teaching, but also significant differences between the weights placed on the various dimensions [25]. Student evaluating the teachers can be dangerous and teachers may perceive performance appraisals as popularity contests affecting their career will treat their students as customers rather than products. Haskell found that the majority of scholars believed that student evaluating teachers may act as an infringement of academic freedom [26].

However, a modified-PhyTES is an instrument which can be used to evaluate teachers of physiotherapy in academic settings. The design of the instrument was to assess the full range of teaching skills such as lecture and problem-based learning and instructional skills. The generalizability is limited to the measurement of teaching as it may not cover some of the clinical-teaching activities that occur in non-academic centers. Moreover, the students' opinions have to be authenticated by the other methods of evaluation of a teacher. This will be a subject of separate study. Though students' ratings of the teachers are a highly valued component of teaching evaluation, it is advisable also to gather other types of data for a complete evaluation of teaching effectiveness which may include peer evaluations, self-evaluations, and observations.

Though this tool is useful for physiotherapy teachers' evaluation it may be applicable for the other health professions where there are both clinical and theory or classroom teaching is the routine for training the skills.

Acknowledgement

The authors are grateful to HRD department of Veer Narmad South Gujarat University, Surat for their valuable guidance and training in statistical analysis of this study.

REFERENCES

1. Baker, J. A., Terry, T., Bridger, R., & Winsor, A. (1997). Schools as caring communities: A relational approach to school reform. *School Psychology Review*, 26, 576-588.

2. Fraser, B. J., & Fisher, D. L. (1982). Predicting students' outcomes from their perceptions of classroom psychosocial environment. *American Educational Research Journal*, 1982, 498- 518.
3. Bryant, J., Comisky, P. W., Crane, J. S., & Zillmann, D. (1980). Relationship between college teachers' use of humor in the classroom and students' evaluations of their teachers. *Journal of educational psychology*, 72(4), 511.
4. Hartmut, J. (1978). Supportive dimensions of teacher behavior in relationship to pupil emotional cognitive processes. *Psychology in Erziehung und Unterricht*, 25, 69-74.
5. Osborne, E.; Salzberger, I.; Wittenberg, G. W. 1999. *The Emotional Experience of Learning and Teaching*. Karnac Books, London.
6. Duncan, P. W. (2003). One grip a little stronger. *Physical therapy*, 83(11), 1014-1021.
7. Knox, J. E., & Mogan, J. (1985). Important clinical teacher behaviours as perceived by university nursing faculty, students and graduates. *Journal of advanced nursing*, 10(1), 25-30.
8. Elnicki, D. M., Kolarik, R., & Bardella, I. (2003). Third-Year Medical Students' Perceptions of Effective Teaching Behaviors in a Multidisciplinary Ambulatory Clerkship. *Academic Medicine*, 78(8), 815-819.
9. Braskamp, L. A., Brandenburg, D. C., & Ory, J. C. (1984). *Evaluating teaching effectiveness: A practical guide*. Corwin.
10. Centra, J. A. (1993). *Reflective Faculty Evaluation: Enhancing Teaching and Determining Faculty Effectiveness. The Jossey-Bass Higher and Adult Education Series*. Jossey-Bass Inc., 350 Sansome St., San Francisco, CA 94104.
11. Angelo, T. A., & Cross, K. P. (1993). Minute paper. *Classroom assessment techniques: A handbook for college teachers*, 148-153.
12. Edgerton, R. (1991). The teaching portfolio: Capturing the scholarship in teaching.
13. Marsh, H. W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of educational psychology*, 76(5), 707.
14. Seldin, P. (1997). *The teaching portfolio* (2nd ed.). Bolton, MA: Anker Publishing Company, Inc.
15. Tabachnick, B. G., Fidell L. S. (2001). *Using multivariate statistics*. Needham Heights, MA, Allyn & Bacon.
16. Comery, A. L., Lee H. B. (1992). *A First Course in Factor Analysis*. Hillsdale, NJ, Erlbaum.
17. BMDP, Statistical software, Inc., 1993.
18. Stevens, J. (2003). *Applied Multivariate Statistics for the Social Sciences* (4th Edition). Mahwah, NJ: Lawrence Erlbaum Associates.
19. Elzubeir, M., & Rizk, D. (2002). Evaluating the quality of teaching in medical education: are we using the evidence for both formative and summative purposes?. *Medical Teacher*, 24(3), 313-319.
20. Emery, C. R., Kramer, T. R., & Tian, R. G. (2003). Return to academic standards: a critique of student evaluations of teaching effectiveness. *Quality assurance in Education*, 11(1), 37-46.
21. Feldman, K. A. (2007). Identifying Exemplary Teachers and Teaching: Evidence from Student Ratings¹. In *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 93-143). Springer Netherlands.
22. Marsh, H. W., & Roche, L. A. (1993). The use of students' evaluations and an individually structured intervention to enhance university teaching effectiveness. *American Educational Research Journal*, 30, 217-251.
23. McKeachie, W.J. (1997). Student ratings: The validity of use. *American Psychologist*, 52, 1218-1225.
24. Greenwald, A.G. (1997). Validity concerns and usefulness of student ratings of instruction. *American Psychologist*, 52, 1182-1186.
25. Buskist, W., Sikorski, J., Buckley T., & Saville, B.K. (2002). Elements of master teaching. In S.F. Davis & W. Buskist (Eds.), *The teaching of psychology: Essays in honor of Wilbert J. McKeachie and Charles L. Brewer* (pp. 27-39).
26. Haskell, R.E. (1997). Academic Freedom, Tenure, and Student Evaluation of Faculty. *Education Policy Analysis Archives*. Vol. 5 No. 6