

RESEARCH

Open Access



Designing and validating a potential formative evaluation inventory for teacher competences

Zohre Mohamadi^{1*}  and Negin Malekshahi²

* Correspondence: [Zohre.mohamadi@kiau.ac.ir](mailto:mohamadi@kiau.ac.ir)

¹English Translation Department, Karaj Branch, Islamic Azad University, Karaj, Iran
Full list of author information is available at the end of the article

Abstract

Background: Inadequacy of authority-based defensive teaching and summative and product-based evaluation such as certification and observation measures in providing information about the actual teaching teachers do was an inspiration in this study to design an inventory for formative and process-based evaluation of teacher competences. This study aimed at designing an inventory for formative and process-based evaluation of teacher competences.

Methods: To this end, teacher competences were theoretically defined and the indicators of competence in practice were derived and operationalized through Competency Framework for Teachers proposed by department of education and training in Australia (2004) by a panel of five EFL (English as a foreign language) teaching experts through focused group discussion. The resulting inventory was 65 items on four teacher competences including critical, clinical, personal and technical competences from three perspectives of student, departmental, learning and growth measured on 5 point likert scale.

Results: Testing the inventory with 216 Iranian EFL teachers indicated that there were high Cronbach's alpha reliability indices for the three main perspectives and their dimensions. This implies that the inventory enjoyed appropriate internal consistency. The results of exploratory factors analysis indicated that there was no construct irrelevant factor and all the indicators were loaded in the related teacher competence and perspective dimension. Four separate structural equation models (SEM) were tested in order to probe the trait structure of the inventory. The first three SEM models targeted the three perspectives individually, while the last model explored the structure of the total data. The results indicated that all items had significant contributions to their respective dimensions.

Conclusions: The potential application of this inventory in teacher education programs and the factors that limit its applicability were discussed.

Keywords: Teacher competences, Teacher balanced scorecard, Process-based evaluation, Formative evaluation, Teacher education

Background

Among many factors influencing student learning, teacher quality is the most determining one (Snook et al. 2013). Teaching quality is an important criterion for quality assessment of education utilized by students, parents, and authorities (Feistauer and

Richter 2016) as it is the most determining factor in students' achievement (Sanders et al. 1997). Research indicates that teachers improve their teaching quality by acting on students and authority evaluations (Dresel and Rindermann 2011). There is little attention on formative teacher assessment, and the existing studies on teacher evaluation are either certificate or accomplishment and product-based evaluation based on students' scoring. The problem with this type of evaluation is that they do not provide any information about the teaching practice teachers do (Bastian et al. 2016; Henry et al. 2010). Although new approaches have been introduced to the field of teacher education and evaluation, it has been a long time that teacher evaluation was through the students' assessment of teachers' teaching. This evaluation has been conducted through teacher evaluation questionnaires (Marsh et al. 2009) which are under question for reliability concerns (Feistauer and Richter 2016). Although later classroom observation, student evaluation questionnaire, teacher individual interviews, teacher self-evaluation, and teacher testing (Santiago and Benavides 2009; Smith et al. 2004) were introduced to the field of teacher evaluation, they provide little insights about how to improve teaching practice (Duckor et al. 2014). The inadequacy of the product-based approaches towards teacher evaluation led practitioners to think of inventories that focus more on process-based teacher evaluation and real teaching act (Navidinia et al. 2015). This study is intended to design an inventory for evaluating teacher competences and its potential in tracking changes in the actual teaching act.

Background and purposes

Teacher competences

Teacher development is defined as teachers' construction of teaching competences (Avalos 2011). Competence is defined as a set of professional skills that underlie successful performances (Blašková et al. 2014). Avalos (2011) stated that teacher competence is the teachers' ability in critical analysis of teaching phenomena and education policies which enables them to design the teaching process and procedure in a way to achieve the objectives. Duță et al. (2014) also state that competence is the ability to use skills and knowledge in a coherent and dynamic way to solve problems efficiently. Accordingly, competence is defined by three dimensions: cognitive dimension (knowledge), functional dimension (skills), and attitudes and value dimension (teacher autonomy and responsibility).

Zimpher and Howey (1987) describe four teacher competences: "(1) clinical competence (practical reading and problem solving), (2) personal competence (understanding of self from multiple perspectives with expertise in interactive capacities in interpersonal interactions), (3) critical competence (disposition to engage in social critique and reconstruction of repressive practices), and (4) technical competence (determining in advance what is to be learned and how it is to be learned and criteria by which success is to be measured".(p. 103). There is a great deal of diversity in terms of what performances are indicators of competences. Lasauskienė et al.'s (2015) action research verified teaching practices and performances that relied on teacher competences. Competency Framework for Teachers proposed by the Department of Education and Training in Australia (2004) is another project in finding the indicators of competences in performance. It has been claimed that although teacher evaluation has received

special attention around the world, teachers have been provided with least support for self-evaluation since educators are unaware of the potential evaluation and support tools (Alamoudi and Troudi 2017). The inefficacy of student evaluation of teachers through questionnaires and alternatives such as one-dimensional classroom observation of teaching practice, teacher interviews and self-assessment and teacher portfolio writing were rather product-based evaluations rather than a process-based one (Imhof and Picard 2009) since it was indicated teachers' focus on immediate performance rather than understanding underlying processes (Mansvelder-Longayroux et al. 2007). Wei's (2015) study of formative (classroom observation) and summative (student survey of teaching quality) also indicated that when there is no clear feedback on and definition of what good teaching practice is, summative and formative assessment are meaningless and less effective for teachers, students, and high stakes.

Developing a measure for teacher competences

Many ways are suggested about how to measure and help teachers further develop their teaching competences. Most of the teacher evaluation programs were based on students' achievement scores, and they provide no information about specific teaching practices teachers do, no information for teachers to identify the problems stem from programs, and provide no evidence for teacher performances (Bastian et al. 2016; Henry et al. 2010). Although recently, other measurement instruments such as classroom observations and questionnaires are advocated by the education researchers (Henry et al. 2010), they all suffer from a problem; they come too late to help teachers improve (Bastian et al. 2016). Student rating process does not show goal attainment, increase teacher effectiveness, and student learning (Hughes and Pate 2012). The validity of using non-academic measures as students rating is under question since studies find a positive relation between students' scores and their rating to the teachers.

Teacher performance evaluation at high stakes measures either for decisions on certification or program completion and adaptations (Duckor et al. 2014). Bastian et al. (2016) compared locally and officially scored performance assessment, and the results of their study indicated that local scores were higher than official scores. However, to make high stake decisions, locally scored performance assessment is not appropriate. It is more logical to have both local and official scoring performance assessments; local scoring performance assessment can provide language, context, and evidence-based evaluation, and official scoring evaluation performance can provide information about if it has construct validity, predictive validity, and reliability.

In a study, Moreno-Murcia et al. (2015) designed and validated a measuring instrument to evaluate the performance of university students, and through factor analysis, they have found that there are three important performances that are considered to be important: (1) planning which refers to previous reflection and designing of the teaching including planning of courses, learning activities, and evaluation criteria; (2) development of the course which is anything related to execution of and compliance with education curriculum; (3) results which refer to the achievement of objectives, achievements of the students, revisions and improvement of teaching activities, and creation of teaching materials.

The inadequacy of product-based and certification approaches led the practitioners to use a more process-oriented evaluation (Imhof and Picard 2009). Among process-

oriented evaluation techniques, portfolio assessment has received good attention. In their study of what makes a portfolio and its effect determining in teacher education classes, they assert that portfolio assessment should be an integral part of the education environment and be valued by supervisors and teachers, and they should be given feedback; otherwise, they will consider that portfolios are tedious, time-consuming, and ineffective.

Admiraal et al. (2011) used video portfolios to assess teacher performances and analyzed the reliability and construct and consequential validity of this instrument. They highlight the qualitative and contextual information they provide for the researchers. In their study of reliability and validity concerns attributed to video portfolio instrument for assessing teacher competences, it was established that although there were problems considering reliability and validity of video portfolio as instrument for data collection, teacher assessors rated them positively, and several techniques of think-aloud sessions and reflection session helped the researchers cater for reliability and validity issues. E-portfolio also indicated an increase in teacher reflection and collaboration (Hooker 2017). Pre- and post-interview, reflective journals, and recoding of professional learning community intervention indicated that experienced teachers' self-efficacy was improved in terms of more use of innovative teaching strategies and language proficiency, and novice teachers were improved in terms of classroom management and autonomy (Zonoubi et al. 2017).

A more recent research conducted by Hughes and Pate (2012) suggests a teacher balanced scorecard as an instrument to evaluate teacher education induction programs. They stated that balanced scorecard is mostly used by organizations to manage their customer services by “translating the organization’s strategy and vision to objectives and measures and targets from financial, customer, internal business processes perspectives” (p. 59). They worked on the possibility of changing classic balanced scorecard into teaching balanced scorecard. Table 1 shows what information can teacher balanced scorecard provide from different perspectives including institutional, departmental/administrative, and learning and growth perspectives—the perspectives introduced by Hughes and Pate (2012).

As it is indicated in Table 1, teaching balanced scorecard (TBSC) is a multiple measures of teacher education from various perspectives. It is a talking paper that helps the teachers and faculty to communicate and for the faculty to convey expectation it has from the teachers and addresses those aspects of teaching that are beyond the students' capacity to rate. The classic balanced scorecards are developed mostly for measuring the adequacy of the functioning of organizations from managerial perspectives for the

Table 1 The classic balanced scorecard (BSC) versus the teaching balanced scorecard (TBSC)

Classic balanced scorecard perspectives	Teaching balanced scorecard perspectives	Addresses the question:
Financial perspective	Institutional perspective	How do we look to providers of financial resources?
Customer perspective	Student perspective	How do students see us?
Internal business process perspective	Departmental/administrative perspective	At what must we excel?
Learning and growth perspective	Learning and growth perspective	Can we continue to improve and create value?

purpose of maximizing product sells and higher income. Therefore, there is a need for a modified balanced scorecard to be used in teacher education agenda tapping how teacher competences and, in turn, teacher performances can be improved through inductions. This study aimed at preparing a teacher balanced scorecard and assessing its reliability and construct validity as a potential instrument for teacher evaluation.

Methods

Participants

Teachers

A randomly selected sample of 216 Iranian EFL male ($n = 98$) and female ($n = 118$) teachers' teaching acts was evaluated by three supervisors. Teachers had more or less the same years of teaching experience ($m = 5$) and they ranged 26–32 in age ($m = 29$). All teachers were MA graduates in EFL. They were duty-paid job English teachers in language College of the Researcher's institution. They were required by the institution to follow the same educational objectives through the same educational materials. This research was a self-funded project. To observe the ethics in research, teachers were informed about the research and were assured that their responses were confidential and would only be used for research purposes, and they signed a consent form for the perusal of their responses in this project. The research deputy of the researcher's institution (Dr. Reza Ezati—the deputy of research and technology) can approve actions on ethical consideration in this research project.

Supervisors

An invitation letter was sent to the three supervisors from the three institutes. They had the same years of supervisory experience ($m = 7$) in teaching English as a foreign language (TEFL) centers. They were also Ph.D. holders in TEFL. Since the supervisors were the students of leading researcher and there was the risk of their compulsory participation in research because of power relation and respects they had towards her, they were assured that their decline to participate would not affect their relationship. Therefore, the supervisors' voluntary participation would assure their motivation and serious endeavor and effort they put into action. They evaluated teachers on TBSC through portfolio writing. The inter-rater reliability Cronbach alpha level of 0.78 indicted reliability of decisions made on TBSC assessment and portfolio writing.

Panel of experts

Five Iranian male ($n = 1$) and female ($n = 4$) assistant professors in TEFL from researchers' institution made the panel of experts. They contributed to the study at two phases: (a) designing the themes and indicators of teacher competences and (b) arranging the competence indicators in teacher balanced scorecard (TBSC).

Instrument

To investigate whether the teacher inventory was effective in detecting teacher competences in tracking competence developments, the researcher asked the teachers to write teacher portfolios in three occasions of the beginning, middle, and end of the semester. Teacher writing portfolio consisted of reflective evaluation of their growth, references to the evidences of growth by providing the best exemplar from the archive of teaching they have, their future vision of the problems they have in teaching and how they are

going to solve them, and their evaluation of feedback they received from the mentors and how they respond to the comments. Portfolio writing had checkpoints for teachers, reflection prompts by which teachers’ reflection is directed to have an appropriate account of their progress, and an area for reviewing portfolios and checking grades of portfolio assessment. Several suggestions on how to interpret the themes and how to provide requested information were provided for each theme.

Procedure in data collection and analysis

To compile items of the TBSC inventory, the panel of experts reviewed the literature on teacher competences. Four teacher competences were identified. The definition of each competence was carefully studied to identify unique characteristics of each competence. Four teacher competences including clinical, technical, personal, and critical from three perspectives of student, departmental/administrative, learning and growth were identified. The indicators of competence in practice were derived from the literature and operationalized through Competency Framework for Teachers proposed by the Department of Education and Training in Australia (2004). In operationalizing the indicators of teacher competences, a focused group discussion was conducted by the panel of experts to assess the appropriateness of each indicator of four teacher competences not only with respect to its transparency and relevance but also in terms of the appropriateness of locating them in the right perspective measures. The deigned TBSC had 65 items rated on five Likert scales of unacceptable, slightly unacceptable, neutral, slightly acceptable, and acceptable points (Additional file 1). Table 2 displays the structure of the TBSC questionnaire.

Three supervisors examined the TBSC inventory with 216 teachers. Their evaluation of the teachers’ portfolio writing on three occasions (beginning, middle, and the end of the semester).The portfolio of teachers was assessed using Bakker et al.’s (2011) schemata which required supervisors to look for negative and positive evidences of teacher

Table 2 Structure of teachers balanced scorecard

Perspectives	Student		Departmental		Learning	
	Items	Example	Items	Examples	Items	Examples
Technical	7	Allowing the students to organize and distribute part of the assignments to be performed in the course	16	Providing the contents following a clear and logical framework, highlighting the important aspects	2	Using of technology when conducting lectures
Clinical	10	Catering for individual student learning styles and needs	4	Providing the contents following a clear and logical framework, highlighting the important aspects	2	Examining what one is doing in the classroom and making needed changes
Personal	10	Facilitating student-student and student-professor interaction	3	Working cooperatively with colleagues	2	Engaging in informal dialog with your colleagues on how to improve your teaching
Critical	1	Explaining own developing approach to teaching and learning	4	Developing and applying and understanding to the curriculum policy and program teamwork	4	Initiating action to promote ongoing professional growth

competence, look for (counter) evidences of what contributes to professional thinking and acting, differentiate less and more important evidences and assign score, specify if entire performance can be attributed to specific level of competence, and write a brief summary in which comments on scores were given and important arguments and evidences are cited and consult follow assessor and discuss if the assigned scores could be compared and discuss the assigned scores and the rational pertained to the scores by providing evidences and arguments and determine whether to hold on to the original score or make adjustments.

The measurement of teacher competences on five Likert scales of 65 item TBSC led to the scores ranging from a minimum score of 4 to a maximum score of 34, and the results were put into SPSS to investigate the reliability and validity of the inventor. In the inter-rater reliability of raters, the rating was reported in the previous section.

Results

The purpose of the present study is to design and validate a teacher inventory called the teacher balanced scorecard (TBSC) by computing its reliability and validity—both exploratory and confirmatory methods in order to enable researchers to employ it in their future studies. The TBSC questionnaire includes 65 items which measure student, departmental, and learning perspective each of which has four aspects. The data were analyzed in order to probe its reliability and exploratory and confirmatory factor analyses. Before discussing the results, it should be mentioned that the assumptions of univariate and multivariate normality were met. As noted by Bae and Bachman (2010), the absolute values of the skewness and kurtosis values (Table 3) were lower than 1.96, indicating univariate normality of the data.

The multivariate normality assumption was also retained. The Mardia index of .009 was lower than ± 3 (Bae and Bachman 2010).

Cronbach's alpha reliability indices

Table 4 displays the Cronbach's alpha reliability indices for the three main perspectives and their dimensions. The reliability indices for the student, departmental, and learning perspectives were .90, .91, and .76, respectively. The latter had only 10 items. The reliability indices of the dimensions ranged from a low of .65 for personal aspect of learning which had only two items to a high of .93 for the technical aspect of departmental perspective.

Exploratory factor analysis

A factor analysis was run to probe the underlying constructs of the 65 items of the TBSC questionnaire. Figure 1 suggested 3 to 12 factors to be extracted. The 12 extracted factors accounted for 53.25% of the total variance. Since the TBSC questionnaire had 12 subsections, it was decided to extract the 12 factors using principal axis factor method and varimax rotation. The 12 extracted factors accounted for 53.25% of the total variance.

Table 5 displays the factor loadings of the 65 items under the extracted factors. Based on these results, it can be concluded that:

Table 3 Testing univariate and multivariate normality assumptions

Items	Min	Max	Skew	Kurtosis
1	0	5	-0.148	0.349
2	0	5	-0.117	0.044
3	0	5	-0.147	0.237
4	0	5	-0.114	0.050
5	0	5	-0.246	0.312
6	0	5	-0.064	0.173
7	0	5	0.050	-0.228
8	0	5	0.111	-0.158
9	0	5	-0.002	-0.057
10	0	5	0.193	-0.184
11	1	5	0.078	-0.572
12	1	5	-0.181	-0.668
13	0	5	-0.184	-0.509
14	0	5	0.005	-0.254
15	1	5	0.196	-0.576
16	0	5	0.118	-0.214
17	0	5	-0.075	-0.104
18	0	5	0.039	0.159
19	0	5	-0.053	0.162
20	0	5	-0.030	0.188
21	1	5	0.237	-0.281
22	1	5	0.282	0.287
23	0	5	-0.015	0.110
24	1	5	0.314	-0.083
25	0	5	-0.143	-0.185
26	0	5	0.183	-0.071
27	0	5	0.059	-0.163
28	0	5	-0.140	-0.099
29	0	5	-0.230	-0.280
30	0	5	-0.105	0.086
31	0	5	0.007	-0.060
32	0	5	0.038	0.142
33	0	5	-0.039	-0.212
34	0	5	-0.120	0.010
35	0	5	-0.156	-0.325
36	1	5	0.106	-0.065
37	0	5	-0.092	-0.130
38	0	5	-0.126	-0.444
39	1	5	-0.062	-0.599
40	0	5	-0.142	0.016
41	0	5	-0.176	0.055
42	0	5	-0.252	0.526
43	0	5	-0.262	-0.166
44	0	5	-0.143	-0.238

Table 3 Testing univariate and multivariate normality assumptions (*Continued*)

Items	Min	Max	Skew	Kurtosis
45	1	5	-0.054	-0.482
46	0	5	0.093	0.026
47	0	5	-0.270	-0.008
48	0	5	-0.306	-0.116
49	0	5	-0.075	-0.128
50	0	5	-0.059	-0.265
51	1	5	-0.051	-0.721
52	1	5	-0.144	-0.547
53	0	5	-0.227	-0.150
54	0	5	-0.145	-0.478
55	0	5	0.103	-0.046
56	1	5	-0.115	-0.386
57	0	5	-0.052	-0.015
58	1	5	0.150	-0.254
59	0	5	0.037	-0.325
60	0	5	0.099	0.043
61	0	5	-0.019	-0.070
62	0	5	0.049	-0.102
63	1	5	0.163	-0.419
64	1	5	-0.067	-0.783
65	0	5	0.014	-0.404
Multivariate			0.112	0.009

Table 4 Reliability statistics

		Cronbach's alpha	No of items
Student	Technical	.860	7
	Clinical	.895	10
	Personal	.901	10
	Critical	1	-
	Total	.904	28
Departmental	Technical	.935	16
	Clinical	.787	4
	Personal	.770	3
	Critical	.834	4
	Total	.913	27
Learning	Technical	.730	2
	Clinical	.750	2
	Personal	.654	2
	Critical	.853	4
	Total	.762	10

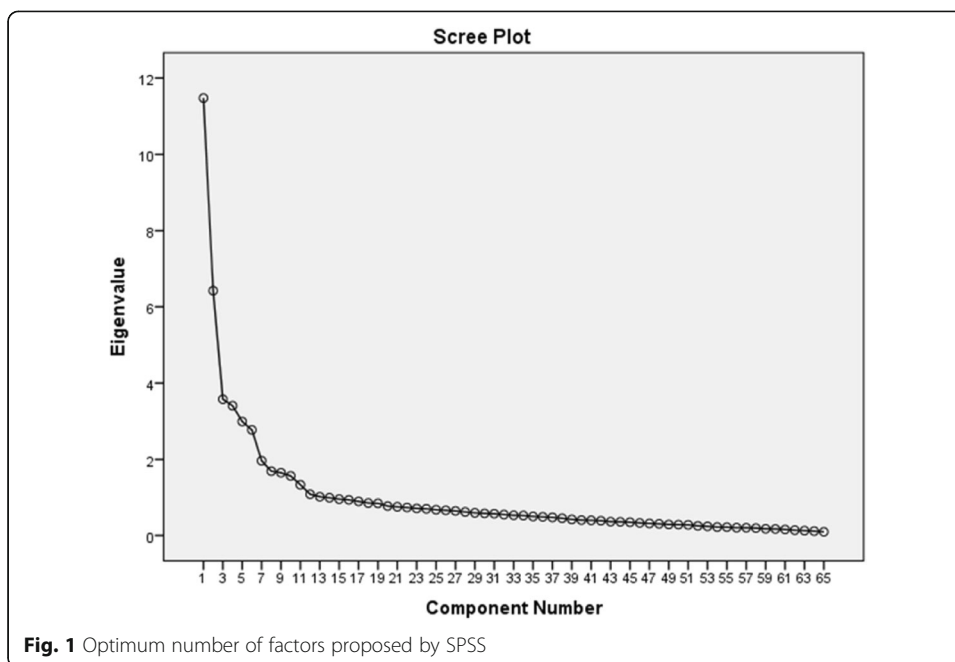


Fig. 1 Optimum number of factors proposed by SPSS

- The first factor includes the 16 items related to the technical aspect of departmental perspective.
- The 10 items related to the personal aspect of student perspective loaded under the second factor.
- The 10 items loading under the third factor were related to the clinical aspect of student perspective.
- The fourth factor includes the 7 items related to the technical aspect of student perspective. Item 28, which was the single indicator of the critical aspect of student perspective, also loaded under the fourth factor.
- The 4 items related to the critical aspect of learning perspective loaded under the fifth factor.
- The 4 items loading under the sixth factor were related to the critical aspect of departmental perspective.
- The seventh factor includes the 4 items related to the clinical aspect of departmental perspective.
- The 3 items related to the personal aspect of departmental perspective loaded under the eighth factor.
- The 2 items loading under the ninth factor were related to the technical aspect of learning perspective.
- The tenth factor includes the 2 items related to the clinical aspect of learning perspective, and finally,
- The 2 items related to the personal aspect of learning perspective were loaded under the second factor. The 12th factor did not include any meaningful ($\geq .30$) loadings.

Based on these results, it can be concluded that the construct validity of the TBSC questionnaire was confirmed employing an exploratory method.

Table 5 Rotated factor matrix (*Continued*)

	Factor											
	1	2	3	4	5	6	7	8	9	10	11	12
Q3				.581								
Q28				.285								
Q62					.795							
Q63					.744							
Q64					.728							
Q65					.721							
Q55						.739						
Q54						.739						
Q53						.699						
Q52						.695						
Q48							.722					
Q47							.685					
Q46							.549					
Q45							.542					
Q49								.735				
Q51								.660				
Q50								.656				
Q57									.750			
Q56									.710			
Q58										.866		
Q59										.604		
Q61											.763	
Q60												.568

Confirmatory factor analysis

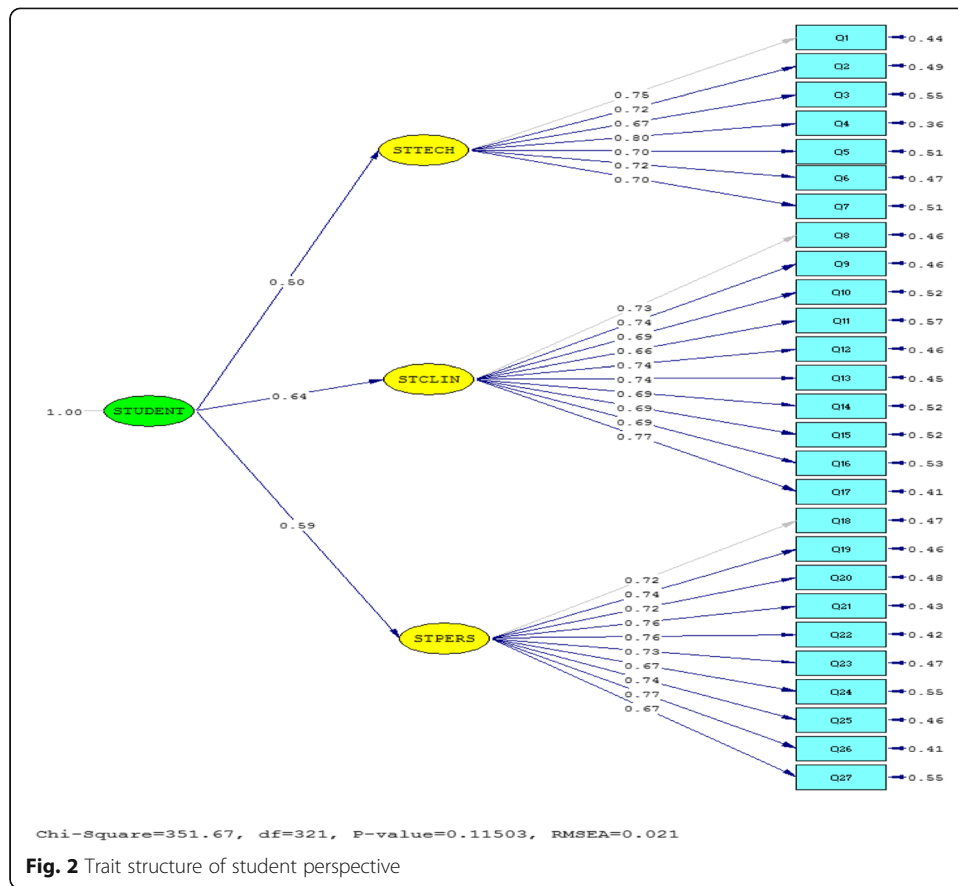
Four separate structural equation models (SEMs) were developed and tested in order to probe the trait structure of the TBSC questionnaire. The three SEM models targeted the three perspectives individually, while the last model explored the structure of the total data.

Confirmatory factor analysis of student perspective

The trait structure of the three components of the student perspective is displayed in Fig. 2. Except for the critical aspect which was dropped from the model, the figure shows the standardized relationships between the items (blue squares) and their related aspects (yellow ovals) which eventually contributed to the “student” perspective (green oval).

All items have significant contributions to their respective dimensions ($\geq .30$), and all three aspects also significantly loaded on the student perspective. The non-significant chi-square statistics ($\chi^2 (321) = 351.67, p = .115$) indicated that the model enjoyed a good fit. The ratios of the chi-square over the degree of freedom, i.e., $351.67/321 = 1.09$, was lower than 3. These results also supported the fit of the model. The RMSEA statistic and its 90% confidence intervals (RMSEA = .021, 90% CI [.000, .034]) were lower than .05.

The model enjoyed a good fit. The PCLOSE statistic of one was higher than .05. All these statistics proved the fit of the model. The indices of NFI, NNFI, CFI, IFI, and RFI were all higher than .90, indicating fit of the model. The critical *N* (CN) value of 235.07



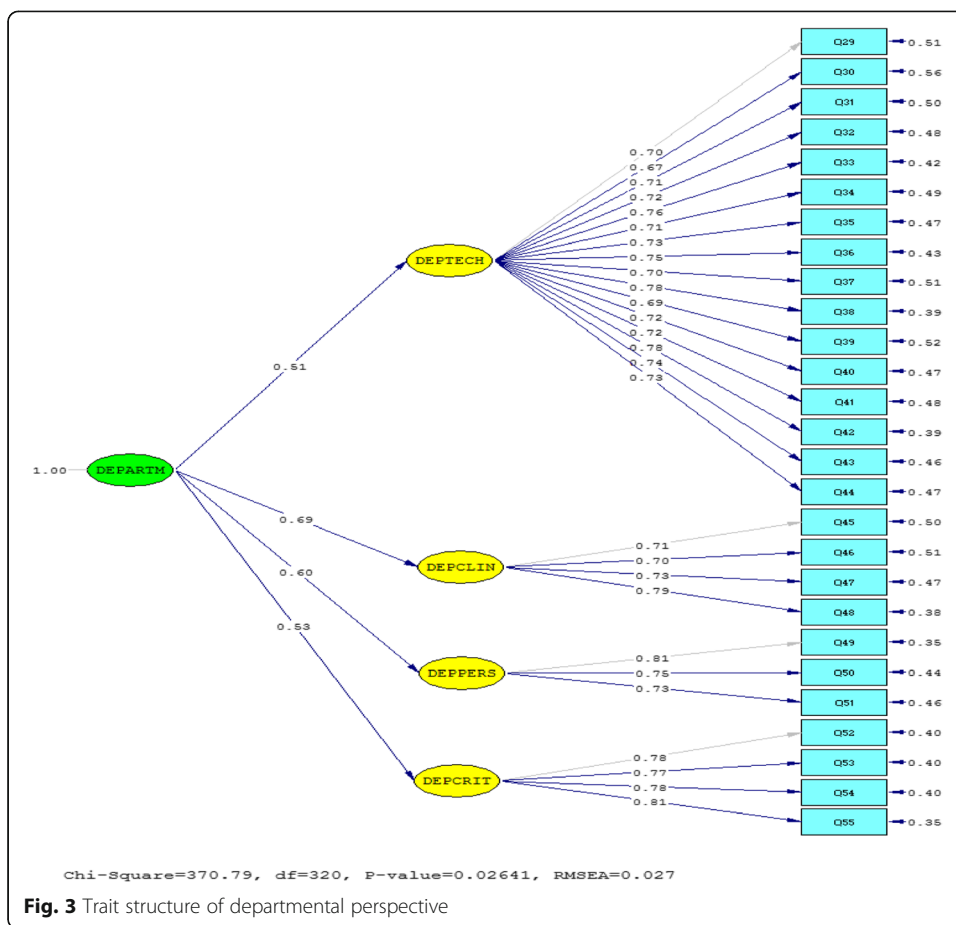
was higher than 200. The CN results proved the sampling adequacy of the present model. Table 6 displays the fit indices related to the student perspective.

Confirmatory factor analysis of departmental perspective

The trait structure of the three components of the departmental perspective is displayed in Fig. 3. Although the chi-square statistic was significant ($\chi^2(320) = 370.79, p = .026$), it indicated that the model did not enjoy a good fit. Since the chi-square statistic is sensitive

Table 6 Fit indices; student perspective

Indices	Model	<i>p</i>	Recommended level
Chi-square	351.67 (321)	.115	Non-significant
Chi-square ratio	1.09	–	≤ 3
NFI	.96	–	≥ .95
NNFI	1	–	≥ .95
RFI	.95	–	≥ .95
CFI	1	–	≥ .95
IFI	1	–	≥ .95
CN	235.05	–	≥ 200
RMSEA	.021	–	≤ .05
95% CI RMSEA	[.000, .034]	–	≤ .05
PCLOSE	1.000	–	> .05



to large sample sizes, its ratio over the degree of freedom should be consulted. The PCLOSE statistic of .24 was higher than .05. All these statistics proved the fit of the model.

All items have significant contributions to their respective dimensions ($\geq .30$), and all four aspects also significantly loaded on the departmental perspective. The ratios of the chi-square over the degree of freedom; i.e., $370.79/320 = 1.15$, was lower than 3. These results also supported the fit of the model. The RMSEA statistic and its 90% confidence intervals (RMSEA = .027, 90% CI [.010, .039]) were lower than .05.

The indices of NFI, NNFI, CFI, IFI, and RFI were all higher than .90, indicating fit of the model. The critical N (CN) value of 222.37 was higher than 200. The CN results proved the sampling adequacy of the present model. Table 7 displays the fit indices related to the departmental perspective.

Confirmatory factor analysis of learning perspective

The trait structure of the three components of the learning perspective is displayed in Fig. 4.

The PCLOSE statistic of .24 was higher than .05. All these statistics proved the fit of the model.

All items have significant contributions to their respective dimensions ($\geq .30$), and all four aspects also significantly loaded on the learning perspective. The model enjoyed a good fit, although the chi-square statistic was significant ($\chi^2(31) = 55.20, p = .004$) indicated that the model did not enjoy a good fit. Since the chi-square statistic is sensitive

Table 7 Fit indices; departmental perspective

Indices	Model	<i>p</i>	Recommended level
Chi-square	370.79 (320)	.026	Non-significant
Chi-square ratio	1.15	–	≤ 3
NFI	.96	–	≥ .95
NNFI	.99	–	≥ .95
RFI	.96	–	≥ .95
CFI	.99	–	≥ .95
IFI	.99	–	≥ .95
CN	222.37	–	≥ 200
RMSEA	.027	–	≤ .05
95% CI RMSEA	[.010, .039]	–	≤ .05
PCLOSE	1.00	–	> .05

to large sample sizes, its ratio over the degree of freedom should be consulted. The ratios of the chi-square over the degree of freedom, i.e., $55.20/31 = 1.78$, was lower than 3. These results also supported the fit of the model. The RMSEA statistic and its 90% confidence intervals (RMSEA = .060, 90% CI [.033, .086]) were between .05 and .08. This range is considered as “reasonable fit” by Byrne (2016).

The indices of NFI, NNFI, CFI, IFI, and RFI were all higher than .90, indicating fit of the model. The critical *N* (CN) value of 204.31 was higher than 200. The CN results proved the sampling adequacy of the present model. Table 8 displays the fit indices related to the learning perspective.

Confirmatory factor analysis of learning perspective

The trait structure of the three components of the TBSC overall model is displayed in Fig. 5.

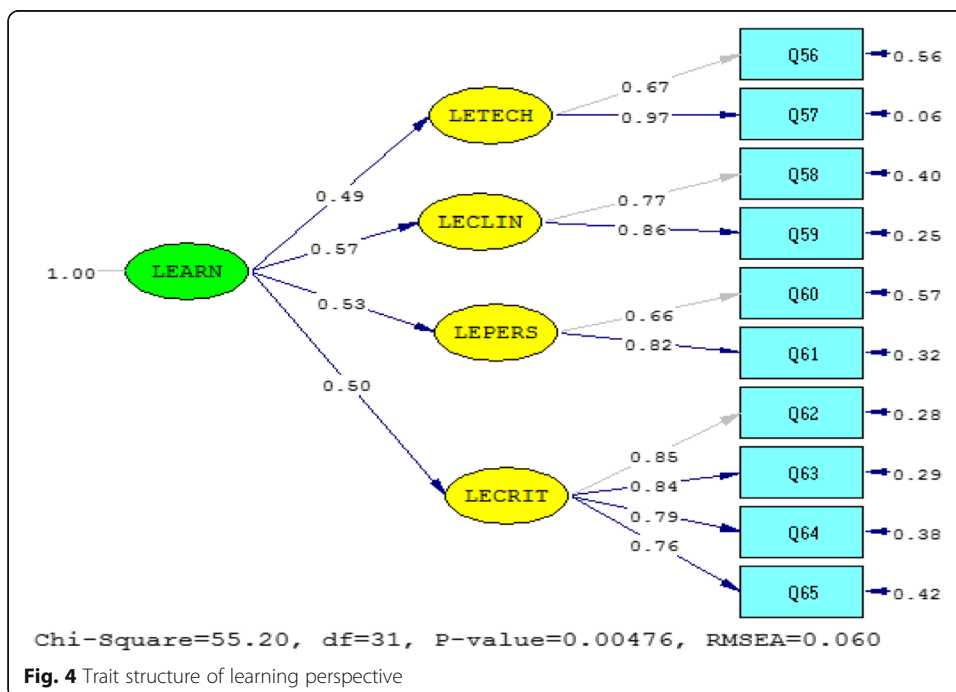


Table 8 Fit indices; learning perspective

Indices	Model	<i>p</i>	Recommended level
Chi-square	55.20 (31)	.004	Non-significant
Chi-square ratio	1.78	–	≤ 3
NFI	.95	–	≥ .95
NNFI	.97	–	≥ .95
RFI	.93	–	≥ .95
CFI	.98	–	≥ .95
IFI	.98	–	≥ .95
CN	204.31	–	≥ 200
RMSEA	.060	–	≤ .05
95% CI RMSEA	[.033, .086]	–	≤ .05
PCLOSE	.24	–	> .05

All aspects had significant contributions to their respective dimensions (≥ .30), and all three perspectives also significantly loaded on the TBSC. The model enjoyed a good fit, although the chi-square statistic was non-significant ($\chi^2(41) = 26.72, p = .958$) indicated that the model enjoyed a good fit. The ratios of the chi-square over the degree of freedom, i.e., $26.72/41 = .65$, was lower than 3. These results also supported the fit of the model. The RMSEA statistic and its 90% confidence intervals (RMSEA = .000, 90% CI [.000, .000]) were all lower than .05 and indicated that the present model enjoyed a good fit.

The PCLOSE statistic of one was higher than .05. All these statistics proved the fit of the model. The indices of NFI, NNFI, CFI, IFI, and RFI were all higher than .90, indicating fit of the model. The critical *N* (CN) value of 523.62 was higher than 200. The CN results proved the sampling adequacy of the present model. Table 9 displays the fit indices related to the overall model.

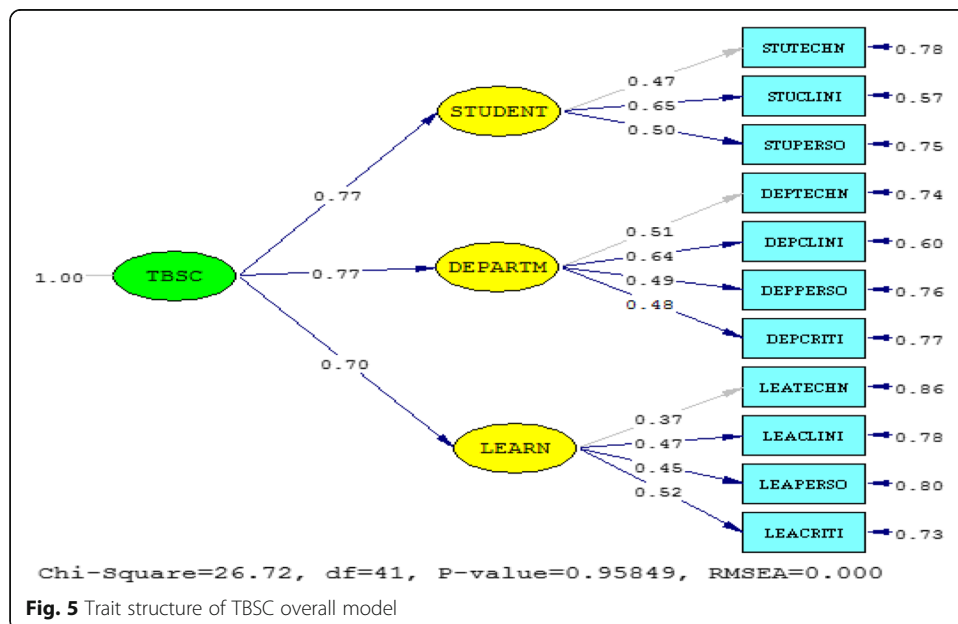


Table 9 Fit indices; TBSC overall model

Indices	Model	p	Recommended level
Chi-square	26.72 (41)	.958	Non-significant
Chi-square ratio	.65	–	≤ 3
NFI	.94	–	$\geq .95$
NNFI	1	–	$\geq .95$
RFI	.91	–	$\geq .95$
CFI	1	–	$\geq .95$
IFI	1	–	$\geq .95$
CN	523.62	–	≥ 200
RMSEA	.000	–	$\leq .05$
95% CI RMSEA	[.000, .000]	–	$\leq .05$
PCLOSE	1	–	$> .05$

Discussion

This study was an attempt to design a teacher evaluation inventory named TBSC which focuses on teacher competences from the three perspectives of student, departmental, and learning and growth ones. Cronbach's alpha reliability indices for the three main perspectives and their dimensions show that the assessment is instrument independent (good internal consistency). The results of the exploratory factor analysis indicated that there was no construct-irrelevant factor, and all the indicators were loaded in the related teacher competence and perspective dimension and they assess what they are supposed to assess. Four separate structural equation models (SEMs) were tested in order to probe the trait structure of the TBSC questionnaire. The first three SEM models targeted the three perspectives individually, while the last model explored the structure of the total data. The results indicated that all items had significant contributions to their respective dimensions. This study can provide insights into how to manage the criticism made to teacher value-added approaches towards teacher education and evaluation.

Werbińska's (2015) review of most approaches towards teacher education highlighted the fact that most appraisal systems are product based which do not provide any information of the teaching reforms taking place in teaching development. He criticizes the teacher induction programs since they are output-based and focuses on teacher certification or student achievements. Later observation-based checklist tick off points which leave the understanding of the context of the observation behind decreases the value of teacher education programs. Mentee observation feedback on critical incidences also leaves no space for teacher themselves to evaluate their own act of teaching. The use of artifacts such as running commentaries and transcribed feedback sessions was introduced as catalysis. The fact is that all the approaches have something in common and that is teachers reform their teaching on the basis of the mentor or supervisors which prevent teachers from forming their own identity. Therefore, this sets a divide between advocators of student-centered progressivist and teacher authority-based defensive teaching. This inventory can direct teachers in their self-evaluation and reflection and help teacher transformation. A transformation which entails a change from a mastery teaching is giving priority to the appropriate act of delivering teaching to learners to

what Richards (2010) calls “learner-focused teaching.” It is a kind of teaching in which the focus is maximizing the potential for learning.

To be more specific with this inventory and its efficacy in teacher evaluation, each competence indicators are reviewed. The first competence is critical competence which requires teachers being engaged with teamwork, maximizing teaching quality via asking and suggesting critical ideas, volunteering in policy and program making tasks and initiating actions, and sharing innovations and developments. These indicators are in line with what Richards (2010) requires all teachers to develop—pedagogical content knowledge. He distinguishes pedagogical content knowledge from disciplinary knowledge. Disciplinary knowledge is the teachers’ knowledge of his discipline. In the case of linguistics, it can be the knowledge of semantics, syntax, and discourse and pragmatics. Whereas pedagogical content knowledge is the knowledge about teaching and learning which helps the teachers to solve problems raised in the actual classroom context, and it is the knowledge acquired through reflective thinking. Reflection includes looking back and forward to teaching experiences and initiating necessary changes and managing the consequences of those changes. As Mezirow (2000) suggests, reflection should be both on content (teaching experiences) and process (how the problems are solved and ongoing development are achieved).

The other competence is personal competence. The indicators of personal competence include personal involvement and establishing a sense of community. Part of teacher development comes from participating with communities having the same goal, interests, and values. The sense of community creates collegiality which provides opportunities for group-oriented activities and joint problem solving, and this helps the learners to play new roles of team leaders, teacher trainer, and mentor and critical friend (Richards and Farrell 2005).

The other competence is clinical competence. The indicators of these competences show that they are related to real-time teaching action. The indicators suggest learner-focused teaching. The development of these competences shows how teachers transit from a survival and mastery stage to a stage where teachers focus more on learners’ learning. At survival stage, teachers act within their comfort zone and focus on their teaching, and at later stages of development, they focus more on the impact of their teaching on student learning (Farrell 2012). The trend of change in critical competence to technical and clinical competence indicates that when mind undergoes changes as a result of reflection through a portfolio, its results can be seen in actions in classes. A transformation which entails a change from a mastery teaching which is giving priority to the appropriate act of delivering teaching to learners to what Richards (2010) calls “learner-focused teaching.” It is a kind of teaching in which the focus is maximizing the potential for learning. Besides learner-focused teaching, teachers gain skills in reasoning, application of pedagogical content knowledge (a knowledge by which they can manage their teaching), anticipate and recognize problems, and take actions for solving them. Besides, as teachers increase their knowledge and experience, they develop improvisational teaching which is moving towards flexibility in teaching. Improvisational teaching is having cognition behind the teaching skills acquired through experiences (Richards 2010).

The other competence is technical competence. A review of indicators of technical competence shows that this competence is related to the metacognition and pre- and

post-planning of teaching act. Teachers need to be aware of the teaching they do which means teachers should develop professionalism. Professionalism is to be technical both at large-scale dimension responding to institutionally prescribed teaching to be accountable in terms of managerial dimensions pertained to ministries of education and teaching organizations and local scale dimension which is called independent professionalism and requires teachers to be consciously aware of ones' teaching practices.

Conclusion

This study aimed at designing and validating an instrument for teacher evaluation. The indicators pertained to four teacher competences including critical, clinical, technical, and personal competences were identified theoretically through literature review and operationally through focused group discussion by the panel of experts. The results of reliability analysis indicated that the assessment on the basis of the newly developed inventory is instrument independent which means that the inventory enjoys internal consistency. The results of the exploratory factor analysis indicated that there was no construct-irrelevant factor, and all the indicators were loaded in the related teacher competence and perspective dimension and they assessed what they were supposed to assess.

The results of this study are of great significance for education research and teacher development. This inventory makes the teachers capable of monitoring their class automatically. The TBSC inventory can also help teachers self-evaluate their teaching ability and performance. TBSC inventory can help teachers to monitor their teaching timely and dynamically. The items of TBSC (can be checked in Table 1 and Additional file 1) are related to teachers' awareness in optimizing teaching quality. Besides, TBSC inventory can help teachers identify their strength and weakness and track their learning and growth.

However, certain caveats apply to the conclusions. First, the sample consists of EFL teachers in one of the districts of Iran, and hence, we cannot claim that the inventory has the same potential in all educational contexts which limits the generalizability of its use. Second, it might be possible that different educational context shows a different compliance with the inventory because education ideology is fostered in that context. Educational systems are primed with certain ideologies that mediate any changes happening in the education ecology. These macro-ideologies are the co-creative and directors of teacher perception and practices (Vasileiadis et al. 2013) that act as filters that legitimize serotypes in teacher perception and teacher practices (Vasileiadis et al. 2013). The macro ideologies have implicit messages for teachers. For example the anti-American attitude Iranian government instigates may imply that EFL teachers' development is not appreciated since learning that the target language may bring the values of the target culture to native one or the government's lack of infrastructure facilities to implement technology-mediated learning and teaching may imply that whether teachers keep up with the latest education technology is necessary and cause teachers not take technology serious in their teaching practice. The invention or implementation of any inventories without paying attention to its origin and context that is formed may jeopardize what Cohen (1995) calls coherence in practice. For any attempts, there should be coherence between teachers and education ecology. Future research can implement the resulting teacher competence inventory in different teaching context to attest the accreditation of this inventory in different

educational contexts since teaching is liable to the charge of different biases and understandings that are culture-bound.

Additional file

Additional file 1: Teacher balanced scorecard (TBSC). (DOCX 45 kb)

Acknowledgements

Gratitude goes to all people especially teachers, research assistants, and panel of experts in designing the inventory and collecting the data.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Availability of data and materials

Data and material are available.

Authors' contributions

To achieve the purpose of the study, NM conceived the study and coordinated the data collection. ZM participated in the design of the study, performed the statistical analysis, and wrote the final draft of the manuscript. Both authors read and approved the final manuscript.

Authors' information

Dr. Zohre Mohamadi is an assistant professor at English Translation Department of Islamic Azad University, Karaj branch, Karaj, Iran, and the head of Young Researchers and Elites club. She has published in the areas of discourse, interaction, and conversation analysis, teaching English as a foreign language and computer-assisted language learning. She has published several articles on writing skill and computer-assisted language learning. Currently, she is working on teacher education and development.

Ms. Negin Malekshahi is a MA graduate, and she has published in task-based instruction and attended many national conferences.

Competing interests

Both authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹English Translation Department, Karaj Branch, Islamic Azad University, Karaj, Iran. ²English Teaching Department, Karaj Branch, Islamic Azad University, Karaj, Iran.

Received: 26 January 2018 Accepted: 13 March 2018

Published online: 17 April 2018

References

- Admiraal, W, Hoeksmas, M, van de Kamp, M-T, van Duin, G. (2011). Assessment of teacher competence using video portfolios: reliability, construct validity, and consequential validity. *Teaching and Teacher Education*, 27(6), 1019–1028.
- Alamoudi, K, & Troudi, S (2017). *EFL teacher evaluation: a theoretical perspective. In evaluation in foreign language education in the Middle East and North Africa* (pp. 29–41). Cham: Springer.
- Avalos, B. (2011). Teacher professional development in teaching and teacher education over ten years. *Teaching and Teacher Education*, 27(1), 10–20.
- Bae, J, & Bachman, LF. (2010). An investigation of four writing traits and two tasks across two languages. *Language Testing*, 27(2), 213–234.
- Bakker, ME, Roelofs, EC, Beijgaard, D, Sanders, PF, Tigelaar, DE, Verloop, N. (2011). Video portfolios: the development and usefulness of a teacher assessment procedure. *Studies in Educational Evaluation*, 37(2), 123–133.
- Bastian, KC, Henry, GT, Pan, Y, Lys, D. (2016). Teacher candidate performance assessments: local scoring and implications for teacher preparation program improvement. *Teaching and Teacher Education*, 59, 1–12.
- Blašková, M, Blaško, R, Kucharčíková, A. (2014). Competences and competence model of university teachers. *Procedia-Social and Behavioral Sciences*, 159, 457–467.
- Byrne, B. M. (2016). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. Routledge.
- Cohen, DK. (1995). What is the system in systemic reform? *Educational Researcher*, 24(9), 11–31.
- Dresel, M, & Rindermann, H. (2011). Counseling university instructors based on student evaluations of their teaching effectiveness: a multilevel test of its effectiveness under consideration of bias and unfairness variables. *Research in Higher Education*, 52(7), 717–737.
- Duckor, B, Castellano, KE, Téllez, K, Wihardini, D, Wilson, M. (2014). Examining the internal structure evidence for the performance assessment for California teachers a validation study of the elementary literacy teaching event for Tier I Teacher Licensure. *Journal of Teacher Education*, 65(5), 402–420.

- Duță, N, Pânișoară, G, Pânișoară, IO. (2014). The profile of the teaching profession—empirical reflections on the development of the competences of university teachers. *Procedia-Social and Behavioral Sciences*, 140, 390–395.
- Farrell, TS. (2012). Novice-service language teacher development: bridging the gap between preservice and in-service education and development. *TESOL Quarterly*, 46(3), 435–449.
- Feistauer, D, & Richter, T. (2017). How reliable are students' evaluations of teaching quality? A variance components approach. *Assessment & Evaluation in Higher Education*, 42(8), 1263–1279.
- Henry, GT, Thompson, CL, Fortner, CK, Zulli, RA, Kershaw, D (2010). *The impact of teacher preparation on student learning in North Carolina public schools*. Chapel Hill: Carolina Institute for Public Policy, University of North Carolina at Chapel Hill.
- Hooker, T. (2017). Transforming teachers' formative assessment practices through ePortfolios. *Teaching and Teacher Education*, 67, 440–453.
- Hughes, K, & Pate, GR. (2012). Moving beyond student ratings: a balanced scorecard approach for evaluating teaching performance. *Issues in Accounting Education*, 28(1), 49–75.
- Imhof, M, & Picard, C. (2009). Views on using portfolio in teacher education. *Teaching and Teacher Education*, 25(1), 149–154.
- Lasauskienė, J, Rauduvaitė, A, Barkauskaitė, M. (2015). Development of general competencies within the context of teacher training. *Procedia-Social and Behavioral Sciences*, 191, 777–782.
- Mansvelder-Longayroux, DD, Beijaard, D, Verloop, N. (2007). The portfolio as a tool for stimulating reflection by student teachers. *Teaching and Teacher Education*, 23(1), 47–62.
- Marsh, HW, Muthén, B, Asparouhov, T, Lüdtke, O, Robitzsch, A, Morin, AJ, Trautwein, U. (2009). Exploratory structural equation modeling, integrating CFA and EFA: application to students' evaluations of university teaching. *Structural Equation Modeling: A Multidisciplinary Journal*, 16(3), 439–476.
- Mezirow, J (2000). *Learning to think like an adult. Core concepts of transformation theory*. In J. Mezirow and Associates (Eds.), *Learning as transformation. Critical perspectives on a theory in progress* (pp. 3–33). San Francisco: Jossey-Bass.
- Moreno-Murcia, J. A., Torregrosa, Y. S., & Pedreño, N. B. (2015). Cuestionario de evaluación de las competencias docentes en el ámbito universitario. Evaluación de las competencias docentes en la universidad.
- Navidinia, H, Reza Kiani, G, Akbari, R, Ghaffar Samar, R. (2015). EFL teacher performance evaluation in Iranian high schools: examining the effectiveness of the status quo and setting the groundwork for developing an alternative model. *The International Journal of Humanities*, 21(4), 27–53.
- Richards, JC. (2010). Competence and performance in language teaching. *RELC Journal*, 41(2), 101–122.
- Richards, JC, & Farrell, TSC (2005). *Professional development for language teachers: strategies for teacher learning*. New York: Cambridge University Press.
- Sanders, WL, Wright, SP, Horn, SP. (1997). Teacher and classroom context effects on student achievement: implications for teacher evaluation. *Journal of Personnel Evaluation in Education*, 11(1), 57–67.
- Santiago, P, & Benavides, F (2009). *Teacher evaluation: a conceptual framework and examples of country practices, Paper for presentation at the OECD Mexico* (pp. 1–2).
- Smith, JS, Szelest, BP, Downey, JP. (2004). Implementing outcomes assessment in an academic affairs support unit. *Research in Higher Education*, 45(4), 405–427.
- Snook, I, O'Neill, J, Birks, K S, Church, J, & Rawlins, P. (2013). The assessment of teacher quality: an investigation into current issues in evaluating and rewarding teachers.
- Vasileiadis, KN, Tsioumis, KA, Kyridis, A. (2013). The effects of dominant ideology on teachers' perceptions and practices towards the "other". *International Journal of Learning and Development*, 3(1), 33–48.
- Wei, W. (2015). Using summative and formative assessments to evaluate EFL teachers' teaching performance. *Assessment & Evaluation in Higher Education*, 40(4), 611–623.
- Werbińska, D (2015). *Teacher evaluation in second language education*, (vol. 52, pp. 149–158) system.
- Zimpher, N, & Howey, KR. (1987). Adapting supervisory practices to different orientations of teaching competence. *Journal of Curriculum and Supervision*, 2(2), 101–127.
- Zonoubi, R, Rasekh, AE, Tavakoli, M. (2017). EFL teacher self-efficacy development in professional learning communities. *System*, 66, 1–12.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
