

Competency-Based Tests as a Tool for Teacher Evaluation in Higher Education Institutions

Kamila Ludwikowska¹

Submitted: 30.03.2021. Accepted: 22.05.2022

Abstract

Purpose: This study aims to verify the quality of the teacher competency inventory as a pedagogical competencies measure of teachers in higher education institutions.

Design methodology: The development and testing of the teacher competency inventory covered three phases: (1) item generation, (2) dimension identification and data reduction (factor analysis), and (3) scale evaluation. The current study tests its construct validity, using an independent samples t-test, variance analysis, and Anova in two different environments.

Findings: Results indicate that the inventory measures unitary constructs and provide its predictive capacity. The developed instrument measures levels of future-oriented pedagogical competencies with comparable reliability and validity.

Practical implications: Teacher competency inventory may be applied to teachers' recruitment, selection, development, and performance evaluation with the goal to identify current and required levels of competencies. The inventory allows one to compare teacher competencies with the expected competency profile. Possible differences may indicate deficits or overruns in relation to the expected profile.

Originality: Teacher competency inventory is a meaningful tool to improve human resource practices – including recruitment, selection, and teacher performance monitoring – and to indicate the current and required level of competencies. The identified competency gap can help plan teacher development to increase job performance.

Keywords: competency test, HR practice, assessment, self-evaluation, teacher evaluation.

JEL: M12

¹ Wrocław University of Science and Technology, Poland; e-mail: kamila.ludwikowska@pwr.edu.pl; <https://orcid.org/0000-0002-2975-7539>.

Introduction

Higher education institutions' (HEIs) policymakers need to continuously analyze the challenges universities face and seek solutions that increase organizational performance. They need to recognize problems in the quality of teaching and learning linked to less effective teachers that become reasons for low-performing educational institutions. Properly implemented and executed human resource practices are vital factors in the success (Cohen, 2015) of HEIs.

One of the most recent challenges is to recruit and retain quality teachers (Adnot et al., 2017). Thus, HR practices need to focus on resourcing and continuous development to stimulate teacher abilities (Keller-Schneider et al., 2020), motivations, and opportunities to grow, while assisting them in reaching higher performance (Pak, 2019). Therefore, we require appropriate measurement and evaluation methods to accurately identify low-performing teachers and replace them with more effective ones (Adnot et al., 2017). Measurements should focus on individual ability in terms of competencies (Ulrich & Dulebohn, 2015). It becomes a key issue to focus attention on making decisions based on sufficient and measurable data (Firestone & Donaldson, 2019).

The role of competency orientation for teaching in educational institutions has recently increased. The high level of teacher competencies informs the quality of teaching regarded as the determinant of any educational institution's quality (Yeşilçınar & Çakır, 2020). The concept of competency is a major focus of each human resource (HR) practice, not just recruitment and selection but also performance evaluation, development, and others. These practices measure critical characteristics of the employee or candidate, which then allows forecasting their workplace performance. Such an approach requires the application of reliable assessment methods to identify competencies meaningful for the institution (Baartman, 2013).

To predict teacher job performance, most higher education institutions (HEIs) still use commercial solutions that lack evidence-based support. They recruit more teachers instead of finding the person who best fits the needs of students and institutions. It may result not only in higher recruitment process costs or increased teacher turnover but also in low-performing teachers (Adnot et al., 2017). Consequently, poor HR practices in teacher assessment and evaluation can result in a decrease in the quality of education. The evidence-based approach promotes accountability and should be required to measure teacher quality (Darling-Hammond, 2014) and performance. Therefore, the critical examination of assessment tools and their use should be the focus of HEIs policymakers (Firestone & Donaldson, 2019) and HR administrators.

Competency tests may measure the level of teacher occupational competencies and capture their development need, thus improving teacher and education quality.

There is a lively discussion in the literature about the assessment of competencies. Several researchers consider the need for validating and assessing measures reliability (Baartman et al., 2013; Muñiz-Rodríguez et al., 2017), but evidence lacks in this regard. Furthermore, the instruments used are often limited to the evaluation of the course or specific competencies with a narrow perspective not adapted to HEI (Bergsmann et al., 2018). Therefore, I developed a test called teacher competency inventory (TCI) that measures teacher competencies in four dimensions: stimulating students for their achievement (SA); using different teaching approaches to improve student learning (TA); preparing students to be globally competent (GC); creating supporting learning environment (LE). Its reliability proved high internal quality (Ludwikowska, 2019). This article describes the second phase of the research aimed at examining the validity of TCI dimensions in two groups of teachers from India and Poland. In this sense, the author contributes to the literature on competency test validation and its application to HR practice in HEI. Validated dimensions of the TCI may be applied to teacher recruitment, selection, and performance evaluation with the goal to identify the current and desired levels of competency. The identified competency gap can be the basis for planning the development of teachers to enhance their performance. Considering the above, the following research questions guided this study:

- Q1.** Do the four TCI dimensions establish a valid assessment tool to measure a teacher's pedagogical competencies?
- Q2.** Does the TCI fits different groups (i.e. gender, education) thus explaining its internal quality?

There is a shift in evaluating the quality of assessments from external toward internal evaluation, along with responsibility placed on organizations (McNamara & O'Hara, 2008; Vanhoof et al., 2010). Therefore, this article aims to investigate the competency inventory of abilities and its application to HR practice as internal organizational practice.

In the first part, I will critically review the literature so as to then define the concept of competence according to various approaches and discuss the importance of competence tests, including self-assessment. Next, I will describe the research methodology, i.e. description of the research group, construction of items, and dimensions. In the next section, I will present in detail the method of validating and testing the quality of the TCI, also by formulating the research hypotheses, and then I will describe the

research results obtained. Finally, I will conclude, discuss research limitations, and suggest further research in this area.

Literature Review

Currently, competencies are the most important unit of educational effectiveness (Ridei et al., 2021). The competency-based approach in HEIs is becoming a priority; it aims to develop all aspects of teachers' competencies (Ridei et al., 2021). According to Armstrong (2009) competency-based human resources management (HRM) involves the use of the concept of competency and its analysis to improve HR practices, including employee recruitment, selection, development, performance evaluation, and remuneration.

Initially, competency has been defined as a formal right to deal with specific matters and make decisions in a set range on behalf of an organization. However, the ability to act effectively and efficiently was generally associated with qualifications. In the 1980s, competencies began to be understood more broadly as the employee's scope of rights, duties, and responsibilities to the assigned job. Boyatzis (1982) describes it as the potential existing in a person, which leads to behavior satisfying work requirements in the parameters of the organization's environment, which in turn provides desired results. Boyatzis establishes there is a range of factors that differentiates a successful performance from the less successful one. These factors include personal qualities, motivations, experience, and behavioral characteristics (Armstrong, 2009). Most authors struggle with the need to determine the competency components (Yeşilçınar & Çakır, 2020). Knowledge and skills are the essential components of competency. Relatively often, competency components include attitudes and behaviors (Muñiz-Rodríguez et al., 2017). Therefore, competency is the ability to use and apply knowledge in professional situations, as a set of capabilities that enable effective job performance, the achievement of goals, and compliance with standards of action. Moreover, the study by Keller-Schneider et al. (2020) revealed that teachers take competency as a resource and can differentiate their sense of competency across different domains.

Hager et al. (1994) recommend an integrated approach, in which competency is not observed directly, but it is inferred from performance. This approach places the assessment of competency in the same situation as other evaluation methods in the frame of available procedures with the goal to maximize its validity and reliability (Hager et al., 1994). This requirement is crucial in any attempt to put employee competencies into practice. The competency test is a tool that allows measuring the level of all components of competencies, namely knowledge, skills, and behavior.

Competency tests belong to the group of standardized methods of measurement and competency assessment. They are reliable tools that provide objective and reliable information about the level of competency. Competency tests are psychometric tools created to test specific competencies. As there are no universal competencies, there are also no comprehensive tests that measure competencies. The key indicator is test reliability and validity, which describes the accuracy of the competency test, along with the parameter that determines what exactly is the subject of measurement. The assessment method validity refers to how well it measures what it is supposed to measure.

A specific form of evaluation is self-evaluation, in which the employee gathers information and evidence to prove the level of competencies. Existing research shows that self-evaluation stimulates reflection and leads to concrete points for improvement (McNamara & O'Hara 2008; Baartman et al., 2013). Nevertheless, the quality of self-evaluation differs significantly (Schildkamp, 2007) concerning factors related to evaluators' attitudes toward self-evaluation, the provision of support, and the view of self-evaluation as a policy action that requires systematic data collection. Moreover, McNamara & O'Hara (2008) revealed that teachers collect evidence of their competencies, which are far from reality. Moreover, teachers find it difficult to conduct a systematic self-evaluation. To overcome these problems, TCI can be used as self-evaluation, but it can also be performed by supervisors who collect and provide appropriate evidence to demonstrate assessment quality (Baartman et al., 2013).

Research Methodology

This study was conducted to verify the proposed hypotheses (section 3.2). The research was conducted among HEI teachers, students, and graduates in Poland and India. Both countries are quickly developing; their educational systems are continuously changed by national regulations; however, the educational context differs. These elements motivated me to study the two different contexts. The sample consisted of 219 respondents from India and 47 from Poland. The difference in sample size was due to the number of higher education institutions in both countries. The education market in India is much larger than in Poland, and so, the sample size was representative of the number of institutions in both countries. In both cases, the scales were shown to be reliable, so the sample size was sufficient for validation.

This study followed the widely authenticated framework outlined by Hinkin (1998) and Churchill (1979) to develop a comprehensive psychometric survey instrument.

Accordingly, the empirical development and validation of the TCI involved the following steps: (1) item generation, (2) dimension identification and data reduction (factor analysis), and finally (3) scale evaluation. The measurement of validity and reliability was done following the guidelines provided by Hair et al. (2010).

Item Generation and Dimensions Identification

First, the study developed the initial pool of items for the questionnaire. Then, the data were analyzed to generate a list of items for each dimension (Turker, 2009). I specified the content of each dimension, identified items specific to each dimension, and then empirically determined how much these items were relevant to each studied domain. Based on the domain conceptualization, I generated 71 items that described teacher competencies and related teacher behavior based on the extensive literature review (Ramsden et al., 2007; Xu, 2011; Dani & Mhunpiew, 2019) and in-depth interviews with five experts from higher education institutions. The in-depth interviews were grounded on questions regarding the basic research question of this study: “What are the future-oriented teacher competencies?”

Following the guidelines that “the new items should be scaled using 5-point Likert scales” (Hinkin, 1998, p. 110), all items were anchored at five points (1=strongly disagree to 5=strongly agree), and participants expressed their level of agreement with each item.

The survey with 71 items was conducted among teachers, students, and graduates of higher education institutions in India and Poland.

The factor analysis was performed to investigate the underlying constructs of the 71 items in the TCI. As a result, five questions were removed from the survey because they did not make a significant contribution to any of the identified factors. Finally, there was high Cronbach’s alpha reliability for the identified dimension, which implied that the inventory enjoyed proper internal consistency. The resulting inventory was reliable and valid with 66 items in the four dimensions regarding future-oriented teaching competencies in HEIs (Appendix 1; Ludwikowska, 2019). A test of the inventory indicated that there was a high Cronbach alpha reliability for both groups from India (from 0.715 to 0.970) and Poland (from 0.749 to 0.979; Ludwikowska, 2019).

The following dimensions of the teacher competency inventory (TCI) emerged in the study: stimulating students for their achievement (SA), using different teaching approaches to improve student learning (TA), preparing students to be globally competent (GC), creating a supportive learning environment (LE).

The results show that the inventory enjoyed adequate internal consistency for four dimensions as general teaching competencies. For selected dimensions, I performed normalization aimed at adjusting the values measured on different dimensions to a standard scale. For each dimension, a sten score has been calculated (Table 1). Normalization ensured that all data look and read the same way across all records.

Table 1. Initial normalisation of the scales measured

Sten	SA	TA	GC	LE	Total
1	< = 2,69	< = 2,70	< = 2,74	< = 2,75	< = 12,76
2	(2,69–2,96]	(2,70–2,96]	(2,74–2,96]	(2,75–3,00]	(12,76–13,39]
3	(2,96–3,22]	(2,96–3,21]	(2,96–3,25]	(3,00–3,25]	(13,69–14,62]
4	(3,22–3,49]	(3,21–3,47]	(3,25–3,51]	(3,25–3,50]	(14,62–15,56]
5	(3,49–3,75]	(3,47–3,72]	(3,51–3,76]	(3,50–3,75]	(15,56–16,48]
6	(3,75–4,02]	(3,72–3,98]	(3,76–4,02]	(3,75–4,00]	(16,48–17,41]
7	(4,02–4,28]	(3,98–4,23]	(4,02–4,27]	(4,00–4,25]	(17,41–18,34]
8	(4,28–4,55]	(4,23–4,49]	(4,27–4,53]	(4,25–4,50]	(18,34–19,27]
9	(4,55–4,81]	(4,49–4,74]	(4,53–4,78]	(4,50–4,75]	(19,27–19,95]
10	> 4,81	> 4,74	> 4,78	> 4,75	> 19,95

Source: own work.

The results of each scale are in the range of 1–5 (average items were counted). The TCI with the final 66 items and four dimensions explained 47.279% of the total variance and indicated a strong factor model (Ludwikowska, 2019).

The development of a valid and reliable instrument is a complex and systematic process. These results gave the foundation for investigating the internal quality of the test in terms of its validity. Therefore, the second phase of the study sought to evaluate the degree to which the TCI dimensions constitute an operational definition of the construct used to determine teachers' pedagogical competencies. This study tested the construct validity of the test by using the independent sample t-test, variance analysis Anova in two different environments (India, Poland), and its application to HEI HR practices. The analysis was conducted to verify the proposed hypotheses.

Validity and Test Quality Evaluation

Further empirical investigation of TCI's psychometric properties is essential in developing the inventory and its constructs, bearing in mind that establishing the psychometric properties of behavioral measures is a complex and lengthy process (Cronbach & Meehl, 1955; Anastasi & Urbina, 1997). Hence, I performed validation to establish the test quality. To verify the research questions and the validity of TCI, the following hypothesis was formulated:

H1: There are differences among groups of respondents from India in perceiving the importance of TCI dimensions.

The following sub-hypotheses were specified as follows:

H1a: There are differences among male and female respondents from India in perceiving the importance of TCI dimensions.

H1b: There are differences between respondents from India who studied/taught abroad and those who did not in perceiving the importance of TCI dimensions.

H1c: There are differences among the Indian respondents who had lectures at their home university with a teacher from abroad and those who did not in perceiving the importance of TCI dimensions.

H1d: There are differences between respondents from India with different current occupations (student, graduate, academic teacher, other) in perceiving the importance of TCI dimensions.

H1e: There are differences among the Indian respondents with different educational backgrounds (engineering, humanities, social sciences, art, management, other) in perceiving the importance of the dimensions of TCI.

H1f: There are differences between respondents from India with different educational levels (undergraduate, postgraduate, Ph.D.) in perceiving the importance of TCI dimensions.

Then I formulated the second hypothesis:

H2: There are differences among respondents of different nationalities (India, Poland,) in perceiving the importance of TCI dimensions.

Validity refers to “the methodological and/or conceptual soundness of research” (Graziano & Raulin, 2000, p. 436). The *Standards for Educational and Psychological Testing*

refers to validity as “the appropriateness, meaningfulness, and usefulness of the specific inferences made from test scores. Test validation is the process of accumulating evidence to support such inferences” (American Psychological Association, American Educational Research Association, & National Council on Measurement in Education, 1985, p. 9).

The determination of a measure’s validity focuses on the extent to which its content captures the construct it intends to measure, as well as the extent to which empirical evidence supports its theoretical structure: “validity is the most important consideration in test evaluation” (American Psychological Association, American Educational Research Association, & National Council on Measurement in Education, 1985, p. 9).

The current conceptualizations of validity emphasize the need to understand what is being assessed, how it can be identified, and what else may be contributing to the scores to verify the content and validity of a measure. Thus, validity must be a primary concern throughout the instrument development process.

There are four types of validation: predictive validity, concurrent validity, content validity, and construct validity (Cronbach & Meehl, 1955). The current study aims to investigate the construct validity of TCI to measure its quality.

The idea of construct validity was introduced by the American Psychological Association with the goal to validate theoretical attributes and qualities (Colliver et al., 2012). Construct validity is central to establishing test quality (Embretson & Gorin, 2001). Construct validation is defined as a process by which researchers provide evidence to establish appropriate inferences from observed scores to conceptualization for a particular construct (Forer & Zumbo, 2011). Construct validation is applied whenever a test seeks to measure some attributes or quality which are not “operationally defined” (Lee, 1955). In terms of TCI, the attribute measure refers to four identified dimensions of teacher competencies. A “construct” is defined as a person’s postulated attribute. We expect that a person will at any time possess (or not) a qualitative attribute, namely competency. Construct validity attempts to validate the theory behind the instrument. The TCI constructs were theoretically defined and described based on the literature review (Ludwikowska, 2019).

The essence of the theoretical validity assessment process is the reference to the theory of the measured characteristic. Cronbach and Meehl (1955) distinguish five ways of estimating theoretical validity. This study focused on the analysis of intergroup

differences (t-test, variation analysis Anova). If the test is an accurate measure of a particular construct, the results obtained by the two groups should vary significantly (Westen & Rosenthal, 2005).

The t-test is applied as a statistical method to assess whether the mean value of the data from an independent sample – which follows a normal distribution – is consistent with or significantly departs from the mean value of a null hypothesis. Furthermore, the t-test evaluates whether the difference between the means of two independent samples following a normal distribution is statistically significant (Rivas-Ruiz et al., 2013). However, there was a limited number of studies analyzing the validity of the test using the t-test. An Anova test is a way to find out if the survey results are significant.

Therefore, to evaluate the validity of the TCI, this study scrutinized the quality of the test in terms of its construct validity by using the independent sample t-test, variance analysis, and Anova in two different environments with context variables (Cronbach & Meehl, 1955).

Results and Discussion: Differences Between Groups

To examine the proposed hypotheses and answer research questions, differences between groups were compared with the t-test and the Anova one- and two-factor difference tests. For significant interactions, Duncan's multiple range test was applied. Descriptive statistics were calculated, including mean and standard deviation, degrees of freedom (df), and the level of statistical significance (p). Descriptive statistics in the t-tests between the following groups were included:

- men and women for the groups from India;
- respondents answering “Yes” and “No” to the question “Have you ever studied/taught abroad?” for the group from India;
- respondents answering “Yes” and “No” to the question “Have you ever attended classes by a foreign teacher at your university?” for the group from India;
- respondents of different nationalities.

Descriptive statistics between the following groups were calculated in the single-factor Anova difference tests:

- respondents of different current positions for the group from India;
- respondents studying different faculties for the group from India;

- respondents with different levels of education for the study group from India;
- respondents of varying gender and nationalities.

Duncan's multiple range tests were calculated for significant interactions between the following groups:

- respondents of different genders and nationalities in the TA dimension;
- respondents of different genders and nationalities in the GC dimension.

To verify hypotheses H1 and H2, I analyzed whether there are statistically significant differences between the two groups of respondents from India. For this purpose, the t-test statistics for independent samples were calculated.

First, descriptive statistics between men and women for the groups from India were analyzed (Table 2).

Table 2. Descriptive statistics and the t-test between male and female for the groups from India

TCI Dimensions	Gender	Descriptive statistics			Statistics for t-test		
		N	Mean	Standard deviation	t (t-test)	df	p
SA	Male	136	4,19	0,57	-2,062	217	0.040
	Female	83	4,34	0,43			
TA	Male	136	4,18	0,55	-1,685	217	0.093
	Female	83	4,30	0,43			
GC	Male	136	4,23	0,54	-0,940	217	0.348
	Female	83	4,30	0,45			
LE	Male	136	4,23	0,54	-0,927	217	0.355
	Female	83	4,29	0,43			
Total	Male	136	16,83	2,00	-1,555	217	0.121
	Female	83	17,23	1,58			

Source: own work.

The results obtained in the SA dimension were statistically significant ($t(217) = -2.062$; $p = 0.040$) and show that women have higher scores (4.23 ± 0.43) than men (4.19 ± 0.57 ; Table 2). Next, the t-test was applied to verify whether respondents answering "Yes"

or “No” to the question “Have you ever studied/taught abroad?” differ statistically significantly in terms of average values of measured variables (Table 3).

Table 3. Descriptive statistics and the t-test between respondents answering “Yes” or “No” to the question “Did you ever study/teach abroad?” for the group from India

TCI Dimensions	Did you ever study/teach abroad?	Descriptive statistics			Statistics for t-test		
		N	Mean	Standard deviation	t	df	p
SA	Yes	87	4,14	0,58	-2,450	217	0.015
	No	132	4,32	0,48			
TA	Yes	87	4,15	0,53	-1,765	217	0.079
	No	132	4,27	0,49			
GC	Yes	87	4,21	0,48	-0,978	217	0.329
	No	132	4,28	0,52			
LE	Yes	87	4,19	0,51	-1,613	217	0.108
	No	132	4,30	0,49			
Total	Yes	87	16,69	1,92	-1,882	217	0.061
	No	132	17,17	1,80			

Source: own work.

Statistically significant results ($t(217) = -2.450$; $p = 0.015$) were observed in the SA dimension, in which respondents from India who answered “No” to the question “Have you ever studied/taught abroad?” had higher results (4.32 ± 0.48) from respondents who answered “Yes” (4.14 ± 0.58 ; Table 3). The analysis with the t-test revealed that respondents who answered “Yes” or “No” to the question “Have you ever attended classes by a foreign teacher at your university?” differed in a statistically significant way in terms of average values of measured dimensions (Table 4).

The results obtained in the SA dimension showed a statistically significant difference ($t(217) = -2.894$; $p = 0.004$): respondents from India who answered “No” to the question “Have you ever attended classes by a foreign teacher at your university?” have higher scores (4.38 ± 0.50) than respondents who answered “Yes” (4.17 ± 0.53 ; Table 4).

Statistically significantly ($t(217) = -2.347$; $p = 0.020$) in the LE dimension respondents from India, who answered “No” to the question “Have you ever attended classes by

a foreign teacher at your university?” had higher scores (4.19 ± 0.5) than people who answered “Yes” (4.35 ± 0.48 ; Table 4).

Table 4. Descriptive statistics and the t-test between respondents answering “Yes” or “No” to the question “Did you ever attend classes by a foreign teacher at your university?” for the group from India

TCI Dimensions	“Did you ever attend classes by a foreign teacher at your university?”	Descriptive statistics			Statistics for t-test		
		N	Mean	Standard deviation	t	df	p
SA	Yes	137	4,17	0,53	-2,894	217	0.004
	No	82	4,38	0,50			
TA	Yes	137	4,18	0,51	-1,693	217	0.092
	No	82	4,30	0,51			
GC	Yes	137	4,23	0,49	-1,105	217	0.270
	No	82	4,30	0,54			
LE	Yes	137	4,19	0,50	-2,347	217	0.020
	No	82	4,35	0,48			
Total	Yes	137	16,77	1,84	-2,218	217	0.028
	No	82	17,34	1,84			

Source: own work.

The further study aimed to verify differences between the four dimensions for respondents of different nationalities (India, Poland; Table 5), which revealed that:

- statistically significantly ($t(264) = 6.024, p < 0.01$), respondents from Poland on the SA dimension had lower scores (3.67 ± 0.86) than Indian respondents (4.25 ± 0.53 ; Table 5).
- statistically significantly ($t(264) = 7.272, p < 0.01$), respondents from Poland on the TA dimension had lower scores (3.58 ± 0.71) than Indian respondents (4.22 ± 0.51 ; Table 5).
- statistically significantly ($t(264) = 7.189, p < 0.01$), respondents from Poland on the GC scale had lower scores (3.61 ± 0.76) than Indian respondents ($4,26 \pm 0.51$; Table 5).
- statistically significantly ($t(264) = 6.956; p < 0.01$), respondents from Poland on the variable LE had lower scores (3.65 ± 0.72) than Indian respondents (4.25 ± 0.50 ; Table 5).

Table 5. Descriptive statistics and the t-test between respondents of different nationalities

TCI Dimensions	Country	Descriptive statistics			Statistics for t-test		
		N	Mean	Standard deviation	t	df	p
SA	India	219	4,25	0,53	6,024	264	<0,01
	Poland	47	3,67	0,86			
TA	India	219	4,22	0,51	7,272	264	<0,01
	Poland	47	3,58	0,71			
GC	India	219	4,26	0,51	7,189	264	<0,01
	Poland	47	3,61	0,76			
LE	India	219	4,25	0,50	6,956	264	<0,01
	Poland	47	3,65	0,72			
Total	India	219	16,98	1,86	7,452	264	<0,01
	Poland	47	14,50	2,86			

Source: own work.

One-factor ANOVA was calculated to analyze the differences between more than two studied groups. Duncan's multiple range tests were calculated for significant interactions.

First, the differences between respondents of different current positions for the group from India were calculated (Table 6).

Table 6. Descriptive statistics and the one-factor Anova difference test between respondents of different Current Position for the group from India

TCI Dimensions	Current Position								One-factor Anova
	student (n = 136)		alumnus/ graduates (n = 35)		academic teacher (n = 18)		other (n = 30)		
	M	SD	M	SD	M	SD	M	SD	
SA	4,35	0,59	4,28	0,54	4,40	0,50	4,39	0,51	F(3;215) = 0,194; p = 0.900
TA	4,29	0,51	4,33	0,38	4,44	0,45	4,35	0,53	F(3;215) = 0,205; p = 0.893
GC	4,20	0,55	4,27	0,48	4,27	0,53	4,27	0,53	F(3;215) = 0,233; p = 0.874
LE	4,13	0,56	4,15	0,53	4,20	0,53	4,20	0,54	F(3;262) = 0,550; p = 0.649
Total	16,97	1,93	17,04	1,68	17,31	1,84	17,21	1,94	F(3;262) = 0,265; p = 0.851

Source: own work.

Table 7. Descriptive statistics and the one-factor Anova difference test between respondents of different Faculty/ Branch for the group from India

TCI Dimensions	Faculty/ Branch you study/ graduate/ teach														One-factor Anova
	Engineering (n = 96)		Humanity (n = 23)		Social Sciences (n = 23)		Ar. (n = 13)		Management Studies (n = 51)		Other (n = 13)		M	SD	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD			
SA	4,27	0,58	4,34	0,55	4,32	0,42	4,42	0,50	4,18	0,45	3,94	0,52			F(5;213) = 1,577; p = 0,169
TA	4,25	0,53	4,35	0,44	4,33	0,42	4,07	0,66	4,20	0,46	3,85	0,59			F(5;213) = 2,218; p = 0,054
GC	4,29	0,50	4,26	0,58	4,28	0,50	4,16	0,67	4,27	0,43	3,96	0,50			F(5;213) = 1,100; p = 0,362
LE	4,28	0,53	4,29	0,47	4,37	0,37	4,14	0,62	4,25	0,44	3,89	0,53			F(5;213) = 1,910; p = 0,094
Total	17,09	1,96	17,24	1,87	17,29	1,48	16,80	2,33	16,90	1,56	15,64	1,95			F(5;213) = 1,715; p = 0,132

Source: own work.

The current position of the respondents revealed that the four groups did not differ statistically significantly ($p > 0.05$) in terms of the average values measured in the four dimensions.

Then, the differences between respondents of different faculty/branch for the group from India were calculated (Table 7).

The “Faculty/Branch you study/graduate/teach” section revealed that the four studied groups did not differ significantly ($p > 0.05$) in terms of average values measured in four dimensions (Table 7).

Third, the differences between respondents of different educational levels for the group from India were calculated (Table 8).

Table 8. Descriptive statistics with the one-factor Anova difference test between respondents of different Educational level for the group from India

TCI Dimensions	Educational level						One-factor Anova
	Undergraduate (n = 84)		Postgraduate (n = 110)		PhD (n = 21)		
	M	SD	M	SD	M	SD	
SA	4,19	0,60	4,28	0,48	4,30	0,42	$F(2;212) = 2,921; p = 0.056$
TA	4,19	0,54	4,24	0,50	4,24	0,46	$F(2;212) = 1,035; p = 0.357$
GC	4,22	0,52	4,26	0,51	4,27	0,45	$F(2;212) = 0,165; p = 0.848$
LE	4,20	0,53	4,28	0,48	4,26	0,46	$F(2;212) = 0,073; p = 0.930$
Total	16,79	2,03	17,06	1,77	17,06	1,58	$F(2;212) = 0,819; p = 0.442$

Source: own work.

The educational level of the respondents revealed that the four studied groups did not differ significantly ($p > 0.05$) in terms of the average values measured in the four dimensions (Table 8).

The two-factor ANOVA test was performed to analyze the differences between respondents of different sex and nationalities in terms of measured variables (Table 8).

Table 9. Descriptive statistics with the two-factor Anova difference test between respondents of different gender and nationality

TCI Dimension	India				Poland				Nationality	Gender	Nationality x Gender
	Male (n = 136)		Female (n = 83)		Male (n = 17)		Female (n = 30)				
	M	SD	M	SD	M	SD	M	SD			
SA	4,19	0,57	4,34	0,43	3,38	0,90	3,83	0,81	F(1;262) = 44,885; p < 0,01	F(1;262) = 9,305; p = 0.003	F(1;262) = 2,326; p = 0.128
TA	4,18	0,55	4,30	0,43	3,24	0,89	3,77	0,50	F(1;262) = 66,125; p < 0,01	F(1;262) = 13,296; p < 0,01	F(1;262) = 5,379; p = 0.021
GC	4,23	0,54	4,30	0,45	3,25	0,98	3,81	0,52	F(1;262) = 63,614; p < 0,01	F(1;262) = 11,490; p < 0,01	F(1;262) = 7,028; p < 0,01
CLE	4,23	0,54	4,29	0,43	3,39	0,89	3,79	0,56	F(1;262) = 55,807; p < 0,01	F(1;262) = 6,777; p < 0,01	F(1;262) = 3,537; p = 0.061
Total	16,83	2,00	17,23	1,58	13,26	3,51	15,21	2,19	F(1;262) = 67,975; p < 0,01	F(1;262) = 11,983; p < 0,01	F(1;262) = 5,190; p = 0.023

Source: own work.

The interaction between respondents' nationality and sex was significant for the TA dimension ($F(1; 262) = 5.337; p = 0.021$) and for the GC ($F(1; 262) = 7,028, p < 0.01$; Table 9). For these variables, post hoc test scores were calculated for specific differences (Table 10).

Table 10. Descriptive statistics of post-hoc Duncan difference test between respondents of different gender and nationalities in the TA dimension

	Group	Gender	{1}	{2}	{3}	{4}
			4,18	4,30	3,24	3,77
1	India	Male		0,348039	0,000011	0,001520
2	India	Female	0,348039		0,000003	0,000067
3	Poland	Male	0,000011	0,000003		0,000033
4	Poland	Female	0,001520	0,000067	0,000033	

Source: own work.

An important conclusion is that respondents of different sex from India in the TA dimension have higher scores than respondents from Poland. An interesting result is also that there are no differences ($p = 0.348$) between women ($M = 4.30$) and men ($M = 4.18$) in the group from India in relation to the dimension mentioned above. In the Polish group, men had significantly ($p < 0.01$) lower scores ($M = 3.24$) than women ($M = 3.77$). Figure 1 presents the mean values with a 95% confidence interval for the mean value broken down by group and sex.

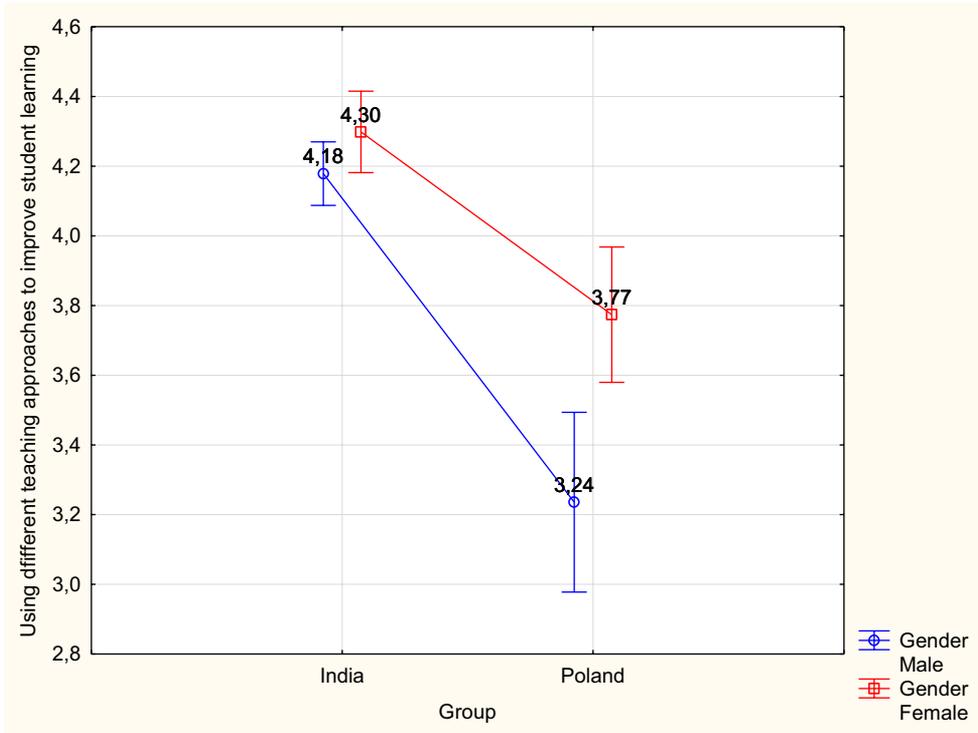
The descriptive statistics of Duncan's multiple range test were calculated for respondents of different sex and nationalities in the GC dimension (Table 11).

Table 11. Descriptive statistics of post-hoc Duncan difference test between respondents of different gender and nationalities in the GC dimension

	Group	Gender	{1}	{2}	{3}	{4}
			4,23	4,30	3,25	3,81
1	India	Male		0,606491	0,000011	0,001188
2	India	Female	0,606491		0,000003	0,000258
3	Poland	Male	0,000011	0,000003		0,000027
4	Poland	Female	0,001188	0,000258	0,000027	

Source: own work.

Figure 1. The mean values with a 95% confidence interval for the average value of the TA dimension by gender and group



Source: own work.

An important conclusion was also that respondents of a different sex from India in the GC dimension have higher scores than respondents from Poland. An interesting result was the lack of differences ($p = 0.606$) between women ($M = 4.30$) and men ($M = 4.23$) in the dimension mentioned above in the Indian group. In the Polish group, men had significantly ($p < 0.01$) lower scores ($M = 3.25$) than women ($M = 3.81$; Table 11).

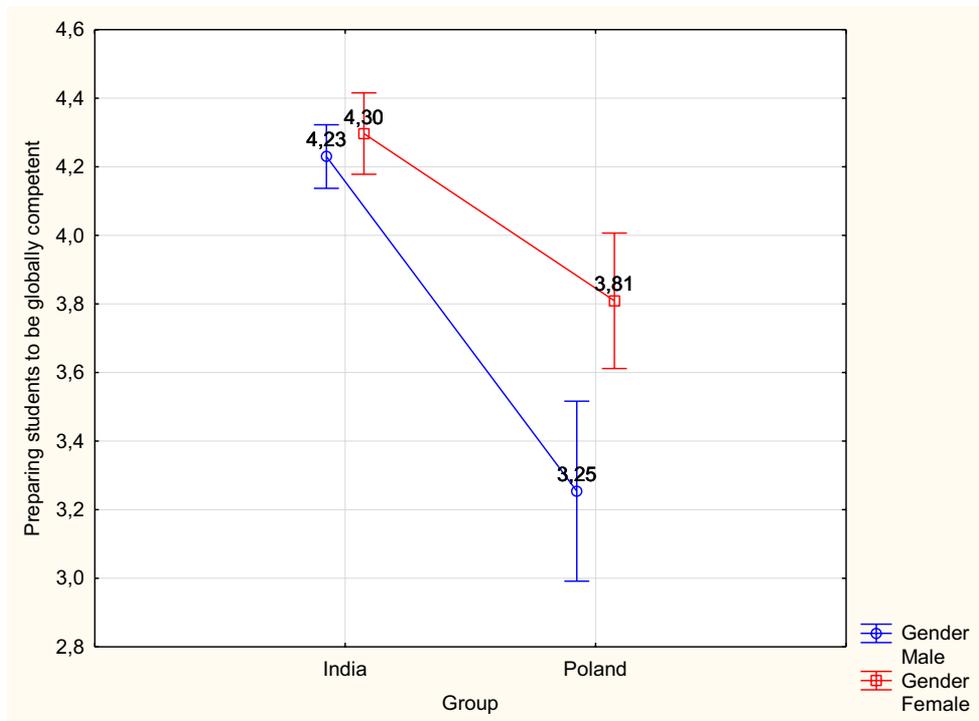
Figure 2 presents the mean values with a 95% confidence interval for the mean value divided by respondents' nationality and sex.

The results obtained in the study confirmed the high internal quality of TCI. The research analysis that applied the t-test and the Anova one- and two-factor difference tests confirmed H1: There are differences between groups of respondents from India in perceiving the importance of the dimensions of TCI. The sub-hypotheses H1a, H1b, and H1c were also evaluated:

- **H1a:** the women in the SA dimension had higher scores (4.23 ± 0.43) than the men (4.19 ± 0.57 ; Table 4);
- **H1b:** statistically significant, respondents from India in the SA dimension who have chosen the answer “No” to the question “Have you ever studied/taught abroad?” (4.32 ± 0.48) had higher results than respondents who answered “Yes” (4.14 ± 0.58 ; Table 3);
- **H1c:** only the SA and LE dimensions were significant (Table 4).

The respondents from Poland had lower scores (3.67 ± 0.86) than the respondents from India (4.25 ± 0.53) in all dimensions (Table 5), which confirmed H2.

Figure 2. The mean values with a 95% confidence interval for the average value of the GC dimension by gender and group



Source: own work.

Furthermore, the results of the two-factor Anova difference test between respondents of different sex and nationalities strengthened H1a and H2. Statistically significantly, the interaction between nationality and gender of the respondents appeared for the TA dimension (Table 9). These differences were supported by the results of Duncan's

multiple range tests for specific differences (Table 10). In two dimensions, TA and GC, respondents of different sex from India have higher scores than respondents from Poland. Moreover, there were differences between women and men in the group from India in both cases. However, in the Polish group, men had significantly lower scores than women (Table 10, Table 11).

However, the sub-hypotheses H1d, H1e, and H1f were not confirmed. The current position – the faculty/branch you study/graduate/teach – and the educational level of respondents revealed that the four groups do not differ statistically significantly ($p > 0.05$) in terms of average values measured in four dimensions (Table 6, Table 7, Table 10). This result gives the basis to accept the HO about the lack of differences between the analyzed groups of respondents.

Discussion

There is a growing number of assessments, testing, and evaluation methods. The ultimate goal of any measurement instrument is to produce enough information to allow the user to make appropriate judgments. The results of the study allowed me to answer the research questions:

Q1: Research analyses confirmed that there was a high Cronbach's alpha reliability for both groups from India and Poland for the developed inventory. Furthermore, the comparison test in most cases confirmed the formulated hypotheses. The internal quality of the test has been established through the validity of the construct. The analysis confirmed the quality of the TCI in terms of its validity as a measure of the pedagogical competencies of academic teachers and its application to HR practices.

Q2: In most analyzed cases, the model fits different groups related to sex, nationality, respondents' experience in studying/teaching abroad, and respondents' experience in interacting with a foreign teacher at home university. Thus, the results confirmed the quality of the TCI in terms of its validity as a measure of pedagogical competencies of academic teachers and its application to HR practices.

Competency tests are standardized methods of competency assessment (Armstrong & Taylor, 2014). These are the tools that provide objective and reliable information about the level of competency a person has and enable planning further development. The objectivity of the results has high psychometric parameters that include accuracy

and reliability. Moreover, a comparison is possible between the respondents considering themselves and the competency profile associated with their positions.

Therefore, the teacher competency inventory (TCI) was revealed to be a meaningful tool to enhance HR practices in HEIs and to indicate the current and desired level of competencies. The validated dimensions of the competency model can be used to control the quality of teaching considered as a determinant of the quality of any educational institution (Yeşilçınar & Çakır, 2020).

Conclusions

The TCI is a four-dimensional scale that explains future-oriented teacher competencies. The findings of the study contribute to the field of assessing teacher competencies. The TCI allows comparing teacher competencies with the expected competency profile. Possible differences may indicate deficits or overruns in relation to the expected profile. The 66-item TCI is an accessible and easily administered measure of teacher competencies.

The identified four competencies and behavioral descriptions provide a framework of competencies that can be used to build the desired competency profile of the teacher in HEIs in India and Poland.

Comparison of the desired and existing teacher competency profile allows identifying the competence gap. The identified competency gap may be the basis for increasing job performance through different HR practices. Competencies are indicators of the quality of education, which results at the microlevel in a higher quality of teaching and learning in the classroom and – at the macrolevel – of education as a whole.

Competence profiles are widely used in HRM, including improving quality and efficiency of work, recruitment, selection, and adaptation of employees, increasing the flexibility of employment, transparency, and clarity of employee roles, training, development of professional skills, planning a career development path, rewarding employees, and creating equal opportunities and possibilities.

However, this empirical study was burdened with certain limitations. First, the hypotheses were tested based on one research sample, limited to educational organizations operating in Poland and India. Nevertheless, it seems to be a solid underlining of the need for further analysis of the validity of tests used for HR practice in HEIs. Second,

the assessment of competences can also be based on other methods: performance samples, natural observation at the workplace, and evidence from previous achievements.

To further investigate the quality of the test, future studies should increase the size of the research sample in both countries. Complementary studies could focus on how to integrate the available methods of competency assessment and combine evidence provided by different stakeholders.

References

- Adnot, M. et al. (2017). Teacher Turnover, Teacher Quality, and Student Achievement in DCPS. *Educational Evaluation and Policy Analysis*, 39(1), 54–76. <https://doi.org/10.3102/0162373716663646>.
- Anastasi, A. & Urbina, S. (1997). *Psychological Testing*, 7th edition. Prentice Hall.
- Armstrong, M. (2009). *Armstrong's Handbook of Human Resource Management Practice*. Kogan Page, Business & Economics.
- Armstrong, M., & Taylor, S. (2014). *Armstrong's Handbook of Human Resource Management Practice*. 13 th edition.
- Baartman, L., Gulikers, J. & Dijkstra, A. (2013). Factors influencing assessment quality in higher vocational education. *Assessment & Evaluation in Higher Education*, 38(8), 978–997. <https://doi.org/10.1080/02602938.2013.771133>.
- Bergsmann, E., Klug, J., Burger, Ch., Först, N. & Spiel, Ch. (2018). The Competence Screening Questionnaire for Higher Education: Adaptable to the needs of a study programme. *Assessment & Evaluation in Higher Education*, 43(4), 537–554. <https://doi.org/10.1080/02602938.2017.1378617>.
- Boyatzis, R (1982). *The Competent Manager*. Wiley.
- Cohen, D.J. (2015). HR past, present and future: A call for consistent practices and a focus on competencies. *Human Resource Management Review*, 25(2), 205–215. <https://doi.org/10.1016/j.hrmr.2015.01.006>.
- Colliver, J.A., Conlee, M.J., & Verhulst, S.J. (2012). From test validity to construct validity ... and back?, *Medical Education*, 46(4), 366–371. <https://doi.org/10.1111/j.1365-2923.2011.04194.x>.
- Cronbach, L.J., & Meehl, P.E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281–302. <https://doi.org/10.1037/h0040957>.
- Dani, A.A., & Mhunpiew, N. (2019). A Development of an Academic Leadership Model for Higher Education in India. *Scholar: Human Sciences*, 11(1), 45–45.
- Darling-Hammond, L. (2014). One Piece of the Whole: Teacher Evaluation as Part of a Comprehensive System for Teaching and Learning. *American Educator*, 38(1), 4–13.
- Embretson, S., & Gorin, J. (2001). Improving construct validity with cognitive psychology Principles. *Journal of Educational Measurement*, 38(4), 343–368. <https://doi.org/10.1111/j.1745-3984.2001.tb01131.x>.
- Firestone, W.A. & Donaldson, M.L. (2019). Teacher evaluation as data use: what recent research suggests. *Educational Assessment, Evaluation and Accountability*, 31(3), 289–314. <https://doi.org/10.1007/s11092-019-09300-z>.

- Forer, B., & Zumbo, B.D. (2011). Validation of multilevel constructs: Validation methods and empirical findings for the EDI. *Social Indicators Research*, 103(2), 231–265. <https://doi.org/10.1007/s11205-011-9844-3>.
- Graziano, A.M., & Raulin, M.L. (2000). *Research Methods: A process of inquiry fourth edition*. Allyn & Bacon.
- Hager, P., Gonczi, A. & Athanasou, J. (1994). General Issues about Assessment of Competence. *Assessment & Evaluation in Higher Education*, 19(1), 3–16. <https://doi.org/10.1080/0260293940190101>.
- Keller-Schneider, M., Zhong, H.F. & Yeung, A.S. (2020). Competence and challenge in professional development: teacher perceptions at different stages of career. *Journal of Education for Teaching*, 46(1), 36–54. <https://doi.org/10.1080/02607476.2019.1708626>.
- Ludwikowska, K.H. (2019). Competence inventory: an empirical study on future-oriented competences of the teaching profession in higher education in India. *Education and Training*, 61(9), 1123–1137. <https://doi.org/10.1108/ET-12-2018-0266>.
- McNamara, G., & O'Hara, J. (2008). The Importance of the Concept of Self-evaluation in the Changing Landscape of Educational Policy. *Studies in Educational Evaluation*, 34, 173–179. <https://doi.org/10.1016/j.stueduc.2008.08.001>.
- Muñoz-Rodríguez, L. et al. (2017). Developing and validating a competence framework for secondary mathematics student teachers through a Delphi method. *Journal of Education for Teaching*, 43(4), 383–399. <https://doi.org/10.1080/02607476.2017.1296539>.
- Pak, K., Dorien, T.A., Kooij, M., De Lange, A.H., & Van Veldhoven, M.J.P.M. (2019). Human Resource Management and the ability, motivation and opportunity to continue working: A review of quantitative studies. *Human Resource Management Review*, 29(3), 336–352. <https://doi.org/10.1016/j.hrmr.2018.07.002>.
- Ramsden, P., Prosser, M., Trigwell, K., & Martin, E. (2007). University teachers' experiences of academic leadership and their approaches to teaching. *Learning and instruction*, 17(2), 140–155. <https://doi.org/10.1016/j.learninstruc.2007.01.004>.
- Ridei, N., Khomenko, O., Ivanenko, I., Filyanina, N., & Poberezhets, H. (2021). Competence of teachers of HEIS in the context of lifelong learning. *Laplace em Revista*, 7(1), 516–530. <https://doi.org/10.24115/S2446-6220202171853p.516-530>.
- Rivas-Ruiz, R., Pérez-Rodríguez, M., & Talavera, J.O. (2013). Clinical research XV. From the clinical judgment to the statistical model. Difference between means. Student's t-test. *Revista medica del Instituto Mexicano del Seguro Social*, 51(3), 300–303.
- Schildkamp, K. (2007). *The utilization of a Self-evaluation Instrument for Primary Education*. PhD diss., Twente University, The Netherlands.
- Standards for educational and psychological testing (1985). American Educational Research Association, American Psychological Association, & National Council on Measurement in Education.
- Turker, D. (2009). Measuring corporate social responsibility: A scale development study. *Journal of Business Ethics*, 85(4), 411–427. <https://doi.org/10.1007/s10551-008-9780-6>.
- Ulrich, D., & Dulebohn, J.H. (2015). Are we there yet? What's next for HR? *Human Resource Management Review*, 25(2), 188–204. <https://doi.org/10.1016/j.hrmr.2015.01.004>.
- Westen, D., & Rosenthal, R. (2005). Improving construct validity: Cronbach, Meehl, and Neurath's ship: Comment. *Psychological Assessment*, 17(4), 409–412. <https://doi.org/10.1037/1040-3590.17.4.409>.
- Vanhoof, J., de Maeyer, S., & van Petegem, P. (2010). Variation in the Conduct and the Quality of Self-evaluations: A Multi-level Path Analysis. *Educational Studies*, 37, 277–287. <https://doi.org/10.1080/03055698.2010.506326>.

- Yeşilçınar, S., & Çakır, A. (2020). Suggesting a Teacher Assessment and Evaluation Model for Improving the Quality of English Teachers. *Education and Science*, 45(202), 363–393. <https://doi.org/10.15390/EB.2020.8463>.
- Xu, K. (2011). An empirical study of Confucianism: Measuring Chinese academic leadership. *Management Communication Quarterly*, 25(4), 644–662. <https://doi.org/10.1177/0893318911405621>.

Appendix 1

TCI Scale Items

1. Encourages students to move forward, think creatively and out of the box.
2. Motivates students to grow not only in their academic but also non-academic endeavors.
3. Understands individual needs of the students.
4. Guides students to contemporary works of other academics relevant to their interests.
5. Sensitive to the student's abilities and life circumstances.
6. Knowledgeable, teaches what is both in and beyond the handbook.
7. Creates a learning environment that allows for alternative styles of learning, new and different ways of participation of students.
8. Takes time to recognize individual abilities of students.
9. Creates an environment in which students can openly express their viewpoints and encourages sharing various opinions in the classroom.
10. Systematically collects and interprets the evaluation results of multiple assessments to monitor, improve, and report individual and group achievements.
11. Provides students with timely, helpful, and accurate feedback on their performance and progress toward achievement goals.
12. Holds and directs the interest of the students.
13. Makes use of interactive and activity-based sessions and lectures, such as group discussion, role playing, group projects, and practical activities.
14. Applies experiential-based teaching where students can learn through experience and reflection.
15. Applies real-life cases as examples during lectures.
16. Shares basic readings to develop a base for the presented concepts.
17. Allows students to present their ideas on different research topics.
18. Allows students to be exposed.
19. Engages in problem-solving and critical thinking.
20. Has high level of self-discipline and prefers to follow a plan rather than act spontaneously.
21. Sensitive to students' needs.

22. Listens to opinions and feedback from students and does not misinterpret students' viewpoints.
23. Is receptive to new ideas and enjoys learning new things.
24. Understands students and their perspective.
25. Shows patience to students of different learning speeds and is responsive to student questions without acting on annoyance or anger in a negative way.
26. Confident in the knowledge of the material taught and in own teaching abilities.
27. Dedicated to students with commitment to their education and well-being.
28. Brings structure and organization to the classroom in a way that students appreciate the value of discipline.
29. Has an optimistic view of human nature, is pleasant to communicate with and willing to interact with students.
30. Easygoing, outgoing, and friendly.
31. Faces challenges and think about problems and tasks in novel and creative ways.
32. Is using technology in teaching that enhance the learning process.
33. Courses are taught from multiple perspectives, referring to theoretical foundations with practical applications.
34. Makes classrooms into discussion forums to enhance or criticize existing views.
35. Has a relationship with students based on empathy and understanding.
36. Keeps oneself and teaching notes up to date with recent findings in the subject field.
37. Teaches learners to expand the applicability of their ideas and what is being taught in class in their practical life.
38. Facilitates student ideas and trains their minds.
39. Can communicate in an attractive, simple, and convincing manner so that students can freely discuss different issues.
40. Motivates and inspires students to strive for growth.
41. Organizes teaching in advance to predict different issues.
42. Can be empathetic and understands student needs and requirements.
43. Acts as a friend to students while maintaining authority.
44. Builds trust and develops rapport.
45. Shows passion and interest in teaching and students.
46. Structures statements depending on the listener so that they understand everything quickly and correctly.
47. Can choose appropriate leadership style, so that sometimes teaching approach is authoritative or democratic.
48. Identifies and brings out the best in individuals, shares credit, and responsibility with the students.
49. Understands and respects diversity without bias and judgement.

50. Is updated for what is happening around the world and incorporates that into classroom settings.
51. Willing to learn from students about their individual cultures.
52. Has an excellent knowledge of cross-cultural ethics and principles.
53. Takes every opportunity to gain cultural competence and broaden own inter-cultural perspective.
54. Respects and accepts individuals belonging to different nationalities and ethnicities.
55. Adopts a teaching style and techniques that can be applied to students from diverse cultures.
56. Creates a self-awareness in students and develops activities that increase self-confidence of minority students.
57. Incorporates cultural contexts in learning wherever appropriate.
58. Sensitive to different cultures and student viewpoints.
59. Incorporates advances in technology in teaching methods.
60. Connects students to the world and exposes them to real life situations.
61. Makes teaching in classes relevant to the changing world.
62. Explains to students what methods are followed in other countries.
63. Updates own knowledge to be interdisciplinary.
64. Looks at things from a variety of viewpoints.
65. Can adjust to different cultural settings.
66. Inspires students to think deeply about global issues.