Pertanika J. Soc. Sci. & Hum. 31 (1): 319 - 338 (2023)

PERTANIKA JOURNALS

SOCIAL SCIENCES & HUMANITIES

Journal homepage: http://www.pertanika.upm.edu.my/

Rasch Analysis for Standards-Setting Appraisal of Competency Level-Based Performance on the Part of Instructors in Higher Education

Chatchawan Nongna[1], Putcharee Junpeng[1]*, Jongrak Hong-ngam[2], Chalunda Podjana[3] and Keow Ngang Tang[4]

[1]Department of Educational Measurement and Evaluation, Faculty of Education, Khon Kaen University, 40002 Khon Kaen, Thailand
[2]Faculty of Economics, Khon Kaen University, 40002 Khon Kaen, Thailand
[3]Department of Research Measurement and Evaluation, Faculty of Education, Suratthani Rajabhat University, 84100 Surat Thani, Thailand
[4]Institute for Research and Development in Teaching Profession for ASEAN, Khon Kaen University, 40002 Khon Kaen, Thailand

ABSTRACT

This research aimed to examine higher education instructors' performance assessment in determining the cut-off point by setting criteria on the Wright map from big data. It is followed by designing performance assessment standards and assessing their quality. A total of 603 instructors from a Thai public university were selected as participants. The researchers employed a design-based research method encompassing four phases: analyzing the results of the performance assessment, formulating the standards-setting appraisal, applying trial and quality inspection, and improving the standards-setting appraisal approach. Data were analyzed using the Rasch model and the Maximum Likelihood Estimation method. The results of the determination of the cut-off point in terms of assessing instructors' performance indicated that there are four cut-off points in ascending order, specifically, -11.67, -2.68, 4.59, and 9.76. The standards-setting appraisal showed that the assessment criteria consisted of five score ranges converted from estimation competency parameters into the scale and raw scores, respectively. Even though the standards-setting appraisal was determined, the researchers found that the transition point with regard to determination will be accurate and consistent in terms of those instructors

ARTICLE INFO

Article history:
Received: 16 February 2022
Accepted: 07 September 2022
Published: 17 March 2023

DOI: https://doi.org/10.47836/pjssh.31.1.17

E-mail addresses:
chatnon@kku.ac.th (Chatchawan Nongna)
jputcha@kku.ac.th (Putcharee Junpeng)
jonaka1@kku.ac.th (Jongrak Hong-ngam)
chalundap@kkumail.com (Chalunda Podjana)
tangng@kku.ac.th (Keow Ngang Tang)
* Corresponding author

ISSN: 0128-7702
e-ISSN: 2231-8534

who are at a moderate to high competency level and not suitable for evaluating those at a low competency level. The standards-setting appraisal approach is relevant for use as a criterion for recruiting and selecting higher education instructors. It can also support the development of sustainable human capital. It implies that instructors must possess high core competencies to match the high demand for quality teaching.

Keywords: Core competencies, higher education instructors, performance assessment, standards-setting appraisal

INTRODUCTION

Higher education instructors represent a specialized group of employees and the intellectual elite of a nation because they are a model of erudition and permanent progress that can be used for the benefit of society as a whole (Blašková et al., 2014; Nongna et al., 2021). It is supported by Tan et al. (2019), who highlighted the challenge currently faced by higher education institutions with regard to improving instructors' teaching and other relevant core competencies to enhance student's learning process. Therefore, the competency level-based performance of this group means that they have to accept many responsibilities that they will find very demanding in terms of their mental conditions and personal core competencies (Lohman, 2021). Turturean (2013) identified the core competencies of higher education instructors as a combination of knowledge, skills, attitudes, and behavior required for effective performance in the academic context. Following this line of reasoning, determining the desirable profile of a higher education instructor means that we must attempt to identify the designed requirements, characteristics, attributes, and elements that contribute to his or her knowledge, skills, and experience with which to assist his or her students. These should be the core competencies model or used to identify a standard (Aliasghar et al., 2017). Each higher education instructor should follow this core competencies model in terms of behaviors and performance with regard to his or her work (Blašková & Blaško, 2012).

Selvi (2010) listed nine core competencies as they apply to a higher education instructor: field competencies, research competencies, curriculum competencies, lifelong learning competencies, socio-cultural competencies, emotional competencies, communication competencies, information and communication technologies competencies, and environmental competencies. Field competencies are strictly related to the content knowledge of their teaching expertise. Research competencies encompass the competencies with regard to research methods and techniques, as well as designing and conducting research. Curriculum competencies include curriculum development competencies and curriculum implementation competencies. Lifelong learning competencies comprise the abilities with learning to learn and instructors' responsibilities for their professional development. Socio-cultural competencies relate to knowledge about the socio-cultural background of students and instructors, local, national, and international

values, democracy and human rights issues, team and collaborative work with others, and social studies that promote learning. Emotional competencies refer to instructors' and students' values, morals, beliefs, attitudes, anxieties, motivation, and empathy. Communication competencies comprise communication models, interactions among instructors, students, the social environment, and learning topics. Information and communication technologies (ICT) competencies are based on tools and technical equipment for approaching, searching, and transferring knowledge. Environmental competencies mean competencies in terms of ecological and environmental safety.

Prasetio et al. (2017) studied the relationship between higher education instructors' professional competency and the impact on student's academic performance. According to Prasetio et al., higher education instructors should have a wide and deep knowledge of the courses they teach and possess the practical experience to make their lessons attractive. Therefore, higher education instructors are expected to provide high-quality education to support their students in obtaining jobs in an increasingly competitive environment when they graduate. It is seen as a key indicator of job performance. Following this line of reasoning, instructors' core competencies are important factors that can contribute to improving the cognitive abilities and mental attitudes of higher education graduates. However, their research showed no significant relationship between instructors' core competencies and students' academic performance in the School of Economics and Business at Telkom University. Although instructors' core competencies do not directly impact students' academic achievement, their core competencies and positive behavior will certainly encourage new insights and capabilities on the part of their students (Prasetio et al., 2017).

The teaching workload of higher education institutions is not limited to their regular teaching activities, such as giving lectures, preparing lesson plans, evaluating scripts, and attending training programs and conferences. In addition, they have to be involved in non-academic activities such as proctoring, coordinating various higher education institution activities, and taking on various administrative posts (Hosain, 2016; Islam et al., 2019). Moreover, higher education instructors' work performance is based on their dedication to the job, success in research and development, student achievement, and active communication involving international collaboration and networking (Le, 2021; Odero & Makori, 2018). As such, we cannot overlook the potential effect of a standards-setting appraisal on the effectiveness of the strategic decision-making process with regard to instructors' performance. Kurtulmus et al. (2016) stated that one of the significant elements of effective decision-making is to be evidence-based on multiple components of the standard-setting appraisal.

According to many past researchers (Laei et al., 2014; Le, 2021), performance

appraisal consists of a system of measuring and assessing the effectiveness of an instructor's characteristics, behaviors, career gains, and level of interest in determining their current performance level. Moreover, higher education institutions must keep track of various performance indicators on the part of their instructors due to the high awareness of the need for quality audits for personal improvement and stakeholder satisfaction (Aliasghar et al., 2017; Alkhafaji, 2013). Gómez and Valdés (2019) found that instructors' performance is assessed by higher education institutions to favor teacher learning over the practice to improve it. Therefore, the analysis and establishment of standards-setting appraisal are important, leading to the need to create a proposal for the professional development of instructors and improvement in education quality (Aliasghar et al., 2017; Gómez & Valdés, 2019; Islam et al., 2019; Prasetio et al., 2017).

Tan et al. (2019) investigated how higher education students in Malaysia evaluated their instructors based on non-instructional factors, namely physical attractiveness and psychological factors, which could affect students' perceptions regarding their instructors' performance. Their results indicated that students' degree of confidence and level of acceptance are the mediators of the relationship between instructors' teaching competencies and their performance assessment. Aliasghar et al. (2017) and Altbach et al. (2009) emphasized the importance of standards-setting appraisal as common practice for higher education institutions to recognize instructors' strengths and weaknesses. As a result, a standards-setting appraisal system should include classroom settings, instructors, and students' context of the syllabus, learning activities, learning outcomes, and any efforts to enhance teaching performance (Aliasghar et al., 2017; Altbach et al., 2009).

Gómez and Valdés (2019) critically reviewed the methods used to evaluate higher education instructors' job performance. They also reviewed the different appraisal models to find discrepancies between objectives and practices. Their result showed a poor relationship between the appraisal model and what was done. Their research into higher education institutions indicated that the focus of performance appraisal is improvement in practice. In addition, they found that most of the higher education institutions used questionnaires as an instrument to evaluate instructors' work performance. However, this appraisal method failed to evaluate instructors' teaching performance even though it is an accurate measurement in the best-case scenario. Therefore, they proposed a constructivist model that encourages the changes required through reflection. It, in turn, results from a review in the form of self-evaluation, hetero-evaluation, and co-evaluation.

Laei et al. (2014) used a close-ended questionnaire with 25 Likert scale items and an open-ended questionnaire as instruments to investigate how to improve their performance and determine the competence that can be achieved from the

higher education instructors' perspective. A total of 114 full-time instructors from Kermanshah Azad University participated in the research. The results revealed that most instructors believed that the currently used job assessment method did not work properly to improve the instructors' level of education and research competencies at a 0.99% confidence level. Most instructors preferred the method of reporting scores relating to assessment as the most appropriate form of feedback for improving instruction quality. Even though their results indicated that the success level of the current method for assessing instructors' performance and identifying the consistency level of the instructors' performance with instructional standards was weak, with a 0.99% confidence level, the current plan was found successful in identifying instructors' performance in terms of professional, ethical standards.

Curzi et al. (2019) surveyed 865 employees to identify specific characteristics of performance appraisal that were more likely to be perceived as promoting individual innovation at work. These researchers utilized the process-based approach to human resource management proposed by Bowen and Ostroff (2004). They employed logit analysis to assess the correlation between data on performance appraisal systems and data on the effectiveness of performance appraisal as a promoter of innovative work behavior. Their research showed that formal performance appraisal is more likely to reduce the perception that performance appraisal promotes individual innovation and creativity at work than informal feedback. Moreover, they also found that a performance appraisal focused on the achievement of pre-set quantitative outcomes is more likely to positively affect innovative work behavior than a performance appraisal focused on pre-defined skills that employees exhibited while performing their work. They concluded that performance assessment should focus on the new competencies that the employees develop because the competency-based appraisal approach has a perceived positive impact and is even stronger than the result-oriented appraisal.

Lohman (2021) reviewed the recent literature that expresses the prominent arguments in research into student assessment of teaching. It is followed by a peer review of teaching and outlining essential performance appraisal and management principles. Those principles are then utilized to analyze representative faculty assessment policies and procedures and illuminate the weaknesses of traditional and recently amended teaching assessment approaches. Lohman's synthesis of past research revealed that behavioral assessment as part of standards-setting appraisal needs a well-constructed rating instrument. It implies that reliable and valid instruments are vital to assess the behavioral competencies of instructors (Caruth & Humphreys, 2008) but that it is particularly important to acknowledge the difficulty of measuring teaching performance and this difficulty is potentially reducing the emphasis on teaching in standards-setting appraisal approaches (Cardoso et al., 2015).

In general, instructor performance assessment in this research university comprises two major components, namely, 70% work achievement and 30% behavioral assessment (Khon Kaen University, 2015, Nongna et al., 2021). Instructor performance assessment mainly focuses on the operational competencies encompassing service mind, expertise, achievement motivation, teamwork, and integrity, as indicated in the guidelines published by the Thailand Civil Service Commission in 2011 (Office of the Higher Education Commission, 2018). Despite decades-old texts offering descriptions of, and guidance on, instructors' performance evaluation, well-established standards-setting appraisal measurement models have yet to be outlined systematically in higher education institutions to address concerns about instructors' core competencies. It requires a measurement model to reveal the full potential of human resources and to act as a tool to achieve a consistent evaluation of quality teaching (Lohman, 2021). Therefore, the researchers aimed to formulate and assess the quality of a standards-setting appraisal measurement model to address the gap and clarify how the measurement model can be utilized as a sound performance appraisal measurement model.

MATERIALS AND METHODS

Research Procedure

A design research method encompassing four phases was employed to develop and refine a standards-setting appraisal measurement model to advance the existing theory that can support and lead to a deepened understanding of competency level-based performance evaluation (Reeves, 2006; Vongvanich, 2020). It is because design-based research is often associated with conducting research in technology-enhanced assessing contexts and has been used in the broader field of research in education (Vongvanich, 2020).

Firstly, the researchers began investigating higher education instructors' performance assessment results from big data. Secondly, they used the performance assessment from the first phase to formulate the standards-setting appraisal of competency level-based performance for instructors in higher education institutions. Thirdly, the researchers tried out the formulated standards-setting appraisal of competency level-based performance for quality inspection. Finally, they reflected, revised, and improved the formulated standards-setting appraisal of competency level-based performance for instructors in the higher education institution. Figure 1 elucidates the research procedure.

Research Participants

The main data source was derived from a total of 603 instructors' performance assessment results from three clusters of educational programs, namely, science and technology, health sciences, and humanities and social sciences of a public university in Khon Kaen province, Thailand. In addition, key assessors' views were considered while researchers were formulating the standards-

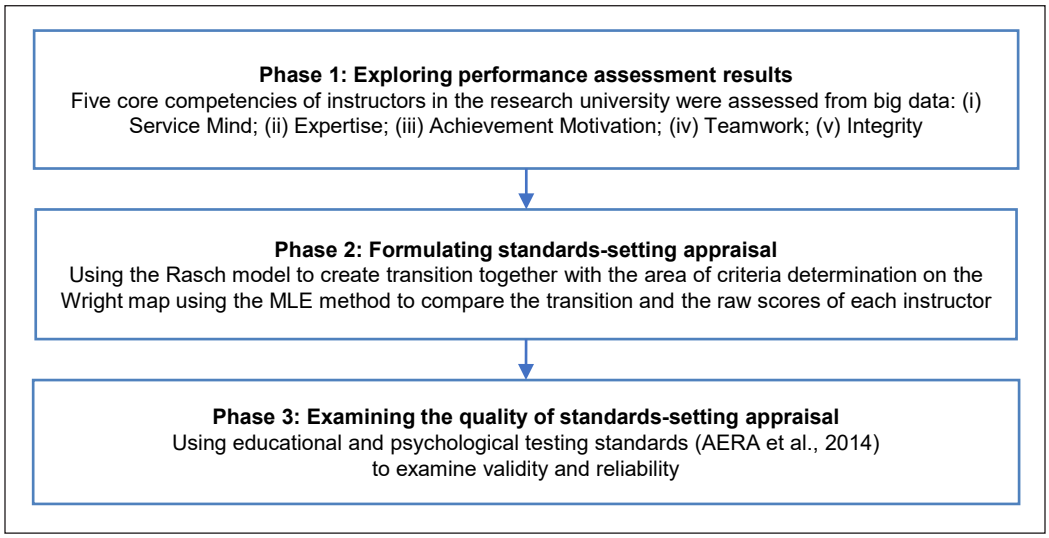

Figure 1. Research procedure

setting appraisal of competency level-based performance for instructors. They were the dean or associate dean from each cluster of educational programs, the director of human resources, and the research university's vice president for Education and Academic Services.

Research Tool

The researchers adopted an official performance appraisal form for civil servants in public higher education institutions from the Office of the Civil Service Commission (2009). The research tool is comprised of five core competencies being evaluated, namely good service (Service Mind: SERV), accumulated expertise in a professional career (Expertise-EXP), achievement and motivation (ACH), teamwork (TW), and adhering to righteousness and ethics (Integrity-ING). These five core competencies are categorized into five competency levels in descending order and used as a rubric with descriptions. Table 1 shows an example of the Expertise-EXP performance assessment rubric. At the same time, Figure 2 illustrates the five core competencies included in the formulated standards-setting appraisal for higher education instructors' measurement model.

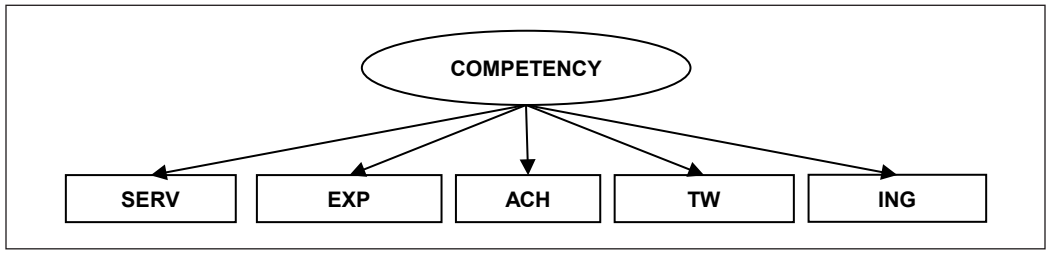

Figure 2. The formulated standards-setting appraisal for higher education instructor measurement model

Table 1
Example of the expertise-EXP performance assessment rubric

Core competency	Accumulated expertise in a professional career (Expertise–EXP)			
Competency definition	Accumulating knowledge and abilities through continuous learning, conducting research, and self-development until applying the learned knowledge and expertise to perform job tasks.			
Competency descriptions				
Level 0: Does not perform at all or unclearly displayed				
Level 1	Level 2	Level 3	Level 4	Level 5
Show interest and pursue new knowledge in their professional or related fields. 1.1 Interested in new knowledge. 1.2 Improve their knowledge and abilities. 1.3 Update new knowledge from various sources.	Demonstrate competency level 1 in their professional or related fields. 2.1 Well-versed in new knowledge that can affect their performance. 2.2 Aware of the cutting-edge technology trends and continuously related to their job tasks.	Demonstrate competency level 2 in their professional or related fields. 3.1 Ability to apply new knowledge. 3.2 Can solve problems by adopting their new knowledge to their job tasks.	Demonstrate competency level 3 deeply in their professional or related fields. 4.1 Have expertise in interdisciplinary subject matter and can apply knowledge in various applications. 4.2 Can apply their integrated knowledge to create a vision for future operations.	Demonstrate competency level 4 by giving support and emphasizing expertise. 5.1 Promoting expertise development through the organization of resources, tools, and equipment. 5.2 Administering new knowledge to perform job tasks continuously.
Criteria for performance evaluation (pass not less than 3 assessment criteria)				
CL1.1: Can perform job tasks without analyzing data and have been performed before.	CL2.1: Pass the assessment criteria of Level 1 and possess the three related behaviors.	CL3.1: Pass the assessment criteria of Level 2 and possess the three related behaviors.	CL4.1: Pass the assessment criteria of Level 3 and possess the three related behaviors.	CL5.1: Pass the assessment criteria of Level 4 and possess the three related behaviors.
CL1.2: Study for knowledge according to supervisor's order.	CL2.2: Follow the related knowledge and be responsible for improvements.	CL3.2: Participate in knowledge development to solve the team's problems.	CL4.2: Use expertise and experience for personnel development.	CL5.2: Promote competence and expertise development at various levels.
CL1.3: Focus on related knowledge but unable to apply the learned knowledge.	CL2.3: Use the related knowledge and apply it in their job tasks.	CL3.3: Presenting new projects beneficial to the job tasks and department.	CL4.3: The research results or innovations are useful for improving the work system.	CL5.3: Promote and support expertise development for personnel within the organization.
CL1.4: Can apply related knowledge to explain their job tasks.	CL2.4: Can apply related knowledge and methods of others to improve their job tasks.	CL3.4: Can apply related knowledge in preparing research for work development.	CL4.4: Being representative to overview the performance of the department.	CL5.4: Performing the knowledge and expertise and being an example of good practice to the department.

Table 1 *(continue)*

Criteria for performance evaluation (pass not less than 3 assessment criteria)				
CL1.5: Applying the new knowledge as recommended by the supervisor.	CL2.5: Self-development to become proficient in their professional career.	CL3.5: Present the research results and for them to be used by their supervisor to make decisions.	CL4.5: Being representative to present in internal and national seminars.	CL5.5: Performing knowledge development and developing guidelines for good performance within the department.

Data Collection and Data Analysis

The researchers utilized the instructors' performance assessment results from the big data with more than two scores using the Partial Credit Model to collect data for the initial phase. The Partial Credit Model is a way of comparing and calibrating items. In applying Rasch's model to instructor performance assessment, every item has an imagined location on the measured variable (Masters, 2005). The quality of the instructor performance assessment results was analyzed using the Rasch model analysis ACER Conquest 2.0 program (Wu et al., 2007). When test data conform to the Rasch model, the relative difficulties of two items can be estimated by using any group of people without regard to their abilities or other characteristics. Making multiple pairwise comparisons of this kind makes it possible to estimate the relative locations of several items on the measurement variable (Masters, 2005). In short, Rasch analysis is an approximation characterized as a logistic function by only considering the ability (θ) and the difficulty value.

On the other hand, Maximum Likelihood Estimation (MLE) was used in this research because it is a method for estimating the parameters of a probability distribution by maximizing a likelihood function so that under the presumed statistical model, the examined data is the most possible. According to Rossi (2018), the point in the parameter space that maximizes the likelihood function is called the MLE. The logic of maximum likelihood is instinctive and adaptable, and the method has become a dominant means of statistical inference (Ward & Ahlquist, 2018). Therefore, the main principle of MLE is that the examined data are generated by randomization through selection from the population with a distribution based on one specific parameter value. As a result, the researchers could maximize the likelihood of being able to select such samples randomly.

The researchers validated the quality of the formulated standards-setting appraisal measurement model by considering its validity and reliability. The quality of the measurement model was tested using educational and psychological testing standards (American Educational Research Association [AERA] et al., 2014). The first evidence of validity is that the test content coverage must be comprehensively included with regard to all the core competencies

of higher education instructors and the levels of such competency. In addition, the researchers continued to examine the reliability of the measurement model using the Expected-A-Posteriori and Separation Reliability (EAP/PV) which is a measurement of the consistency of the Rasch analysis and is equal to Cronbach's alpha in terms of precision. The method of analysis was for this research because the researchers intended to examine not only the validity and suitability of the standards-setting appraisal measurement model in terms of whether it met the acceptable criteria according to the testing standards but also to identify the evidence for the validity and reliability of the model. Finally, the threshold values were used to estimate the quality of the model.

RESULTS AND DISCUSSION

Results of Determination of the Cut-Off Point in Assessing Instructors' Competency Level

The average performance threshold of each core competency was used to formulate a standards-setting appraisal measurement model. The researchers formulated the assessment standards by calculating the transition and considering the criteria area on the Wright map for each core competency, which was determined by the average threshold at the same level for the five core competencies. Table 2 demonstrates the results of calculating the transition in each core competency to determine the effectiveness of the standards-setting appraisal measurement model.

The results of the determination of the cut-off point in assessing the core competencies of instructors' performance assessment from the big data revealed that the transition in performance assessment could be divided using four cut-off points to create five levels in ascending order, specifically -11.67, -2.68, 4.59, and 9.76. Figure 3 shows the use of the Wright map to determine the transition point by setting the criteria area so that researchers can compare instructors and items, to understand better how appropriately the performance assessment approach measured the core competencies (Lunz, 2010). It leads to formulating the standards-setting appraisal in the second phase.

Table 2
Results of determination of the cut-off point in assessing core competencies of instructors

Core competency	Threshold			
	1	2	3	4
Service Mind	-11.71	-5.33	5.90	10.65
Expertise	-7.98	1.54	6.17	10.75
Achievement Motivation	-13.34	1.55	5.91	10.41
Teamwork	-11.47	-5.54	2.51	7.29
Integrity	-13.87	-5.60	2.48	9.68
Mean of thresholds for standards-setting	-11.67	-2.68	4.59	9.76

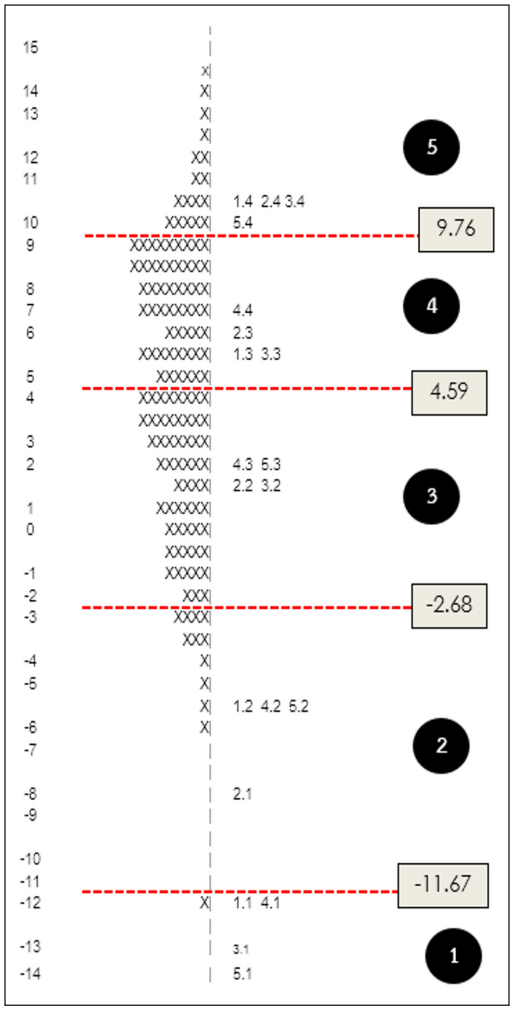

Figure 3. Determining the transition point by setting area criteria on the Wright map

When the researchers considered the transition point, it was found that the descending order of the average threshold was reasonable. It is important to note that the nature of the transition can be determined by considering the average threshold for each competency level over a wide range. Typically, in the latent trait (θ) parameter estimation, the normal range is -3 to +3 (Baker & Kim, 2017). Conceptually, some researchers have set the range θ from -4 to +4 (Embretson & Reise, 2000; Junpeng et al., 2020). However, in this research, the distribution range θ went from -0.85 to 13.24 with a threshold of -13.87 to 10.75, as shown in Table 2. When a comparison is made between thresholds at Level 1, the five core competencies assessment items are in the very low range of -13.89 to -7.58. Therefore, when calculating the transition using the mean of the five thresholds, the threshold value is -13.89 to -7.58. Nevertheless, it was found that there was a low point of -11.67 on the left-hand side of the Wright map. It was found that there was no instructor with parameters lower than -11.67, which reflected the transition points in Level 1. The result showed no instructors at that level because the actual θ value started at -0.85. It can be concluded that instructors' competency levels can be assessed in real-world contexts should the transition be greater than or equal to -0.85.

Furthermore, the results of the assessment criteria indicated that there are five levels of instructors' performance. Therefore, instructors with the lowest competency level will be terminated if they do not improve their performance. It is because instructors in the research university are required to improve their job performance in accordance with the criteria set by the government of Thailand. In other words, instructors are considered an asset for higher education and a driving force of the learning process, as well as key elements in determining the success of higher education (Anwar et al., 2017).

Results of Determination of Instructors' Performance Assessment Standards

The researchers continued to design the standards-setting appraisal approach for instructors' performance based on the assessment criteria results from the first phase. Hence, the researchers identified five score ranges converted from estimation competency parameters into the scale and raw scores. The results showed that instructors with the lowest competency level need to improve their performance urgently, given that they revealed a performance level lower than -11.67 or with a scale score lower than -66.70. Subsequently, the researchers used this range to compare with the raw score from 0 to 1 point. Likewise, if the instructor has a score of θ higher than 9.760 or has a scale score higher than 147.60, the instructor is considered to possess the highest competency level. In other words, the raw scores range from 18 to 20 points. It can be concluded that the highest competency level is the best practice and a role model for other instructors. Table 3 details the determination of instructors' performance assessment standards.

The above results show that five intervals correspond to the scale and raw score range while the researchers formulated the standards-setting appraisal measurement model. As a result, the formulated model successfully indicates the real context of 603 instructors from three clusters of educational programs. It provides an overview of the instructors' competence in the research of higher education institutions (Figure 4). The overall results revealed that most instructors

Table 3
Results of determination of performance assessment standards

Competency level	Transition θ	θ range	Scale scores	Raw scores
5	9.76	>9.76	>147.60	18-20
4	4.59	4.59<θ≤9.76	95.90 – 147.60	13-17
3	-2.68	-2.68<θ≤4.59	23.20 – 95.80	8-12
2	-11.67	-11.67<θ≤-2.68	-66.70 – 23.10	2-7
1		<-11.67	<-66.70	0-1

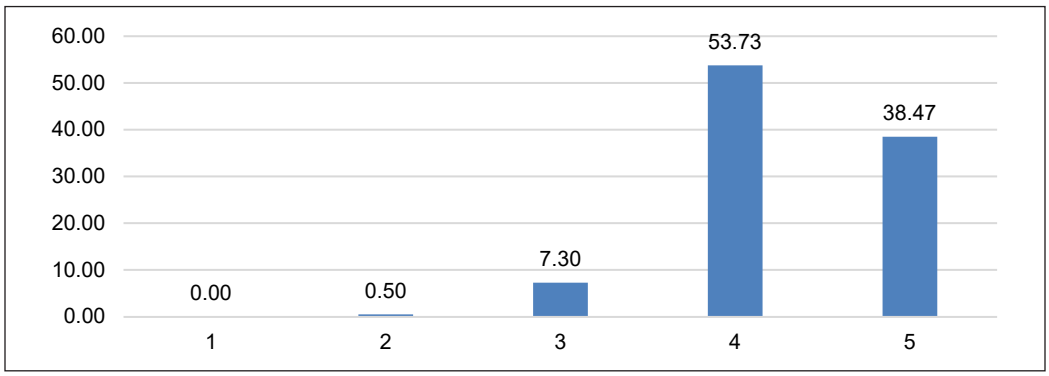

Figure 4. Overall results of 603 instructors' performance assessment

(53.73%) are at Level 4. It is followed by 38.47% of instructors who are at Level 5. Only 7.80% of the 603 instructors have a competence level at Level 3 or lower.

Figure 5 to Figure 7 elucidate the situation specifically in terms of the three clusters, namely, science and technology, health sciences, and humanities and social sciences. The results indicate that instructors from the science and technology cluster possess more outstanding competency levels compared to the other two clusters. Figure 5 shows that most instructors from science and technology (74.36%) are at Level 5 (Figure 5), while the majority of instructors from humanities and social sciences (64.36%) and health sciences (53.78%) are at Level 4 (Figures 6 and 7). It reflects the current trend in higher education whereby academics in science and technology have made significant contributions. For example, science has given an immense body of knowledge, while technology has made education easier

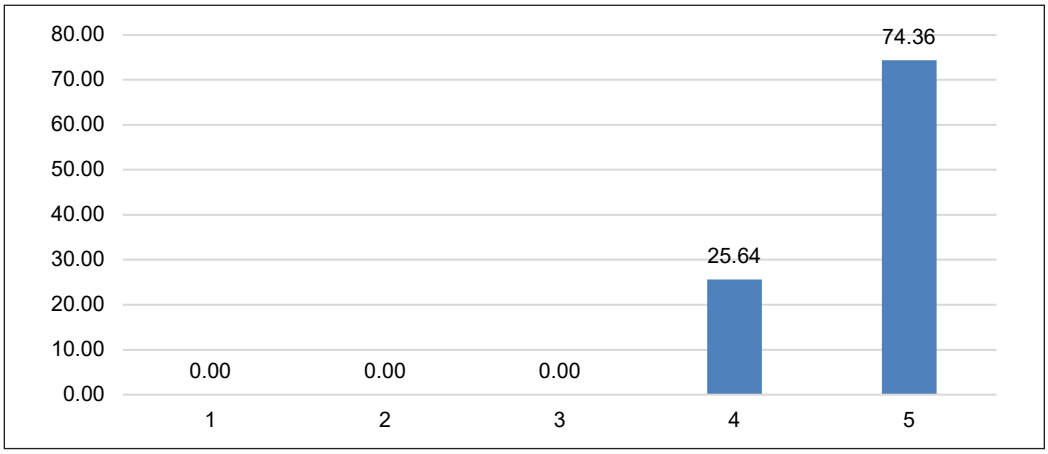

Figure 5. Performance assessment results of instructors from the Science and Technology cluster

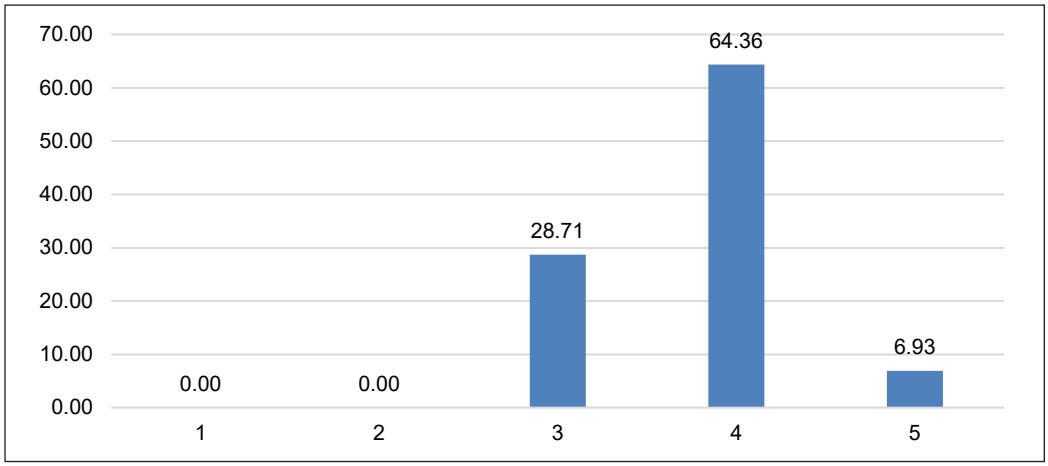

Figure 6. Performance assessment results of instructors from the Humanities and Social Sciences cluster

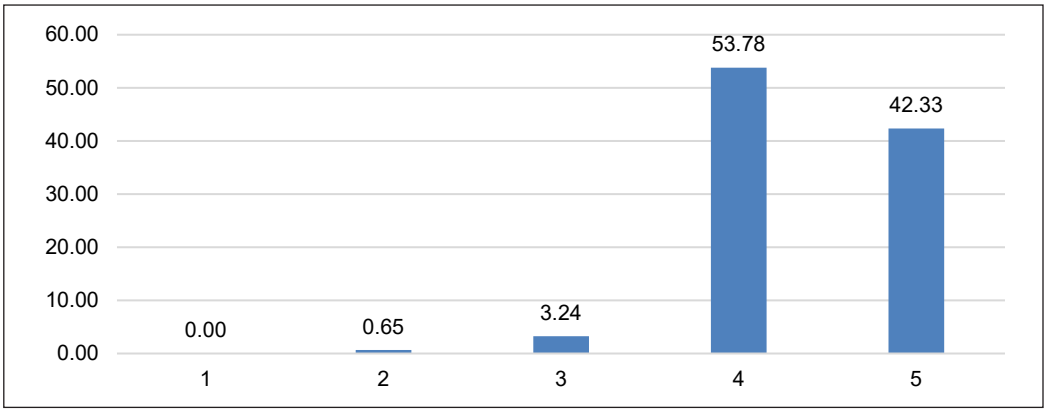

Figure 7. Performance assessment results of instructors from the Health Sciences cluster

through smart classes, multimedia devices, e-libraries, and e-books.

Results of Quality Inspection of Standards-Setting Appraisal Measurement Model

Results of Validity Evidence. The internal structural validity of the standards-setting appraisal measurement model was determined using a Wright map with a cascaded manner of scoring according to the milestone grading guidelines, from lowest to highest level (Likelihood Ratio Chi-Squared, $G^2 = 4295.31$, Akaike Information Criterion, AIC = 4337.31, Bayesian Information Criterion, BIC = 4489.27). In addition, the standards-setting appraisal measurement model is accurate and consistent in assessing instructors who possess moderate and high competency levels because all items of the measurement model have covered all the necessary competency levels. Moreover, the researchers considered whether the transition should be collapsed in terms of the assessment standards at Levels 1 and 2 or whether the position of the transition point should be modified by adjusting the assessment level from five to four levels.

The internal structural integrity check results showed that the standards-setting appraisal measurement model is not suitable for annual assessment for salary advancement or academic position promotion. Rather, it is more suitable for use as a criterion for evaluating individual work performance. As a result, this measurement model can be used for actual assessment after actions such as collapsing Levels 1 and 2, modifying the expected value, and defining the key indicators of the expected performance level.

The final validity evidence was found when there were significant correlations between the formulated assessment standards and the external criteria by using the actual assessment work achievement in the real context through their supervisors. Table 4 shows significant correlations between the Wright map score and the actual assessments because the r values ranged from 0.21 to 0.55 at a significant level of 0.01. It implies that most instructors' core

Table 4
Results of correlation between formulated standards assessment and actual assessment

Clusters	Correlations between Wright map scores and work achievement scores by their supervisors
Humanities and Social Sciences	0.55**
Science and Technology	0.47**
Health Sciences	0.21**
Overall	0.27**

competencies are in the same direction as the standards-setting appraisal measurement model. This result suggests that Level 1 and Level 2 can be collapsed into one, thus having four competency levels.

Results of Reliability Evidence. The researchers began by using the standard deviation graph SEM (standard error of measurement) to assess the reliability of the standards-setting appraisal measurement model. The results showed that the standards-setting appraisal measurement model has the same characteristics as those obtained from the actual assessments, as illustrated in Figure 8.

The results suggest that the standards-setting appraisal measurement model is more appropriate for instructors with moderate to high competency levels than those with low ones. It is because the lowest competency level of instructors showed the highest error in terms of SEM value. Table 5 demonstrates that the mean score at a high

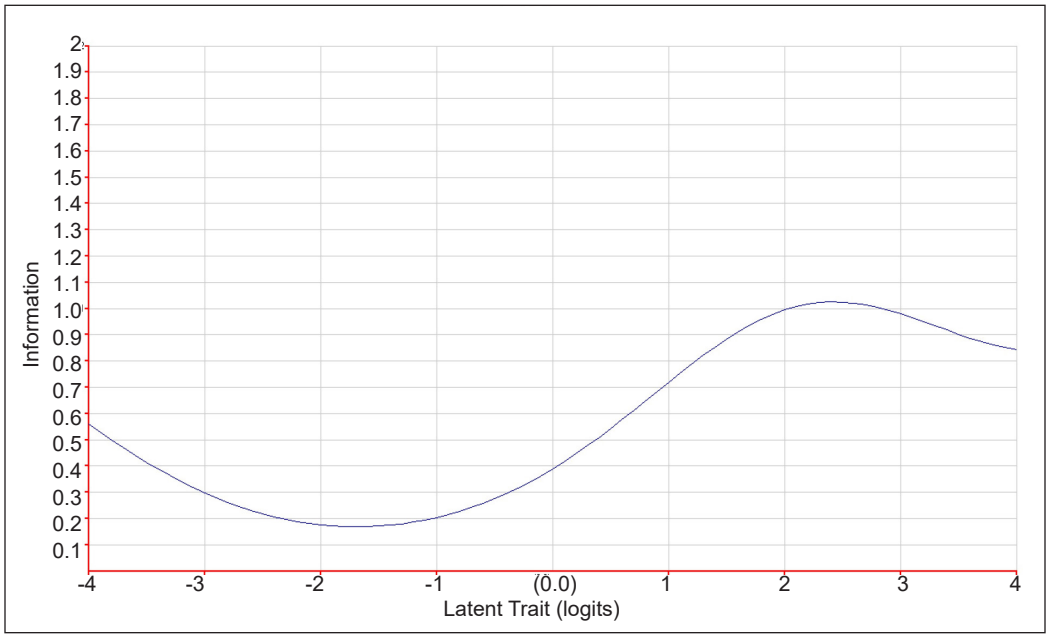

Figure 8. Test information function to reflect the consistency of standards-setting appraisal measurement model

competency level is 4.75, with a θ value of -0.85 to 13.24. It implies that instructors have a wide range of competencies and tend to perform well. Moreover, the SEM value indicated that the errors in estimating instructors' competency levels were between 0.24 and 1.75, considered a small discrepancy when the measurement model was applied to instructors with moderate to high competency levels. On the other hand, the SEM value was high, or there was a lack of consistency in the estimation when instructors possess a low level of competence. It can be concluded that the standards-setting appraisal measurement model is accurate and consistent with those instructors who possess moderate to high competency levels. Table 5 shows the details of the SEM results.

Table 5
Results of basic values of SEM

Basic statistical value	θ	SEM
Mean score	4.57	1.16
Standard deviation	4.15	0.26
Maximum	13.24	1.75
Minimum	-0.85	0.24

The EAP/PV and separate reliability values were found to be 0.94 and 0.99, respectively, indicating that it is of high quality and can be used to classify the competency levels of performance assessment separately. Finally, researchers examined the suitability of each item of the measurement model using the INFIT MNSQ value. According to Adam and Khoo (1996) and Wilson et al. (2006), the INFIT MNSQ value should be between 0.75 and 1.33. The researchers obtained the INFIT MNSQ value of the measurement model and found it in the range of 0.84 to 1.06. Consequently, it can be concluded that each item in the standards-setting appraisal measurement model is suitable.

CONCLUSION

The main aim of this research is to formulate and assess the quality of a standards-setting appraisal measurement model in terms of validity and reliability developed to evaluate the instructors' work performance in a public university in Thailand. As we know, the general aims of the standard-setting appraisal measurement model are to assess the core competencies of instructors, given that these are regarded as being crucial to the performance management process (Aliasghar et al., 2017; Le, 2021; Tan et al., 2019). Therefore, this model will provide a sound performance evaluation tool because it has undergone a comprehensive and scientific research procedure to address instructors' five core competencies identified by setting criteria on a Wright map from big data. To this end, the results of calculating the transition in each core competency to determine the standards-setting appraisal measurement model indicated that instructors' competency levels had been assessed in a real-world context. The results are in parallel with those of past researchers' (Blašková & Blaško, 2012; Islam et al., 2019; Le, 2021; Prasetio et al., 2017; Turturean, 2013).

Moreover, this model can contribute significantly to the educational measurement

and evaluation field because it has raised the opportunity to address significant concerns by drawing on the basic principles of performance appraisal and management, thus establishing itself in the human resources literature, as emphasized by Lohman (2021). It corresponds to Molefe's (2010) ideas that the priorities with regard to some core competencies, such as 'subject mastery' and 'research' were perceived as significantly more important than 'change management' and 'project management' in their standard-setting appraisal measurement model of the five best higher education institutions in South Africa, the United States of America, the United Kingdom, Australia, and Nigeria.

The initial results with regard to the determination of the cut-off point in assessing core competencies of instructors from the big data revealed that the transition in performance assessment could be divided into four cut-off points relating to five levels in ascending order, specifically -11.67, -2.68, 4.59, and 9.76, as derived from the Wright map. The use of the Wright map to determine the transition points by setting criteria areas has been supported by Lunz (2010), which helps to compare instructors and items. Following this reasoning, the researchers confirmed that performance assessment could be measured by investigating the core competencies while instructors carry out their job tasks. Moreover, the initial results also show that, in this study, there were no instructors whose performance evaluation was below -0.85 in the actual context. Hence, the initial results seem to fulfill the criteria set by the Thai government, as those instructors who cannot perform to this level have to be terminated. It implies that higher education instructors are valuable assets and an essential driving force when upgrading the quality of teaching in higher education institutions, as emphasized by Anwar et al. (2017) and Islam et al. (2019).

In the second phase, the researchers formulated the standards-setting appraisal measurement model after identifying the score ranges converted from estimation competency parameters into the scale and raw scores. It was followed by trying out the standards-setting appraisal measurement model in the third phase. The results of the third phase indicated that most instructors from the research university are at Level 4 (53.73%) and Level 5 (38.47%). It implies that 92.20% of the 603 instructors surveyed are assessed as being highly competent. However, the science and technology cluster instructors are more highly competent than the instructors in the other two clusters. Even though instructors from humanities (64.36%) and social sciences and health sciences (53.78%) demonstrate having lower competency levels, the instructors' core competencies and positive behavior still enhance student learning in terms of new insights and capabilities (Prasetio et al., 2017).

In the final phase of this research, the developed standards-setting appraisal measurement model has undergone a thorough quality inspection in terms of its validity and reliability. The researchers provided three pieces of evidence regarding the validity: (2) internal structural validity,

(2) an internal structural integrity check, and (3) correlations between the formulated assessment standards and actual assessment in the real context. The results in terms of internal structural validity were determined using a Wright map (G^2 = 4295.31; AIC = 4337.31; BIC = 4489.27). In addition, the results imply that this measurement model can be used for actual assessment after conducting the following actions: collapsing Levels 1 and 2, modifying the expected value, and defining the key indicators of the expected performance level. The final validity evidence was the significant correlations between the Wright map score and actual assessment ($0.21<r<0.55$), which indicates a 0.01 significant level. It implies that most instructors' core competencies are in the same direction as the standards-setting appraisal measurement model. Last but not least, the SEM was used to assess the reliability of the standards-setting appraisal measurement model and showed that it has the same characteristics as those obtained from the actual assessment of work achievement. In conclusion, the standards-setting appraisal measurement model can provide rich and accurate information regarding instructors with moderate to high competency levels in their work performance (Curzi et al., 2019; Gómez & Valdés, 2019).

ACKNOWLEDGMENT

The authors gratefully acknowledge the use of service and facilities of the Faculty of Education, Khon Kaen University, Khon Kaen 40002, Thailand. The contents of this manuscript are derived from the first author's doctoral dissertation, thus fulfilling the Ph.D. requirement of Khon Kaen University.

REFERENCES

Adam, R., & Khoo, S. (1996). *Quest: Interactive test analysis system version 2.1. The Australian council for educational research.* https://research.acer.edu.au/cgi/viewcontent.cgi?article=1002&context=measurement

Aliasghar, M., Seyed, A. H., & Kamran, M. K. (2017). A model for competency assessment of the faculty in Islamic Azad University. *International Education Studies, 10*(3), 17-23.

Alkhafaji, S. (2013). Instructor's performance: A proposed model for online evaluation. *International Journal of Information Engineering and Electronic Business, 4*, 34-40. https://doi.org/10.5815/ijieeb.2013.04.05

Altbach, P. G., Reisberg, L., & Rumbley, L. E. (2009). *Trends in global higher education: Tracking an academic revolution*. UNESCO Publishing.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing* (6th ed.). American Educational Research Association.

Anwar, M., Chandrarin, G., Darsona, J. T., Respati, H. (2017). Lecturer job performance study: Motivation, emotional intelligence, organizational culture, and transformational leadership as antecedents with job satisfaction as an intervening. *IOSR Journal of Business and Management, 19*(6), 1-9.

Baker, F. B., & Kim, S. (2017). *The basics of item response theory using R.* Springer.

Blašková, M., & Blaško, R. (2012). Dimensions and attributes of the university teacher quality. In

Human potential management in a company (pp. 32-43). University of Matej Bel.

Blašková, M., Blaško, R., Jankalová, M., & Jankal, R. (2014). Key personality competencies of university teacher: Comparison of requirements defined by teachers and/versus defined by students. *Procedia – Social and Behavioral Sciences, 114*, 466-475.

Bowen, D. E., & Ostroff, C. (2004). Understanding HRM-firm performance linkages: The role of the 'strength' of the HRM system. *The Academy of Management Review, 29*, 203-221. https://doi.org/10.5465/amr.2004.12736076

Cardoso, S., Tavares, O., & Sin, C. (2015). The quality of teaching staff: Higher education institutions' compliance with the European standards and guidelines for quality assurance – The case of Portugal. *Educational Assessment Evaluation and Accountability, 27*(3), 205-222. https://doi.org/10.1007/s11092-015-9211-z

Caruth, D. L., & Humphreys, J. H. (2008). Performance appraisal: Essential characteristics for strategic control. *Measuring Business Excellence, 12*(3), 24-32. https://doi.org/10.1108/13683040810900377

Curzi, Y., Fabbri, T., Scapolan, A. C., & Boscolo, S. (2019). Performance appraisal and innovative behavior in the digital era. *Frontiers in Psychology, 10,* 16-59. https://doi.org/10.3389/fpsyg.2019.01659

Embretson, S. E., & Reise, S. P. (2000). *Item response theory of psychologists.* Erlbaum.

Gómez, L. F., & Valdés, M. G. (2019). The evaluation of teacher performance in higher education. *Propósitosy Representaciones, 7*(2), 479-515. https://doi.org/10.20511/pyr2019v7n2.255

Hosain, M. S. (2016). Teaching workload and performance: An empirical analysis on some selected private universities of Bangladesh. *International Journal of English and Education, 5*(3), 1-11.

Islam, R., Haidoub, I. M., & Tarique, K. M. (2019). Enhancing quality of education: A case study on an international Islamic university. *Asian Academy of Management Journal, 24*(1), 141-156.

Junpeng, P., Marwiang, M., Chiajunthuk, S., Suwannatrai, P., Chanayota, K., Pongboriboon, K., Tang, K. N., & Wilson, M. (2020). Validation of a digital tool for diagnosing mathematical proficiency. *International Journal of Evaluation and Research in Education*, *9*(3), 665-674. https://doi.org/10.11591/ijere.v9i3.20503

Khon Kaen University. (2015). *Criteria and methods for evaluating the performance of personnel.* Khon Kaen University printing house.

Kurtulmus, B. E., Warner, B., & Özari, Ç. (2016). Research or teaching oriented? Game theory models for the strategic decision-making process of universities with the external environment held neutral. *Electronic Journal of Applied Statistical Analysis, 9*(3), 469-490.

Laei, S., Abdi, A., Karamaerouz, M. J., & Shirkhani, N. (2014). Instructors' evaluation as an instrument to improve performance and determine competence. *Universal Journal of Educational Research, 2*(2), 110-118. https://doi.org/10.13189/ujer.2014.020202

Le, L. C. (2021). Assessing lecturer competence: A case study of public universities in Ho Chi Minh City. *Academy of Strategic Management Journal, 20*(S2), 1-12.

Lohman, L. (2021). Evaluation of university teaching as sound performance appraisal. *Studies in Educational Evaluation, 70*, Article 101008. https://doi.org/10.1016/j.stueduc.2021.101008

Lunz, M. E. (2010). *Measurement research associates test insights.* https://www.rasch.org/mra/mra-01-10.htm

Masters, G. N. (2005). Partial credit model. *Encyclopedia of Social Measurement, 7-17. https://doi.org/10.1016/BO-12-369398-5/00462-X*

Molefe, G. N. (2010). Performance measurement dimensions for lecturers at selected universities: An international perspective. *SA Journal of Human Resource Management, 8*(1), Article 243. https://doi.org/10.4102/sajhrm.v8i1.243

Nongna, C., Junpeng, P., Hong-ngam, J., Podjana, C., & Tang, K. N. (2021). Creating core competencies and workload-based outcome indicators of university lecturers' performance assessment: Functional analysis. *Journal of Education and Learning, 10*(6), 82-91. https://doi.org/10.5539/jel.v19n6p82

Odero, J. A., & Makori, E. M. (2018). Employee involvement and employee performance: The case of part time lecturers in public universities in Kenya. *International Journal of Management and Commerce Innovations, 5*(2), 1169-1178.

Office of the Civil Service Commission. (2009). *Guide to core competencies.* P. A. Living.

Office of the Higher Education Commission. (2018). *Guidelines for enhancing the quality of teaching and learning management of instructors in higher education institutions*. Prints.

Prasetio, A. P., Azis, E., Fadhilah, D. D., & Fauziah, A. F. (2017). Lecturers' professional competency and students' academic performance in Indonesia higher education. *International Journal of Human Resource Studies, 7*(1), 86-93.

Reeves, T. C. (2006). Design research from a technology perspective. In J. V. D. Akker, K. Gravemeijer, S. McKenney, & N. Nieveen (Eds.), *Educational design research* (pp. 52-66). Routledge.

Rossi, R. J. (2018). *Mathematical statistics: An introduction to likelihood-based inference*. John Wiley & Sons.

Selvi, K. (2010). Teachers' competencies. *Cultura International Journal of Philosophy of Culture and Axiology, 7*(1), 167-175. https://doi.org/10.5840/cultura20107133

Tan, S., Lau, E., Ting, H., Cheah, J. H., Simonetti, B., & Tan, H. L. (2019). How do students evaluate instructors' performance? Implication of teaching abilities, physical attractiveness and psychological factors. *Social Indicators Research, 146*, 61-76. https://doi.org/10.1007/s11205-019-02071-6

Turturean, M. (2013). University trainees' key competencies – a global profile. *Procedia – Social and Behavioral Sciences, 76*, 801-805.

Vongvanich, S. (2020). *Design research in education.* Chulalongkorn University Printing House.

Ward, M. D., & Ahlquist, J. S. (2018). *Maximum likelihood for social science: Strategies for analysis.* Cambridge University Press.

Wilson, M., Allen, D. D., & Li, J. C. (2006). Improving measurement in health education and health behavior research using item response modelling: Comparison with the classical test theory approach. *Health Education Research, 2*(1), i19-i32.

Wu, M. L., Adams, R. J., Wilson, M. R., & Haldane, S. A. (2007). *ACER Con Quest 2.0: Generalized item response modelling software.* ACER Press.