

DEVELOPMENT AND VALIDATION OF INSTRUMENTATION TO ASSESS UNIVERSITY ACADEMICS' RESEARCH AND TEACHING PERFORMANCE IN PUNJAB, PAKISTAN

Aziz Ahmad and Muhammad Saeed

*Department of Advanced Studies in Education (DASE), Institute of Education and Research,
University of the Punjab, Lahore, Pakistan.*

ABSTRACT

The study was aimed to develop and validate instruments of academic performance in terms of research and teaching performance of university teachers having foreign and domestic doctoral education from an internationalization perspective. The study background was to develop and validate these instruments through literature, content validation, reliability, and validity estimates before commencing a large-scale Ph.D. study of assessment of academic practices of academics. As research is scarce in this area in Pakistan and consequently, the comprehensive measures to capture research and teaching performance of academics are also minimal. Adcock and Collier (2001) standard shared framework for instrument development was followed to devise research and teaching instruments. The validation process included expert panel review to refine instruments and ensure content validity. Moreover, composite factor analysis was conducted for the teaching-related instrument to ensure construct reliability and construct validity using SmartPLS 3. The 'academic research performance instrument (ARPI) was finalized through content validation across three dimensions of academic research. The second instrument, 'students' evaluation of teaching quality (SETQ), comprised seven teaching-related dimensions and was found to be reliable and valid after assessing psychometric properties. Both instruments may be used to evaluate the research and teaching performance of academics with foreign PhDs concerning domestic PhDs in the context of social and physical sciences. The instruments and resultant information may also further inform the extent to which the internationalization of research and teaching is taking place at universities.

Keywords. Content validation, foreign returnees, higher education

Introduction

University academics are required to perform several roles and functions to support and achieve institutional objectives and missions. Among them, the research, teaching, and service roles are the most expected. Globally, quality universities prioritize knowledge production and knowledge dissemination as embedded in their mission statements (Cadez et al., 2017). More indicative is the universities' effort to attract competent academics and introduce favorable research policies to produce research and win grants in favor of upgrading their ranking and image (Coggburn & Neely, 2015; Douglas, 2013; Mazzucchelli et al., 2018; Shin & Cummings, 2010; Ter-Bogt & Scapens, 2012). Moreover, research in a particular academic field advances the knowledge base and leads to academic awards and honors (Hardre et al., 2011; Smith et al., 2013). Therefore, expertise is considered a critical factor in knowledge production and can create enriched instructional opportunities for students (Griffin et al., 2018).

Both components of (research & teaching) involve numerous activities and practices that require to be delivered strategically and competently. Research and teaching have become the essential part of a globalized competitive system partly based on evaluating academic competencies (Thornton et al., 2018; Jenkins et al., 2007). In this context, scholarly productivity has gained significant attention among stakeholders in knowledge-based societies (Shin & Cummings, 2010; Webber, 2011). Policymakers and administrators are crafting better research measures and factors that best represent essential scholarly products (Allen et al., 2018).

Recently a body of research has been rapidly growing, focusing on returnee scholars' academic contributions. This growing body of research may be partly attributed to amidst increase of return migration to countries of origin to gauge knowledge and technology transfer, reintegration process, and academic performance. In Pakistan, investigations involving returned academics are somehow a neglected area. Recently, few studies have focused on individual dimensions such as research performance (Baloch et al., 2020) and reverse cultural shock (Aktar et al., 2018). Still, overall, research is scarce on a scholarly investigation to gauge the performance of academic returnees in different dimensions. Consequently, instruments to gauge different dimensions of academic performance in the Pakistani university context are also almost non-existent. Therefore, to fill this gap and bring new insights, the current study is designed to develop and validate instruments' regarding academics' performance, especially in research and teaching domains, before conducting a large-scale Ph.D. study concerning the assessment of academic performance. In this regard, different indicators and factors are included during the development process to gauge academic research and teaching performance to obtain a more enriched snapshot of performance.

In this context, the purpose of this research is to develop the measures of academic performance in the domains of research and teaching quality in Punjab, Pakistani context that may be applicable across social and natural sciences. So that academic performance in these domains could be gauged mainly to compare the academic performance of foreign returnees having doctoral education with their domestic counterparts. Initially, Adcock and Collier (2001) standard shared framework for instrument development was followed to devise instruments. This model carefully considers that indicators and resulting scores greatly represent the essence of concepts formulated in the systematized definition (Adcock & Collier, 2001). Further, the development process was informed and refined by guidance produced by leading experts in the field to ensure that instruments are valid in various dimensions and able to elicit the required information.

In this study, content validity is greatly emphasized as an initial study. Content validity is an important aspect and provides evidence regarding the degree to which the instrument's elements are relevant and representative of the construct under question in connection with the purpose of the instrument (Almanasreh et al., 2019). Content validity plays a central role in the development of any instrument and prerequisite for evaluating other validity pieces of evidence (Slocumb & Cole, 1991). This validation mostly requires a panel of experts to review the constructed instruments to judge relevance, representativeness, and other essential aspects. In the current study, all the measures along with objectives and operational definitions of constructs were sent to a panel of international experts (10 experts) specialized in higher education through emails for content validation of the instruments. They were requested to review and provide feedback regarding instruments and items' relevance, representativeness, clarity, and overall alignment. Based on the invaluable experts' feedback, instruments were modified to be more relevant, clear, and representative. The study was conducted to prepare the instruments related to academic performance in the domains of research and teaching of university teachers having foreign and domestic doctoral education. More specifically, the following objectives were addressed:

- To prepare an instrument to assess the research performance of university academics associated with general public-sector universities.
- To prepare an instrument to assess the teaching performance of university academics as perceived by students

Literature Review

Academic performance may be assessed to include several academic domains, such as research, teaching, services both internal and external, and contribution to society, to name a few. Eventually, performance may be determined based on single or combining multiple areas for various purposes in the given context. As the current research includes academic domains of research and teaching, literature regarding these domains is presented to develop or identify factors to measure research and teaching performance keeping in view research objectives.

Firstly, research performance is judged on various dimensions based on several indicators ranging from research in different recognized publication outlets, funded projects/studies, presentations (Altbach, 2015; Cadez et al., 2017; Iqbal & Mahmood, 2011; Porter & Umbach, 2001; Shin, 2011), memberships to various academies (White et al., 2012), nature of research collaborations to research dissemination in top tier journals, citations and overall impact (Harvey et al., 2010; Long et al., 2009). Research activities also include gathering and analyzing data, supervising postgraduate students and their class projects, obtaining patents and licenses (Nafukho et al., 2019), getting research grants, performing editorial duties (Creswell, 2012), and engagement in various academic activities (Hug et al., 2013). Abramo and D'Angelo (2014) asserted that research activity is a production process in which the inputs consist of human, tangible, and intangible resources, and where output, in this case, comprises the new knowledge. Knowledge production has a complex character of both tangible natures (publications, patents, conference presentations, databases, etc.) and intangible nature (tacit knowledge, consulting activity, etc.). Among these indicators, some are given more importance for different purposes in judging the research performance at various levels and in different contexts (Zhang & Shin, 2015).

Moreover, Wilder and Walters (2018) assert that measuring scholarly outputs generally involves two approaches: contribution studies and productivity studies. Contribution studies evaluate the scholarly contributions of researchers, universities, departments/centers, or any other contributor to a well-defined body of literature. While productivity studies assess the scholarly outputs of particular contributors holistically, often for comparative purposes. Although, no study can include all the research outputs due to, to name a few, feasibility, discipline variations, and databases coverage. In essence, research assessment may include a range of indicators across various dimensions of research to allow evaluation as per the study's purpose. The assessment process may involve technologies, bibliometrics, self-reported data and academics CVs, etc., with unique combinations and weightage given to each scholarly aspect. Self-reported research data is also seen as a credible source and often used in the survey due to its simplicity despite some inconsistency due to recall errors (Aiston & Jung, 2015; Allison & Stewart, 1974; Creswell, 1985; Xie & Shauman, 1998; Zhang & Shin, 2015).

In the current study, different indicators are included to gauge the research performance of academics. In this regard, firstly, the construct of research productivity is considered due to its broad encompassing nature to embrace a variety of indicators. As the literature suggests that research performance may be evaluated by a variety of indicators ranging from research publications, including sponsored research, through presentations to lending your services both within the institution and external academic engagement. Moreover, in addition to these academic research activities, collaboration in research in the form of co-authorship with domestic colleagues and with the international community is positively viewed. Therefore, research collaboration is considered an important dimension and also linked with increased productivity (Engels & Ruschenburg, 2008). Research collaboration greatly facilitates the

capacity building of researchers and improves the chances to access skills, funding, and facilities, along with increasing research visibility (Katz & Martin, 1997). The dimensions of collaboration are also incorporated in the assessment of academics in this study. Both domestic and transnational research collaborations are considered to enrich the evaluation process. Domestic collaboration indicates the degree of engagement with the domestic system and colleagues while undertaking different research activities and contributing to healthy academic culture. While transnational collaboration and joint publications are essential aspects of internationalization. Foreign education from leading institutions is usually associated with forming international academic links and collaboration in the future ((Eduan, 2019) and also contributes to the endogenous growth of source countries (Dustmann et al., 2011; Saxenian, 2005). As a result, in addition to research productivity, dimensions of collaborations are considered in the current study as the assessment involved returned academics having a foreign doctorate.

Based on literature insights, the study operationalizes research performance as various academic research outcomes in terms of quantity and quality of research and the nature of research collaborations. Specifically, the 'research performance' instrument is designed to assess key research performance areas, such as the overall productivity of research outputs, domestic research collaborations, and transnational research collaboration regarding the academic activity. Overall research performance encompasses the key research outputs such as peer-reviewed articles produced, books/chapters published, presentations given and funded projects conducted, etc. while domestic research collaborations aspect concerned with academics' co-authorship at the same university or anywhere in Pakistan and their memberships/engagements for research purposes in Pakistan between the years 2018-till date. Finally, the transnational research collaborations dimension captures the detail of global academic connectivity (Internationally collaborative work) for various research and academic purposes between the years 2018-till the date. This information suggests the international dimension of their academic products and services. Among the different types of publications and nature of collaborations, publishing in international journals and transnational scholarly collaborations are generally highly recommended in developing contexts, especially in the Pakistani context, in which this study is based. Overall, three different facets of research performance are conceptualized in the current study as bibliometrics and research evaluation literature growingly consider that construct of research performance has multiple facets and relevant indicators (Moed, 2017).

Further, the teaching function disseminates the latest knowledge and skills to equip students with 21st-century skills. The globalization of education encourages institutions to deliver academic excellence and maintain quality instructional standards (Thornton et al., 2018). Students' evaluation of teaching (SET) is one of the widely used ways to gauge teaching quality at universities for various purposes ranging from teaching improvement, appraisal, and institutional accountability (Spooren *et al.*, 2017). Although there is some debate over the

reliability and validity of students' evaluations of teaching, SET has become institutionalized to a large extent. Several studies consider SET a relatively reliable and logical source (Marsh, 2007; Nasser & Fresko, 2002; Zhao & Gallant, 2012) and are also associated with students' learning gains (Wachtel, 1998). Students' evaluations have yielded reliable results across various studies, partly because students have direct and extended teaching experience (Jimenez, 2008 as cited in Manrique, 2016). Teaching encompasses several dimensions and activities; consequently, its measures are generally multidimensional (Spooren et al., 2017). Effective teaching is conceptualized in various ways in different contexts, resulting in various conceptual frameworks guiding instrument development (Devlin & Samarawickrema, 2010; Penny, 2003).

Based on the previous studies and literature (Al-Hinai, 2012; Baliyan & Moorad, 2018; Bedggood & Donovan, 2012; Ching, 2018; Feldman, 1976; Hsu & Chiu, 2009, Lu & Wu, 2018; Manrique, 2016; Marsh, 2007; Richardson, 2005), important dimensions of teaching were identified, and relevant specific items were devised in the current study to develop teaching quality measure to be used in Pakistani context across different universities and disciplines. Specifically, instruction (using effective teaching strategies with enthusiasm and encouragement of participation), utilization of learning aids (De Neve & Janssen, 1982), the impact of instruction or students' outcomes, organization of the course, subject mastery, assessment and feedback, and interpersonal skills (Feldman, 2007; Marsh, 2007). These dimensions and related aspects facilitate the student learning process leading to enhanced learning outcomes as found in previously mentioned investigations. Among these dimensions, some dimensions were identified that seemed suitable in the studied context.

Moreover, Vermunt and Verschaffel (2000) asserted that, based on a substantive literature review, three domains of activity namely, cognitive, affective, and regulative play an important role in students' learning. The cognitive aspect is concerned with processing the content of learning, such as understanding concepts, making relationships between concepts, etc. While affective domain encompasses the different emotions and feelings that may arise during the learning process and may affect the learning process, for instance, motivating oneself during a learning task. Finally, the regulative domain deals with regulating the cognitive and affective aspects of learning and indirectly facilitating the learning process (Knol et al., 2016; Vermunt & Vermetten, 2004). Therefore, these aspects are also incorporated in the instrument about various dimensions of teaching, and some items correspond to these domains of activity. For instance, a) explaining subject matter coupled with real-life applications of concepts in an organized fashion facilitate effective cognitive functioning, b) meaningful connecting with students through interpersonal behaviors to promote positive emotional climate and to encourage participation/initiative, c) providing guidance in the form of feedback and addressing learning issues encountered by students.

The current study conceptualizes teaching quality as the extent to which various instructional processes have been successfully executed by employing a set of competencies/behaviors and resources with the effect of improved student learning and enhanced instructional practice. In the current study, based on literature and operational definition of teaching quality, seven important dimensions were identified perceived as centered at the heart of the teaching process. Moreover, the instrument construction process is predominantly guided by constructivist teaching principles. To this end, in light of the study's purpose, important factors and conceptualizations are realized about research and teaching performance areas as the basis to develop relevant specific indicators capturing the constructs. A depiction model is created in figure 1, followed by factors and definitions in a summarised form (Table 1).

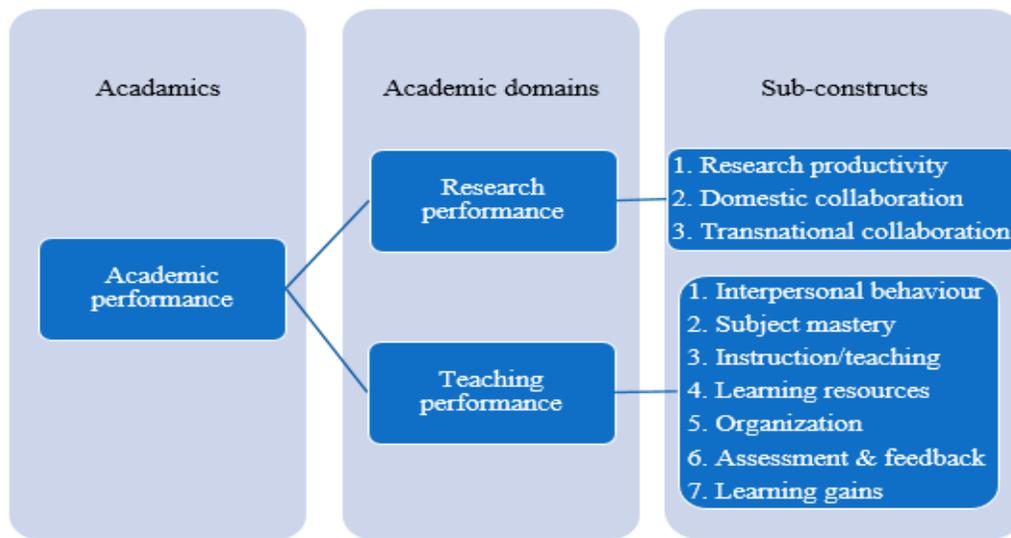


Figure 1: Research Model Involving Key Constructs and Sub-Constructs

Table 1: Description of the Constructs and Sub-Constructs as Conceptualised in the Study

Constructs and sub-constructs	Description
A. Research performance	Academic research outcomes in terms of quantity and quality of research and nature of research collaborations
i. Research productivity	Research productivity encompasses the key research outputs such as peer-reviewed articles produced, books/chapters published, presentations given and funded projects conducted, etc.
ii. Domestic research collaborations	This aspect concerned with academics' co-authorship at the same university or anywhere in Pakistan and their memberships/engagements for research purposes in Pakistan between the years 2018-till date.
iii. Transnational research collaborations	This dimension captures the detail of global academic connectivity (Internationally collaborative work) for various research and academic purposes between the years 2018-till the date.
B. Teaching performance	The extent to which various instructional processes have been successfully executed by employing a set of competencies/behaviours and resources with the effect of improved student learning and enhanced instructional practice.
i. Interpersonal behaviour	The degree to which interpersonal skills are exercised to connect with students meaningfully.
ii. Subject mastery	The degree to which teacher exhibits a state of art knowledge, understandings and applications related to subject.
iii. Instruction	The degree to which teacher employs various instructional methods and techniques in line with overarching learning objectives.
iv. Learning resources	The degree to which teacher uses the available learning resources and facilitates students in introducing and using various resources.
v. Organization	The degree to which subject matter is dealt with in an organized fashion within the given time bound.
vi. Assessment/feedback	The degree to which assessment and feedback procedures are employed in such a way to foster students learning.
vii. Learning gains	The degree to which students have achieved learning outcomes in different accounts.

Methods

Instrumentation Development

Adcock and Collier (2001) standard shared framework for instrument development was followed to devise instruments. This standard shared framework for instrument design carefully considers that dimensions, relevant indicators, and resulting scores should greatly represent the essence of concepts formulated in the systematized definition. The process involves literature searches, identifying dimensions and relevant indicators, items generation, content validation from experts, and field test to conduct factor analysis leading to the assessment of construct reliability and construct validity. Following the guidelines, the current study was based on the development of academics' research performance instrument (ARPI) and student evaluation of teaching quality instruments (SETQ) based on content validation, construct reliability, and validity. The overall development process is mirrored in the second

figure below. Further details of the process for each instrument are provided in the proceeding sections.



Figure 2: Instrument Development Process

Academics Research Performance Instrument (ARPI)

Overall, during the development of the questionnaire, various aspects were included that are deemed most important, globally accepted, and widely used in most of the universities worldwide for multiple purposes from research assessment, individual promotions, institutional ranking to awards and honors. Therefore, based on literature review and research evaluations criteria employed by leading universities across the globe, specific research-related aspects were included in the research performance to fully design the questionnaire fulfilling the aim of the current study. Specifically, the instrument of research performance of academics was devised to include three dimensions such as overall research productivity, domestic research collaborations, and transnational research collaboration concerning academic activity. Specifically, research performance encompassed the key research outputs such as peer-reviewed articles produced, books/chapters published, presentations given and funded projects conducted, etc. while domestic research collaborations aspect concerned with academics' co-authorship at the same university or anywhere within the country and their memberships for research purposes in Pakistan. Finally, the transnational research collaborations dimension captured the details of global academic connectivity or internationally collaborative work for various research and academic purposes between specified periods.

After developing a pool of items about the above-mentioned three research dimensions, an expert panel review process was employed to refine and generate new items sequentially to ensure the instrument's content validity. During the process, the request for feedback was sent to numerous experts with common specialization in higher education research via emails. During the process, ten experts agreed to provide feedback; among them, eight were international academics, while two were domestic university academics in Punjab, Pakistan. All of the academics had doctorates serving different universities and were accessed through emails after visiting faculty profiles of different universities. An instrument draft was sent to experts along with instrument purpose and operational definitions of constructs to yield

judgments about relevance, representativeness, clarity, and overall alignment. Especially, they were requested to rate each item's degree of relevance or representativeness regarding research dimensions (research productivity, domestic collaborations & transnational collaborations). Overall, 23 indicators were retained, some items seemed redundant, while a few indicators were found ambiguous and irrelevant leading to deletion of these items.

They also provided valuable feedback apart from judging item relevance to increase response rate and ease of providing research information on the part of university teachers. For instance, some reviewers asserted that some specified time should be mentioned in the instrument while guiding the university teachers so that research outcomes may be comparable across foreign and domestic Ph.D. holders and disciplines, etc. Moreover, while administering the instrument, a few experts suggested asking people to estimate or approximate the publication record that occurred between specified periods rather than asking teachers to mention the *exact publication record*. Therefore, the wording in guiding the respondents was changed to include *estimation* so that respondents may feel better while providing their publication record, leading to improve response rate. Finally, two experts highlighted that information about publications as first or secondary authors should be amended to include all publications, whether primary or supporting. They believed that academics often think of their work as publications and do not divide it so distinctly between the first and second authors. To this end, these suggestions were incorporated in the different sections of the instrument. Table 2 indicates the experts' feedback and retained indicators.

Table 2: Content Validation of 'Academics Research Performance Instrument' (ARPI) Based on Experts' Opinion

Research Dimensions and Definitions	Preliminary Indicators with Improved Version in Bracket Based on Experts' Feedback	Experts' Suggestions
1. Research Productivity. Research productivity refers to the various academic research outcomes in terms of quantity and quality of research either independently or collaboratively realised between specified period	Articles published in national journals	Retained
	Articles published in international journals	Retained
	Books Published (Books & edited books published in national or internal publication outlets	Suggested to improve
	Chapters authored for national publisher	Retained
	Chapters authored for international publishers	Retained
	Research presentations in national conferences & seminars	Retained
	Research presentations in international conferences & seminars	

	Funded projects conducted, sponsored by national and internal bodies (split into two distinct items: Funded projects conducted, sponsored by national bodies; Funded projects conducted, sponsored by international bodies	Suggested to improve
	Citations across all scholarly work published between specified period (Google scholar citations across all scholarly work published between specified period)	Suggested to improve
	Scientific distinctions and awards received at university/national level (split into two distinct items: Scientific distinctions and awards received at university level; Scientific distinctions and awards received at national level	Suggested to improve
<p>2. Domestic Research Collaborations. This dimension concerns with academics' co-authorship at the same university or anywhere in Pakistan as primary or supporting contributor and memberships/engagements for research purposes in Pakistan between the specified period.</p>	Academic articles published with domestic co-authors	Retained
	Books published with domestic co-authors (Books & edited books published with domestic co-authors)	Suggested to improve
	Edited books published with domestic co-authors	Suggested to merged in previous item
	Research memberships of national journal editorial boards	Retained
	Memberships of various national scientific organisations/associations	Retained
	Memberships of different expert panels that evaluate national projects/proposals	Suggested to remove
<p>3. Transnational Research Engagement This aspect relates to the researcher's international co-authorship of academic work as primary/supporting contributor (including in-press publication) & academic memberships for research purposes between specified period.</p>	Co-authorship (foreign) as primary and supporting authors in the following research outlets: <ul style="list-style-type: none"> a. Journal articles b. Books/Edited books c. Chapter d. Reports (this sub-item is excluded) 	Suggested to improve
	Memberships of international journal editorial boards	Retained
	Memberships of international scientific organisations/associations	Retained
	Member of review teams/panels for articles/research work	Suggested to remove
	Inviting foreigners for conferences or for other academic purposes	Suggested to remove
	Participation in expert panels that evaluate international projects/proposals	Retained
	Translations of foreign work	Retained

Students' Evaluations of Teaching Quality (SETQ)

Based on the previous studies and literature, important dimensions of teaching were identified, and relevant specific items were devised in the current study to develop teaching quality measures to be used in the Pakistani context across different universities and disciplines. Initially, a collection of about 90 assessment items was compiled from a primary literature review and grouped into seven latent factors or variables with multiple indicators. Moreover, the construction process, compatible with operational definition, follows a broader framework in developing the instrument and includes instructional features of design, implementation, facilitation, effects, and teacher personal and interpersonal skills. These features and relevant items have been developed predominantly through the lens of constructivist principles of teaching and learning.

Similarly, an expert panel review process was employed to refine and generate new items sequentially to ensure the content validity of the SETQ as employed in the ARPI instrument's content validation. Based on experts' reviews and suggestions, factors and relevant items were refined. Some items were removed, which were seen as irrelevant and mismatched to the overarching focus of the instrument. Moreover, a few words were replaced with more meaningful words that convey the intended meaning. In conclusion, based on experts' views, the instrument was made more representative, clear, and congruent with the instrument focus and overarching research purpose.

Afterward, to assess the measurement model, the instrument was administered from 333 university students (200 females; 133 males) studying in two different general public sector universities in Lahore. Students were associated with natural and social sciences subjects. Among them, nearly 53% were enrolled in honors-level programs, about 26% were in master's study, around 14% were doing MPhil, and the remaining 1% were Ph.D. scholars. The data collection was primarily completed with the assistance of Google Forms, and about 20% of respondents filled the instrument as paper-based.

The measurement model of the SETQ was composed of seven latent variables with multiple indicators for each latent variable. The latent variables include interpersonal behavior (6 items), subject mastery (5 items), teaching/instruction (7 items), learning resources (5 items), organization (6 items), assessment/feedback (8 items), and learning gains (6 items). The items were measured on a five-point Likert scale ranging from never (1) to always (5). CFA was run on the final data set using SmartPLS 3 software to verify further items and dimensions hypothesized earlier based on literature and expert opinion.

Firstly, to evaluate the measurement model, the constructs' loading, reliability, and convergent validity are assessed. Results are shown in Table 3. Loadings should be above .70 to consider the item as reliable (Garson, 2016; Hair et al., 2020). For the data, all the loadings are above

the acceptable cut-off. The composite reliabilities for all the constructs are also above the generally recommended threshold of .70 (Garson, 2016), indicating good reliability (0.947-0.905). Further, Cronbach's Alpha was used to assess the extent to which a different set of questions taps a single underlying construct. Allen et al. (2018) and Fayers and Machin (2016) recommend that Cronbach's Alpha values should be around .9 ideally and anything above .7 or around .7 is considered appropriate for most research purposes Results exhibits that Cronbach's Alpha values for the constructs vary from 0.936 to 0.868 suggesting the indicators are consistent. Convergent and discriminant validity were also evaluated for the constructs. Convergent validity may be assessed using Average Variance Extracted (AVE), the criterion for which is AVE values should be 0.5 or greater to establish this validity (Hair et al., 2020). For this study, the AVE values for the constructs are above 0.5, thus establishing convergent validity.

Table 3: Construct Reliability and Construct Convergent Validity

Dimensions and related items	Loadings	Cronbach's Alpha	CR	AVE
I. Interpersonal Behaviour		0.904	0.926	0.677
The teach had shown concern to students' problems.	0.848			
The teacher had flexibility to accept diversity of opinions.	0.799			
The teacher treated students with kindness.	0.818			
This teacher was helpful to students.	0.867			
The teacher cared about how students feel.	0.828			
The teacher inspired me in this course	0.774			
II. Subject Mastery		0.868	0.905	0.656
The teacher presented up-to-date knowledge of the latest developments in the subject/field.	0.841			
The teacher provided a syllabus that covered all necessary contents of the course.	0.764			
The teacher shared knowledge beyond the curriculum.	0.771			
The teacher combined theoretical concepts with real world applications	0.844			
The teacher answered competently to every question asked by students	0.825			
III. Instruction/Teaching		0.917	0.934	0.669
The aims and objectives of the course were clear to me	0.791			
The teacher utilized a variety of teaching techniques (quizzes, group work, projects etc.) in the course.	0.794			
The teacher accommodated the students with different learning abilities	0.847			
The teacher provided meaningful opportunities for active student participation in learning activities.	0.846			
The teacher taught students how to independently derive knowledge rather than showing it straightway	0.870			
The teacher displayed enthusiasm and humour.	0.749			
The teaching methods were in line with the course objectives.	0.825			

IV. Learning Resources		0.878	0.912	0.674
The books/materials suggested for the course were relevant to course objectives.	0.727			
The teacher provided guidance about using online learning resources	0.859			
Online learning resources (e.g. journals, articles, e-books) suggested by the teacher were easily accessible	0.847			
The teacher utilized available learning resources in the classroom.	0.882			
The teacher helped students to utilize lab or library resources	0.781			
V. Organization		0.888	0.915	0.642
The teacher presented content in an organized manner	0.748			
The teacher managed class time effectively.	0.774			
The teacher created an environment in which students felt comfortable asking questions and expressing their views.	0.817			
The workload of the subject (e.g. Assignments, projects) given to students was appropriate	0.805			
The subject matter was taught at an appropriate pace	0.828			
The teacher helped students for learning other than class time.	0.834			
VI. Assessment and Feedback		0.936	0.947	0.692
The teacher communicated the assessment and evaluation criteria for the course.	0.840			
Assessment and evaluation methods were in line with course objectives.	0.823			
Students were assessed from a variety of assessment techniques.	0.838			
Assessment techniques were interesting to promote deep learning beyond memory recall.	0.863			
The teacher provided examples of excellent work.	0.827			
The teacher provided timely feedback on assessment to help me learn.	0.854			
Enough time were provided after assessment to improve the learning	0.817			
The teacher was fair in assessment.	0.788			
VII. Learning Gains		0.930	0.945	0.743
I improved my understanding in this subject course.	0.834			
I understood the concepts, theories or important ideas presented in this course.	0.869			
The information/materials provided for a course help me to learn independently.	0.894			
The teacher taught the students to use high level thinking in this course (e.g. analyzing ideas, synthesizing ideas, making judgments about information, applying information to new situations).	0.873			
My perspective expanded (e.g. intellectually, culturally) about the material studied in this course.	0.887			

Finally, discriminant validity assessment is carried out using the heterotrait-monotrait ratio of correlations (HTMT), which is one of the recommended methods for this purpose (Henseler

et al., 2015). Researchers may employ cut-off points of 0.85 and 0.90 to assess results, and the more lenient threshold of 0.90 may be used to assess the discriminant validity of similar concepts (Hair et al., 2020). In this study, the constructs are similar as all are related to the one higher-order construct of teaching quality; therefore, the more liberal cut-off point 0.90 is used. Results signify that HTMT scores are below 0.90 thresholds, satisfying the discriminant validity (table 4).

Table 4: Discriminant Validity (HTMT)

Variables	A	IB	LG	LR	O	SM	T
Assessment							
Interpersonal behaviour	0.802						
Learning gains	0.855	0.708					
Learning resources	0.895	0.748	0.800				
Organization	0.891	0.838	0.845	0.891			
Subject mastery	0.867	0.847	0.839	0.877	0.897		
Teaching	0.886	0.818	0.843	0.880	0.895	0.870	

Discussion

The purpose of this research was to develop the measures of academic performance in terms of academic research and quality of academics' teaching in campus-based higher institutes of central Punjab that also may be economical and multidisciplinary. The intent is to compare foreign scholarship returnees (having foreign Ph.D.) with the reference group of non-returnees (having domestic Ph.D.) in the future by employing these measures. As there is a need to develop more comprehensive tools to assess academic performance in various domains especially academic research and teaching in the Pakistani institutional context. The main intent behind the development is to design such instruments that would enable the assessment of research and teaching performance of academic returnees (having foreign PhDs) and non-returnees (having domestic PhDs) serving in public-sector universities of Punjab. The reference group (non-returnees) is added to facilitate comparative assessment and interpretation of performance as performance assessment is mostly dependent upon the reference group (Daumiller et al., 2019). Firstly, the academic research performance instrument (ARPI) was developed based on three sub-constructs (research productivity; domestic research collaboration; transnational research collaboration), capturing different dimensions of research keeping in view academic returnees and internationalization of academic activity. Literature review greatly facilitated identifying these research constructs and various relevant indicators. Overall. This instrument encompassed the key research outputs such as peer-reviewed articles produced, books, edited books, chapters published, presentations given and funded projects conducted, etc. while domestic research collaborations aspect concerned with academics' co-authorship at the same university or anywhere within the country and their memberships for research purposes in Pakistan between specified periods. Finally, transnational research collaborations dimension captured

the detail of transnational academic connectivity (internationally collaborative work) for various research and academic purposes. In this effort, multiple indicators are included, which is considered an appropriate approach in research performance assessment especially involving different disciplines (Hicks et al., 2015; Gogolin & Stumm, 2014). Overall, the emphasis is largely on production, participation, associations, and collaboration patterns among academics, with the focus on comparing returnees and non-returnees. There is a growing consensus in research evaluation literature that research performance is a multifaceted construct (Moed, 2017) and studies have growingly been conducted by incorporating this multitude of dimensions approach.

Keeping in view the instrument,s nature, experts' judgments were obtained to evaluate its representativeness, clarity, readability, and degree of correspondence with the overarching purpose of the research purpose. In the instructions section of this instrument, approximately and estimation words are incorporated into direction and statements to facilitate academics/respondents to remember or indicate the information with ease rather than asking to write the exact record in paper-based data collection. As many do not remember offhand the answers without referring to the database or CVs. Further, publication in the national language is also considered in assessing research records in addition to the English language to avoid a biased account where publishing in English counts more than publishing in Pakistani or any other language. This instrument may be employed in a variety of different ways. For instance, the data can be gathered from digital databases, peer reviews, academics' CVs, or self-report questionnaires based on instrument information. Every source has its strengths, limitations, coverage, and ease of administration in light of the research aims (Fangmeng, 2016; Fontes, 2007). The self-report questionnaire is widely employed in social sciences to assess workplace performance (Garcia & Gustavson, 1997 as cited in Daumiller et al., 2019) and is considered economical (Daumiller et al., 2019). Anonymity is critical to increase the likelihood of participation and elicit reliable information on the parts of the respondents.

Secondly, institutions and academicians do widely employ students' ratings of teaching to gauge the teaching quality of teachers in the current globalized world, largely as a component of quality assurance (Barth, 2008; Ulker, 2021). This is an important component of how students feel about the teaching-learning process, ultimately informing improvement and supporting students' learning. The student's evaluation of teaching instruments may also be used for numerous purposes, including comparisons among academics and institutions as per the purposes of the study in question. In the current study, students' evaluation of teaching quality (SETQ) instrument encompasses a range of teaching-related aspects or indicators such as interpersonal skills, subject mastery, teaching/instruction, learning resources, organization, assessment, feedback, and learning gains. The items related to these constructs were measured on a five-point scale ranging from always true to never true. This multidimensional scaling approach is seen as more appropriate in literature keeping in view the complex nature of the

teaching process. Spooren et al. (2007) assert that the scaling approach is likely to be more robust against social desirability, vague interpretations, and accidental fluctuations in responses compared to singly-item teaching assessment.

While administering the instrument, students' personal identifying information is not included to increase the probability of honest or genuine responses; as Macklain et al. (2018) state that format that will protect the identity of students should be preferred either online or paper-based. The final version questionnaire was found to be reliable and valid based on 333 responses at the initial stage at this point. The psychometric assessment included item and construct level reliability, convergent validity, and discriminant validity. Item loadings, composite reliabilities, and Cronbach's Alpha values were within acceptable thresholds. Convergent validity results confirmed that indicators extracted the amount of variance deemed enough to explain relevant sub-constructs. Moreover, homogeneity of indicators and strong associations among all constructs pointed towards the assumed one-dimensional structure of higher-order factor titled teaching quality comprising seven teaching aspects. Therefore, a more liberal cut-off point of 0.90 (Hair et al., 2020) was used to assess discriminant validity and results confirmed the discriminant validity through this process (HTMT).

Instruments like these that employ various aspects of teaching may provide feedback on areas that requires improvement leading to teaching excellence (Wilson et al., 1997). More importantly, the feedback is crucial for institutions to engage in the learning process to improve and provide quality provision of teaching (Bowden, 2011). In addition, if a questionnaire like this one is used with peer coaching, experts' consultation and formative feedback or any other relevant assistance are more likely to produce teaching quality and consequently improved learning outcomes and instructional practice (Knol et al., 2013; Knol, et al., 2016; Penny & Coe, 2004; Ulker, 2021). The results yielded from such student evaluations should be used primarily for developmental or formative purposes rather than for summative purposes (Hedges & Webber, 2014; Wolbring & Treischl, 2016), in this way, instructors will be more likely to engage in this feedbacks to inform continuous development.

Implications and Future Recommendations

The study offers several implications for various stakeholders or consumers of research. Firstly, the instrument ARPI may be used to assess university academics' research performance in terms of overall research productivity, research collaborations within the university or country, and collaborations with international researchers and the community. The information may inform the research practices, performance levels, and degree of internationalization of academics' research efforts. The research tool can also be applied to compare academics having foreign PhDs and those with domestic PhDs to ascertain how both groups differ in research dimensions and whether foreign qualifications affect their

transnational research collaborations. This tool is primarily developed to compare foreign returnees' research performance with domestic academics. The comparisons may also be made within the returnees' group to see, for instance, whether returnees from European countries have different research contributions from returnees having doctoral education from Asia or other countries. The resultant information may inform about the functionality of returnees in different domains to produce evidence in the Pakistani context regarding whether scholarships provided to students or faculty members to gain education abroad have any significant effects on their contributions upon return. As the literature indicates, the most influential factor in producing research outcomes is the teachers' higher education and training (Chepkorir, 2018). This achievement is partly conditional on healthy working conditions and an overall support system.

Moreover, the second instrument, SETQ, may also be used for similar purposes in Punjab, Pakistan. The instrument may provide evidence whether the teaching quality of academics having foreign PHDs differs from academics with domestic PhDs. As the instrument is multidimensional, therefore, resultant data will be enriched to compare specific instructional aspects across university status and discipline, in addition to the nature of higher qualifications as perceived by university students. The instrument may also inform university teachers' instructional practice and what areas they need to improve their teaching. Both measures may be applicable across social and natural science disciplines. Based on literature guidelines and a range of indicators included allowed the supposed applicability across these two disciplines. Indicators included in these measures may be considered as shared criteria for both fields in which some indicators could be weighted more in a specific discipline. Therefore, various indicators are included to ensure multidisciplinary applicability (social & natural sciences) and economic administration.

The current study also recommends some research directions in the future based on the limitations. The first instrument, 'APRI,' is only content validated. Therefore, the instrument requires a field test to gauge the ease of administration and respondents' views. Further, largely the focus in this instrument is on assessing the production of documents, participation, associations for journals, and collaboration patterns. Therefore, future studies may include others indicators capturing the impact and quality of the research activity, knowledge, internal and external services of academics, to name a few, as variables in the analysis of academics at cluster or institutional levels. While about the second instrument, the sample is limited to only two public sector universities in Lahore. Thus, results may not be representative of universities located in diverse locations. Therefore, to enhance the generalizability of the instrument, multiple universities situated in diverse locations and a larger number of students may be considered to collect data to confirm or refine the psychometric properties of SETQ.

Moreover, concurrent validity may be assessed of this SETQ instrument with other instruments measuring similar constructs. Finally, to ensure measurement invariance of both

measures across social and natural sciences, studies may confirm whether data confirmed the identical structure in different disciplines. In this way, the generalizability of the measures could be gauged across various disciplines.

Conclusion

The research has developed two instruments, labeled ARPI and SETQ, to assess the academic performance of academics in terms of academic research and teaching quality. Key constructs and relevant indicators were included with the assistance of literature review and experts' suggestions. The ARPI instrument is content validated and comprises 17 items across dimensions of research productivity, domestic research collaboration, and global research collaboration, while SETQ consists of 43 items encompassing seven teaching aspects. Results attested that this instrument has embodied desirable and basic psychometric properties. Taken together, this research underscores the importance of developing instruments about research and teaching performance of academics in the university context and contributes to standard practices of instrument development to assess the academic performance of academics' especially academic returnees (having foreign Ph.D.) in domains of research and teaching in comparison with their domestic counterparts (having domestic Ph.D.). The application and resultant information may additionally contribute to cover research and knowledge gaps that exist in the area of assessment of academics' research and teaching performance in central Punjab, Pakistan context.

References

Abramo, G., & D'Angelo, C. (2014). How do you define and measure research productivity? *Scientometrics*, 101(2), 1129-1144. <https://doi.org/10.1007/s11192-014-1269-8>

Aiston, S., & Jung, J. (2015). Women academics and research productivity: An international comparison. *Gender and Education*, 27(3), 205-220. <https://doi.org/10.1080/09540253.2015.1024617>

Allen, J. L., Huggins-Hoyt, K. Y., Holosko, M. J., & Briggs, H. E. (2018). African American social work faculty: Overcoming existing barriers and achieving research productivity. *Research on Social Work Practice*, 28(3), 309-319. <https://doi.org/10.1177/1049731517701578>

Allen, P., Bennett, K., & Heritage, B. (2018). *SPSS Statistics: A practical guide with student resource access 12 months* (4th ed.). Cengage Learning Australia.

Al-Hinai, S. N. (2011). *Effective college teaching and students' ratings of teachers: What students think, what faculty believe, and hat actual ratings show implications for policy and practice in teaching quality assurance and control in higher education in Oman*, (Doctoral thesis, Durham University, UK). <http://etheses.dur.ac.uk/649/>

Almanasreh, E., Moles, R., & Chen, T. F. (2019). Evaluation of methods used for estimating content validity. *Research in Social and Administrative Pharmacy*, 15(2), 214-221. <https://doi.org/10.1016/j.sapharm.2018.03.066>

Altbach, P. (2015). What counts for academic productivity in research universities?. *International Higher Education*, (79), 6-7. <https://doi.org/10.6017/ihe.2015.79.5837>

Barth, M. M. (2008). Deciphering student evaluations of teaching: A factor analysis approach. *Journal of Education for Business*, 84(1), 40-46. <http://dx.doi.org/10.3200/JOEB.84.1.40-46>

Bedggood, R. E., & Donovan, J. D. (2012). University performance evaluations: what are we really measuring? *Studies in Higher Education*, 37(7), 825-842. <https://doi.org/10.1080/03075079.2010.549221>

Cadez, S., Dimovski, V., & Zaman Groff, M. (2017). Research, teaching and performance evaluation in academia: The salience of quality. *Studies in Higher Education*, 42(8), 1455-1473. <https://doi.org/10.1080/03075079.2015.1104659>

Bowden, J. L. (2011). Engaging the student as a customer: A relationship marketing approach. *Marketing Education Review*, 21(3), 211–228. <https://doi.org/10.2753/MER1052-8008210302>

Chepkorir, K. R. (2018). Effect of academic staff qualification on research productivity in Kenyan Public universities; Evidence from Moi university. *International Journal of Economics, Commerce and Management*, 6(2), 609–620. <http://ijecm.co.uk/>

Ching, G. (2018). A literature review on the student evaluation of teaching. *Higher Education Evaluation and Development*, 12(2), 63-84. <https://doi.org/10.1108/HEED-04-2018-0009>

Cogburn, J. D., & Neely, S. R. (2015). Publish or perish? Examining academic tenure standards in public affairs and administration programs. *Journal of Public Affairs Education*, 21(2), 199–214. <https://doi.org/10.1080/15236803.2015.12001828>

Creswell, J. W. (2012). *Qualitative inquiry and research design: Choosing among five approaches*. Sage Publishing.

Daumiller, M., Siegel, S. T., & Dresel, M. (2019). Construction and validation of a short multidisciplinary research performance questionnaire (SMRPQ). *Research Evaluation*, 28(3), 241–252. <https://doi.org/10.1093/reseval/rvz009>

De Neve, H. M., & Janssen, P. J. (1982). Validity of student evaluation of instruction. *Higher Education*, 11(5), 543–552. <https://doi.org/10.1007/BF00194419>

Devlin, M., & Samarawickrema, G. (2010). The criteria of effective teaching in a changing higher education context. *Higher Education Research & Development*, 29(2), 111-124. <http://dx.doi.org/10.1080/07294360903244398>

Douglas, A. (2013). Advice from the professors in a university Social Sciences department on the teaching-research nexus. *Teaching in Higher Education*, 18(4), 377-388. <https://doi.org/10.1080/13562517.2012.752727>

Dustmann, Fadlon, I., & Weiss, Y. (2011). Return migration, human capital accumulation, and the brain drain. *Journal of Development Economics*, 95(1), 58–67. <https://doi.org/10.1016/j.jdeveco.2010.04.006>

Eduan. (2019). Influence of study abroad factors on international research collaboration: Evidence from higher education academics in sub-Saharan Africa. *Studies in Higher Education* (Dorchester-on-Thames), 44(4), 774–785. <https://doi.org/10.1080/03075079.2017.1401060>

Engels, & Ruschenburg, T. (2008). The uneven spread of global science: Patterns of international collaboration in global environmental change research. *Science & Public Policy*, 35(5), 347-360. <https://doi.org/10.3152/030234208X317160>

Fangmeng, T. (2016). Brain circulation, diaspora, and scientific progress: A study of the international migration of Chinese scientists, 1998–2006. *Asian and Pacific Migration Journal*, 25(3), 296-319. <https://doi.org/10.1177/0117196816656637>

Fayers, P., & Machin, D. (2016). *Quality of life: The assessment, analysis, and reporting of patient-reported outcomes* (3rd ed.). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118758991>

Feldman, K. A. (1976). The superior college teacher from the students' view. *Research in Higher Education*, 5(3), 243-288. <https://doi.org/10.1007/BF00991967>

Feldman, K. A. (2007). Identifying exemplary teachers and teaching: evidence from student ratings¹. In R. P. Raymond & J. C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 93–143). Springer Netherlands. https://doi.org/10.1007/1-4020-5742-3_5

Fontes, M. (2007). Scientific mobility policies: how Portuguese scientists envisage the return home. *Science & Public Policy (SPP)*, 34(4), 284–298. <https://doi.org/10.3152/030234207X214750>

Garson, G. D. (2016). *Partial least squares: Regression & structural equation models*. Statistical Associates Publishing.

Gogolin, I., and Stumm, V. (2014). The EERQI peer review questionnaire. from the development of 'intrinsic indicators' to a tested instrument. In I. Gogolin & F. Astrom & A. Hansen (Eds.), *assessing quality in European educational research. indicators and approaches*, (pp. 107–20). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-05969-9_8

Griffin, D., Bolkan, S., & Dahlbach, B. (2018). Scholarly productivity in communication studies: Five-year review 2012-2016. *Communication Education*, 67(1), 88-101. <https://doi.org/10.1080/03634523.2017.1385820>

Hair, J. F., Howard, M. C., & Nitzl, C. (2020). Assessing measurement model quality in PLS-SEM using confirmatory composite analysis. *Journal of Business Research*, 109, 101–110. <https://doi.org/10.1016/j.jbusres.2019.11.069>

Hardré, P. L., Beesley, A. D., Miller, R. L., & Pace, T. M. (2011). Faculty motivation to do research: across disciplines in research-intensive universities. *Journal of the Professoriate*, 5(1), 35–69. www.researchgate.net/publication/

Harvey, C. E., Kelly, A., Morris, H., & Rowlinson, M. (2010). *Academic journal quality guide*. London: The Association of Business Schools.

Hedges, M. R., & Webber, D. J. (2014). Using student evaluations to improve individual and department teaching qualities. *Research in Post-Compulsory Education* 19(3), 323–339. <http://dx.doi.org/10.1080/13596748.2014.920584>

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, 43(1), 115–135. <https://doi.org/10.1007/s11747-014-0403-8>

Hsu, J. L., & Chiu, Y. H (2009). Perceived differences in teaching performance from viewpoints of lecturers and students. *International Journal of Educational Management*, 23(7), 564–573. <https://doi.org/10.1108/09513540910990807>

Hug, S. E., Ochsner, M., & Daniel, H. D. (2013). Criteria for assessing research quality in the humanities: A Delphi study among scholars of English literature, German literature, and art history. *Research Evaluation*, 22(5), 369–383. <https://doi.org/10.1093/reseval/rvt008>

Iqbal, M., & Mahmood, A. (2011). Factors related to low research productivity at the higher education level. *Asian Social Science*, 7(2), 188-193. <http://dx.doi.org/10.5539/ass.v7n2p188>

Jenkins, A., Healey, M., & Zetter, R. (2007). *Linking teaching and research in disciplines and departments*. York: The Higher Education Academy.

Katz, & Martin, B. R. (1997). What is research collaboration? *Research Policy*, 26(1), 1–18. [https://doi.org/10.1016/S0048-7333\(96\)00917-1](https://doi.org/10.1016/S0048-7333(96)00917-1)

Knol, M. H., Veld, R., Vorst, H. C. M., van Driel, J. H., & Mellenbergh, G. J. (2013). Experimental effects of student evaluations coupled with collaborative consultation on college professors' instructional skills. *Research in Higher Education*, 54(8), 825-850. <https://doi.org/10.1007/s11162-013-9298-3>

Knol, M. H., Dolan, C. V., Mellenbergh, G. J., van der Maas, H. L. J. (2016). Measuring the quality of university lectures: Development and validation of the instructional skills Questionnaire (ISQ). *PLoS ONE* 11(2), 1-25. <http://dx.doi.org/10.6084/m9.figshare.1278889>

Long, R., Crawford, A., White, M., & Davis, K. (2009). Determinants of faculty research productivity in information systems: An empirical analysis of the impact of academic origin and academic affiliation. *Scientometrics* 78(2), 231–60. <https://doi.org/10.1007/s11192-007-19907>

Lu, Y., & Wu, C. (2018). An integrated evaluation model of teaching and learning. *Journal of University Teaching and Learning Practice*, 15(3), 1-19. <https://ro.uow.edu.au/jutlp/vol15/iss3/8/>

Manrique, M. E. R (2016). Relationship between Social Skills and Teaching Performance from the perception of adult students in a private university in Lima, Peru. *Revista Digital de Investigación en Docencia Universitaria*, 10(2), 17-30. <http://dx.doi.org/10.19083/ridu.10.465>

Marsh, H. W. (2007). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and usefulness. In R. P. Perry & J. C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 319-383). <https://www.semanticscholar.org/>

Mazzucchelli, T., Burton, E., & Roberts, L. (2019). Scholarly productivity and citation impact of Australian academic psychologists. *Australian Journal of Psychology*, 71(3), 305-311. <https://doi.org/10.1111/ajpy.12248>

Moed, H. F. (2017). *Applied evaluative informetrics*. Springer International Publishing

Nafukho, F., Wekullo, C., & Muyia, M. (2019). Examining research productivity of faculty in selected leading public universities in Kenya. *International Journal of Educational Development*, 66, 44-51. <https://doi.org/10.1016/j.ijedudev.2019.01.005>

Nasser, F. & Fresko, B. (2002), Faculty views of student evaluation of college teaching. *Assessment & Evaluation in Higher Education*, 27(2), 187-198. <https://doi.org/10.1080/02602930220128751>

Penny, A. R., & Coe, R. (2004). Effectiveness of consultation on student rating feedback. *Review of Educational Research*, 74(2), 215-253. <https://doi.org/10.3102/00346543074002215>

Penny, A. R. (2003). Changing the agenda for research into students' views about university teaching: Four shortcomings of SRT research. *Teaching in Higher Education*, 8, 399-411. <http://dx.doi.org/10.1080/13562510309396>.

Porter, S. R., & Umbach, P. D. (2001). Analyzing faculty workload data using multilevel modeling. *Research in Higher Education*, 42(2), 171-196. <https://doi.org/10.1023/A:1026573503271>

Richardson, J. T. E. (2005). Instruments for obtaining student feedback: A review of the literature. *Assessment & Evaluation in Higher Education*, 30(4), 387-415. <https://doi.org/10.1080/02602930500099193>

Saxenian. (2005). From brain drain to brain circulation: Transnational communities and regional upgrading in India and China. *Studies in Comparative International Development*, 40(2), 35-61. <https://doi.org/10.1007/BF02686293>

Shin, J. C., & Cummings, W. (2010). Multilevel analysis of academic publishing across disciplines: Research preference, collaboration, and time on research. *Scientometrics*, 85(2), 581-594. <https://doi.org/10.1007/s11192-010-0236-2>

Shin, J. C. (2011). Teaching and research nexuses across faculty career stage, ability, and affiliated discipline in a South Korean research university. *Studies in Higher Education*, 36(4), 485-503. <https://doi.org/10.1080/03075071003759052>

Spooren, P., Mortelmans, D., & Denekens, J. (2007). Student evaluation of teaching quality in higher education: Development of an instrument based on 10 Likert-scales. *Assessment and Evaluation in Higher Education*, 32(6), 667-679. <https://doi.org/10.1080/02602930601117191>

Slocumb, E. M., & Cole, F. L (1991). A practical approach to content validation. *Applied Nursing Research*, 4(4), 192-195. [https://doi.org/10.1016/S0897-1897\(05\)80097-7](https://doi.org/10.1016/S0897-1897(05)80097-7)

Smith, K. M., Crookes, E., & Crookes, P. A. (2013). Measuring research 'impact' for academic promotion: Issues from the literature. *Journal of Higher Education Policy and Management*, 35(4), 410-420. <https://doi.org/10.1080/1360080X.2013.812173>

Spooren, P., Vandermoere, F., Vanderstraeten, R., & Pepermans, K. (2017). Exploring high impact scholarship in research on student's evaluation of teaching (SET). *Educational Research Review*, 22, 129-141. <https://doi.org/10.1016/j.edurev.2017.09.001>

Ter-Bogt, H., & Scapens, R. (2012). Performance Management in Universities: Effects of the Transition to More Quantitative Measurement Systems. *European Accounting Review*, 21(3), 451-497. <https://doi.org/10.1080/09638180.2012.668323>

Thornton, K., Walton, J., Wilson, M., & Jones, L. (2018). Middle leadership roles in universities: Holy Grail or poisoned chalice. *Journal of Higher Education Policy and Management*, 40(3),208–223. <https://doi.org/10.1080/1360080X.2018.1462435>

Ulker, N. (2021). How can student evaluations lead to improvement of teaching quality? A cross-national analysis. *Research in Post-Compulsory Education*, 26(1), 19-37. <https://doi.org/10.1080/13596748.2021.1873406>

Vermunt, J. D., & Vermetten, Y. J. (2004). Patterns in student learning: relationships between learning strategies, conceptions of learning, and learning orientations. *Educational Psychology Review*, 16(4), 359–384. <https://doi.org/10.1007/s10648-004-0005-y>

Vermunt J. D., & Verschaffel L. (2000) Process-oriented teaching. In R. J. Simons, J. Linden & T. Duffy (Eds.), *New learning* (pp. 209-225). Springer, Dordrecht. https://doi.org/10.1007/0-306-47614-2_11

Wachtel, H. (1998). Student Evaluation of College Teaching Effectiveness: A brief review. *Assessment & Evaluation in Higher Education*, 23(2), 191-212. <https://doi.org/10.1080/0260293980230207>

Webber, K. L. (2011). Measuring Faculty Productivity. In J. C. Shin, R. K. Toutkoushian & U. Teichler (Eds.), *University rankings: Theoretical basis, methodology, and impacts on global higher education* (pp. 105–121). https://doi.org/10.1007/978-94-007-1116-7_6

White, C., James, K., Burke, L., & Allen, R. (2012). What makes a "research star"? Factors influencing the research productivity of business faculty. *International Journal of Productivity and Performance Management*, 61(6), 584-602. <https://doi.org/10.1108/17410401211249175>

Wilson, K. L., A. Lizzio, & Ramsden, P. (1997). The development, validation, and application of the course experience questionnaire. *Studies in Higher Education* 22(1), 33–53. <https://doi.org/10.1080/03075079712331381121>

Wilder, E., & Walters, W. (2019). Quantifying scholarly output: Contribution studies and productivity studies in sociology since 1970. *The American Sociologist*, 50(3), 430-436. <https://doi.org/10.1007/s12108-018-9400-6>

Wolbring, T., & Treischl, E. (2016). Selection bias in students' evaluation of teaching: Causes of student absenteeism and its consequences for course ratings and rankings. *Research in Higher Education* 57, 51–57. <https://doi.org/10.1007/s11162-015-9378-7>

Xie, Y., & Shauman, K.A. (1998). Sex differences in research productivity: new evidence about an old puzzle. *American Sociological Review* 63(6), 847–870. <https://doi.org/10.2307/2657505>

Zhang, L., & Shin, J. C. (2015). The research–teaching nexus among academics from 15 institutions in Beijing, Mainland China. *Higher Education*, 70(3), 375-394. <https://doi.org/10.1007/s10734-014-9836-8>

Zhao, J., & Gallant, D. J. (2012). Student evaluation of instruction in higher education: Exploring issues of validity and reliability. *Assessment & Evaluation in Higher Education*, 37, 227–235. <https://doi.org/10.1080/02602938.2010.523819>