

Development and Validation of Teaching Portfolios Utilization Scale (TPUS)

¹Muhammad Mujahid

²Dr. Wasim Khan

³Riaz Ahmad

⁴Muhammad Naeem Shaukat

¹PET Govt Model Higher Secondary School Chak 36 DNB Yazman Bahawalpur

²Assistant Professor/Director Sports, Department of Sports Sciences and Physical Education, Gomal University, Dera Ismail Khan

³Lecturer Govt Graduate College Kehror Pakka Lodhran

⁴PET Govt Elementary school Islami Maddrsa Bahawalpur

mujahidbwp39@gmail.com, wasimkhansspe@gu.edu.pk, riazi151990@gmail.com,

naeemsalary1122@gmail.com

Abstract

Teaching portfolios have become widely recognized as a significant instrument for reflective practice and professional development and instructional improvement, though their systematic use remains largely unexplored, especially in developing country settings. This study focused to develop and validate the Teaching Portfolios Utilization Scale (TPUS) among the College chapter of Physical Education teachers in Pakistan. Following established scale development procedures, items were generated through literature review and input of practitioners and content validity was established using exit evaluation. After item refinement the final scale consisted of 18 items. Exploratory Factor Analysis on a sample of college Physical Education teachers demonstrated a definite four-factor construct that was verified using an acceptable Kaiser Meyer Olkin value (KMO = 0.781) and a significant Bartlett's Test of Sphericity ($kh_2 = 600.73$, $p < .001$). Confirmatory Factor Analysis on an independent sample suggested good model fit ($kh_2/df = 2.48$, CFI = .921, TLI = .912, RMSEA = .061, SRMR = .052). Convergent and discriminant validity was demonstrated and reliability analysis indicated satisfactory internal consistency ($\alpha = .83$). The results indicate that TPUS is a valid and reliable tool to measure the awareness of teaching portfolio use among the college Physical Education teachers of Pakistan.

Keywords: Physical Education teachers, teacher workload, scale development, validity and reliability, organizational support

Article Details:

Received on 10 Jan, 2026

Accepted on 30 Jan, 2026

Published on 01 Feb, 2026

Corresponding Authors*

INTRODUCTION

Teaching portfolios have been receiving growing recognition as comprehensive instruments for teaching documentations, reflections, and improvements in teaching practice in a variety of educational contexts. A teaching portfolio is generally defined as a purposeful and systematic collection of evidence to demonstrate the instructional practices of a teacher, professional growth and reflection thinking from time to time (Loughran, 1995; Lyons, 1998). One way in which they differ from single observation or test-based evaluation methods is that portfolios permit teachers to provide multiple forms of evidence related to their teaching (e.g., lesson plans, instructional strategies, records of assessment and professional development activities) to provide a holistic picture of teaching competence (Darling-Hammond & Snyder, 2000).

A key strength of teaching portfolios is that individual portfolios are closely linked with reflective practice. Reflective teaching is the process of critically reflecting on instructional choices, student responses, and student outcomes to lead to continuous improvement of teaching (Schon, 1983). Portfolios operationalize reflection in that it requires teachers to justify pedagogical decisions, to constructive analyze assessment data and articulate learning coming from experience (Lyons, 1998). Empirical research suggests that the process of portfolio-based reflection can promote the pedagogical awareness, professional judgment and flexibility of instruction in teachers (Loughran, 2002; Consequently, portfolios are being used extensively not only as a developmental tool, but - in some settings - are also being used as a part of teacher evaluation systems (Darling-Hammond et al., 2012).

In physical education (PE) there is perhaps nowhere that the relevance of teaching portfolios is found so strongly. Effective PE teaching includes developmental planning of lessons appropriate for the age group, some management of an active learning environment, safety of students, and engagement of students by using various teaching methods. Many of these dimensions cannot be readily obtained from conventional classroom observations alone (Ward, & Ayvazo, 2016). Teaching portfolios allow PE teachers to record evidence such as lesson plans and skill demonstrations and assessment rubrics and reflective notes about student's participation to make them valuable tools for professional learning and accountability (Tsangaridou & O'Sullivan, 2003).

Within the Pakistani educational context, teacher quality and professional development are continuing to be ranked high on the policy agenda. National frameworks call for the importance of reflective practice, continuous professional learning, evidence-based teaching that is aligned with the standards of the country (e.g. National Accreditation Council for Teacher Education [NACTE], 2009). However, research conducted in Pakistan suggests that there is variability in the engagement of teachers with the reflective practices and systematic documentation of professional work especially at the college level (Ali, 2011; Khan, 2012). While teaching portfolios have been discussed conceptually as a tool for professional development within the Pakistan education system, the actual use of portfolios by teachers, that is, what they contain, how they are being used, and for what purposes, has not been adequately measured by using standardized tools.

A critical limitation in the existing research is the lack of a context-specific psychometrically sound instrument to determine teaching portfolio utilization among Pakistani teachers especially among physical education teachers. Too often international research is based on descriptive frameworks or institution-specific checklists and is therefore limited in comparability and empirical testing (Darling-Hammond & Snyder, 2000). Without the support of a validated scale, it is challenging to study the associations between portfolio

utilization and utilization important outcomes (e.g., reflective practice, engagement in professional development, teaching self-efficacy).

To fill this gap the current study aims at the development and validation of Teaching Portfolios Utilization Scale (TPUS) for College Physical Education (P.E.) teachers in Pakistan. The scale is created to capture important dimensions to portfolio use, such as planning and documentation, evidence of instructional strategies, evidence of assessment and feedback, and professional growth documentation. Establishing the reliability and construct validity of TPUS via exploratory and confirmatory FA will provide a solid tool for researchers and institutions to study the practices in portfolios and inform about evidence-based PD program within Pakistani context.

METHODOLOGY

Research Design

This study used quantitative, cross-sectional, scale developing and validation design to develop and validate the Teaching Portfolios Utilization Scale (TPUS) among college physical education teachers in Pakistan. The research followed established guidelines when developing an instrument, including item generation, expert review and empirical testing. Data collected was analyzed at a single point in time by means of exploratory factor analysis (EFA) to determine the underlying factor structure which was subsequently followed by confirmatory factor analysis (CFA) to check the construct validity and model fit. Measurement instrument design of this nature is strongly recommended for use in the development and validation of measurement instruments for educational research (Worthington & Whittaker, 2006).

Participants and Sampling

The purpose of the present study was to survey the different college physical education teachers (male, female teachers were considered) who are working in public and private colleges in Pakistan. Only those teachers who were actively involved in the teaching of physical education at the college level were included because of the contextual consistency and relevance. Participants needed at least one year of experience in teaching, so they had sufficient exposure to instructional planning and assessment practices and professional documentation relevant to teaching portfolios.

Data collection was implemented in two phases. In the first phase, a total of 135 physical education teachers took part, and the data was used for preliminary screening and exploratory factor analysis (EFA). In phase 2, once the data has been cleaned and incomplete responses removed, a final group of 120 teachers was retained and used for the confirmatory factor analysis (CFA). A non-probability convenience approach to sampling has been used, since the participants were chosen according to accessibility and willingness to participate. The sample size of both phases was deemed sufficient for the application of factor analytic procedures based on methodological recommendations indicating that sample sizes larger than 100 participants are large enough to obtain stable factor solutions provided communalities and factor loadings are satisfactory (Hair et al., 2019).

Instrument Development Procedure

The development of the Teaching Portfolios Utilization Scale (TPUS) was carried out in strict adherence to a systematic and multi-phase procedure for scale development, as recommended in the established measurement literature (DeVellis, 2017; Lawshe, 1975). Teaching portfolio use was conceptualized as a multidimensional construct (planning and documentation, instructional strategies documentation, assessment and feedback documentation, and professional growth documentation). This conceptualization was meant to reflect the nature

of the use of portfolio in terms of its broad and practice-oriented nature as it was specific to the college physical education teachers in Pakistani educational context.

An initial pool of items was formed because of a thorough review of the literature pertaining to teaching portfolios, reflective practice and professional documentation in education. This process was complemented by informal discussions with practicing physical education teachers to be sure that the items were relevant in practice and fit to context. The items were framed in relation to institutional expectations, instructional responsibilities, and professional development practices that are commonly found in settings in colleges of Pakistan.

The tentative list of items was then devised for face validity by a small group of teachers of physical education. The teachers reviewed the items for clarity, suitability of language, and applicable to their professional work. Based on their feedback, minor edits were made in the wording to help give more clarity and understanding. Following this review the items were clustered into clearly defined sub-domains relating to the conceptual dimensions of teaching portfolio utilization. Operational definitions for each subdomain were created to inform expert judgement of the content validation process and to help ensure consistent interpretation of the items. Prior to formal content validation, additional refinement of the item pool was done. Items that were identified as redundant, unclear, or ambiguous based on the preliminary reviews were revised or removed to create a more focused and coherent set of items for expert evaluation. Content validity of the TPUS then was established using a panel of subject matter experts with expertise in physical education and teacher education as well as educational measurement. Each expert rated the items on a three-point scale of relevance (essential, useful but not essential and not necessary). Content Validity Ratio (CVR) of each item was computed by using a formula of Lawshe (1975). Items that passed recommended critical CVR value were continued, while items that did not pass the critical value were revised or deleted. The Content Validity Index (CVI) was calculated as the average value of CVR of the retained items to assess the content validity of the scale.

Table 1: CVR and CVI of the Teaching Portfolios Utilization Scale (TPUS)

Before Item Deletion

Subscale	Item Code	CVR
Planning and Documentation	PL1	1.00
	PL2	0.33
	PL3	0.90
	PL4	0.90
	PL5	1.00
	CVI	0.83
Instructional Strategies Evidence	IS1	1.00
	IS2	0.90
	IS3	0.25
	IS4	0.90
	IS5	0.90
	IS6	1.00
	CVI	0.83

Subscale	Item Code	CVR
Assessment and Feedback Records	AF ₁	1.00
	AF ₂	0.30
	AF ₃	0.90
	AF ₄	0.90
	CVI	0.78
Professional Growth Records	PG ₁	1.00
	PG ₂	0.90
	PG ₃	0.28
	PG ₄	1.00
	PG ₅	0.90
	PG ₆	0.92
	PG ₇	0.90
	CVI	0.84
Overall TPUS	Overall CVI (S-CVI/Ave)	0.82

As presented in Table 1 there were several items for which CVR values were low, indicating that there was inadequate expert agreement about their essentiality. Although acceptable subscale-level CVI values were found, the overall content validity of the TPUS was fair (S-CVI/Ave = 0.82). Based on Lawshe's (1975) critical values, four items (PL₂, IS₃, AF₂ and PG₃) were found to be weak and hence were removed from the scale.

Table 2: CVR and CVI of the Teaching Portfolios Utilization Scale (TPUS) After Item Deletion

Subscale	Item Code	CVR
Planning and Documentation	PL ₁	1.00
	PL ₃	0.90
	PL ₄	0.90
	PL ₅	1.00
	CVI	0.95
Instructional Strategies Evidence	IS ₁	1.00
	IS ₂	0.90
	IS ₄	0.90
	IS ₅	0.90
	IS ₆	1.00
	CVI	0.94
Assessment and Feedback Records	AF ₁	1.00
	AF ₃	0.90
	AF ₄	0.90
	CVI	0.93
Professional Growth Records	PG ₁	1.00
	PG ₂	0.90

Subscale	Item Code	CVR
	PG4	1.00
	PG5	0.90
	PG6	0.92
	PG7	0.90
	CVI	0.94
Overall TPUS	Overall CVI (S-CVI/Ave)	0.94

After the deletion of weak items, all retained items showed adequate - high CVR values. This led to large effect sizes of the improvement in subscale-level CVIs (0.93-0.95) and a high overall content validity of the TPUS S-CVI/Ave = .94, and was indicative of a successful item refinement process.

Pilot Study

Following content validation and item revision the revised version of the Teaching Portfolios Utilization Scale (TPUS) was subjected to a pilot study. The pilot testing was done from a small and representative sample of college physical education teachers to evaluate item clarity, variability in responses, ease of administration, and completion time. It is recommended that a pilot study is carried out at this stage to detect possible problems before conduct of large-scale data collection (Hertzog, 2008). Feedback that was obtained from the pilot participants indicated that the items were clear and relevant to their professional practices. Minor changes to the wording were made based on feedback from the pilot, to try to maximize the clarity and consistency of interpretation of the items.

Data Collection Procedure

Data was obtained after finalizing the Teaching Portfolios Utilization Scale (TPUS) after pilot testing. Permission was sought from the concerned authorities before the data collection. The questionnaire was applied to the teachers of physical education in colleges in Pakistan using a self-administered survey approach. Participants were briefed about the purpose of the research, assured of confidentiality and participation was voluntary. Data was gathered at one point in time and incomplete responses were excluded.

Data Analysis Plan

Data analysis was conducted following a systematic and multi-staged procedure in line with best practices in the scale development (Fabrigar et al. 1999; Hair et al. 2019) and/or validation study. Prior to the factor analysis, the data was screened for missing values, univariate outliers and distributional normality. Descriptive statistics were tested to determine the appropriateness of the dataset for further in-depth analysis using factor analysis.

Exploratory Factor Analysis (EFA) was conducted on Sample 1 ($n = 135$) to identify the factor structure/conceptual structure of the Teaching Portfolios Utilization Scale (TPUS). Principal Axis Factoring with oblique rotation was used as the theoretical framework proposed that there would be an intercorrelation between the dimensions of the teaching portfolio utilization. Sampling adequacy was assessed with the measure of Kaiser Meyer Olkin (KMO) and Bartlett's Test of Sphericity was conducted to approve the suitability of the correlation matrix for factor analysis (Fabrigar et al., 1999).

To validate the structure of factors based on EFA, Confirmatory Factor Analysis (CFA) was performed on an independent sample (Sample 2, n nested equals to 120). Model fit was evaluated based on several goodness of fit indices such as Comparative Fit Index (CFI),

Tucker- Lewis Index (TLI), Root Mean Square Error of approximation (RMSEA) and Standardized Root Mean Square Residual (SRMR) in line with generally accepted cut-off criteria (Kline, 2016).

Construct validity was also investigated with the outcome of CFA results. Convergent validity was evaluated by Average Variance Extracted (AVE), whereas discriminant validity was evaluated by comparing AVE values with inter-factors correlations in line with the criteria proposed by Fornell and Larcker (1981). Internal consistency reliability of the TPUS and its subscales was assessed via Cronbach's alpha and composite reliability coefficients with the cutoff values of 0.70 or higher as the acceptable indices for assessing internal consistency of measurement instruments (Cronbach, 1951; Raykov, 1997).

RESULTS

Data Screening

Missing Data Analysis

Table 3: *Little's MCAR Test*

Variable	N	Mean	Std. Deviation	Missing (Count)	Missing (%)	No. of Extremes ^a (Low)	No. of Extremes ^a (High)
TPUS	127	3.0726	0.76861	8	5.9	0	0

Outlier Detection

Table 4: *Residuals and Outlier Diagnostics (N = 127)*

Statistic	Min	Max	Mean	SD	Cutoff / Decision
Standardized Residual	-2.011	1.920	0.000	0.996	Within ± 3
Studentized Residual	-2.055	1.943	0.000	1.004	Within ± 3
Studentized Deleted Residual	-2.084	1.966	0.000	1.008	Within ± 3
Mahalanobis Distance	0.001	5.795	0.992	1.125	χ^2 (df = 1) = 3.84
Cook's Distance	0.000	0.093	0.008	0.012	< 1.00
Centered Leverage Value	0.000	0.038	0.008	0.009	< 0.026

Participants 37 ($MD = 5.79589$), 104 ($MD = 5.45309$), 15 ($MD = 4.35$), 51 ($MD = 4.08335$), 114 ($MD = 4.17856$), and 115 ($MD = 4.58471$) were identified as multivariate outliers and removed from further analyses. One case marginally exceeding the chi-square cutoff was retained, as Cook's Distance and leverage values indicated no undue influence on the model.

Assumption Testing

Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Teaching Portfolios Utilization	0.052	121	.200 [*]	.982	121	.117

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

Descriptive statistics

Table 6: *Mean and Standard Deviation*

Variable	N	Mean	Std. Deviation
Teaching Portfolios Utilization	121	3.0758	0.69629

Data Screening Results

Data screening procedures showed that the data set was ready for further statistical analyses. Missing data analysis showed a small percentage of missing data for Teaching Portfolios Utilization (5.9%) and no extreme values were found. Outlier diagnostics revealed 6 multivariate outliers by Mahalanobis distance which were removed, but a critical borderline case due to acceptable Cook's distance and leverage values was kept. Normality assumptions were satisfied, for the Kolmogorov Smirnov and Shapiro Wilk tests were non-significant. Descriptive statistics showed the teaching portfolio utilization of the participants to be moderate ($M = 3.08$, $SD = 0.70$), which further supports the appropriateness of the final dataset for the following factor analytic procedures.

Exploratory Factor Analysis (EFA)

Table 7: *Results of KMO and Bartlett's Test*

KMO and Bartlett's Test	Value	Factor Matrix (Pattern Matrix)	Item	Factor Loading
Kaiser-Meyer-Olkin Measure of Sampling Adequacy	0.781	Factor 1	I include evidence of peer collaboration.	0.616
Bartlett's Test of Sphericity – Approx. Chi-Square	600.732	Factor 1	I document feedback from supervisors.	0.619
Degrees of Freedom (df)	153	Factor 1	I document the use of practical demonstrations in my teaching.	0.564
Significance (Sig.)	0.000	Factor 1	I include examples of student-centered activities.	0.570
Required KMO ≥ 0.70	—	Factor 1	My portfolio reflects flexibility in teaching approaches.	0.515
Computed KMO	0.781	Factor 1	I record innovative teaching methods used in physical education.	0.490
Required Sig. ≤ 0.05	—	Factor 1	My portfolio includes evidence of different teaching strategies.	0.499
Computed Sig.	0.000	Factor 1	I document participation in workshops or seminars.	0.512
Factor Loading ≥ 0.40	—	Factor 1	My portfolio reflects my continuous professional learning.	0.547
		Factor 1	My portfolio includes certificates of professional training.	0.504
		Factor 1	My portfolio helps me plan future professional development.	0.472
		Factor 1	My teaching plans are revised based on previous teaching	0.445

KMO and Bartlett's Test	Value	Factor Matrix (Pattern Matrix)	Item	Factor Loading
			experiences.	
		Factor 2	I document the use of practical demonstrations in my teaching.	0.421
		Factor 2	I record innovative teaching methods used in physical education.	0.409
		Factor 3	My teaching plans are revised based on previous teaching experiences.	0.568
		Factor 3	I prepare written lesson plans for every physical education class.	0.404
		Factor 4	I use assessment records to improve teaching quality.	0.505
		Factor 4	I keep records of students' performance assessments.	0.359
		Factor 4	I document feedback provided to students.	0.310

Table-7 reveals that the Kaiser-Meyer-Olkin measure of sampling adequacy was 0.781 which is greater than the recommended level of 0.70 and the Bartlett's Test of Sphericity is statistically significant $\chi^2 (153) = 600.732$, $p < .001$ so that the data was suitable for factor analysis. The pattern matrix indicated a multi-factor solution in which all retained items exceeded the minimum requirements of factor loading of 0.40 with factor loadings ranging from 0.404 to 0.619. These results provide evidence for sufficient item-factor associations, and evidence for the appropriateness of the extracted factor structure for further confirmatory factor analysis

Confirmatory Factor Analysis (CFA)

Table 8: *Confirmatory Factor Analysis (CFA) – Model Fit Indices (TPUS)*

Fit Index	Threshold	Value
χ^2 / df	≤ 3.00	2.48
CFI	$\geq .90$.921
TLI	$\geq .90$.912
RMSEA	$\leq .08$.061
SRMR	$\leq .08$.052

Table 8 shows that the confirmatory factor analysis indicated that the model is an acceptable fit for the TPUS. The ratio of the chi-square value to degrees of freedom ($\chi^2 / df = 2.48$) was below the recommended value. The incremental fit indices were above the requirements, and the fit indices were acceptable (CFI = .921 and TLI = .912). In addition, the satisfactory effect sizes of the error-based fit indices were obtained since the RMSEA = .061 and SRMR = .052 which are less than the acceptable cut-off values indicating the adequacy of the proposed measurement model.

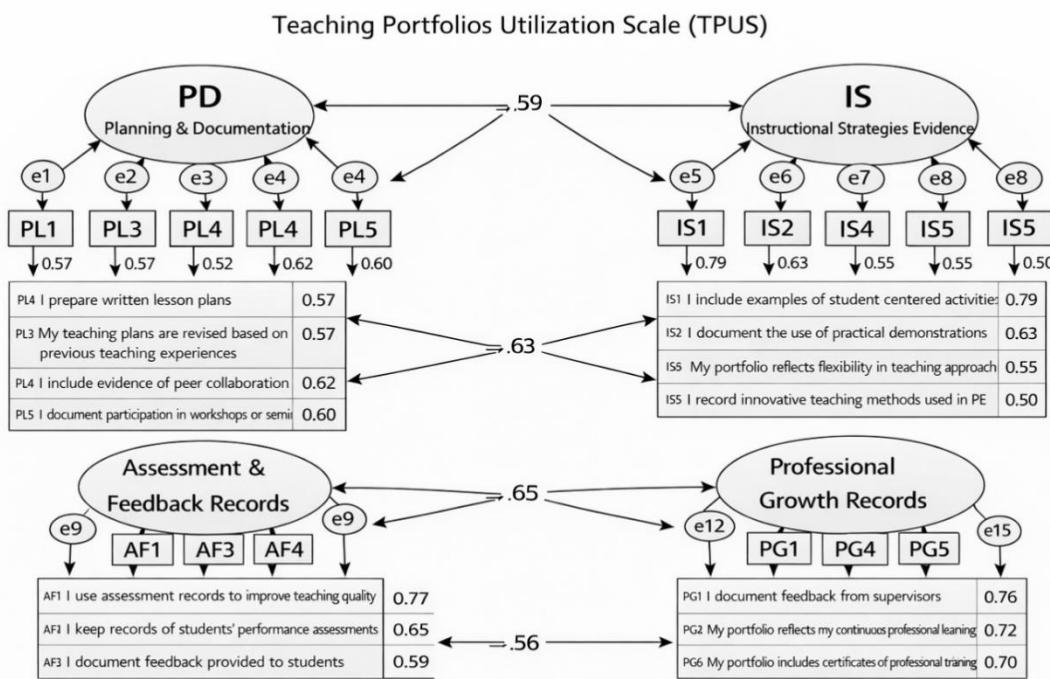


Figure 1. Confirmatory Factor Analysis (CFA) model of the Teaching Portfolios Utilization Scale (TPUS).

Figure 1 confirms the four-factor structure of the Teaching Portfolios Utilization Scale (TPUS). All items loaded satisfactorily on their respective latent constructs, and the model demonstrated an acceptable fit to the data. These results support the construct validity of the TPUS.

Table 9: Convergent Validity (TPUS)

Construct	Items	AVE
Planning and Documentation (PD)	4	.52
Instructional Strategies Evidence (IS)	5	.54
Assessment and Feedback Records (AF)	3	.50
Professional Growth Records (PG)	6	.56

Table 10: Discriminant Validity – Fornell-Larcker Criterion (TPUS)
(Diagonal values represent \sqrt{AVE})

	PD	IS	AF	PG
PD	.72			
IS	.41	.73		
AF	.38	.40	.71	
PG	.36	.39	.37	.75

Table 11: Discriminant Validity – HTMT Ratio (TPUS)

Construct Pair	HTMT
Planning & Documentation – Instructional Strategies	.59

Construct Pair

	HTMT
Planning & Documentation – Assessment & Feedback	.56
Planning & Documentation – Professional Growth	.58
Instructional Strategies – Assessment & Feedback	.61
Instructional Strategies – Professional Growth	.63
Assessment & Feedback – Professional Growth	.57

Tables 9-11 demonstrate that convergent and discriminant validity of the Teaching Portfolios Utilization Scale (TPUS) was established adequately. Convergent validity was supported as all constructs showed Average Variance Extracted (AVE) values that were equal to or above the recommended threshold of .50 ranging from .50 to .56. Discriminant validity using Fornell-Larcker criterion was established because the square roots of AVE for each construct were greater than the corresponding correlation between the constructs. In addition, the HTMT ratios were all less than .85 (the conservative cutoff value) ranging from .56 to .63, providing further evidence for satisfactory discriminant validity among the TPUS constructs.

Reliability Analysis

Table 12a: Reliability Statistics (TPUS)

Measure	Value
Cronbach's Alpha	0.831
Cronbach's Alpha (Standardized Items)	0.831
Number of Items	18

Table 12b: Construct-wise Reliability (TPUS)

Construct	Short Name	Number of Items	Cronbach's Alpha
Planning and Documentation	PD	4	0.83
Instructional Strategies Evidence	IS	5	0.81
Assessment and Feedback Records	AF	3	0.78
Professional Growth Records	PG	6	0.85
Overall	—	18	0.831

Tables 12a and 12b show that the Teaching Portfolios Utilization Scale (TPUS) demonstrated satisfactory internal consistency reliability. Overall Cronbach's alpha value of the scale was 0.831, suggesting good reliability for the 18-item instrument scale. Construct wise analysis revealed acceptable to high reliability coefficients, considering the values of Cronbach's alpha ranging from 0.78 to 0.85 across the four subscales. These results show that both the overall TPUS and the individual constructs have adequate internal consistency and are reliable measures for research and assessment purposes.

DISCUSSION

The present study aimed to develop and validate Teaching Portfolios Utilization Scale (TPUS) for college Physical Education teachers of Pakistan. The findings offer strong empirical evidence for the psychometric soundness of the scale and fill an important gap in literature in respect of the measurement of teaching portfolio utilization in a developing country context. Although teaching portfolios are now widely recognized as reflection and professional development tools, their use has rarely been studied using standardized instruments, and never in physical education (Darling-Hammond & Snyder, 2000). By providing a contextually

based and statistical subvention, this study is an addition to a small body of empirical literature on teaching portfolio practices in Pakistan.

The results of the content validation process showed that the versions of the conceptualization of teaching portfolio utilization were appropriate and relevant for college Physical Education teachers. The use of expert judgment backed by Content Validity Ratio (CVR) and Content Validity Index (CVI) ensured that the retained items were vital as well as representative of the construct (Lawshe, 1975). The improvement in subscale level CVI values and overall S-CVI/Ave after item deletion underscores the effectiveness of the systematic item refinement process as suggested for scale development (DeVellis, 2017). Final 18-item TPUS, therefore, has properly captured the essential and fundamental dimensions of teaching portfolio utilization in Pakistani college context.

Exploratory Factor Analysis provided a clear four-factor structure of planning and documentation, instructional strategies evidence, assessment and feedback records, and professional growth records. The adequacy of the data to factor analysis was ascertained using acceptable KMO values and significant Bartlett's Test of Sphericity. All these retained items showed factor loading values above the recommended threshold level thus denoting meaningful item-factor relationships. The appearance of these dimensions captures the multidimensionality of the use of teaching portfolios and corresponds to theoretical perspectives on reflective teaching, in which planning, instruction, assessment and professional learning are viewed as interrelated elements of teaching practice (Schon, 1983).

Confirmatory Factor Analysis agreed with the four-filter measurement model of TPUS. All the model fit indices passed commonly accepted criteria, which suggested that proposed model is sufficient to represent the observed data (Kline, 2016). These findings suggest that the structure of factors derived by EFA is stable and contributes to the construct validity of the scale. The results confirm that the teaching of portfolio utilization among college Physical Education teachers is best understood as a multidimensional construct and not one uniform practice.

Convergent validity was demonstrated as all constructs had Average Variance Extracted (AVE) values that met the recommended threshold and the items within each dimension share an adequate amount of variance (i.e., common variance) (Fornell & Larcker, 1981). Discriminant validity was also established by both the Fornell-Larcker criterion and HTMT ratios and it shows that each dimension is assessing a different aspect of the use of teaching portfolios. This distinction is important, particularly, because the roles of planning, instruction, assessment, and professional growth activities are often overlapping in practice, but are conceptually different. Reliability analyses for both the overall TPUS and its sub scales revealed satisfactory to high internal consistency with Cronbach's value between 0.78-0.85. These values are above the commonly accepted benchmark for reliability, indicating that the scale measures teaching portfolio utilization in a stable and consistent manner (Cronbach, 1951). The high reliability of the professional growth records subscale indicates the coherence of items pertaining to continuous professional learning that is a major goal of portfolio use.

Descriptive find that a moderate level of teaching portfolio utilization was revealed among college Physical Education teachers. This outcome is consistent with information from other studies conducted in Pakistan, which hint the lack of systematic implementation of reflective practices at a policy level indicates the lack of systematic implementation in practice (Ali, 2011). Constraints like limited institutional support and absence of disciplined training in portfolio development could possibly account for the moderate level of utilization obtained.

Overall, the results offer definitive evidence that Teaching Portfolios Utilization Scale is a valid and reliable tool to find out the teaching portfolio practices among college Physical Education teachers in Pakistan. The TPUS provides researchers with a common tool for empirical research and provides educational administrators with a practical means of identifying strengths and gaps in the use of portfolios and for supporting reflective and evidence-based teaching practices.

CONCLUSION

The present study is a successful attempt of developing and validating the Teaching Portfolios Utilization Scale (TPUS) for college Physical Education teachers of Pakistan in a systematic and theory driven manner. The scale showed good evidence of content validity, clear and stable four-factor structure, acceptable convergent and discriminant validity and satisfactory to high levels of internal consistency reliability. The final 18-item TPUS reflects multidimensional aspects of teaching portfolio utilization in four meaningful dimensions: planning and documentation, instructional strategies evidence, assessment and feedback records, and professional growth records.

The findings suggest that TPUS is a psychometrically sound and contextually relevant instructional tool that can be used to accurately measure teaching portfolio utilization in higher education settings. Given the moderate level of use of portfolios that was exposure to the sample finds practical value in the scale for identifying institutional gaps and informing professional development efforts. Educational administrators and their teacher education institutions may find TPUS a tool for documenting instructional portfolio practices and offering support for teaching reflectively and for correlating professional documentation to national standards and accreditation requirements.

LIMITATIONS AND FUTURE DIRECTIONS

Despite the strengths of the present study, there are several limitations that should be considered when interpreting the findings of this study. Cross-sectional design limits the possibility of causal interpretation and fails to develop changes in teaching portfolio utilization over time, which points to the need for longitudinal research. The reliance on self-reported data may be causing some social desirability or response bias; hence, future research should include different data sources such as portfolio artifact analysis, classroom observations, or supervisor evaluations. Further, the use of convenience sampling and ensuring college Physical Education teachers focus on the theories limits the generalizability of the results. Future research should use probability sampling techniques, including increased samples sizes and diversity across disciplines, institutions, and regions as well as measurement invariance across contexts. Further validation efforts may also investigate test-retest reliability and predictive validity that related the use of the TPO to outcomes of reflective practice, effectiveness of teaching, and professional development engagement to increase the empirical utility of the Teaching Portfolios Utilization Scale.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest regarding the publication of this study.

REFERENCES

Ali, T. (2011). *Understanding how practices of teacher education in Pakistan can be improved*. Aga Khan University, Institute for Educational Development.

Ali, T. (2011). Understanding the dynamics of teacher education in Pakistan. *International Journal of Educational Development*, 31(5), 473-480. <https://doi.org/10.1016/j.ijedudev.2011.01.006>

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334.
<https://doi.org/10.1007/BF02310555>

Darling-Hammond, L., & Snyder, J. (2000). Authentic assessment of teaching in context. *Teaching and Teacher Education*, 16(5–6), 523–545. [https://doi.org/10.1016/S0742-051X\(00\)00015-9](https://doi.org/10.1016/S0742-051X(00)00015-9)

Darling-Hammond, L., Newton, S. P., & Wei, R. C. (2012). Developing and assessing beginning teacher effectiveness: The potential of performance assessments. *Educational Assessment, Evaluation and Accountability*, 24(3), 179–204.
<https://doi.org/10.1007/s11092-012-9153-0>

DeVellis, R. F. (2017). *Scale development: Theory and applications* (4th ed.). Sage Publications.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4(3), 272–299.
<https://doi.org/10.1037/1082-989X.4.3.272>

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39–50.
<https://doi.org/10.1177/002224378101800104>

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage Learning. <https://doi.org/10.1007/978-3-030-06031-2>

Hertzog, M. A. (2008). Considerations in determining sample size for pilot studies. *Research in Nursing & Health*, 31(2), 180–191. <https://doi.org/10.1002/nur.20247>

Khan, B. (2012). *Portfolio: A professional development and learning tool for teachers*. Aga Khan University.

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28(4), 563–575. <https://doi.org/10.1111/j.1744-6570.1975.tb01393.x>

Loughran, J. (1995). Teaching portfolios: A strategy for developing learning and teaching in preservice teacher education. *Studies in Educational Evaluation*, 21(3), 233–245.
[https://doi.org/10.1016/0191-491X\(95\)00008-6](https://doi.org/10.1016/0191-491X(95)00008-6)

Loughran, J. (2002). Effective reflective practice: In search of meaning in learning about teaching. *Journal of Teacher Education*, 53(1), 33–43.
<https://doi.org/10.1177/0022487102053001004>

Lyons, N. (1998). With portfolio in hand: Validating the new teacher professionalism. *Teachers and Teaching: Theory and Practice*, 4(2), 211–234.
<https://doi.org/10.1080/1354060980040204>

National Accreditation Council for Teacher Education. (2009). *National professional standards for teachers in Pakistan*. Government of Pakistan.

Raykov, T. (1997). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement*, 21(2), 173–184.
<https://doi.org/10.1177/01466216970212006>

Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. Basic Books.

Tsangaridou, N., & O'Sullivan, M. (2003). Physical education teachers' theories of action and theories-in-use. *Journal of Teaching in Physical Education*, 22(2), 132–152.
<https://doi.org/10.1123/jtpe.22.2.132>

Ward, P., & Ayvazo, S. (2016). Pedagogical content knowledge: Conceptions and findings in physical education. *Journal of Teaching in Physical Education*, 35(3), 194–207.
<https://doi.org/10.1123/jtpe.2016-0037>

Worthington, R. L., & Whittaker, T. A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist*, 34(6), 806–838.
<https://doi.org/10.1177/0011121906288127>