International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18 (2026) pp. 1-10

https://doi.org/10.70917/ijcisim-2026-4944
Article

Local Adaptation and CFA Validation of the Professional Development Programme Design (PDPD) Scale for Chinese Lower-Secondary Schools

Yajie Zou [1,*], Jamil Bin Ahmad [1] and Bity Salwana Alias [1]

[1] Faculty of Education, Universiti Kebangsaan Malaysia, Malaysia
* Correspondence author: p147269@siswa.ukm.edu.my

Abstract: Valid measurement of professional development (PD) programme design is essential for identifying how programme features support teacher learning and for accumulating comparable evidence across studies. However, instruments that retain the classic distinction between core and structural PD features remain insufficiently adapted and structurally validated for Chinese secondary-school contexts. To address this gap, This study adapted the Professional Development Programme Design (PDPD) scale for Chinese secondary schools and evaluated its structural validity. After translation–back translation, expert review, and cognitive pretesting, 333 in-service lower-secondary teachers from Hubei Province, China, completed the Chinese PDPD. Using confirmatory factor analysis in a theory-driven workflow and competing-model comparisons, we tested the original five-factor model (Content Focus, Active Learning, Coherence, Duration, and Collective Participation) against an alternative model that merged Coherence and Duration. Both models showed acceptable global fit; however, Coherence and Duration were nearly indistinguishable (latent $r \approx 0.95$) and two coherence items displayed unstable loadings. After removing four problematic items and combining Coherence and Duration, the final 24-item, four-factor model (Content Focus, Active Learning, Structural Features, and Collective Participation) demonstrated good fit and strong internal consistency ($\alpha \geq 0.80$). Composite reliability exceeded 0.86 and average variance extracted exceeded 0.50 for all factors, supporting convergent validity, while inter-factor correlations highlighted the need for further discriminant and invariance testing. The study contributes context-specific evidence that Coherence and Duration may function as an integrated Structural Features dimension in Chinese secondary schools, while providing a parsimonious instrument for future research on PD quality and outcomes.

Keywords: teacher professional development; scale adaptation; confirmatory factor analysis; Professional Development Programme Design (PDPD) scale; Chinese lower-secondary teachers

1. Introduction

Teacher professional development is commonly viewed as a key mechanism for enhancing teachers' knowledge, instructional practice, and student learning; however, its effectiveness often hinges on interactions among programme design quality, teachers' motivation, and school environments [1-2]. Early large-scale empirical studies suggested that core features such as content focus, active learning, and coherence are significantly associated with teacher learning outcomes, whereas structural features such as activity format, collective participation, and duration tend to influence teacher learning indirectly through core features [3-4]. Building on this evidence, Desimone proposed a conceptual framework linking PD features to teacher learning, instructional practice, and ultimately student outcomes, and argued that stronger conceptualisation and measurement are essential for improving the explanatory power and cumulative nature of PD impact research [5].

In recent years, scholarship has also problematised the putative “consensus list” of effective PD features. First, studies vary in how they define and operationalise particular features. Coherence, for example, has been conceptualised as external alignment, internal alignment, or negotiated goal coherence, among other forms, raising questions about the robustness of the claimed consensus [6]. Second, several reviews have noted limitations in evidence synthesis methods and inferential logic, suggesting that conclusions about “which features work” are not always reliable [7-8]. At the same time, emerging theoretical and meta-analytic work emphasises that effective PD may consist of multiple combinations of causally active components, and that mechanisms often operate through processual pathways such as motivating teachers, developing practice, and embedding learning in workplace contexts [9]. Collectively, these developments highlight a central issue: even when a framework has broad explanatory utility, its measurement tools must be validated in specific cultural and institutional settings [5] [10].

In Chinese secondary schools, teacher learning frequently occurs through organisational vehicles such as teaching and research groups, lesson planning teams, and school-based professional learning initiatives. Such professional learning communities may foster collaboration, yet they are simultaneously shaped by leadership practices, collegial trust, and institutional arrangements [11-12]. Recent shifts in policy and governance logics may also reshape teachers’ time structures for work and learning, thereby influencing how teachers perceive design elements such as duration, frequency, and coherence as an integrated whole [13]. For instance, teachers have been found to be particularly dissatisfied with poorly organised PD in terms of “time, duration, and frequency”, often co-occurring with issues related to appraisal and management as well as goal setting and coordination [14]. These findings imply that, in Chinese school management practice, teachers’ experience of coherence may not always manifest as a distinct dimension; instead, it may be intertwined with structural conditions such as duration, frequency, and organisational articulation, forming a more composite perception of structural features [4] [6].

Although PD quality scales for routine evaluation have been developed and validated in large samples internationally [15], instruments that are appropriate for Chinese secondary-school contexts while explicitly aligning with the classic core/structural feature distinction remain scarce. Moreover, recent methodological discussions emphasise that scale localisation should be supported by a transparent chain of validity evidence—covering construct definition and translation equivalence, content validity, and structural validity testing with clear reporting—while minimising researcher degrees of freedom [16-19].

Against this backdrop, the present study adapts and validates the Professional Development Programme Design (PDPD) scale and evaluates its structural validity among lower-secondary teachers in Hubei Province, China. Specifically, it aims to: (1) test the structural fit of the original five-factor measurement model, including overall model fit and construct distinctiveness; (2) compare competing models (including an alternative specification that merges Coherence and Duration) and determine the final factor structure and item set based on model diagnostics and item performance; and (3) report reliability and structural validity evidence for the final model (e.g., internal consistency, CR/AVE, and latent correlations), and outline further validation needs such as measurement invariance testing for group comparisons.

2. Method

2.1. Participants and Procedure

Participants were in-service lower-secondary-school teachers recruited through convenience sampling in Hubei Province, China. Paper questionnaires were administered face-to-face. Part of the data was collected during a teacher professional development training meeting attended by more than 100 teachers from different schools, while the remaining questionnaires were administered during teacher meetings at several lower-secondary schools in Hubei Province with the assistance of school coordinators. Completed questionnaires were collected on site. A total of 376 questionnaires were distributed and collected. After screening the questionnaires for completeness and response quality, 43 questionnaires were excluded because they were incomplete or unsuitable for further analysis, leaving 333 valid responses and an effective response rate of 88.6%. The final sample of 333 participants was considered adequate for CFA estimation and the comparison of competing measurement models [20]. Before questionnaire administration, all participants were informed of the study purpose, the voluntary nature of participation, their right to withdraw, and the intended use of the data. Informed consent was obtained, no personally identifiable information was collected, and the findings were reported only in aggregate form.

2.2. Measures

2.2.1. Instrument Source and Construct Framework

The PDPD is grounded in the classic distinction between core and structural features of professional development [4-5]. Item wording was adapted from the Characteristics of Teacher Professional Development (CTPD) instrument developed by Soine and Lumpe (2014), which operationalises these features for survey-based measurement [21]. For consistency with the focus on programme design, the adapted instrument is referred to as the PDPD scale in the present study.

2.2.2. Translation and Cultural Adaptation

To support cross-language measurement equivalence, the study adopted a translation–back translation procedure and iteratively refined items with the goal of semantic and functional equivalence [22]. After translation, domain experts reviewed the items for content validity, focusing on relevance, conceptual clarity and contextual appropriateness, and provided suggestions for revision. Based on this feedback, wording was adapted to better reflect organisational contexts and everyday expressions in Chinese secondary schools. This evidence-building process follows established perspectives on content and construct validity, which emphasise alignment between item content and theoretical definitions prior to statistical testing [18] [23]. Consistent with this approach, a separate study has reported the instrument's content validity evidence for secondary education [24].

2.2.3. Cognitive Pretesting and Item Refinement

Before the main survey, a small-scale cognitive pretest was conducted in addition to expert content review to examine item comprehensibility, semantic equivalence, and response feasibility among the target population [25-26]. Six in-service secondary-school teachers from different subject areas and career stages were invited to read and complete the questionnaire, after which semi-structured feedback was collected item by item. The pretest focused on: (1) whether any items were ambiguous or unclear; (2) whether wording matched teachers' everyday work language; (3) whether response options were easy to understand and differentiate; and (4) whether the overall length and response burden were acceptable.

Overall, most items were perceived as clear and the response process was smooth. A small number of items required further cultural adjustment in terms of wording and contextual fit. For example, some items related to coherence were considered abstract when translated literally; teachers tended to interpret coherence through the time structure of PD arrangements, continuity of frequency, and organisational coordination. Accordingly, the research team made minor wording adjustments to improve comprehensibility and face validity in the Chinese secondary-school context.

Notably, this stage aimed to iteratively improve item wording rather than to conduct a psychometric pilot test or statistical item screening; therefore, pretest responses were not retained as analytic data. Together, the cognitive pretest and expert review constitute complementary evidence at the levels of response processes and content validity: the former illuminates how the target group actually understands and responds to items, whereas the latter evaluates alignment between item content and theoretical definitions [18] [23]. After dual revisions based on expert review and cognitive pretesting, the final Chinese version of the scale was prepared for the main survey.

2.3. Data Analysis

2.3.1. Analytical Procedure and Item Refinement

To develop a parsimonious yet psychometrically robust PDPD scale for Chinese secondary schools, this study followed a theory-informed sequential workflow comprising data screening, auxiliary exploratory diagnostics and preliminary item screening, comparison of competing CFA models, iterative model evaluation, and final reliability and validity assessment. Figure 1 summarises the six-step analytical workflow and the iterative refinement logic used in this study. Exploratory analyses were used only as auxiliary diagnostics to identify clearly misfitting items and to check structural complexity, rather than as the primary basis for structural conclusions [16] [19-20]. The overall procedure comprised six steps:

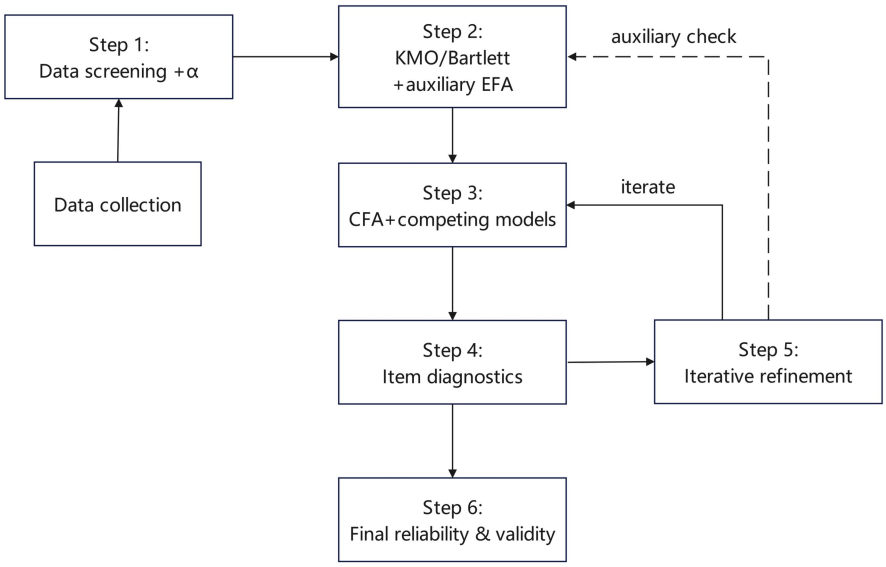

Figure 1. Analytical procedure and item refinement

Step 1: Data screening and preliminary reliability: Data quality was examined through basic distributional checks (missing values, skewness, kurtosis, and extreme values). Cronbach's alpha was calculated to assess internal consistency. Although alpha values of 0.70 or above are commonly considered acceptable [27], alpha is sensitive to the number of items and dimensionality; therefore, it was used only as a preliminary screening index and interpreted together with structural validity evidence [28].

Step 2: Factorability check and auxiliary exploratory diagnostics: To ensure that the data were suitable for latent-structure modelling, the Kaiser-Meyer-Olkin (KMO) measure and Bartlett's test of sphericity were reported. Where an initial diagnosis of structural complexity was necessary, auxiliary exploratory checks (e.g., principal axis factoring with oblique rotation) were conducted to identify items with low loadings, obvious cross-loadings, or unclear factor membership, thereby providing references for subsequent CFA specification and item refinement [29]. Importantly, structural conclusions in this study were based primarily on CFA and competing-model evidence; exploratory results were used only as supplemental diagnostics.

Step 3: CFA and competing-model comparison: Confirmatory factor analysis (CFA) was conducted in AMOS to evaluate model fit, beginning with theory-based specifications. Because the chi-square statistic is sensitive to sample size, model fit was judged using a combination of indices, including chi-square/df, RMSEA, CFI, and TLI [30-31]. Model specification, parameter estimates, and fit conclusions were reported in line with SEM reporting recommendations [19]. In competing-model comparisons, beyond global fit, latent-factor correlations and potential threats to discriminant validity were examined to avoid unnecessary dimensional splitting among highly redundant constructs [32].

Step 4: Item contribution and model diagnostics: When local misfit or weak construct separation emerged, item- and model-level diagnostics were conducted. Decisions were guided by standardised factor loadings (with low-loading items prioritised for deletion or revision), standardised residuals and local fit information, potential cross-loading risks, and item–total correlations, while also considering theoretical coherence and content appropriateness to preserve clear construct representation and discriminant validity [29] [33].

Step 5: Identify misfitting items and iterate: Under the premise of maintaining theoretical alignment and controlling researcher degrees of freedom, a small number of candidate items were deleted or structural adjustments were made. After each modification, CFA and competing-model comparisons were repeated until a model with a clear structure, stable parameter estimates, and acceptable fit was obtained [16] [20]. Auxiliary exploratory checks were used only when necessary to corroborate clearly misfitting items and were not treated as the primary basis for model decisions.

Step 6: Final reliability and validity assessment: After the final model was established, internal consistency (Cronbach's alpha) was reported for each dimension. Based on CFA results, additional evidence for convergent and discriminant validity was provided (e.g., composite reliability [CR], average variance extracted [AVE], and latent correlations) [32]. For subsequent studies involving group comparisons (e.g., gender or teaching experience), measurement invariance should be tested in an independent or larger sample to ensure valid interpretation of between-group differences [34-36].

2.3.2. Statistical Software

SPSS was used for data screening, descriptive statistics, internal consistency estimates, and factorability tests and auxiliary exploratory diagnostics (e.g., KMO, Bartlett's test, and exploratory structure checks when needed). AMOS was used for CFA and competing-model comparisons. Model evaluation relied on commonly used SEM fit indices and methodological guidance regarding their interpretation [30-31]. Bias-corrected bootstrap (2,000 resamples) was used to obtain 95% confidence intervals for standardised factor loadings.

3. Results and Discussion

3.1. Reliability Analysis

After establishing the final measurement structure, internal consistency was examined for the revised PDPD scale (24 items; four dimensions: Content Focus, Active Learning (Table 1), Coherence+Duration, and Collective Participation). All Cronbach's alpha coefficients indicated good reliability ($\alpha \geq 0.80$), suggesting strong homogeneity and stability within each dimension. It should be noted, however, that alpha is influenced by the number of items and the dimensional structure and should therefore be interpreted in conjunction with structural validity evidence [27].

Table 1. Reliability of the revised PDPD scale

Dimension	**Items**	**Cronbach's α**
Content Focus (A1_1–A1_5)	5	0.894
Active Learning (A2_1–A2_5)	5	0.866
Coherence+Duration (A3_3–A3_5 + A4_1–A4_6)	9	0.937
Collective Participation (A5_3–A5_7)	5	0.887

Overall, the four subscales demonstrated high internal consistency, meeting basic requirements for subsequent validity testing and model comparisons.

3.2. Results

3.2.1. Construct validity of the PDPD scale

EFA was used only as an auxiliary diagnostic to flag clearly problematic items and to gauge structural complexity; substantive conclusions were drawn primarily from CFA and competing-model evidence. Accordingly, an exploratory factor analysis (principal axis factoring with Promax rotation) was conducted to support item screening. Sampling adequacy was excellent (KMO = 0.966–0.970), and Bartlett's test of sphericity was significant across solutions ($p < 0.001$), indicating that the data were suitable for factor analysis [29].

Diagnostic variants focused on items A3.1, A3.2, A5.1, and A5.2, based on three criteria: (a) low or non-salient primary loadings and/or substantial cross-loadings; (b) interpretability and stability of the rotated solution; and (c) alignment with the intended theoretical dimensions. Across rotations, A3.1 and A3.2 failed to exhibit stable primary loadings at commonly used thresholds (≥ 0.40/0.50), suggesting limited representation of the intended Coherence construct in this sample. In addition, A5.1 and A5.2 increased structural complexity and blurred factor boundaries. After removing these four items, the exploratory structure became clearer and more stable. In the resulting four-cluster solution, items from the original Coherence subscale (A3_3–A3_5) and the original Duration subscale (A4_1–A4_6) consistently loaded together, indicating an integrated Structural Features cluster. The remaining item sets formed coherent clusters corresponding to Content Focus (A1_1–A1_5), Active Learning (A2_1–A2_5), and Collective Participation (A5_3–A5_7).

For the final diagnostic solution, KMO remained high (0.966) and Bartlett's χ^2 (276) = 5671.287 ($p < 0.001$). The four clusters accounted for approximately 62.361% of the total variance (extraction sums of squared loadings, unrotated). Taken together, these auxiliary results supported proceeding to CFA with a theoretically informed competing-model comparison, including an alternative specification in which Coherence and Duration were modelled as a single Structural Features factor [20] [29].

3.2.2. Model fit of hypothesised models

Building on the theory-based model specification and item-level diagnostics, CFA was conducted in AMOS to compare two competing measurement models (Table 2). M5 was specified as a five-factor model comprising Content Focus, Active Learning, Coherence, Duration, and Collective Participation. M4 specified a four-factor alternative in which Coherence and Duration were combined into a single factor (Coherence+Duration), with the remaining factors unchanged. Given the sensitivity of χ^2 to sample size, model fit was evaluated using multiple indices, including χ^2/df, RMSEA, CFI, and TLI [19] [30-31]. Both CFA models were estimated on the refined 24-item set identified in the diagnostic stage.

Table 2. Fit indices for the five-factor and four-factor models (N = 333).

Model	χ^2	df	χ^2/df	CFI	TLI	RMSEA	GFI	RMR
Five-factor (M5)	406.142	242	1.678	0.970	0.966	0.045	0.909	0.026
Four-factor (M4: A3 + A4)	434.590	246	1.767	0.966	0.962	0.048	0.904	0.028

Both models met commonly used criteria for acceptable fit (CFI/TLI close to or above 0.90, RMSEA < 0.06, χ^2/df < 3). Notably, however, the five-factor model yielded an extremely high latent correlation between Coherence and Duration (r ≈ 0.95), suggesting limited construct separation and a substantial discriminant validity concern. In this situation, retaining two near-redundant factors is difficult to justify; therefore, Coherence and Duration were combined into a single Structural Features factor in the competing four-factor specification, which provides a more parsimonious and defensible measurement structure [32].

3.2.3. Convergent and discriminant validity (CR/AVE and latent correlations)

To examine the measurement quality of the final four-factor model, Table 3 reports the standardised factor loadings (λ), bias-corrected bootstrap 95% confidence intervals, and squared multiple correlations (R^2) for all retained items.Latent factor correlations ranged from 0.70 to 0.86, indicating substantial overlap among dimensions.

Table 3. Standardised factor loadings for the final four-factor CFA model

Factor	Indicator	λ	95% CI	R^2 (SMC)
Factor 1: Content Focus	A1_1	0.799	[0.723, 0.856]	0.639
	A1_2	0.775	[0.701, 0.835]	0.600
	A1_3	0.85	[0.804, 0.890]	0.723
	A1_4	0.772	[0.705, 0.822]	0.596
	A1_5	0.779	[0.718, 0.830]	0.607
Factor 2: Active Learning	A2_1	0.704	[0.598, 0.778]	0.496
	A2_2	0.741	[0.668, 0.803]	0.549
	A2_3	0.791	[0.732, 0.839]	0.626
	A2_4	0.822	[0.767, 0.864]	0.676
	A2_5	0.711	[0.637, 0.773]	0.506
Factor 3: Structural Features (Coherence+Duration)	A3_3	0.806	[0.752, 0.851]	0.650
	A3_4	0.768	[0.710, 0.815]	0.590
	A3_5	0.794	[0.738, 0.836]	0.631
	A4_1	0.799	[0.748, 0.845]	0.639
	A4_2	0.833	[0.790, 0.872]	0.694
	A4_3	0.804	[0.737, 0.858]	0.647
	A4_4	0.785	[0.729, 0.829]	0.616
	A4_5	0.752	[0.688, 0.809]	0.566
	A4_6	0.767	[0.701, 0.816]	0.589
Factor 4: Collective Participation	A5_3	0.788	[0.732, 0.832]	0.621
	A5_4	0.742	[0.651, 0.810]	0.550
	A5_5	0.752	[0.685, 0.808]	0.566
	A5_6	0.8	[0.741, 0.847]	0.641
	A5_7	0.832	[0.778, 0.873]	0.693

Note. Standardised loadings are from AMOS Standardized Regression Weights. Confidence intervals are bias-corrected bootstrap 95% CIs (2,000 resamples). R^2 values are squared multiple correlations (SMC).

Based on these standardised loadings, composite reliability (CR) and average variance extracted (AVE) were calculated to evaluate convergent validity (Table 4).

Table 4. CR and AVE for the four-factor model (based on standardised loadings).

Dimension	**CR**	**AVE**
Content Focus	0.896	0.633
Active Learning	0.869	0.570
Coherence+Duration	0.937	0.624
Collective Participation	0.888	0.614

As shown in Table 4, all CR values exceeded 0.80 and all AVE values exceeded 0.50, indicating adequate convergent validity. At the same time, latent correlations among dimensions were generally high, which is consistent with the theoretical reality that PD design features are often tightly coupled and tend to co-vary [4-5]. Nevertheless, these results also suggest that future studies should further scrutinise discriminant validity using stricter evidence and should test measurement invariance across groups when group comparisons are intended [34-36].

3.3. Discussion

The central finding of this study is that, in the present sample of lower-secondary teachers from Hubei Province, China, the PDPD scale yielded a four-factor structure after item screening, and that the originally theorised Coherence and Duration factors were empirically highly integrated. This pattern is both logically interpretable and consistent with psychometric principles for at least three reasons.

Empirical evidence supports the need for merging. In the five-factor model, the correlation between Coherence and Duration approached unity ($r \approx 0.95$), suggesting that teachers did not clearly distinguish the two constructs in practice. Retaining them as separate factors under such conditions would likely produce poor discriminant validity and model instability, weakening interpretability. By contrast, the four-factor model maintained acceptable fit while avoiding the measurement risk of juxtaposing two highly collinear latent variables [30-32].

Theoretically, coherence allows multiple definitions and contextual variation. Prior syntheses have shown that coherence in PD research is defined in multiple ways; thus, an apparent "consensus" does not imply that coherence will always emerge as a stable, standalone dimension across contexts [6]. In Chinese school organisational practice, teachers often experience coherence together with the continuity, frequency, and time structure of PD arrangements, making coherence more likely to be perceived as a structural experience that blends with duration.

The research design followed a theory-informed sequential model-evaluation process. Auxiliary exploratory diagnostics were first used to identify items with unstable or ambiguous performance. CFA was then used to compare a refined five-factor model retaining the original dimensional structure with a four-factor alternative that combined Coherence and Duration. This evidence-based process maintained theoretical alignment while limiting researcher degrees of freedom, thereby improving transparency and reproducibility [16] [20]. Although participants were recruited from several lower-secondary schools rather than a single institution, the use of convenience sampling within one province limits the representativeness of the sample. Accordingly, the findings should be interpreted as preliminary structural-validity evidence for the sampled context and should not be generalised directly to all lower-secondary schools in China.

In sum, the revision from five to four factors is supported by both statistical evidence and theoretical reasoning: it responds to the data structure (discriminant validity risk due to extremely high correlations) and aligns with the integrative nature of PD design features in specific institutional and organisational contexts [4-5]. Future research should independently cross-validate the four-factor structure using larger and more diverse samples from different provinces, school types, and institutional contexts, and should examine measurement invariance before group comparisons or subsequent structural modelling are undertaken [30-32].Overall, the refined 24-item PDPD shows promise as a parsimonious tool for evaluating PD programme design among Chinese lower-secondary teachers, although its broader application requires independent validation across more diverse educational contexts.

References

1. Borko, H. (2004). Professional Development and Teacher Learning: Mapping the Terrain. Educational Researcher, 33(8), 3–15. https://doi.org/10.3102/0013189X033008003

2. Opfer, V. D., Pedder, D. G., & Lavicza, Z. (2011). The role of teachers' orientation to learning in professional development and change: A national study of teachers in England. Teaching and Teacher Education, 27(2), 443–453. https://doi.org/10.1016/j.tate.2010.09.014
3. Desimone, L. M., Porter, A. C., Garet, M. S., Yoon, K. S., & Birman, B. F. (2002). Effects of professional development on teachers' instruction: Results from a three-year longitudinal study. Educational Evaluation and Policy Analysis, 24(2), 81–112. https://doi.org/10.3102/01623737024002081
4. Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. American Educational Research Journal, 38(4), 915–945. https://doi.org/10.3102/00028312038004915
5. Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. Educational Researcher, 38(3), 181–199. https://doi.org/10.3102/0013189X08331140
6. Lindvall, J., & Ryve, A. (2019). Coherence and the positioning of teachers in professional development programs. A systematic review. Educational Research Review, 27, 140–154.
7. Asterhan, C. S. C., & Lefstein, A. (2024). The search for evidence-based features of effective teacher professional development: A critical analysis of the literature. Professional Development in Education, 50(1), 11–23. https://doi.org/10.1080/19415257.2023.2283437
8. Sims, S., & Fletcher-Wood, H. (2021). Identifying the characteristics of effective teacher professional development: A critical review. School Effectiveness and School Improvement, 32(1), 47–63. https://doi.org/10.1080/09243453.2020.1772841
9. Sims, S., Fletcher-Wood, H., O'Mara-Eves, A., Cottingham, S., Stansfield, C., Goodrich, J., Van Herwegen, J., & Anders, J. (2025). Effective Teacher Professional Development: New Theory and a Meta-Analytic Test. Review of Educational Research, 95(2), 213–254. https://doi.org/10.3102/00346543231217480
10. Kennedy, M. M. (2016). How does professional development improve teaching? Review of Educational Research, 86(4), 945–980. https://doi.org/10.3102/0034654315626800
11. Sargent, T. C., & Hannum, E. (2009). Doing More With Less: Teacher Professional Learning Communities in Resource-Constrained Primary Schools in Rural China. Journal of Teacher Education, 60(3), 258–276. https://doi.org/10.1177/0022487109337279
12. Zheng, X., Yin, H., & Li, Z. (2019). Exploring the relationships among instructional leadership, professional learning communities and teacher self-efficacy in China. Educational Management Administration & Leadership, 47(6), 843–859. https://doi.org/10.1177/1741143218764176
13. Gupta, A., & Zhao, X. (2024). Teachers' work under responsibilising policies: An analysis of educators' views on China's 2021 educational reforms. Journal of Education Policy, 39(4), 622–639. https://doi.org/10.1080/02680939.2023.2236067
14. Fan, L., Xie, S., Luo, J., Li, L., Tang, J., & Li, S. (2023). Teachers' perceptions of less successfully organized professional development practices in mathematics: A study of nine secondary schools in Shanghai, China. Journal of Mathematics Teacher Education, 26(5), 667–697. https://doi.org/10.1007/s10857-023-09591-6
15. Richter, E., & Richter, D. (2024). Measuring the quality of teacher professional development: A large-scale validation study of an 18-item instrument for daily use. Studies in Educational Evaluation, 81, 101357. https://doi.org/10.1016/j.stueduc.2024.101357
16. Flake, J. K., & Fried, E. I. (2020). Measurement Schmeasurement: Questionable Measurement Practices and How to Avoid Them. Advances in Methods and Practices in Psychological Science, 3(4), 456–465. https://doi.org/10.1177/2515245920952393

17. Hinkin, T. R. (1995). A Review of Scale Development Practices in the Study of Organizations. Journal of Management, 21(5), 967–988. https://doi.org/10.1177/014920639502100509
18. Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. American Psychologist, 50(9), 741.
19. Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting Structural Equation Modeling and Confirmatory Factor Analysis Results: A Review. The Journal of Educational Research, 99(6), 323–338. https://doi.org/10.3200/JOER.99.6.323-338
20. 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/10.1177/0011000006288127
21. Soine, K. M., & Lumpe, A. (2014). Measuring characteristics of teacher professional development. Teacher Development, 18(3), 303–333. https://doi.org/10.1080/13664530.2014.911775
22. Brislin, R. W. (1970). Back-Translation for Cross-Cultural Research. Journal of Cross-Cultural Psychology, 1(3), 185–216. https://doi.org/10.1177/135910457000100301
23. Haynes, S. N., Richard, D., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. Psychological Assessment, 7(3), 238.
24. Zou, Y., Ahmad, J. B., & Alias, B. S. (2026). Establishing Content Validity of the Scale for Evaluating Professional Development Programme Design in Secondary Education. Journal of Posthumanism, 6(1), 128–143. https://doi.org/10.63332/joph.v6i1.3864
25. Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). Internet, phone, mail, and mixed-mode surveys: The tailored design method. Indianapolis, Indiana, 17. https://www.ingentaconnect.com/contentone/cis/reis/2016/00000154/00000001/art00009
26. Willis, G. B. (2004). Cognitive interviewing: A tool for improving questionnaire design. sage publications. https://books.google.com.pk/books?hl=zh-CN&lr=&id=On1LBQAAQBAJ&oi=fnd&pg=PT9&dq=Willis,+G.+B.+(2005+).+Cognitive+interviewing:+A+tool+for+improving+questionnaire+design.+Sage+Publications.&ots=ALVHkxRKfh&sig=SM0X0LwgQEhkZHbcJ5pqMB1ADdg
27. Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. Journal of Applied Psychology, 78(1), 98.
28. Hayes, A. F., & Coutts, J. J. (2020). Use Omega Rather than Cronbach's Alpha for Estimating Reliability. But…. Communication Methods and Measures, 14(1), 1–24. https://doi.org/10.1080/19312458.2020.1718629
29. 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.
30. Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal, 6(1), 1–55. https://doi.org/10.1080/10705519909540118
31. Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing hu and bentler's (1999) findings. Structural Equation Modeling: A Multidisciplinary Journal, 11(3), 320–341. https://doi.org/10.1207/s15328007sem1103_2
32. 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

33. Clark, L. A., & Watson, D. (2016). Constructing validity: Basic issues in objective scale development. https://psycnet.apa.org/record/2015-32022-012
34. Chen, F. F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. Structural Equation Modeling: A Multidisciplinary Journal, 14(3), 464–504. https://doi.org/10.1080/107055107 01301834
35. Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. Developmental Review, 41, 71–90.
36. Vandenberg, R. J., & Lance, C. E. (2000). A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research. Organizational Research Methods, 3(1), 4–70. https://doi.org/10.1177/109442810031002

About the Author

Yajie Zou is a PhD candidate in Educational Management at Universiti Kebangsaan Malaysia (UKM), with research interests in teacher professional development, teachers' digital literacy, and training transfer.

Jamil Bin Ahmad received the Ph.D. in Measurement and Assessment from Universiti Kebangsaan Malaysia (UKM). He served as a Senior Lecturer at the UKM Faculty of Education from 2006, his expertise focuses on educational measurement and evaluation, quantitative research methodology, statistics, program evaluation, and SPSS-based data analysis.

Bity Salwana Alias is an Associate Professor and Chair of the Research Center of Leadership and Educational Policy, Faculty of Education, Universiti Kebangsaan Malaysia. She received the Ph.D. and M.Ed. in Educational Administration from Universiti Kebangsaan Malaysia. She began her career as a teacher in 1995, served as an assistant director at the Ministry of Education Malaysia in 2009, and has been a lecturer at UKM since 2018. Her research interests include educational leadership, management, and policy.