Journal of Posthumanism
2026
Volume: 6, No: 1, pp. 128–143
ISSN: 2634-3576 (Print) | ISSN 2634-3584 (Online)
posthumanism.co.uk

DOI: https://doi.org/10.63332/joph.v6i1.3864

Establishing Content Validity of the Scale for Evaluating Professional Development Programme Design in Secondary Education

Yajie Zou[1], Jamil Bin Ahmad[2], Bity Salwana Alias[3]

Abstract

Existing professional development instruments are widely used to evaluate programme quality, yet many rely on broad satisfaction indicators and lack sensitivity to the organisational realities of secondary education. This study establishes the content validity of the Scale for Evaluating Professional Development Programme Design, adapted from Desimone's five-feature framework. A panel of six experts from academic, administrative, policy, and school-based backgrounds reviewed 28 items across five dimensions—Content Focus, Active Learning, Coherence, Duration, and Collective Participation—using a four-point relevance scale. Item-Level Content Validity Index (I-CVI) values met accepted thresholds (≥0.83), and scale-level indices (S-CVI/Ave) ranged from 0.93 to 1.00. Quantitative results were complemented by qualitative feedback, leading to refinement of four items to improve clarity and alignment with secondary-school organisational structures. Overall, the findings support the SEPDD as a content-valid instrument for evaluating PD programme design in secondary education and informing evidence-based administrative decision-making.

Keywords: *Teacher professional development; Content validity; Scale development; Secondary education; Programme design; Evaluation instrument.*

Introduction

Effective professional development (PD) is widely regarded as a key lever for improving instructional practice and supporting student learning (Darling-Hammond et al., 2017; Kennedy, 2016). Empirical research has repeatedly shown that well-designed PD is associated with improvements in teaching quality, teacher retention, and student achievement, particularly in contexts marked by rapid educational change (Garet et al., 2001; Mulaimović et al., 2025). At the same time, evidence suggests that poorly designed PD often produces limited instructional transfer, inefficient use of resources, and declining teacher engagement in subsequent professional learning opportunities (Sims & Fletcher-Wood, 2021). These contrasting outcomes have prompted increasing attention to the design features of PD rather than to participation or delivery alone.

Over the past two decades, this shift in emphasis has been accompanied by a move away from short-term, workshop-based training toward models that foreground sustained engagement, collaboration, and coherence in teacher learning (Amemasor et al., 2025; Desimone, 2023). Within this body of work, Desimone's (2009) framework has played a particularly influential role by synthesising empirical findings into five core features of effective PD: content focus, active learning, coherence, duration, and collective participation. These features have been

[1] Universiti Kebangsaan Malaysia (UKM), Email: p147269@siswa.ukm.edu.my
[2] Universiti Kebangsaan Malaysia (UKM), Email: jamil3191@yahoo.co.uk
[3] Universiti Kebangsaan Malaysia (UKM), Email: bity@ukm.edu.my

examined across multiple educational systems and have demonstrated consistent associations with improvements in instructional practice and learning outcomes (Darling-Hammond et al., 2017; Garet et al., 2001). Subsequent studies have refined these dimensions conceptually and empirically while largely affirming their theoretical robustness (Richter & Richter, 2024; Soine & Lumpe, 2014). As a result, Desimone's framework has become a widely accepted reference for conceptualising PD quality.

Despite this theoretical consolidation, the measurement of PD design features remains uneven. Many evaluations continue to rely on participation counts, satisfaction surveys, or ad hoc indicators that offer limited insight into the multidimensional quality of programme design (Kennedy, 2016; Sims & Fletcher-Wood, 2021). This reliance constrains the capacity of researchers and educational leaders to make evidence-based judgements about which design elements meaningfully support teacher learning. Without validated instruments, it is difficult to disentangle which features of PD contribute most strongly to instructional improvement and under what conditions (Meyer et al., 2023; Mulaimović et al., 2025).Recent instruments have assessed aspects of professional development quality in specific delivery modes, such as online learning platforms (Zhang et al., 2022), yet these tools often prioritise user experience or system quality rather than the underlying design features theorised to support instructional change.

Several attempts have been made to address this gap. Soine and Lumpe (2014), for example, developed the Characteristics of Teacher Professional Development (CTPD) instrument to operationalise Desimone's five features within a multidimensional scale, providing initial evidence of construct validity and internal consistency. However, their validation sample was limited to elementary school teachers participating in subject-specific PD programmes. More recently, Richter and Richter (2024) proposed the Teacher Professional Development Monitor, which assesses general quality dimensions of PD but does not explicitly measure the design features specified in Desimone's framework. Other strands of research have focused on related constructs such as teacher engagement (Kelly et al., 2022) and agency in professional learning (Erdem et al., 2025). Taken together, these efforts highlight continued interest in PD quality while underscoring the limited availability of instruments that comprehensively capture PD design features across educational contexts. As recent reviews caution, instruments validated at one educational level may lack sufficient context sensitivity when applied to another, increasing the risk of measurement error (Amemasor et al., 2025; Sims & Fletcher-Wood, 2021).

This concern is particularly pronounced in secondary education. Unlike elementary settings, where teachers often work within integrated curricular structures, secondary schools are typically organised around subject specialisation, departmentalisation, and differentiated pedagogical demands (Berbegal Vázquez et al., 2024). Such organisational arrangements shape how PD features—especially collective participation and coherence—are experienced in practice. Secondary teachers' professional identities are frequently anchored more strongly in subject departments than in whole-school or grade-level teams, which can alter both the form and function of collaborative professional learning (Berbegal Vázquez et al., 2024). Consequently, PD programmes designed with elementary contexts in mind may not fully reflect the organisational realities of secondary schools (Darling-Hammond et al., 2017).

A prerequisite for addressing this gap is the establishment of content validity, defined as the extent to which scale items adequately represent the conceptual domain of the construct under investigation (Polit & Beck, 2006). Content validity is typically examined through systematic

expert judgement and constitutes a critical early stage in scale development prior to large-scale empirical testing (Boateng et al., 2018; Zamanzadeh et al., 2015). The Content Validity Index (CVI), which quantifies expert agreement at both item and scale levels, remains a widely endorsed approach in educational and behavioural research (Lynn, 1986; Polit et al., 2007; Stefana et al., 2025). Nevertheless, many PD instruments report limited evidence of such validation procedures, despite their importance for ensuring conceptual clarity and representational adequacy (DeVellis & Thorpe, 2021; Roebianto et al., 2023).

In response, the present study aims to develop and establish the content validity of the Scale for Evaluating Professional Development Programme Design (SEPDD), an instrument specifically adapted for secondary school teachers. Grounded in Desimone's (2009) five core features and informed by established scale development protocols (Boateng et al., 2018; Stefana et al., 2025), this study employs a quantitative content validation approach based on expert review and CVI analysis. The guiding research question is: To what extent does the SEPDD demonstrate adequate content validity for evaluating the design quality of professional development programmes for secondary teachers? By providing systematic evidence of content validity, this study seeks to contribute a measurement tool that aligns theoretical coherence with contextual specificity, supporting more precise evaluation and improvement of PD design in secondary education.

Methodology

Content validity was established through expert judgement, a necessary step for examining whether scale items adequately represent the intended construct before large-scale empirical testing (Boateng et al., 2018; Haynes et al., 1995). Expert judgement procedures in content validation share conceptual affinities with structured consensus techniques, such as Delphi-based approaches, which emphasise systematic aggregation and analysis of expert ratings rather than informal agreement (Schmidt, 1997). In line with this purpose, a purposive expert panel consisting of six members was convened. Rather than maximising panel size, the selection focused on achieving an appropriate balance between theoretical expertise, methodological competence, and practical familiarity with secondary education contexts.

Experts were selected based on three criteria commonly recommended in scale development research (Roebianto et al., 2023; Stefana et al., 2025): (a) formal training in education or a closely related discipline; (b) sustained professional experience of at least eight years in educational research, administration, or school practice; and (c) demonstrated engagement with teacher professional development or psychometric work. These criteria were applied to ensure that panel members could evaluate item relevance not only from a theoretical perspective but also with reference to real-world professional development practices.

Attention was also given to diversity in institutional roles and professional backgrounds, as heterogeneous panels have been shown to strengthen content validation by introducing multiple evaluative lenses (Zamanzadeh et al., 2015). From an academic standpoint, one expert was a university-based researcher with extensive experience in scale construction and psychometric validation, while another expert from a provincial education research institute contributed insights grounded in empirical studies of teacher learning. Policy- and system-level perspectives were represented by experts with professional experience in municipal teacher development institutes and education policy research centres, which informed the alignment of items with national teacher standards and training frameworks.

To ensure sensitivity to the organisational realities of secondary schools, the panel also included experts with direct operational experience. One expert was a district-level training officer with hands-on experience in programme implementation and evaluation. In addition, a subject department head from a senior high school, with over two decades of teaching experience, reviewed the items with particular attention to whether they reflected the department-based and subject-specialised structures characteristic of secondary education (Berbegal Vázquez et al., 2024). This combination of academic, administrative, and school-based expertise was intended to support a balanced appraisal of item relevance across policy, research, and practice.

The final panel size of six experts aligns with established methodological guidance indicating that panels of this size can yield stable and interpretable Content Validity Index (CVI) estimates while allowing for meaningful variation in judgement (Lynn, 1986; Polit & Beck, 2006). A summary of the panel members' institutional affiliations and professional experience is presented in Table 2.

Table 2. Profile of the Expert Panel

No.	Expert	Role	Experience
1	Expert 1	Municipal-level teacher development institute	14 years
2	Expert 2	District-level education administration	9 years
3	Expert 3	Provincial education research institute	11 years
4	Expert 4	University-based researcher	17 years
5	Expert 5	Education policy research centre	8 years
6	Expert 6	Senior high school subject department head	21 years

Note(s): Experience refers to years engaged in teacher education, professional development research, or school administration.

Key Constructs and Survey

This study addresses the design quality of professional development programs among secondary school teachers. The design features of professional development will be measured using the "Scale for Evaluating Professional Development Programme Design (SEPDD)" adapted from the Characteristics of Teacher Professional Development (CTPD) instrument by Soine and Lumpe (2014). There are 28 items in total. Researchers have modified specific terminologies to align with the "siloed" departmental structures and "teaching and research group" norms typical of secondary education. All the items are shown in Table 3. There are five dimensions in total: "Content Focus" includes 5 items (Q1 to Q5). "Active Learning" includes 5 items (Q6 to Q10). "Coherence" includes 5 items (Q11 to Q15). "Duration" includes 6 items (Q16 to Q21). "Collective Participation" includes 7 items (Q22 to Q28). The Form for content verification is shown as Table 3 as below.

Table 3 Form used for verifying the content of measured constructs (SEPDD)

Test Items	**Expert Agreement Level**	**Expert Feedback**

Content Focus					
Q1. The professional development used curriculum-aligned teacher resources to deepen my content knowledge.	1	2	3	4	
Q2. The professional development taught me how to recognise and address common student misconceptions about the content.	1	2	3	4	
Q3. The professional development expanded my understanding of how students learn particular subjects or topics.	1	2	3	4	
Q4. The professional development taught me teaching methods specifically for my subject content.	1	2	3	4	
Q5. The professional development guided me in using appropriate instructional representations and resources to convey key concepts (including, where appropriate, digital resources).	1	2	3	4	
Active Learning					
Q6. During the professional development, I participated in a coaching cycle (planning, observation, feedback).	1	2	3	4	
Q7. During the professional development, I analysed student work samples.	1	2	3	4	
Q8. During the professional development, I designed and taught demonstration lessons incorporating new PD ideas.	1	2	3	4	
Q9. During the professional development, I rehearsed new instructional strategies with peers before trying them with students.	1	2	3	4	
Q10. During the professional development, I wrote assessments aligned to learning standards.	1	2	3	4	
Coherence					
Q11. The professional development was designed to build upon each session as the year progressed.	1	2	3	4	
Q12. The professional development was planned based on student-performance data.	1	2	3	4	
Q13. The professional development was consistent with my professional-growth goals.	1	2	3	4	
Q14. The professional development aligned with my school's or district's improvement plans.	1	2	3	4	
Q15. The professional development was consistent with the National Teacher Development Strategy.	1	2	3	4	
Duration					

Q16. The amount of contact time in the professional development was adequate for my learning needs.	1	2	3	4	
Q17. The professional development sessions occurred frequently enough to meet my learning needs.	1	2	3	4	
Q18. The professional development was experienced as a continuous process rather than a one-off event.	1	2	3	4	
Q19. The professional development provided enough time for me to digest and apply what I learnt.	1	2	3	4	
Q20. The professional development's scheduling was reasonable—neither overly intensive nor overly scattered.	1	2	3	4	
Q21. The professional development schedule provided me with ample time for discussion and exchange.	1	2	3	4	
Collective Participation					
Q22. During the professional development, I collaborated with colleagues to design flexible student groups based on need.	1	2	3	4	
Q23. During the professional development, I followed established team norms to maximise group effectiveness.	1	2	3	4	
Q24. During the professional development, my colleagues and I conducted peer reviews of lesson plans and provided actionable feedback.	1	2	3	4	
Q25. During the professional development, I engaged in lesson study or peer observation cycles.	1	2	3	4	
Q26. During the professional development, I shared practices and resources with colleagues in a professional learning community (e.g., a teaching and research group).	1	2	3	4	
Q27. During the professional development, I planned lessons collaboratively with my grade-level or subject-area team.	1	2	3	4	
Q28. During the professional development, I analyzed student achievement data with colleagues to refine instruction.	1	2	3	4	

Note(s): SEPDD = Scale for Evaluating Professional Development Programme Design

Source. Adapted from Desimone (2009).

Table 3 presents the five key constructs through which professional development programme design was examined in this study. Content Focus: Emphasises subject-specific content and pedagogical knowledge, highlighting the importance of discipline-based learning in professional development (Darling-Hammond et al., 2017; Desimone, 2009). Active Learning: Concerns teachers' direct engagement through hands-on activities, practice, and reflection, emphasising experiential and participatory approaches to professional learning (Garet et al., 2001; Kennedy,

2016). Coherence: Examines the alignment between professional development activities and teachers' goals, student needs, and institutional priorities, indicating the role of systemic integration in effective professional learning (Desimone, 2009; Richter & Richter, 2024). Duration: Pertains to the sufficiency and continuity of professional development over time, addressing the temporal conditions necessary for sustained teacher learning (Garet et al., 2001; Sims & Fletcher-Wood, 2021). Collective Participation: Looks at collaborative learning among teachers from the same school, grade, or subject area, highlighting the social and collegial dimensions of professional development (Berbegal Vázquez et al., 2024; Darling-Hammond et al., 2017).

These constructs are based on established models of effective professional development in educational settings, as discussed in seminal works by Desimone (2009). These constructs collectively cover a spectrum of factors affecting professional development programme design, from content-related aspects like subject focus and active engagement to structural elements like duration and collaborative participation. This comprehensive approach allows for a detailed understanding of what influences the quality of professional development programme design.

Measurement Instruments

To ensure the instrument's content validity, two indices were employed: the Item-Level Content Validity Index (I-CVI) and the Scale-Level Content Validity Index (S-CVI/Ave). This dual-index approach remains the widely endorsed standard in psychometric research for ensuring both item relevance and scale representativeness (Polit & Beck, 2006; Stefana et al., 2025). The I-CVI evaluates the relevance of individual items, while the S-CVI/Ave assesses the overall validity of each dimension and the entire scale.

Following established methodological guidelines (Lynn, 1986; Boateng et al., 2018), a four-point ordinal scale was used to assess item relevance: 1 = not relevant, 2 = somewhat relevant, 3 = quite relevant, and 4 = highly relevant. This scale format was implemented to prevent neutral midpoint bias and to improve the reliability of expert judgment by providing a clearer distinction between items. (Lynn, 1986; Roebianto et al., 2023). The evaluation criteria are illustrated in Figure 1.

Relevance Rating Scale:

1 = Not relevant: The item is not relevant to the measured domain.
2 = Somewhat relevant: The item is somewhat relevant but requires major revision to be suitable.
3 = Quite relevant: The item is relevant but may need minor alteration.
4 = Highly relevant: The item is extremely relevant and succinct.

Note: Ratings of 1 and 2 are considered "irrelevant" for CVI calculation; ratings of 3 and 4 are considered "relevant".

Figure 1. Content Validation Evaluation Criteria

Criteria for Inclusion and Exclusion

Items were subjected to rigorous content validity assessment using pre-established quantitative criteria. An I-CVI score of less than 0.78 indicated that the item was not sufficiently endorsed as relevant by the panel, warranting its exclusion or revision. Similarly, an S-CVI/Ave of less than 0.90 would suggest inadequate overall content validity, prompting revisions to the instrument.

This exclusion threshold follows the standards set by Polit and Beck (2006), ensuring that only items meeting stringent content validity criteria were included in the final survey instrument.

The decision-making process was systematic and involved reviewing panel feedback, ensuring that all retained items accurately reflected the constructs of interest and would likely yield valid measures of professional development design quality. Items failing to meet these criteria underwent iterative refinement or were discarded to maintain the integrity and focus of the survey.

Computing the CVI

The Item-Level Content Validity Index (I-CVI) was determined by dividing the number of experts rating an item as 3 or 4 by the total number of experts. Following established guidelines, an I-CVI of 0.78 or above was considered sufficient for item retention. Meanwhile, the Scale-Level Content Validity Index (S-CVI) was assessed using the S-CVI/Ave (average of I-CVIs for all items) to evaluate the scale's overall validity. An S-CVI/Ave of 0.90 or higher was considered satisfactory.

According to Lynn (1986), researchers often calculate two types of Content Validity Indices (CVI) to evaluate research objectives. The first is the I-CVI, determined by the proportion of experts giving a rating of 3 or 4 (relevant). The second is the S-CVI, reflecting the overall validity of the instrument. Polit et al. (2007) provided well-regarded guidelines for acceptable I-CVI values based on the number of experts involved. They suggested that for panels consisting of three to five experts, an I-CVI of 1.00 is necessary. However, for panels of six or more experts, an I-CVI of 0.78 is acceptable, allowing for some variation while maintaining consensus.

Consequently, for this study, which used six experts, any item with an I-CVI less than 0.78 (representing agreement from fewer than five experts) was excluded from the questionnaire. Table 4 outlines the criteria for acceptable cutoff values based on panel size.

Table 4. Comparison between the number of experts and the threshold value

Number of experts	Acceptable CVI	References
3–5 experts	Must be 1.00	Polit et al. (2007)
6–9 experts	At least 0.78	Polit et al. (2007); Lynn (1986)

Results

Outcomes of the Content Validity Index (CVI) Assessment

The CVI analysis aimed to assess the content validity of the Scale for Evaluating Professional Development Programme Design (SEPDD). A panel of six experts independently evaluated each item for relevance using a four-point scale, allowing for the calculation of the Item-Level Content Validity Index (I-CVI) and the Scale-Level Content Validity Index (S-CVI). The item-level content validity results for each of the five dimensions are presented in Tables 4 to 8.

Table 4. Relevance Ratings and Content Validity Indices for Content Focus

Item	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	Experts in Agreement	I-CVI
Q1	4	4	4	4	4	3	6	1
Q2	4	3	3	4	3	3	6	1
Q3	4	4	4	4	4	4	6	1
Q4	4	4	4	3	4	4	6	1
Q5	3	3	4	3	3	3	6	1
							S-CVI/Ave	1

Note. Ratings were provided on a four-point scale (1 = not relevant, 4 = highly relevant). I-CVI represents the proportion of experts rating an item as either 3 or 4. S-CVI/Ave = mean of the item-level CVIs across items

Content Focus Results for the Content Focus dimension, presented in Table 4, demonstrate that all items met or exceeded the recommended I-CVI threshold of 0.78. Specifically, the panel reached unanimous consensus on items Q1 through Q5. The resulting S-CVI/Ave value of 1.00 confirms strong overall content validity for this domain.

Table 5. Relevance Ratings and Content Validity Indices for Active Learning

Item	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	Experts in Agreement	I-CVI
Q6	4	4	4	3	3	3	6	1
Q7	4	4	4	4	4	3	6	1
Q8	4	3	3	4	4	4	6	1
Q9	3	3	4	3	3	3	6	1
Q10	3	3	3	3	2	3	5	0.83
							S-CVI/Ave	0.966

As shown in Table 5, items measuring Active Learning were largely endorsed as relevant by the expert panel. Most items achieved a perfect I-CVI of 1.00, while a single item (Q10) met the acceptable threshold of 0.83. Consequently, the S-CVI/Ave of 0.966 reflects a high level of content validity for this dimension.

Table 6. Relevance Ratings and Content Validity Indices for Coherence

Item	Expert 1	Expert 2	Expert 3	Expert 4	Expert 5	Expert 6	Experts in Agreement	I-CVI
Q11	4	4	3	3	3	4	6	1
Q12	3	3	4	3	4	4	6	1

Q13	4	3	4	3	4	4	6	1
Q14	3	4	3	3	2	3	5	0.83
Q15	4	3	4	3	3	2	5	0.83
							S-CVI/Ave	0.932

The Coherence dimension (Table 6) achieved an overall S-CVI/Ave of 0.932. Unlike previous dimensions, this category showed slight variability: while three items achieved perfect consensus (I-CVI = 1.00), items Q14 and Q15 were rated at the acceptable threshold of 0.83. These specific ratings signaled the need for the qualitative review discussed in the refinement section.

Table 7. Relevance Ratings and Content Validity Indices for Duration

Item	**Expert 1**	**Expert 2**	**Expert 3**	**Expert 4**	**Expert 5**	**Expert 6**	**Experts in Agreement**	**I-CVI**
Q16	4	3	4	3	4	3	6	1
Q17	3	4	3	3	4	3	6	1
Q18	4	4	3	4	4	3	6	1
Q19	4	3	4	3	3	4	6	1
Q20	4	4	4	3	3	3	6	1
Q21	4	3	4	3	4	3	6	1
							S-CVI/Ave	1

Duration Consistently high content validity was observed for the Duration dimension. As indicated in Table 7, most items attained perfect I-CVI values, with the remainder exceeding the minimum requirements. The S-CVI/Ave of 1.00 provides strong evidence of expert consensus regarding the relevance of time-related design features.

Table 8. Relevance Ratings and Content Validity Indices for Collective Participation

Item	**Expert 1**	**Expert 2**	**Expert 3**	**Expert 4**	**Expert 5**	**Expert 6**	**Experts in Agreement**	**I-CVI**
Q22	3	3	4	3	3	4	6	1
Q23	3	3	4	2	3	3	5	0.83
Q24	4	4	4	3	3	3	6	1
Q25	4	3	4	4	3	3	6	1
Q26	4	3	4	4	4	4	6	1
Q27	4	4	3	4	3	3	6	1
Q28	4	3	4	3	4	4	6	1

	S-CVI/Ave	0.976

Finally, expert endorsement for Collective Participation was robust (see Table 8). Six out of seven items achieved perfect consensus (I-CVI = 1.00), with item Q23 meeting the acceptable threshold of 0.83. The resulting S-CVI/Ave of 0.976 indicates satisfactory content validity for this dimension.

Further Analysis and Refinement

While the initial quantitative analysis yielded satisfactory I-CVI scores (≥0.83) across all dimensions of the SEPDD, a rigorous refinement process was undertaken to address items where consensus was not unanimous. Specifically, four items (Q10, Q14, Q15, and Q23) receiving an I-CVI of 0.83 were subjected to closer qualitative examination. As recommended by methodological guidelines, relying solely on quantitative cut-offs is insufficient; qualitative expert feedback provides critical insights into structural and semantic nuances that statistics alone cannot capture (Boateng et al., 2018; Stefana et al., 2025).

In the Coherence dimension, experts raised specific concerns regarding item construction and respondent accessibility. Feedback on Q14 ("aligned with my school's or district's improvement plans") highlighted a "double-barreled" issue: the use of the conjunction "or" combined two distinct organizational levels, potentially confusing teachers if the PD aligned with one but not the other. Similarly, for Q15 ("consistent with the National Teacher Development Strategy"), experts noted that frontline teachers might lack familiarity with specific national-level policy documents, potentially leading to measurement error.

Qualitative feedback also targeted operational clarity in the Active Learning and Collective Participation dimensions. For Q10 ("wrote assessments aligned to learning standards"), experts pointed out that the term "learning standards" was vague in the training context; participants might be unclear whether this referred to national curriculum standards or local examination criteria. Regarding Q23 ("followed established team norms"), feedback indicated that "team norms" was an abstract concept for secondary school teachers. Experts recommended using more concrete terminology resonant with local Jiaoyanzu (teaching research group) practices to ensure the item accurately reflected the structured nature of secondary school cooperation.

Consequently, these items underwent targeted refinement—Q14 was split or simplified to focus on the immediate school context, Q15 was modified to reference general policy requirements, and terms in Q10 and Q23 were concretised to enhance clarity. This iterative process ensures that the instrument is not only statistically valid but also cognitively accessible and methodologically sound (Roebianto et al., 2023).

Overall Content Validity Assessment

Overall, the content validity analysis demonstrated the robust psychometric properties of the SEPDD. The quantitative results revealed that all 28 items met or exceeded the strict I-CVI retention threshold of 0.78, with the vast majority achieving perfect consensus (I-CVI = 1.00). Furthermore, the scale-level indices (S-CVI/Ave) for the five dimensions ranged from 0.932 to 1.00. Notably, all five dimensions surpassed the recommended standard of 0.90 for excellent content validity, exceeding the minimum acceptable threshold of 0.78 (Polit & Beck, 2006; Zamanzadeh et al., 2015).

These findings confirm that the SEPDD items are highly relevant and adequately represent the five core features of effective professional development—Content Focus, Active Learning, Coherence, Duration, and Collective Participation—within the secondary education context. The systematic combination of high expert agreement and targeted qualitative refinement provides strong evidence that the SEPDD is content-valid and ready for empirical field testing and psychometric evaluation (Stefana et al., 2025).

Discussion

Interpretation of Results

The content validity results indicate that the SEPDD provides a generally adequate representation of professional development programme design in secondary education. The consistently high I-CVI and S-CVI/Ave values suggest that, at a broad level, experts agreed on the relevance of the proposed items and dimensions. From a measurement perspective, this level of agreement supports the use of the five-feature framework as a basis for operationalising PD design quality in secondary school contexts. However, the validation process also revealed that high statistical agreement alone does not fully capture how clearly items are understood or interpreted by practitioners.

This issue became apparent in the items that achieved acceptable but not unanimous consensus (I-CVI = 0.83). For example, expert feedback on Q14 and Q15 suggested that alignment across multiple organisational levels (school, district, national policy) may be conceptually sound but cognitively demanding for frontline teachers. Similarly, Q23 highlighted the tension between theoretically meaningful constructs (e.g., collective norms) and the language teachers use to describe their everyday collaborative practices. These observations indicate that content validation is not merely a process of confirming theoretical alignment but also one of identifying where conceptual precision may conflict with practical interpretability.

Importantly, these findings prompted item-level refinements that would not have been evident through CVI statistics alone. In this sense, the validation process functioned as an iterative diagnostic exercise rather than a simple threshold-based screening. The results reinforce methodological arguments that content validity should be treated as a mixed judgement process in which quantitative indices and qualitative feedback play complementary roles, particularly when instruments are adapted to context-specific educational settings.

Comparison with Previous Studies

The structure of the SEPDD is grounded in Desimone's (2009) five-feature framework, which has been widely used to conceptualise effective professional development. Previous instruments, such as the CTPD developed by Soine and Lumpe (2014), provided important foundations for operationalising these features, yet their validation was conducted primarily in elementary education contexts. During the expert review process in the present study, differences between elementary and secondary settings emerged not as abstract theoretical distinctions but as practical concerns related to departmental organisation, subject specialisation, and patterns of collaboration.

In particular, feedback on Collective Participation items underscored how secondary teachers' professional interactions are often structured around subject-based units rather than grade-level teams. The need to revise items to reflect Jiaoyanzu practices illustrates how constructs that

appear transferable across educational levels may require contextual recalibration at the item level. This finding echoes observations in prior research that secondary teachers' professional identities and collaborative routines are more tightly linked to disciplinary boundaries than to whole-school structures (Berbegal Vázquez et al., 2024).

Compared with more recent instruments that emphasise general perceptions of PD quality or participant satisfaction, the SEPDD maintains a deliberate focus on design features that are theoretically linked to instructional change. The refinement of items related to learning standards (Q10) and policy coherence (Q15) further suggests that the usefulness of design-focused instruments depends not only on conceptual coverage but also on whether item language aligns with teachers' everyday frames of reference. From this perspective, the contribution of the SEPDD lies less in proposing new dimensions and more in demonstrating how established dimensions can be operationalised in ways that are sensitive to secondary school contexts.

Future Research Directions

While this study provides initial evidence of content validity, further empirical work is needed to examine the internal structure and stability of the SEPDD. Exploratory and confirmatory factor analyses will be necessary to test whether the five-dimensional model is empirically supported when applied to larger samples of secondary teachers. Particular attention should be given to items that required refinement during the content validation stage, as these items may be more sensitive to contextual variation.

Beyond construct validity, future research should also address the temporal stability of the instrument. Given that professional development programmes often extend over prolonged periods, test–retest reliability would provide valuable information about the consistency of teachers' responses over time. In addition, because the SEPDD incorporates context-specific elements such as Jiaoyanzu, cross-cultural validation studies would help clarify which aspects of PD design are transferable across systems and which are shaped by local organisational arrangements. Examining measurement invariance across subject areas may further illuminate how disciplinary subcultures influence teachers' perceptions of PD design quality.

Conclusion

The development and content validation of the Scale for Evaluating Professional Development Programme Design (SEPDD) represent a step toward more systematic evaluation of teacher learning in secondary education. By addressing the measurement gap noted at the outset, the study offers an instrument that operationalises Desimone's (2009) five core features—Content Focus, Active Learning, Coherence, Duration, and Collective Participation—for secondary school contexts.

The validation process highlighted an important limitation: while Desimone's five features provide a robust conceptual framework, their operationalization requires context-specific adaptation. Items that effectively capture PD design in elementary settings may not translate directly to secondary contexts without linguistic and structural revision (Richter & Richter, 2024). Items that read naturally to a primary-school teacher discussing grade-level collaboration may baffle a physics teacher whose daily professional life centres on a subject department. The changes we made to items about Coherence and Collective Participation arose directly from this tension. Experts flagged phrases like "district improvement plans" and "team norms" as misaligned with how Chinese secondary schools actually organise teacher work—through

Jiaoyanzu and discipline-based meetings, not cross-curricular teams. Adjusting the language was not about elegance; it was about making sure respondents understand what they are being asked (Amemasor et al., 2025; Berbegal Vázquez et al., 2024).

From a practical standpoint, the SEPDD gives administrators something most satisfaction surveys cannot: a way to ask whether a PD programme was designed along lines the research literature associates with teacher growth (Mulaimović et al., 2025). A post-workshop rating form might show that 85 percent of participants "agreed" or "strongly agreed" that the session was useful—but it cannot tell you whether the programme built on teachers' existing knowledge, whether sessions were spaced to allow practice, or whether colleagues from the same department had time to plan together. These are the kinds of design questions the SEPDD is built to surface. Answers to them can inform concrete decisions: extending the duration of a coaching cycle, for instance, or restructuring sessions so that subject teams meet rather than whole-staff assemblies (Meyer et al., 2023; Sims & Fletcher-Wood, 2021).

We should be clear about what a content-validated instrument can and cannot do. It can confirm that items belong on the scale and cover the intended domain; it cannot, on its own, prove that programmes scoring highly will improve instruction or raise student achievement. Those are empirical questions that require field data, and future studies will need to examine whether SEPDD scores predict meaningful outcomes (Darling-Hammond et al., 2017; Erdem et al., 2025). Still, description precedes diagnosis. If schools lack a shared vocabulary for talking about PD design, conversations about quality remain impressionistic—"the workshop felt rushed," "teachers seemed engaged." The SEPDD offers one way to move past impressions, grounding evaluation in features that research has consistently linked to effective professional learning. Whether the instrument survives large-scale testing and proves useful across different secondary contexts remains to be seen, but the content-validity evidence reported here suggests it is worth finding out.

Reference

Amemasor, S. K., Oppong, S. O., Ghansah, B., Benuwa, B.-B., & Essel, D. D. (2025). A systematic review on the impact of teacher professional development on digital instructional integration and teaching practices. *Frontiers in Education*, *10*, 1541031. https://www.frontiersin.org/journals/education/articles/10.3389/feduc.2025.1541031/abstract

Berbegal Vázquez, A., Daza Pérez, L., Carapeto Pacheco, L., & Rivas Flores, I. (2024). Becoming a secondary school teacher: Keys to a meaningful professional identity. *Teaching and Teacher Education*, *148*, 104697. https://doi.org/10.1016/j.tate.2024.104697

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health*, *6*, 149.

Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). Effective teacher professional development. *Learning Policy Institute*. https://eric.ed.gov/?id=ED606743

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

Desimone, L. M. (2023). Rethinking teacher PD: A focus on how to improve student learning. *Professional Development in Education*, *49*(1), 1–3. https://doi.org/10.1080/19415257.2023.2162746

DeVellis, R. F., & Thorpe, C. T. (2021). *Scale development: Theory and applications*. Sage publications. https://books.google.com/books?hl=en&lr=&id=QddDEAAAQBAJ&oi=fnd&pg=PA1&dq=DeVellis,+R.+F.+(2017).+Scale+development:+Theory+and+applications+(4th+ed.).+Sage.&ots=OFiHEPEN6i&sig=vDxVV09FNBHN75VXPJnu39NO1TA

Erdem, C., Toptaş, H. T., Kuzu, H. A., & Varol, B. (2025). The scale of teacher agency in professional development: The validity and reliability study. *International Journal of Turkish Education Sciences*, *13*(1), 345–378.

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

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. https://doi.org/10.1037/1040-3590.7.3.238

Kelly, C. L., Brock, L. L., Swanson, J. D., & Russell, L. W. (2022). Teacher engagement scale for professional development. *Journal of Educational Issues*, *8*(1), 261–278. https://doi.org/10.5296/jei.v8i1.19636

Kennedy, M. M. (2016). How does professional development improve teaching? *Review of Educational Research*, *86*(4), 945–980. https://doi.org/10.3102/0034654315626800

Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research*, *35*(6), 382–386. https://doi.org/10.1097/00006199-198611000-00017

Meyer, A., Kleinknecht, M., & Richter, D. (2023). What makes online professional development effective? The effect of quality characteristics on teachers' satisfaction and changes in their professional practices. *Computers & Education*, *200*, 104805. https://doi.org/10.1016/j.compedu.2023.104805

Mulaimović, N., Richter, E., Lazarides, R., & Richter, D. (2025). Comparing quality and engagement in face-to-face and online teacher professional development. *British Journal of Educational Technology*, *56*(1), 61–79. https://doi.org/10.1111/bjet.13480

Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? critique and recommendations. *Research in Nursing and Health*, *29*(5), 489–497. https://doi.org/10.1002/nur.20147

Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing and Health*, *30*(4), 459–467. https://doi.org/10.1002/nur.20199

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

Roebianto, A., Savitri, S. I., Aulia, I., Suciyana, A., & Mubarokah, L. (2023). Content validity: Definition and procedure of content validation in psychological research. *Testing, Psychometrics, Methodology in Applied Psychology (TPM)*, *30*(1), Article 1.

Schmidt, R. C. (1997). Managing Delphi Surveys Using Nonparametric Statistical Techniques*. *Decision Sciences*, *28*(3), 763–774. https://doi.org/10.1111/j.1540-5915.1997.tb01330.x

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

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

Stefana, A., Damiani, S., Granziol, U., Provenzani, U., Solmi, M., Youngstrom, E. A., & Fusar-Poli, P. (2025). Psychological, psychiatric, and behavioral sciences measurement scales: Best practice guidelines for their development and validation. *Frontiers in Psychology*, *15*, 1494261.

Zamanzadeh, V., Ghahramanian, A., Rassouli, M., Abbaszadeh, A., Alavi-Majd, H., & Nikanfar, A.-R. (2015). Design and implementation content validity study: Development of an instrument for measuring patient-centered communication. *Journal of Caring Sciences*, *4*(2), 165.

Zhang, J., Wang, B., Yang, H. H., Chen, Z., Gao, W., & Liu, Z. (2022). Assessing quality of online learning platforms for in-service teachers' professional development: The development and application of an instrument. *Frontiers in Psychology*, *13*, 998196.