

Teacher Self-Efficacy for Dialogic Teaching: Scale Development and Psychometric Validation

Samarnh Pang

pang54@purdue.edu

Purdue University West Lafayette <https://orcid.org/0000-0003-1263-218X>

Virak Chan

Purdue University West Lafayette <https://orcid.org/0000-0001-5376-3122>

Anne Traynor

Purdue University West Lafayette <https://orcid.org/0000-0003-2662-9184>

Research Article

Keywords: self-efficacy scale, dialogic teaching, preservice English language teachers, scale development, exploratory factor analysis

Posted Date: September 2nd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10234751/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Abstract

Dialogic teaching (DT) is widely used in English language teaching (ELT), but its success depends on what teachers believe they can do with classroom dialogues. However, there is no task-specific scale to measure preservice teachers' self-efficacy for using DT in ELT. As the first phase of a three-paper research project, this study develops and provides initial validation of the Self-Efficacy Scale for Dialogic Teaching (SESdT). The items were based on dialogic teaching theory and Bandura's guidelines, and were refined through expert review and pilot testing. The scale was then given to 367 preservice teachers at the Royal University of Phnom Penh. Exploratory factor analysis showed a four-factor structure: Promoting Critical Dialogue, Facilitating Dialogue, Assessing Dialogue, and Dialogic Questioning. The scale showed high reliability and a clear structure, with related but distinct aspects of dialogic teaching self-efficacy. The current SESdT provides psychometric evidence and can serve as a useful tool for future research and teacher development.

Introduction

Self-efficacy measures, when properly validated, can enhance the reliability and validity of research studies (Bandura, 1997; DeVellis & Thorpe, 2021; Sherer et al., 1982; Tschannen-Moran & Hoy, 2001). Bandura (1977) introduced the first self-efficacy scale to measure coping with stress and phobias, using rating-specific tasks and confidence items. Studies in this area continued to expand with the Generalized Self-Efficacy Scale, developed by Sherer et al. (1982) and later refined by Schwarzer and Jerusalem (1995). This scale has been verified in several countries and translated into other languages (Scholz et al., 2002). The concept of self-efficacy (SE) is linked to some aspects of education, including student engagement, classroom dynamics, learning outcomes, and instructional decision-making (Bandura, 2006; Tschannen-Moran & Hoy, 2001). In other words, the ways teachers address instructional challenges and refine their teaching methods over time are largely related to their beliefs in their own abilities to do so (Klassen & Tze, 2014; Tschannen-Moran & Hoy, 2001; Zee & Koomen, 2016). However, most popular self-efficacy instruments currently available are domain-wide and do not adequately address dialogic teaching tasks in ELT classrooms.

To date, dialogic teaching (DT) is regarded as a robust approach for classroom practice (Lee, 2016; Skidmore & Murakami, 2016; Wegerif, 2013). Mercer (2000) grounded DT in sociocultural theory, Alexander (2008, 2020) framed it as purposeful classroom discussion for intellectual and democratic ends, and Nystrand et al. (1997) and Freire and Macedo (1995) showed how authentic uptake and critical inquiry make dialogue educational. A growing body of dialogic-teaching research in language education has reported many benefits, including improvements in students' reasoning, participation, and language knowledge and skills, as well as socio-emotional benefits (Chow et al., 2023; Cui & Teo, 2023; Lee, 2016; Resnick et al., 2010). Despite this evidence, preservice teachers often report that sustaining cumulative, reciprocal talk is demanding in every classroom (Abd Elkader, 2016; Arslan & Whitehead, 2025; Calcagni et al., 2024). This difficulty makes dialogic self-efficacy an analytically useful construct.

Purpose of the study

Alexander (2018, 2020) explained that DT commonly employs purposeful questioning, uptake, and cumulative discussion; however, these moves are often perceived as demanding for preservice teachers in Cambodian ELT classrooms. To date, no task-specific instrument has been developed to address preservice teachers' confidence in implementing these dialogic practices. This study, therefore, aimed to design and obtain initial validity evidence for the Self-Efficacy Scale for Dialogic Teaching (SESdT) among preservice English teachers in Cambodia. The items were written in accordance with dialogic teaching principles and moves within this tradition and followed Bandura's guidance on the construction of task-specific self-efficacy scales (Bandura, 2006). Therefore, our SESdT instrument is intended for use in preservice coursework, practicum mentoring, and later confirmatory validation studies.

This study was guided by the following research questions:

1. What is the underlying factor structure of the self-efficacy scale for dialogic teaching (SESdT)?
2. To what extent do the scale items demonstrate internal consistency and initial evidence of construct validity?

Literature Review

Defining Self-Efficacy

According to Bandura (1997), self-efficacy primarily refers to teachers' judgments about their capability to organize and execute instructional actions that lead to desired classroom outcomes. According to Pajares (2002), teachers with stronger SE beliefs persist with demanding pedagogies and adjust their instruction when students struggle. Later studies stressed that teacher SE should be task- and context-bound rather than a general sense of confidence (Klassen & Tze, 2014; Zee & Koomen, 2016; Schunk & DiBenedetto, 2020). Maddux (2002) noted that SE is a distinct construct that sets it apart from other concepts, such as *self-esteem* (a general evaluation of one's worth) and *locus of control* (beliefs about whether outcomes are controlled internally or externally). A study by Kim and Morita-Mullaney (2020) found that teachers with stronger preparation and ESL licensure reported higher self-efficacy in addressing the needs of English learners. However, since DT involves complex interactional moves, preservice teachers' efficacy needs to be measured at that level of specificity, which motivates the development of the SESdT.

Dialogic Teaching among Preservice Teachers in Language Teaching Contexts

Dialogic teaching regards classroom talks as a means for learners to explore ideas and collectively reach sensible understandings. Alexander (2020) put forward six principles for the successful implementation of DT: collective, reciprocal, supportive, cumulative, deliberate, and purposeful. These principles aim to help teachers create an intellectual classroom by facilitating learners' contributions in a

consistent pedagogical direction, rather than allowing the conversation to fragment into recitation. However, Mercer (2000) and Nystrand et al. (1997) noted that this type of learning is more likely to occur when teachers consistently use authentic questions and closely follow learners' responses to integrate them into a single line of inquiry. In ELT classrooms, DT is often perceived as posing extra challenges for teachers, as language is commonly seen as both the means and ends of participation, with flexible, multivoiced resources included to keep talk flowing (Lee, 2016; Lopez-Hernandez, 2025; Uştuk & Yazan, 2024).

Furthermore, recent research showed that preservice teachers in ELT settings are often seen to struggle to effectively implement various talk moves due to several factors, including learners' limited language knowledge and skills (Barahona et al., 2021), content challenges (Barahona et al., 2021; Jocuns, 2021), time constraints (Jocuns, 2021), curriculum demands (i.e., exam-oriented syllabi and textbook-driven lessons) (Rungwarapong, 2019; Jocuns, 2021), and discussion facilitation in dialogic inquiry (Sanchez & Athanases, 2023). Despite such challenges, Mercer and Littleton (2007) demonstrated that modeling exploratory talk can help novices build a concrete picture of productive dialogue in their classrooms. Edwards-Groves and Hoare (2012) recommend video-based reflection to help candidates notice missed uptake, and recent work using video, simulations, and short intensive programs reports higher perceived ability to run dialogic lessons when supports target concrete practices such as questioning and revoicing (Calcagni et al., 2024; Arslan & Whitehead, 2025). Wilkinson et al. (2017) similarly found that inquiry-dialogue professional development can be useful for teachers' dialogic practices and students' argumentation, although teachers' epistemological beliefs remained an issue.

Development and Validation of Teacher Self-Efficacy Scales

The early self-efficacy measures were domain-broad (i.e., Bandura, 1977, 2006; Sherer et al., 1982; Schwarzer & Jerusalem, 1995; Scholz et al., 2002). Later, Bandura (2006) provided guidelines for constructing self-efficacy scales and noted that teachers' efficacy must be measured as can-do judgments of capability, not will-do intentions (p. 309). He further stated that items should be tailored to the specific domain and to the actual performances that define success in that domain (pp. 309–311). He also proposed that researchers arrange the tasks from least to most difficult and use a 0–100 confidence scale in order to account for variation within the group (pp. 311–315).

To validate a teacher's self-efficacy scale, researchers commonly follow a sequential approach that begins by clarifying the construction and then scrutinizing the statistics. For example, Tschannen-Moran and Hoy (2001) validated their Teacher Sense of Efficacy Scale using these steps: 1) theory-driven item generation from prior efficacy and classroom management scales, 2) pilot field-testing and revision of a 36-item pool with practicing teachers, and 3) principal-axis factoring with varimax rotation. As a result, their analysis yielded three subscales: Instructional Strategies, Classroom Management, and Student Engagement. However, they continued to perform second-order factor analysis to confirm a general efficacy factor. They also reported internal consistency for the 18-item and 24-item forms and analyzed

construct and discriminant validity through correlations with the RAND items developed in early evaluations of teacher effectiveness.

Equally rigorous, Ma et al. (2023) validated their comprehensive teacher self-efficacy scale among preservice teachers by first generating a 40-item pool from the TSES and other theoretical sources. They then used content validity ratios with 25 in-service teachers to discard weak items and surveyed the final items to 854 Chinese pre-service teachers. However, they decided to split their sample to perform both an exploratory factor analysis using maximum likelihood with promax and parallel analysis to identify a two-factor structure (called Ethos and Teaching). They then ran confirmatory and second-order CFAs on both the 40-item and 20-item short forms; they mainly reported estimated reliability using Cronbach's α , composite reliability, and McDonald's ω . They tested convergent, discriminant, and known-groups validity via correlations with generalized self-efficacy and agreeableness, plus t-tests across practicum-effectiveness and intention-to-teach groups. Nonetheless, Yu et al. (2018) argued that traditional psychometric approaches may overlook item-level performance, and demonstrated that item response theory can provide more precise information about how well individual items function across different levels of a latent construct.

More recently, Palomino et al. (2025) validated their dialogic learning questionnaire through 1) Delphi review, 2) pilot testing, 3) exploratory factor analysis with Varimax rotation, and 4) psychometric reliability analysis using Cronbach's alpha. However, Costello and Osborne (2005) contended that, after achieving content adequacy, researchers should use common factor extraction methods like maximum likelihood or principal axis instead of PCA in exploratory factor analysis (pp. 3–4), and they advised using oblique rotation when the factors are likely to be correlated (p. 5). They also suggested that scaling validation should retain factors on the joint basis of statistical evidence and theoretical coherence (pp. 4–6).

Even though there are many teacher self-efficacy scales, most of them are quite general or do not really address what happens in dialogic teaching, especially in ELT classrooms. Some recent studies (e.g., Ma et al., 2023) have used strong validation methods, but they still tend to focus on broader aspects of teaching rather than the small, moment-to-moment practices we see in dialogic interaction, like uptake, probing, or building ideas together over time. Because of this, we still lack a measure that reflects how dialogic teaching actually unfolds in language classrooms. In this study, we take a step toward filling this gap by developing and providing initial validation evidence for a dialogic teaching self-efficacy scale designed specifically for preservice teachers in an ELT context.

Methodology

This study employed a sequential instrument development and validation protocol to design and test the Self-Efficacy Scale for Dialogic Teaching (SESdT) with preservice teachers in a single cross-sectional administration ($n = 312$). The process comprised five pre-specified phases: (1) scale design grounded in dialogic pedagogy and Bandura's 0 to 100 efficacy scale; (2) a Delphi review to gather content-validity

evidence and refine items; (3) pilot testing with one intact class to examine clarity, cultural fit, and usability; (4) field administration to the full cohort at the research site; and (5) psychometric analysis, including data screening and Maximum-Likelihood EFA with Promax rotation for structure discovery and item reduction (see Fig. 1).

Phase 1: Scale Design

We developed all initial items based on an extensive review of dialogic teaching literature and Bandura's guidelines for constructing task-specific self-efficacy measures. We did not directly adapt items from existing scales; however, we drew on the wording and structure of established self-efficacy instruments to maintain consistency with prior research. At the outset, we generated a pool of 29 items organized around four theoretically informed subdomains of dialogic teaching: (1) creating an environment that supports open and organized discussion; (2) asking questions that promote dialogue, including opening, probing, and uptake-oriented questions; (3) encouraging critical dialogue, such as reasoning with evidence, making connections, and challenging assumptions; and (4) evaluating dialogue, including monitoring engagement, uptake, contribution quality, and alignment with lesson goals. We based these subdomains on dialogic teaching principles and empirical research on classroom discourse (e.g., Alexander; Mercer; Nystrand et al.; Reznitskaya; Wegerif; Resnick et al.), and used them to organize the initial item pool for subsequent analysis.

For SESDT scale development, the current study was guided by Bandura (2006) and adopted a 0 to 100 confidence scale in 10-point intervals, rating from 0 ("cannot do at all"), 50 ("moderately can do"), and 100 ("highly certain can do"). He also argued that good, well-graded unipolar response options should more sensitively target the strength of efficacy beliefs than few-category scales and tend to predict performance better than shorter formats such as a 5-point scale (Bandura, 2006, p. 314). For this reason, we organized the questions on a continuum of grading their complexity and prepared them for in-class completion so that students may complete them at their own pace and with the highest level of accuracy (see Appendix B for details).

Phase 2: Expert Review (content-validity evidence)

To assess the original item pool ($k = 29$), we collaborated with fifteen professors (12 doctorates; 3 master's) with expertise in TESOL, dialogic teaching, and educational measurement. The panel included professors and teacher-trainers with at least 5 years of teacher training experience at the Royal University of Phnom Penh (5), Paññāsāstra University of Cambodia (1), Charles Darwin University (1), Indiana University (1), Purdue University (6), and Salem State University (1). Using Qualtrics, we conducted a Delphi review with blinded ratings on clarity, relevance, and cultural fit for Cambodian preservice teachers. We set decision rules in advance. For each criterion, an item was acceptable when at least seventy-five percent of raters selected the top two categories on a four-point scale (1 = bad; 2 = usable; 3 = good; 4 = excellent). Our analysis of Item-Level Content Validity Indices (S-CVI) supported these decisions as 28 of 29 items reached an I-CVI of at least 0.78. The Scale-Level Content Validity

Index/Average (S-CVI/Ave) for the retained set was approximately 0.93, and showed modified kappa substantial to almost perfect agreement (Polit & Beck, 2006). One item was removed.

In response to panel input, we revised seven items (items 2, 4, 5, 9, 12, 18, 22, and 26). The suggestion included using *plain, student-friendly language, removal of double-barrel phrasing, and alignment with common classroom practices such as pair work and whole-class discussion*. Some technical words were also localized during the revision; these phrases include *classroom language, condensing verbs to fit the 0–100 confidence scale, and group discussion and managing classrooms*. Thus, we refined these items to fit concrete student behaviors (Items 2, 4), specify inclusion across proficiency levels with clarity on tasks (Items 5, 9), use neutral questioning language (Item 12), clarify subjects in assessment statements (Item 18), replace technical terms with classroom vocabulary (Item 22), and improve assessment phrasing for usability (Item 26). We decided to remove 11 due to overlap with Items 5 and 9, as flagged by five reviewers. Appendix A documents these actions in a before-and-after log that links the reviewer's rationale to the final wording.

Phase 3: Pilot Testing

After expert reviews, we piloted the revised 28-item questionnaire using Qualtrics with one intact methods class ($n = 30$; female = 16) at the research site. Our goal was to extend the expert review with real preservice teachers' feedback before field administration. We performed Clarity Index (CI) on a 4-point scale (1 = not clear at all; 2 = usable; 3 = very clear; 4 = excellent) and found responses clustered in the top categories, "Not clear at all" was very infrequent, and most items exceed the $CI \geq .70$ criterion for field use, with a few low CI items flagged for minor rewording; no difficulties were reported with the 0 to 100 response scale. Based on their comments, we made only minor edits to improve the wording on three items and added brief parenthetical cues that match common practicum routines. We also adjusted the instructions (bolded continuum and a one-line reminder above the item list) to minimize the risk of confusion. The pilot confirmed functional delivery in Qualtrics, good face clarity, and strong acceptability of the 0 to 100 format. Therefore, we proceeded to field testing without making any structural changes.

Phase 4: Main Administration

We administered the paper-based questionnaire in person during regular class time at a public university in Cambodia. Data were collected between 05/16/2025 and 06/15/2025 during the 2024–2025 academic year. The target was the entire cohort ($N = 367$) enrolled in the senior year of the Teaching English as a Foreign Language (TEFL) program. For this study, participants were informed verbally at the time of questionnaire administration. Before distributing the survey in class, we explained the purpose of the study, the voluntary nature of participation, the anonymous handling of responses, and participants' right not to participate without any penalty or consequence. Participants were informed that completion and submission of the questionnaire would indicate their agreement to participate in the study. Hence,

students who did not wish to participate were free to decline and were not required to complete or return the questionnaire.

This study involved an anonymous questionnaire and posed minimal risk to participants. Under the approved study protocol, separate written or signed consent was not required. The verbal consent procedure described above was reviewed and approved as part of the study's ethics approval process. No verbal responses were recorded or stored, and no identifying information was collected.

The survey took approximately 20 minutes of class time to complete. In the TEFL program, preservice teachers spend their first three years building English language proficiency, literacy, cultural knowledge, and research foundations. The fourth year concentrates on methodology and teaching preparation, including a 10-week practicum, which this cohort had just completed. 320 preservice teachers returned questionnaires (85.6% response rate from the cohort). After removing forms with missing or invalid responses, 312 cases remained for analysis (83.4% of the population; 97.5% of returns). Our sample of 312 preservice teachers (11 participants per item) met common guidelines for factor analysis (5–10 participants per item; Costello & Osborne, 2005). No incentives were offered as participation was voluntary, and surveys were anonymous. Item-level missingness was minimal and handled as described in Phase 5.

Phase 5: Data Screening and Analysis

We screened the dataset in SPSS 29 prior to factor analysis. Among $n = 312$ usable forms, there were five item-level missing cells (0.06% of the item-by-person matrix). However, we skipped Little's Missing Completely at Random (MCAR) test due to the negligible rate, and imputed each missing cell to the item median to preserve distributional shape and sample size. We also performed a sensitivity check using listwise deletion; it yielded the same number of factors and a pattern of good loadings that was materially similar.

We examined univariate skewness and kurtosis for each item and screened multivariate outliers via Mahalanobis distance ($\alpha = .001$); no cases were removed. To confirm factorability, we evaluated the Kaiser–Meyer–Olkin (KMO) index, item-level Measures of Sampling Adequacy (MSA) from the anti-image matrix, and Bartlett's test of sphericity, and we inspected the correlation-matrix determinant to rule out multicollinearity (target $\det R > .00001$). We also reviewed corrected item–total correlations as a diagnostic of discrimination, using values around .30 as a minimum cue over a strict decision rule (Nunnally & Bernstein, 1994; DeVellis & Thorpe, 2021).

We conducted exploratory factor analysis (EFA) using Maximum-Likelihood (ML) extraction because common-factor models estimate shared variance and, in this context, allow model fit indices and χ^2 comparisons across neighboring solutions (Costello & Osborne, 2005; Henson & Roberts, 2006; Schmitt, 2011). Given the theoretical expectation of correlated dimensions in DT, we used Promax ($\kappa = 4$) rotation. Parallel analysis (PA) (1,000 permutations; 95th percentile criterion) was used as the primary retention

rule, with scree plot inspection and theory providing converging evidence. After each candidate's deletion, we reran PA to avoid over-extraction.

Items were flagged for removal when they showed (a) weak primary loading ($< .40$), (b) cross-loading ($\geq .30$) or primary–secondary loading gap ($< .20$), or (c) low communalities ($h^2 < .30$); removals were one at a time with conceptual fit in view (Thompson, 2004; Worthington & Whittaker, 2006). We inspected the pattern matrix for interpretation, provided the structure matrix and factor correlation matrix to document inter-factor relations, and also checked for Heywood cases and large residuals between observed and reproduced correlations. We also examined ML diagnostic fit, reporting the χ^2 test and the residual profile, specifically the share of off-diagonal residuals below .05 and the largest absolute residual, to further decide whether to proceed with item reduction.

Although confirmatory factor analysis (CFA) is commonly recommended following exploratory factor analysis (EFA), the present study is positioned as an initial scale development and psychometric validation effort. Our sample size ($n = 312$), despite being adequate for EFA, would require splitting into independent subsamples to conduct both EFA and CFA, which would reduce statistical power and stability of the factor solution. Therefore, we prioritized a rigorous exploratory approach in this phase. Future studies should conduct CFA using independent samples to test the stability and replicability of the proposed factor structure.

Results

Demographics

Participants were 312 preservice English teachers (see Table 1). Most were 18–24 years old (94.9%). Gender was primarily female (63.5%), with 34.5% male and 1.9% preferring not to say. In the Cambodian TEFL context, 'preservice teachers' refers to teacher candidates still in training, although many of whom undertake paid tutoring or private-school teaching during their degree, and our experience measure includes this prior work. Before their practicum, 48.1% reported no teaching experience, while 51.9% reported having teaching experience for at least one year (valid $n = 308$). Preferred teaching grade levels were higher education (33.7%), high school (32.7%), no preference (17.9%), primary (8.3%), and secondary (7.1%). Their self-rated English proficiency was mostly advanced (65.9%), followed by intermediate (26.4%), near-native (5.5%), and native-like (2.3%). During their 10-week practicum, all participants were assigned to teach intermediate-level English classes.

Table 1
Sample characteristics of preservice English teachers ($n = 312$).

Characteristic	Category	<i>n</i>	%
Age (valid $n = 311$)	18–24	295	94.9
	25–34	16	5.1
Gender (valid $n = 310$)	Female	197	63.5
	Male	107	34.5
	Prefer not to say	6	1.9
Teaching experience prior to practicum (valid $n = 308$)	None	148	48.1
	< 1 year	81	26.3
	1–2 years	64	20.8
	≥ 3years	15	4.5
Preferred grade level to teach (valid $n = 312$)	Higher education	105	33.7
	High school	102	32.7
	No preference	56	17.9
	Primary school	26	8.3
	Secondary school	22	7.1
	Unspecified	1	0.3
English proficiency (self-rated) (valid $n = 311$)	Advanced	205	65.9
	Intermediate	82	26.4
	Near-native	17	5.5
	Native-like	7	2.3

Assumption Checks for Factor Analysis

Our data was suitable for exploratory factor analysis. The overall KMO = .942 (good) with MSAs = .872–.979 and Bartlett's $\chi^2(378) = 5635.71, p < .001$ are provided in Table 2. This result suggests that there is enough shared variation to extract factors from the sample ($n = 312$). A factorable matrix with no issue with multicollinearity was explained with the correlation-matrix determinant = 7.31×10^{-9} . The univariate form was considered acceptable since the greatest skew is .89 and the biggest kurtosis, 2.06. We screened for multivariate outliers using Mahalanobis distance based on the 28 SESDT items, using a chi-square criterion of $\chi^2(28) = 56.89, p < .001$. No cases exceeded this threshold, so all 312 cases were retained for factor analysis. The moderate to high item–total correlations ($r = .54$ –.75) mean that each item maintained internal coherence as they make a significant contribution to the overall scale.

Therefore, these diagnostics support the use of ML extraction, which provides initial model-fit indices. They also support the use of an oblique Promax rotation for interpretable factor correlations (Fabrigar & Wegener, 2012).

Table 2
Descriptive statistics and assumption check for factor analysis ($n = 312$).

Items	<i>M</i>	<i>SD</i>	Skewness	Kurtosis	MSA	Item-Total <i>r</i>
1	74.31	14.87	-0.453	0.221	.952	0.60
2	75.87	15.78	-0.892	1.076	.931	0.60
3	78.91	14.24	-0.443	-0.269	.944	0.58
4	71.22	14.32	-0.600	0.782	.941	0.57
5	65.29	16.75	-0.394	0.087	.934	0.57
6	67.52	16.59	-0.596	0.467	.926	0.61
7	70.50	13.96	-0.584	1.095	.924	0.64
8	76.62	15.49	-0.773	1.198	.936	0.68
9	76.12	14.93	-0.723	1.303	.949	0.62
10	77.73	14.41	-0.619	0.734	.944	0.67
11	73.97	14.08	-0.686	0.924	.979	0.70
12	75.26	14.18	-0.739	1.517	.956	0.69
13	70.59	16.11	-0.187	-0.309	.945	0.59
14	70.46	14.29	-0.331	0.250	.952	0.66
15	70.07	13.77	-0.552	2.062	.949	0.71
16	71.63	14.74	-0.288	0.368	.951	0.67
17	67.29	15.76	-0.556	0.864	.957	0.72
18	65.59	15.62	-0.467	0.147	.949	0.72
19	75.70	14.70	-0.789	1.288	.938	0.69
20	68.10	15.53	-0.667	0.804	.942	0.72
21	72.25	14.96	-0.512	0.796	.927	0.73
22	78.20	13.27	-0.885	1.982	.872	0.54
23	73.57	13.54	-0.460	0.313	.898	0.57
24	68.37	13.85	-0.569	1.041	.948	0.75
25	70.99	14.55	-0.843	1.395	.951	0.72
26	74.91	14.42	-0.747	0.718	.965	0.68
Determinant = 7.31×10^{-9}						

Items	<i>M</i>	<i>SD</i>	Skewness	Kurtosis	MSA	Item-Total <i>r</i>
27	70.62	14.40	-0.698	1.249	.962	0.73
28	69.20	16.61	-0.449	0.438	.965	0.58
Overall KMO = .942						
Bartlett's Test of Sphericity: $\chi^2(378) = 5635.71, p < .001$						
Determinant = 7.31×10^{-9}						

NOTE

Scale: 0-100 (0 = cannot do at all, 50 = Moderately can do, 100 = perfectly can do; see Appendix A for details)

Factor Structure, Reliability, and Validity of the SESDT

For factor retention, we examined parallel analysis, eigenvalues, convergence criteria, and theoretical considerations (Hayton et al., 2004). In a factor-analysis-based parallel analysis, the first four eigenvalues (12.67, 0.96, 0.84, 0.60) were slightly higher than the corresponding simulated values (0.66, 0.53, 0.47, 0.41). In a principal-components-based PA, only the first two eigenvalues (13.20, 1.54) clearly exceeded their simulations (1.60, 1.51). However, the third and fourth lay just below. Taking these results together with eigenvalues greater than one, the scree plot, and conceptual interpretability, we retained a four-factor SESDT structure.

Additionally, we looked at neighboring solutions and extraction-rotation techniques (such as PC-Oblimin and PAF-Equamax). Five-factor solutions had a weak factor with less than three items and many cross-loadings, whereas three-factor solutions did not fulfill content domains and had decreased interpretability. Therefore, for parsimony and interpretability, we retained a four-factor solution that ML-Promax generated. After extraction, the initial eigenvalues for Factors 1–4 in the final 24-item solution were 13.20, 1.54, 1.41, and 1.15 (all > 1.0), explaining 61.78% of the variance. The factors' respective extraction sums of squared loadings showed that they explained 45.55%, 3.85%, 3.49%, and 2.57% of the variance (55.46% total; Table 3). Although the ML goodness-of-fit test was significant, $\chi^2(272) = 736.21, p < .001$ (as expected with larger samples), residual diagnostics showed acceptable fit: 84.0% of the 378 off-diagonal residual correlations were < .05, 16.0% were > .05, and the largest absolute residual was .146.

Table 3
Total Variance Explained (ML–Promax EFA; $n = 312$, 28 to 24 items).

Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	Var %	Cum %	Total	Var %	Cum %
1	13.2	47.13	47.13	12.75	45.55	45.55
2	1.54	5.5	52.63	1.08	3.85	49.41
3	1.41	5.03	57.66	0.98	3.49	52.89
4	1.15	4.11	61.78	0.72	2.57	55.46

The four-factor solution that ML-Promax produced (Table 4) suggested agreement with the DT theoretical ground. The rotation converged after 11 iterations, and 61.78% of the variance was explained by the four components that were recovered collectively. For items to be retained, the factor loading level had to be at least .40. Items with loadings below this threshold suggest that they contribute poorly to the latent construct and may increase measurement error (Costello & Osborne, 2005; Hair et al., 2019). Thus, we decided to eliminate five items with cross-loading and loadings below this criterion: 7, 13, 21, 26, and 28. Items 7 and 21 both had low loadings (.375 on Factor 1 and .435 on Factor 2), but there was no obvious main factor for item 21, which had equally low loadings (.350, .310, .307). Items 26 and 28 have no obvious factor home and were below the .40 loading requirement. The loadings for item 26 were consistently low (.234, .207, .288, .078). A similar pattern was seen in item 28 (.271, .114, .285, .008). Since item 13's primary loading on F1 was low (.423) and it also loaded on F4 (.327), we also removed it. Appendix C provides further details in this regard.

As a result, 24 items from the four factors were retained for the final scale structure. A stable and understandable structure is suggested by the final solution's obvious factor loadings above .40 and lack of troublesome cross-loadings above .30. To improve readability and interpretability, we arrange the indicators in Table 4 in each factor in decreasing order of their factor loadings. Factor 1, named Promoting Critical Dialogue (CD), included six items (17, 19, 16, 20, 15, 18) with loadings ranging from .615 to .848 and communalities between .586 and .662. This factor focuses on teachers' SE in facilitating students to reason with evidence as well as analyze complex issues while reflecting critically on ideas. It accounted for 45.55% of the variance and demonstrated high internal consistency (Cronbach's $\alpha = .895$).

Factor 2, called Facilitating Dialogue (FD), consisted of six items (2, 4, 1, 3, 6, 5) with loadings between .435 and .857 and communalities ranging from .477 to .600. This factor mainly reports teachers' confidence in organizing inclusive and respectful talk environment in their classroom. It explained 49.41% of the cumulative variance (eigenvalue = 1.54) and showed strong reliability ($\alpha = .857$). Factor 3, named Assessing Dialogue (AD), comprised five items (23, 22, 24, 25, 27) with loadings from .483 to .848 and communalities between .473 and .743. This factor mainly evaluates teachers' perceived ability to evaluate the quality of students' engagement, reasoning, and contribution during classroom talk. It

explained 52.89% of the variance with high internal consistency ($\alpha = .875$). Factor 4, called Dialogic Questioning (DQ), included six items (10, 8, 11, 12, 14, 9) with loadings from .547 to .722 and communalities between .479 and .629. This factor evaluates teachers' confidence in performing open-ended and probing questions to encourage learner thinking and connect prior knowledge with new ideas. It accounted for 55.46% of the variance and displayed strong reliability ($\alpha = .878$).

To further examine the scale, we calculated composite reliability (CR) and average variance extracted (AVE) for each factor. The CR values ranged from .83 to .86, all well above the .70 threshold; these suggest strong internal consistency. The AVE values ranged from .47 to .53; two factors met the .50 criterion, while the others fell slightly below. However, since all CR values exceeded .80, the overall convergent validity can still be considered acceptable (Fornell & Larcker, 1981; Hair et al., 2019).

Table 4
Four-factor scale structure, communality, loading coefficients, and reliability ($n = 312$).

SESDT Subscale and Item	Factor Loading
F1: Promoting Critical Dialogue (CD): Communalities = .586-.662; Cronbach's α = .895; CR = .86; AVE = .50	
17. Ask questions that help students evaluate the strengths and weaknesses of different ideas.	0.848
19. Encourage students to make connections between the discussion topic and real-world issues.	0.765
16. Guide students to use evidence to support their arguments during discussions.	0.713
20. Help students analyze complex problems and break them into manageable parts.	0.638
15. Extend student contributions by prompting them to think more deeply.	0.634
18. Identify biases or gaps in students' own reasoning or in the information presented.	0.615
F2: Facilitating Dialogue (FD); Communalities = .477 – .600; Cronbach's α = .857; CR = .83; AVE = .49	
2. Help students respect each other's ideas when they discuss and work in groups.	0.857
4. Help students stay focused and engaged on the topic during group discussions.	0.709
1. Encourage students to actively listen and take turns in discussion.	0.671
3. Encourage students to share their thoughts and opinions during class discussions.	0.670
6. Engage students who seem uninterested, shy, or hesitant to participate.	0.559
5. Design group discussions that work for different language proficiency levels.	0.478
F3: Assessing Dialogue (AD); Communalities = .473–.743; Cronbach's α = .875; CR = .85; AVE = .55	
23. Evaluate the quality of student contributions during class discussions.	0.848
22. Identify whether students are actively engaged in classroom talk activities.	0.724
24. Assess how well students build on each other's ideas in dialogue.	0.719
25. Identify moments when students demonstrate critical thinking in discussions.	0.573
27. Assess the overall depth and relevance of discussions in achieving lesson objectives.	0.483
F4: Dialogic Questioning (DQ); Communalities = .479 – .629; Cronbach's α = .878; CR = .84; AVE = .47	

SESDT Subscale and Item	Factor Loading
F1: Promoting Critical Dialogue (CD): Communalities = .586-.662; Cronbach's α = .895; CR = .86; AVE = .50	
10. Ask follow-up questions to help students clarify or expand on their responses.	0.722
8. Use open-ended questions to encourage students to think critically about a topic.	0.615
11. Ask questions that encourage students to connect new ideas with their prior knowledge.	0.600
12. Use guiding questions to help students justify their answers with evidence.	0.595
14. Pose questions that encourage students to consider multiple perspectives on a topic.	0.584
9. Make sure my questions are clear for students with different English skills.	0.547

The four SESDT variables showed relatively high positive correlations ($r = .56$ to $r = .71$) in the ML-Promax solution ($n = 312$), as shown in Table 5. Connecting Promoting Critical Dialogue (F1) with Assessing Dialogue (F3) ($r = .71$) and with Facilitating Dialogue (F2) ($r = .71$) were the two most closely related relationships. This trend makes sense since educators who are comfortable steering class discussions toward criticism are also likely to be comfortable maintaining student engagement and accurately evaluating their work. The weakest relation appeared between Assessing Dialogue (F3) and Dialogic Questioning (F4) ($r = .56$), which suggests that monitoring the depth and uptake of talk can be exercised with some independence from posing dialogic questions. Since none of the inter-factor correlations approached the .85 level often taken to signal poor discriminant validity (Hair et al., 2019, p. 134; Kline, 2016, p. 73), an oblique rotation was theoretically expected and empirically warranted. The four factors can therefore be interpreted as related yet distinct dimensions of preservice teachers' dialogic SE.

Table 5
Factor Correlation Matrix ($n = 312$).

Factor	1	2	3	4
1	1.000	0.706	0.713	0.681
2	0.706	1.000	0.652	0.648
3	0.710	0.652	1.000	0.562
4	0.593	0.567	0.562	1.000

Discussion

The current study developed and validated the Self-Efficacy Scale for Dialogic Teaching (SESDT) for preservice English teachers in Cambodia. Prior to factor analysis, the assumption diagnostics support the suitability of our dataset before extraction since item–total correlations were consistently moderate

to high ($r = .54-.75$), and sampling adequacy and factorability were excellent ($KMO = .942$; Bartlett's $\chi^2(378) = 5635.71, p < .001$). The EFA analysis produced a four-factor of 24-item solution with a structure and conceptual-fit results. Specifically, the four correlated dimensions collectively explained 55.46% of the variation in the extraction sums of squared loadings and were kept by using ML extraction with Promax rotation. The decision of using Promax as an oblique rotation was further warranted by the moderate inter-factor correlations ($.56-.71$). Item retention was determined by the normal EFA choice guidelines (i.e., loading $\geq .40$; cross-loading $< .30$; Δ loading $\geq .20$; conceptual fit), and we eliminated five underperforming items (7, 13, 21, 26, 28). However, we also examined the internal coherence with dialogic teaching to verify the scale structure prior to the deletion of any indicators.

Indicators from F1 and F3 both describe the cognitive exploration of DT as Alexander (2018, 2020) explained that dialogic lessons value talks as cumulative, reasoned, and accountable. The items that loaded on F1 underline these principles because they ask teachers to help students compare viewpoints, evaluate evidence, connect issues to real-life situations, and also interrogate biases in arguments. Mercer and Littleton (2007) further argued that guided reasoning is what makes educational dialogue and conversation different. The link with F4 ($r = .68$) also suggests that preservice teachers do not see reasoning and asking questions as two unrelated competencies but as one cycle in which questions open a space and then critical dialogue fills it. Nystrand et al. (1997) noted that open and uptake-oriented questions are important for keeping student ideas rolling, although Alexander (2018) regards such questions as a repertoire to encourage the teacher to raise awareness to promote verbal and cognitive tasks through means such as interrogating, redirecting, or expanding the inquiry.

Our F4 items mimicked this repertoire by including questions that invite students to perform tasks such as explanation and connections to previously learned material by the use of multiple-choice and other reasoning techniques. Teachers in our sample reported feeling confident in addressing such confrontation and asking the questions that make it urgent. In this regard, Abd Elkader's (2016) noted that DT is a potential approach to build democratic venues for facing difference. However, the fact that F1 still correlates slightly more with F3 ($r = .71$) also shows that preservice teachers may have been aware of the reasoning for monitoring students' learning, yet their responses show that critical dialogue and questioning can be a single intellectual pathway knowledge-building.

We note that the first factor explains a relatively large share of the variance (45.55%), which may suggest a broader underlying dimension of dialogic teaching self-efficacy. Nonetheless, the four-factor solution provides clear and meaningful distinctions that reflect key aspects of dialogic teaching. In this study, we focused on identifying these dimensions as a first step in scale development. However, this pattern also suggests that the SESDT could be represented using a higher-order or bifactor model. We see this as an important direction for future research, where confirmatory approaches can be used to examine the structure of the scale in more detail.

In Cambodian preservice classes, F2 and F3 focused on the regulative aspect of dialogic teaching. According to Mercer and Howe (2012), discussion can only be effective if a teacher can ensure all

students are participating and taking turns appropriately. They also need to make sure that students get scaffolding in the process. This is precisely what the items in our F2 do. Typically, the teachers must safeguard polite conversation and engage reluctant or timid students; they also create group discussions for mixed-proficiency courses by encouraging students to listen and take turns. According to the correlation between F2 and F3 ($r = .65$), preservice instructors who believe they can keep students talking also believe they can determine whether or not the conversation is advancing learning. According to Mercer and Howe (2012), our F3 indicators are in good agreement with the "teacher's work of making thinking visible." When students actively participate, build on one other's ideas, exhibit critical thinking, and tie the whole interaction to the learning objectives, they want teachers to take notice. In this context, dialogic SE aims to promote responsible small talk that adheres to assessment and practical criteria rather than idle small chat.

Finding multidimensional, interpretable SE structures utilizing common-factor approaches in their respective fields, as has been done in Cambodian validation work, seems to be in accord with our EFA patterns. In a national sample of primary teachers, Pov et al. (2024) validated an SE scale for inclusive practice and used PAF with Equamax; they reported KMO = .94, a significant Bartlett test, and 56.77% variance with one item trimmed as they explored and confirmed a three-factor TEIP configuration. Similarly, Em (2025) found cumulative variance compatible with a stable latent structure and large loadings (.45-.90) by using EFA with parallel analysis before further modeling. Another study pointing to multifactor effectiveness profiles in Cambodian settings was Seng et al. (2020), which went beyond teacher samples and retrieved four PCA components explaining 70.18% of variation in research SE.

The work that extends beyond Cambodia also followed a similar validation pattern for the self-efficacy scale and achieved similar results. For example, Hoang and Wyatt (2021) validated an SE measure for preservice EFL teachers in Vietnam and retained six task factors and a KMO of .85 and 71.05% explained variance. Gan et al. (2022) split the sample for EFA and CFA in order to develop the SEESS for higher education EFL speaking. Kircali and Cesur (2024) validated a 20-item cultural and linguistic diversity SE scale for preservice teachers and retained five factors in EFA before CFA. Palomino et al. (2025) retained four factors with high internal consistency in their self-efficacy scale development and validation in dialogic learning among college students. Meanwhile, Slater and Main (2020) used Rasch modeling and made the case for a short, unidimensional, highly reliable classroom management scale. However, these validations in ELT and related fields frequently produce related components, with a unidimensional form only in cases when the construct is strictly confined. Building on Yu et al. (2018), future studies may apply item response theory to further examine the SESDT, particularly to evaluate how well the scale distinguishes between different levels of dialogic teaching self-efficacy and to identify potentially weak or redundant items.

Implications and Limitations

We view the SESDT scale as a potentially useful tool for ELT preservice teachers and beyond. Thus, mentors and program supervisors should treat the four subscales as separate targets instead of using

the complete score for assessment. Through this scale, each preservice teacher may obtain a brief profile at the start of their practicum to make them aware of their weakness and strengths in delivering DT on the four dimensions in the classroom. For training purposes, mentors and their teacher candidates may choose, depending on suitability, to focus on one or two related dimensions or some talk moves for practice in classrooms for the initial evaluation of DT. We also suggest the use of a mid-placement check or pre-post evaluation to help teachers see their progress during the mentoring process. For evidence-based practice, we strongly recommend the use of observation rubrics and microteaching assignments to directly address the four strands because this approach may be useful for mentors to give targeted feedback about individual moves. Teacher education program managers may use cohort results, considering their specific contexts, to further improve their courses and preservice support initiatives.

Future studies should extend our psychometric validation of the SESDT through confirmatory factor analysis and measurement invariance testing using independent samples. Moreover, further validation should examine how SESDT scores relate to observed classroom practices. This may include the use of classroom observation rubrics, discourse analysis of teacher–student interaction, and measures of student engagement or learning outcomes. We suggest using well-known SEM packages, including IBM SPSS AMOS, Mplus, JASP or the R-based lavaan environment (and its web GUI versions), to conduct subsequent CFA and invariance analyses. Future SESDT validation research should also look into how the tool performs in various practicum or field placement settings, and investigate its relationship with dialogic observation ratings and student achievement results.

This SESDT is designed as a measure of preservice teachers' perceived capability to implement dialogic teaching; therefore, it is important to recognize that self-efficacy beliefs do not always correspond with actual classroom performance. As noted in prior research, teachers' responses may be influenced by confidence, prior experiences, or even sensitivity to social expectations rather than demonstrated competence. In this sense, the SESDT should be used as a measure of perceived readiness for dialogic teaching, and not as a direct indicator of instructional quality in practice. Nonetheless, self-efficacy remains a meaningful construct, as it influences teachers' instructional decisions and their willingness to engage in complex pedagogical practices such as dialogic teaching.

Declarations

Funding Statements and declarations

This work was prepared as part of the author's doctoral dissertation at Purdue University. The author received Ross Fellowship and graduate assistantship support (e.g., tuition and stipend) from the Purdue University College of Education. The funder had no role in study design, data collection, analysis, the decision to submit the work for publication, or manuscript preparation.

Compliance with ethical standards

The study protocol was reviewed and approved by the Institutional Review Board at Purdue University (IRB-2025-228; approval date: April 3, 2025). Written gatekeeper approval was also obtained from the Royal University of Phnom Penh, Cambodia. Participants were adult preservice teachers. Participation was voluntary and anonymous. Because the study involved an anonymous questionnaire and posed minimal risk, signed consent was not required; completion and submission of the questionnaire indicated informed consent. The manuscript reports only de-identified, aggregate data and contains no identifying personal information.

AI Use Statement

During the preparation of this work, the authors did not use any artificial intelligence (AI) tools to generate or write the manuscript text. AI-assisted tools were used only for limited support tasks. Specifically, tools such as Google Scholar and Google Scholar-related AI features (e.g., Scholar Lab) and Gemini/GPT-based tools were used to help identify relevant literature and to better understand concepts related to dialogic teaching and self-efficacy. Grammarly was used solely for proofreading purposes, including grammar, spelling, and minor language corrections. All interpretations, analyses, and written content were developed by the authors, who take full responsibility for the integrity and accuracy of the work.

Competing interests

The authors declare that there is no conflict of interest.

Data availability

The dataset for this study, including de-identified responses to the Self-Efficacy Scale for Dialogic Teaching (SESdT), is not publicly archived in order to protect participant confidentiality and comply with institutional ethics requirements. However, the data are available from the corresponding author upon reasonable request for scholarly purposes.

Acknowledgments

The authors are grateful to the reviewers and editor of the journal for their constructive feedback on an earlier version of this article.

Authors' biography

Samarnh Pang is a PhD candidate and graduate teaching assistant in literacy and language education at Purdue University's College of Education, USA. He teaches courses in English language teaching and research. He previously worked for ten years as a lecturer and teacher trainer at the Royal University of Phnom Penh. His research interests include teacher education, TESOL methodology, and educational technology.

Virak Chan is a Clinical Associate Professor in literacy and language and an undergraduate ELL pathway coordinator at Purdue University. He obtained a Ph.D. in Culture, Literacy, and Language at the University of Texas at San Antonio. He has extensive English language teaching and research experience, and has published in the areas of language planning and policies, linguistic landscape, teacher education, and teaching writing. He has also facilitated professional development activities for English language teachers, and developing and coordinating educational programs at universities in Cambodia, California, Texas, and Indiana.

Anne Traynor is an Associate Professor of Educational Psychology at Purdue University. She studies content validation methods, which provide evidence to establish the meaning of an educational test's scores. Her current research evaluates methods for quantifying test-to-curriculum content alignment and refining scoring rubrics.

References

1. Abd Elkader, N. (2016). Dialogic Pedagogy and Educating Preservice Teachers for Critical Multiculturalism. *SAGE Open*, 6(1), Article 2158244016628590. <https://doi.org/10.1177/2158244016628592>
2. Alexander, R. (2018). Developing dialogic teaching: genesis, process, trial. *Research Papers in Education*, 33(5), 561–598. <https://doi.org/10.1080/02671522.2018.1481140>
3. Alexander, R. J. (2008). *Essays on pedagogy*. Routledge.
4. Alexander, R. J. (2020). *A dialogic teaching companion*. Routledge.
5. Arslan, S., & Whitehead, G. E. K. (2025). Fostering pre-service teachers' perceived ability to implement dialogic teaching in Turkey: Examining the contributing factors of an intensive short-term teacher education program from the teacher-learners' vantage point. *Language Teaching Research: LTR*, 29(4), Article 13621688221093755. <https://doi.org/10.1177/13621688221093755>
6. Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 191–215. <https://doi.org/10.1037/0033-295X.84.2.191>
7. Bandura, A. (1997). *Self-efficacy: The exercise of control*. W.H. Freeman.
8. Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (Vol. 5, pp. 307–337). Information Age Publishing.
9. Barahona, M., Delaporte-Raurich, C., & Ibaceta-Quijanes, X. (2021). "It is impossible to teach English in English": Preservice teachers' struggles to facilitate L2 comprehensibility in English. *Tesol Journal*, 12(2), e578. <https://doi.org/10.1002/tesj.578>
10. Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quíñonez, 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. <https://doi.org/10.3389/fpubh.2018.00149>
11. Calcagni, E., Gröschner, A., & Klauf, S. (2024). How to Promote Dialogic Teaching During a Field Experience: A Quasi-Experimental Study of Preservice Teachers' Perceptions of Video- and

- Simulation-based Learning. *Pensamiento Educativo*, 61(2), Article 70741.
<https://doi.org/10.7764/PEL.61.2.2024.6>
12. Chow, B. W.-Y., Hui, A. N.-N., Li, Z., & Dong, Y. (2023). Dialogic teaching in English-as-a-second-language classroom: Its effects on first graders with different levels of vocabulary knowledge. *Language Teaching Research: LTR*, 27(6), Article 1362168820981399.
<https://doi.org/10.1177/1362168820981399>
 13. Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*, 10(1), 1–9. <https://doi.org/10.7275/jyj1-4868>
 14. Cui, R., & Teo, P. (2023). Thinking through talk: Using dialogue to develop students' critical thinking. *Teaching and Teacher Education*, 125, 104068. <https://doi.org/10.1016/j.tate.2023.104068>
 15. DeVellis, R. F., & Thorpe, C. T. (2021). *Scale development: Theory and applications* (5th ed.). SAGE.
 16. Edwards-Groves, C., & Hoare, R. (2012). "Talking to learn": Focusing teacher education on dialogue as a core practice for teaching and learning. *Australian Journal of Teacher Education*, 37(8), 82–99.
<https://doi.org/10.14221/ajte.2012v37n8.8>
 17. Em, S. (2025). Effectiveness of school directors' transformational leadership on teachers' self-efficacy and job performance in Cambodia. *CJESS Online Library*, 1–344.
 18. Fabrigar, L. R., & Wegener, D. T. (2012). *Exploratory factor analysis* (1st ed.). Oxford University Press.
 19. Freire, P., & Macedo, D. (1995). *A dialogic pedagogy: Culture, language, and race*. Routledge.
 20. Gan, Z., Yan, Z., & An, Z. (2022). Development and Validation of an EFL Speaking Self-Efficacy Scale in the Self-Regulated Learning Context. *Journal of Asia TEFL*, 19(1), 35–49.
<https://doi.org/10.18823/asiatefl.2022.19.1.3.35>
 21. Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (8th ed.). Cengage.
 22. Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in EFA. *Organizational Research Methods*, 7(2), 191–205. <https://doi.org/10.1177/1094428104263675>
 23. Henson, R. K., & Roberts, J. K. (2006). Use of EFA in published research. *Educational and Psychological Measurement*, 66(3), 393–416. <https://doi.org/10.1177/0013164405282485>
 24. Hoang, T., & Wyatt, M. (2021). Exploring the self-efficacy beliefs of Vietnamese pre-service teachers of English as a foreign language. *System (Linköping)*, 96, Article 102422.
<https://doi.org/10.1016/j.system.2020.102422>
 25. Jocuns, K. F. (2021). Dialogic teaching as a way to promote students' English language use in EFL classroom. *PASAA*, 62(1), 173-203. DOI: 10.58837/CHULA.PASAA.62.1.7
 26. Kim, S. A., & Morita-Mullaney, T. (2020). When preparation matters: A mixed method study of in-service teacher preparation to serve English learners. *Mid-Western Educational Researcher*, 32(3), 231-254. <https://scholarworks.bgsu.edu/cgi/viewcontent.cgi?article=1057&context=mwer>
 27. Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.

28. Kircali, M., & Cesur, K. (2024). Development and Validation of a Cultural and Linguistic Diversity Scale to Measure Pre-service EFL Teachers' Self-Efficacy Beliefs. *SAGE Open*, 14(1), Article 21582440241236589. <https://doi.org/10.1177/21582440241236589>
29. Klassen, R. M., & Tze, V. M. C. (2014). Teachers' self-efficacy, personality, and teaching effectiveness: A meta-analysis. *Educational Research Review*, 12, 59–76. <https://doi.org/10.1016/j.edurev.2014.06.001>
30. Lee, R. (2016). Implementing Dialogic Teaching in a Singapore English Language Classroom. *RELC Journal*, 47(3), 279–293. <https://doi.org/10.1177/0033688216631171>
31. Lopez-Hernandez, A. (2025). The Multivoiced Classroom: Microteaching in EFL and CLIL Teacher Training. *Theory and Practice in Language Studies*, 15(7), 2097–2103. <https://doi.org/10.17507/tpls.1507.01>
32. Ma, K., Luo, J., Cavanagh, M., Dong, J., & Sun, M. (2023). Measuring teacher self-efficacy: Validating a new comprehensive scale among Chinese pre-service teachers. *Frontiers in Psychology*, 13, 1063830. <https://doi.org/10.3389/fpsyg.2022.1063830>
33. Maddux, J. E. (2002). *Self-efficacy: The power of believing you can*. In C. R. Snyder & S. J. Lopez (Eds.), *Handbook of positive psychology* (pp. 277–287). Oxford University Press.
34. Mercer, N., & Littleton, K. (2007). *Dialogue and the development of children's thinking: A sociocultural approach*. Routledge.
35. Mercer, N. (2000). *Words and minds: How we use language to think together*. Routledge.
36. Mercer, N., & Howe, C. (2012). Explaining the dialogic processes of teaching and learning: The value and potential of sociocultural theory. *Learning, Culture and Social Interaction*, 1(1), 12–21. <https://doi.org/10.1016/j.lcsi.2012.03.001>
37. Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.
38. Nystrand, M., Gamoran, A., Kachur, R., & Prendergast, C. (1997). *Opening dialogue* (Vol. 37). New York: Teachers College Press.
39. Pajares, F. (2002). Overview of social cognitive theory and of self-efficacy. Retrieved from <http://www.emory.edu/EDUCATION/mfp/eff.html>
40. Palomino, M. del C. P., Montes, D. A., & Valdivia, E. M. (2025). Evaluation of Dialogical Learning in Higher Education: Psychometric Validation of a Questionnaire. *International Journal of Instruction*, 18(1), 619–634. <https://doi.org/10.29333/iji.2025.18134a>
41. 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 & Health*, 29(5), 489–497. <https://doi.org/10.1002/nur.20147>
42. Pov, S., Kawai, N., & Murakami, R. (2024). Exploring factors influencing teacher self-efficacy in implementing inclusive education in Cambodia: A two-level hierarchical linear model. *Studies in Educational Evaluation*, 83, Article 101415. <https://doi.org/10.1016/j.stueduc.2024.101415>

43. Resnick, L. B., Michaels, S., & O'Connor, C. (2010). How (well-structured) talk builds the mind. In D. Preiss & R. Sternberg (Eds.), *Innovation and creativity in education and training* (pp. 163–194). Routledge.
44. Reznitskaya, A. (2012). Dialogic teaching: Rethinking language use during literature discussions. *The reading teacher*, 65(7), 446–456. <https://doi.org/10.1002/TRTR.01066>
45. Rungwarapong, P. (2019). Factors affecting Thai students' participation in dialogic talks in EFL classes: students' perspectives. *Social Science Asia*, 5(1), 12–21. DOI: 10.14456/ssa.2019.3
46. Sanchez, S., & Athanases, S. (2023). Dialogic teacher inquiry: The case of a preservice teacher learning to facilitate class discussion. *Dialogic Pedagogy: An International Journal for Studies of Dialogic Education*, 11, A1–A38. <https://doi.org/10.5195/dpj.2023.482>
47. Schmitt, T. A. (2011). Current methodological considerations in exploratory and confirmatory factor analysis. *Journal of Psychoeducational assessment*, 29(4), 304–321. <https://doi.org/10.1177/0734282911406653>
48. Scholz, U., Gutiérrez-Doña, B., Sud, S., & Schwarzer, R. (2002). Is general self-efficacy a universal construct? *European Journal of Psychological Assessment*, 18(3), 242–251. <https://doi.org/10.1027//1015-5759.18.3.242>
49. Schunk, D. H., & DiBenedetto, M. K. (2020). Motivation and social cognitive theory. *Contemporary educational psychology*, 60, 101832. <https://doi.org/10.1016/j.cedpsych.2019.101832>
50. Schwarzer, R., & Jerusalem, M. (1995). Generalized Self-Efficacy Scale. In J. Weinman, S. Wright, & M. Johnston (Eds.), *Measures in health psychology: A user's portfolio. Causal and control beliefs* (pp. 35–37). NFER-NELSON.
51. Seng, S., Carlon, M.K.J. & Cross, J.S. (2020). Research Self-Efficacy of Cambodian Undergraduate Students at Province-Based Universities, *International Journal of Sociology of Education*, 9(2), 154–190. Doi: 10.17583/rise.2020.4267
52. Sherer, M., Maddux, J. E., Mercandante, B., Prentice-Dunn, S., Jacobs, B., & Rogers, R. W. (1982). The Self-Efficacy Scale: Construction and validation. *Psychological Reports*, 51(2), 663–671. <https://doi.org/10.2466/pr0.1982.51.2.663>
53. Slater, E. V., & Main, S. (2020). A measure of classroom management: validation of a pre-service teacher self-efficacy scale. *Journal of Education for Teaching: JET*, 46(5), 616–630. <https://doi.org/10.1080/02607476.2020.1770579>
54. Skidmore, D., & Murakami, K. (2016). *Dialogic pedagogy: The importance of dialogue in teaching and learning*. Multilingual Matters.
55. Thompson, B. (2004). Exploratory and confirmatory factor analysis: Understanding concepts and applications. *Washington, DC, 10694*(000), 3. <https://doi.org/10.1037/10694-000>
56. Tschannen-Moran, M., & Hoy, A. W. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education*, 17(7), 783–805. [https://doi.org/10.1016/S0742-051X\(01\)00036-1](https://doi.org/10.1016/S0742-051X(01)00036-1)
57. Uştuk, Ö., & Yazan, B. (2024). Tensions in an Identity-Oriented Language Teaching Practicum: A Dialogic Approach. *TESOL Quarterly*, 58(1), 363–393. <https://doi.org/10.1002/tesq.3234>

58. Wegerif, R. (2013). *Dialogic: Education for the Internet age*. Routledge.

59. Wilkinson, I. A. G., Reznitskaya, A., Bourdage, K., Oyler, J., Glina, M., Drewry, R., Kim, M.-Y., & Nelson, K. (2017). Toward a more dialogic pedagogy: changing teachers' beliefs and practices through professional development in language arts classrooms. *Language and Education*, 31(1), 65–82. <https://doi.org/10.1080/09500782.2016.1230129>

60. Worthington, R. L., & Whittaker, T. A. (2006). Scale development research. *The Counseling Psychologist*, 34(6), 806–838. <https://doi.org/10.1177/0011000006288127>

61. Yu, S., Traynor, A., & Levesque-Bristol, C. (2018). Psychometric examination of the short version of the learning climate questionnaire using item response theory. *Motivation and Emotion*, 42(6), 795–803. <https://doi.org/10.1007/s11031-018-9704-4>

62. Zee, M., & Koomen, H. M. (2016). Teacher self-efficacy and its effects on classroom processes, student academic adjustment, and teacher well-being: A synthesis of 40 years of research. *Review of Educational Research*, 86(4), 981–1015. <https://doi.org/10.3102/00346543156268>

Figures

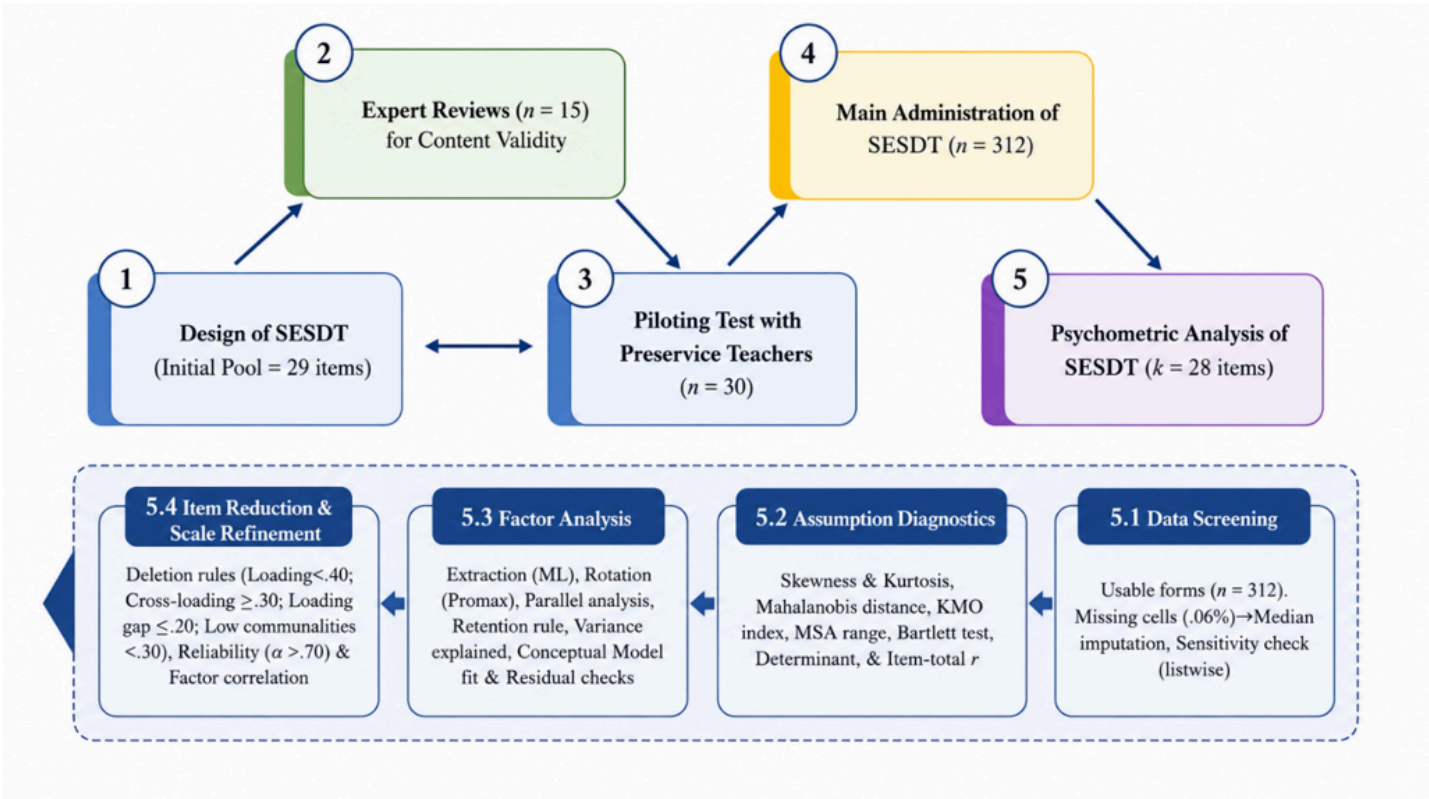


Figure 1

Phases in the validation of the Self-Efficacy Scale for Dialogic Teaching (SESdT).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Appendix.docx](#)