RESEARCH PAPER

The Influence of Digital Pedagogical Competency Training on EFL Teachers' Classroom Instructional Behaviors: A Teacher Professional Development Perspective

Ziye Li[1*], Luh Putu Artini[1], IGA Lokita Purnamika Utami[1], Ni Wayan Surya Mahayanti[1]

[1]Universitas Pendidikan Ganesha (Undiksha), Bali 81116, Indonesia

Emails: liziye5689@163.com (Corresponding Author), putu.artini@undiksha.ac.id, lokitapurnamika@undiksha.ac.id, surya.mahayanti@undiksha.ac.id

ABSTRACT

Background: With regard to the digital revolution in education, there is currently a worldwide strategic focus on the development of teachers' digital pedagogical competency. Yet, current studies tend to employ self-report measures alone to assess the impact of training programs, failing to consider behavioral data collected through classroom observation and neglecting the investigation of the distinct training effect transfer paths in relation to different behavioral aspects of instruction.

Purpose: The purpose of the study was to assess the impact of an 8-week-long digital pedagogical competency training program on the classroom instructional behavior of EFL teachers in China, using the interconnection between the TPACK model and the Interconnected Model of Teacher Professional Growth.

Methods: The design employed a quasi-experimental pretest-posttest nonequivalent control group strategy, complemented by an explanatory sequential mixed methods design. In total, sixty English as Foreign Language (EFL) teachers participated in the experiment; thirty of them were placed into the experimental group, and another thirty made up the control group.

Results: A two-way mixed ANOVA indicated significant group-by-time interactions on the total score (partial $\eta^2 = .442$, $p < .001$) and all three dimensions. The experimental group was found to significantly improve in their TI Behavior ($d = 1.48$), DA Behavior ($d = 1.05$), and CI Behavior ($d = 0.66$) and the effect size gradients were directional and in line with the predicted differential improvement. A thematic analysis highlighted three behavioral change pathways in relation to the Interconnected Model's domain transmission process, where peer collaboration proved to be the pivotal factor, with strong pedagogical beliefs impeding CI Behavior.

Conclusion: This study has provided behavioral evidence based on class observation and illustrated that the effects of training transfer to instructional behavior dimensions differently, leading to further implications for EFL digital training programs.

Keywords: digital pedagogical competency; TPACK; teacher professional development; classroom instructional behavior; EFL teachers; quasi-experimental design

How to cite this article: Li Z, Artini LP, Utami IGALP, Mahayanti NWS. The Influence of Digital Pedagogical Competency Training on EFL Teachers' Classroom Instructional Behaviors: A Teacher Professional Development Perspective. Int J Drug Deliv Technol. 2026;16(51s): 71-88. DOI: 10.25258/ijddt.16.51s.8

Source of support: Nil.

Conflict of interest: None

1. Introduction

The digital transformation of education has become one of the key foci in educational reforms at the international level. As mentioned by the OECD Digital Education Outlook released in 2023, "effective construction of digital education ecosystem calls for deliberate effort in promoting teachers' digital pedagogical competence" [1]. At the same time, the UNESCO Global Education Monitoring Report in the same year further points out that digital technology shall be regarded as an appropriate supplement to face-to-face teaching but not a replacement of traditional teaching, which relies greatly on the proper digital pedagogical training provided to teachers [2]. From the Chinese perspective, in order to ensure successful digital transformation of English as a foreign language teachers, the Ministry of Education published the Teachers' Digital Literacy Standard (JY/T 0646-2022) in 2022. Specifically, there are five aspects involved in digital literacy, including digital awareness, digital technical knowledge and skills, digital application, digital social responsibility, and professional development [3]. There have been several findings from empirical research conducted within the domain of CALL, indicating that digital technologies are able to contribute to the improvement of the quality of language teaching by means of creating multimodal input and giving instant feedback to students. More importantly, digital technologies excel in training oral production and delivering personalized instruction [4]. Although there has been much effort devoted to digital training in the area of TPD, it is still necessary to investigate whether or not teachers' newly gained knowledge and skills are successfully transferred to classroom teaching.

A large body of empirical knowledge has already been gathered within the field of research on the effectiveness of digital competencies training through TPD. According to the results of the meta-analysis including 59 quasi-experimental and randomized experimental studies, teacher education interventions have a significant positive influence on the growth of TPACK (Technological Pedagogical Content Knowledge) ($d = 0.839$, $p < 0.0001$), while longer duration of interventions leads to stronger effects [5]. Even though the results confirm the effectiveness of digital training at the general level, there still remain three important research gaps. First, with regard to measuring the impact of training on changes in the target behavior, a systematic review of the literature on this subject conducted during the last five years showed that most papers were based on the use of self-report measures of changes in instruction behavior among teachers and, therefore, contained indirect and unreliable information due to their vulnerability to social desirability bias and subjective distortions of cognitive processes [6]. Second, with regard to understanding the mechanisms of knowledge translation into action, the results of a longitudinal study among higher education teachers demonstrated a persistent gap between the intention of implementing the digital technology and its subsequent implementation in the classroom when the opportunity for practical experience arose [7]. Third, concerning how the knowledge was transferred into changes in different behavior dimensions, a systematic review of the literature indicated that training effects on instructional behavior change was predominantly studied as a single construct, while little attention was paid to identifying specific pathways [8]. The latter problem is especially relevant to EFL environments, because language teaching per se entails higher interaction and output feedback demands compared to other disciplines.

Building upon these three gaps in the existing literature, the current study adopts a quasi-experimental design with the help of

classroom observation coding in order to empirically investigate the impact of a digital pedagogical competency training program that lasted for 8 weeks on the classroom instructional behaviors of EFL teachers working in Chinese universities, where semi-structured qualitative interviews will be employed to explore cognitive processes underlying behavioral changes. In RQ1, we ask the question about whether the training program makes any significant difference in the classroom instruction behaviors of EFL teachers, thereby addressing the measurement-level gap by substituting self-report measures with behavioral observations. In RQ2, we ask the question about whether changes in instructional behaviors manifest different effects on technology integration behaviors, classroom interaction behaviors, and digital assessment behaviors, thus directly tackling the dimensional differentiation gap.

Contributions of this study can be identified in terms of three different aspects. First, in terms of theory, the study merges the TPACK model and Interconnected Model of Teacher Professional Development to form a theoretical framework that includes training inputs, cognitive restructuring, and behavioral transfers in order to create a new theoretical perspective on behavioral transformation mechanisms for digital training programs. Second, in terms of methodology, the application of classroom observation code sheets as the principal measurement tool to measure instructional behaviors improves ecological validity and objectivity because the actual behaviors performed by teachers are measured. Finally, in terms of practice, the findings offer empirical insights regarding the design and evaluation of digital training programs for EFL teachers.

2. Literature Review

2.1 Digital Pedagogical Competency in EFL Contexts

There has been a shift in the paradigm towards the understanding of teachers' digital competency from ICT literacy to digital competence and then to digital pedagogical competency. According to a systematic review conducted on 38 studies in the literature between 2014 and 2024, early works focused on teachers' capabilities regarding ICT literacy whereas current works have a focus on teachers' capabilities to integrate digital technologies for attaining subject goals [10]. The DigCompEdu framework is an extensively used structure by scholars for analyzing digital competencies, according to which there are six competence areas. The DigCompEdu framework consists of six competence areas, two of which are related directly to classroom instructional behaviors. These are 'Teaching and Learning' which involves the use of digital technology to enhance instructional approaches and 'Assessment', which refers to digital formative and summative assessment [11]. In EFL settings, however, digital pedagogical competence is to be further elaborated due to the specific nature of EFL instruction requiring multimodal inputs, instant interaction, and personalized feedback [12]. For that reason, in this paper, EFL teachers' digital pedagogical competency refers to the capability of EFL teachers to purposefully select and integrate digital technologies in language instruction to facilitate target language learning.

2.2 TPACK as a Framework for Digital Teaching Competency

The TPACK model sees technology-enhanced learning as an interaction of technological knowledge (TK), pedagogical knowledge (PK), content knowledge (CK), their intersections, and TPACK as the intersection of these three domains [13]. Compared to DigCompEdu and other general frameworks for digital skills, TPACK is a better lens through which one can see the relationship between teachers' knowledge about technologies, about pedagogy, and about subject content in shaping actual behavior. Meta-analysis shows that there exists a

significant correlation between the self-reported TPACK of the teachers and their actual technology use behaviors (correlation coefficient r = 0.23-0.49). This indicates both the moderate predictability of this construct and its relatively low correspondence with teachers' self-evaluations [14]. In this paper, TPACK will be the lens by which the classroom observation instrument would be divided into dimensions. First, the Technology Integration Behavior dimension refers to teachers' TK and TPK, including behaviors related to selection and use of technologies to deliver language content. Second, the Classroom Interaction Behavior dimension is based on PK and TPK related to digital mediation of teaching interaction. Finally, the Digital Assessment Behavior dimension is based on TCK and TPACK in action during classroom assessment and formative evaluation.

2.3 The Interconnected Model and the Training-to-Behavior Transfer Pathway

The Interconnected Model of Teacher Professional Growth sees professional learning as a non-linear process of change taking place along four interconnected dimensions. The External Dimension represents sources of stimulation, while the Personal Dimension represents teachers' knowledge, beliefs, and attitudes; the Practice Dimension involves teachers' classroom behaviors, whereas the Consequences Dimension concerns students' performances, all of which interact in two directions via enactment and reflection [15]. A contemporary empirical study conducted within a lesson study framework provided an adaptive transformation of the original model where the Group Dimension replaced the Personal Dimension and demonstrated even greater complexity and individual differences in sequences of change during professional learning in collaboration [16]. This dissertation research project will explore specifically the transfer of knowledge between the External Dimension and the other two dimensions through the Personal Dimension. The training program represents the input into the External Dimension, classroom observation codes directly assess changes in classroom behaviors (Practice Dimension), and interviews measure cognitive changes (Personal Dimension).

2.4 Empirical Evidence on Digital Training and Classroom Behaviors

Research studies into the characteristics of professional development programs have shown that such factors as collaborative learning environments, applicability opportunities, and supportive follow-up processes are positively related to meaningful shifts in teachers' instructional behavior and digital competences [17]. Yet, there exist several barriers limiting the transfer from attitudes towards adoption into actual classroom behavior change. One basic dichotomy within the literature on technology integration differentiates between the so-called first order barriers that are extrinsic to teachers, comprising problems such as insufficient infrastructure, institutional lack of support, and disparities in opportunities for receiving appropriate training, and second order barriers being intrinsic and comprising deeply held pedagogical beliefs and technologies attitudes, along with willingness to change [18][19]. The aforementioned dichotomy is important in a theoretical sense because it shows that while addressing the former type of barriers implies resource allocation, dealing with the latter requires longer term and continuous interventions aimed at challenging beliefs [20]. Such barriers are especially pronounced in English as Foreign Language (EFL) settings. An analysis employing an ecological systems theory perspective and covering 159 EFL teachers during five years uncovered multilevel obstacles varying from individual-level cognitive overload related to multitasking to institution-level misalignment between curricula and technology use, leading to much slower rates of digital integration amongst EFL teachers compared to

other fields [21]. Importantly, it appears that different types of instructional behavior are not equally susceptible to change induced by training. A comprehensive critical review of the evidence on TPD has shown that professional development oriented towards implementation of concrete behaviors yielded better results in terms of behavioral outcomes than did PD oriented towards broader educational philosophy and interactions [22]. Evidence supporting the differential transferability was also provided by the findings of a longitudinal study on the role of value beliefs in technology use by Chinese K-12 teachers: their importance decreased over time in favor of TPACK self-efficacy, which demonstrates different nature of these variables in terms of development and change [23]. Another longitudinal investigation based on a mixed-method approach and covering 11 months of a teacher education program found uneven TPACK development in terms of knowledge domain levels, with TPK and integrative knowledge domains developing considerably while TCK knowledge showed limited progress [24].

2.5 Conceptual Framework and Hypotheses

According to the above literature review, an integrative conceptual framework for this study will be developed as shown in Figure 1. This framework identifies TPACK as the descriptive layer of EFL teachers' digital pedagogical competency structural dimensions, while the Interconnected Model acts as the explanatory layer that reveals the process through which training input is converted to classroom behavior as a result of cognitive restructuring. H1 states that the post-intervention average TIB score will be significantly higher for experimental group teachers than that of the control group. In terms of H2, there will be a differential effect pattern with TIB > DAB > CIB. The reason behind this prediction is based on the varying degrees of cognitive complexity and belief dependence involved in each behavioral dimension. First, TIB entails specific tool manipulation skills that correspond the most with training modules and are easy to apply and transfer through the learning process, since such behaviors can be learned directly and immediately without the need for radical changes to teachers' existing orientations. At the other end of the continuum lies CIB, where teachers will have to make deep interactions that necessitate radical pedagogical orientation change due to being constrained by their longstanding pedagogical beliefs as well as students' habitual responses to them. DAB will fall somewhere between these two extremes in the sense that, although manipulating a digital assessment tool involves a similar procedure to that of TIB, the interpretation of students' responses and adapting instruction based on such interpretations involves more assessment literacy and pedagogical decision-making skills than simply operating tools.

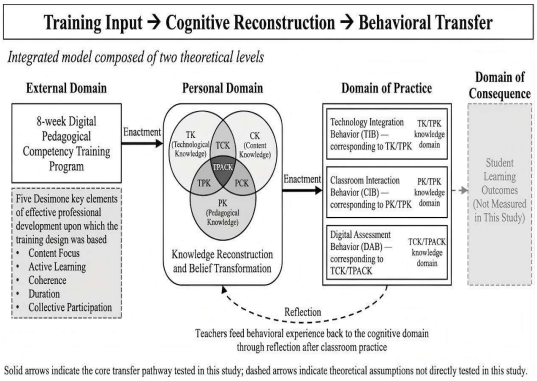


Figure 1. Conceptual Framework Integrating TPACK and the Interconnected Model of Teacher Professional Growth

3. Methodology

3.1 Research Design

For this study, a pretest-posttest nonequivalent control group quasi-experiment design [25] was used where the experimental group was subjected to the 8-week training intervention on digital pedagogical competencies whereas the control group continued teaching and research activities over the same period of time. A mixed-methods approach [26] consisting of a combination of quantitative and qualitative

methods was chosen whereby the quantitative strand involved conducting a pre- and post-test of the observed scores in classrooms which helped address Research Questions 1 and 2 while the qualitative strand consisted of conducting semi-structured interviews to explain the cognitive processes that led to the findings obtained through the quantitative phase and to address RQ3. Figure 2 presents the overall research design.

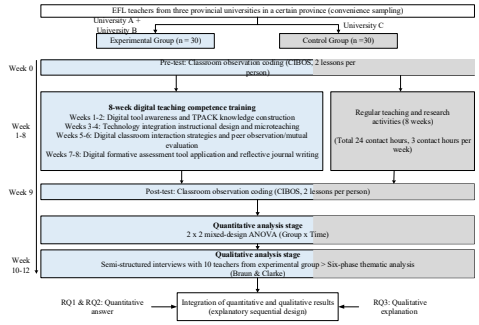


Figure 2. Research Design and Procedural Flowchart

3.2 Participants

A sample of sixty EFL teachers drawn from three provincial universities of a Chinese province participated in the study via convenient sampling. Universities were randomly selected based on feasibility and willingness to participate in the research, and all eligible EFL teachers in those institutions who fit the inclusion criteria were selected to take part in the study. Assignment into the experimental condition and control condition was done at the institution level in order to reduce contamination in between both conditions. Participants from two out of three selected institutions were designated to the experimental condition (n = 30), and teachers from the other one to the control condition (n = 30). This non-random assignment was a feature of the applied quasi-experimental design that arose due to practical concerns regarding contamination of the training activities at the same institution in terms of informal learning process. Demographic characteristics of the two groups as well as equivalency on the basis of gender, highest education degree, years of teaching, and CIBOS pre-test total mean score were measured and compared using independent samples t-tests and chi-square tests. No statistically significant differences were found on the measures between the two groups (all $p > .05$).

Using G*Power 3.1 software program, an a priori power analysis showed that 52 participants (26 participants per group) were necessary to test medium-to-large interaction effect ($f = 0.30$) with 80% statistical power at the significance level $\alpha = .05$ in 2 × 2 mixed ANOVA analysis. The sample size used in the study was larger than the required number.

3.3 Training Intervention

The training program was developed according to the effective professional development model [27], which considers content emphasis, active learning, coherence, sustained time, and collaborative engagement as five essential characteristics of successful teacher professional development programs. The model is grounded theoretically and empirically and currently represents the most popular framework for assessing professional development effectiveness. The program lasted for eight weeks with three weekly contacts, adding up to 24 hours, which corresponds to the minimum 20-hour requirement suggested by the model. In its general design, the program implemented a cognitive, instructional, and reflective spiraling progression process divided into four main modules.

Modules 1–2 (Weeks 1–2), involved awareness building and TPACK knowledge construction where teachers carefully analyzed the functional and pedagogical features of popular digital tools used in EFL teaching, reflecting the content focus component of the model. Modules 3–4 (Weeks 3–4) involved technology integration into instructional design and microteaching where teachers designed 15-minute TPACK-informed technology-integrated lesson

plans and gave mini-presentations to their peers [28], thus reflecting the active learning component of the model. Modules 5–6 (Weeks 5–6) included digital classroom interaction strategy development involving peer observation and structured reciprocal feedback where teachers observed their peers' digitally facilitated interactions and provided them with feedback, thereby reflecting the collective participation component of the model. Finally, modules 7–8 (Weeks 7–8) emphasized implementation of digital formative assessment strategies and reflective journal writing where teachers utilized their digitally oriented instructional approaches developed during the previous phases and reflected on their experiences in an in-depth way, thereby reflecting the coherence component of the model. Each of the modules consisted of one hour of theory input, one hour of practical workshop activities, and one hour of reflection on those activities. Fidelity of implementation was assessed based on attendance data and structured facilitator checklists filled out by facilitators after every training session. All 30 experimental group teachers attended at least seven out of eight weekly sessions (attendance rate: M = 93.8%, range = 87.5%–100%) and did not miss more than one session. Based on facilitator checklists, all components and activities intended for delivery were implemented as planned in both institutions without significant deviation from the planned approach.

3.4 Instruments

3.4.1 Classroom Instructional Behavior Observation Scale (CIBOS)

For the current study, the Classroom Instructional Behavior Observation Scale (CIBOS) was created based on the general principles of observational methodology and the philosophy of rating employed within the ICALT framework [29]. Considering the radical deviation from the initial ICALT dimension construction, the CIBOS can be described as an intentionally designed tool based on the ICALT ideas but not a modified version. The scale consists of three dimensions with 19 indicators overall, including Technology Integration Behaviors (TIB, 8 indicators), Classroom Interaction Behaviors (CIB, 6 indicators), and Digital Assessment Behaviors (DAB, 5 indicators). The level of implementation of each behavior is measured by means of a 5-point Likert-type scale, ranging from 'not observed at all' (code 1) to 'observed with high frequency and quality' (code 5). To ensure the content validity of the CIBOS, a two-round peer review was conducted. At the first round, the set of items (n = 27) was derived based on a combination of dimensions of the TPACK framework, DigCompEdu competence areas, and available observation tools. At the second round, a panel of five experts, consisting of two TPACK scholars, two EFL teacher educators, and one educational measurement expert, independently rated each item according to its relevancy, clarity, and appropriateness for the selected dimension. Items that scored an item-level content validity index (I-CVI) below 0.80 were excluded or corrected, leaving the final set of 19 items with a scale-level CVI (S-CVI/Ave) of 0.93.

Before the formal data collection procedure, a pilot study was conducted among a sample of 120 EFL teachers, who did not participate in the present study. The exploratory factor analysis (EFA) based on the principal axis factoring approach with Promax oblique rotation provided evidence for the appropriateness of a three-factor solution, explaining 67.8% of the total variance, in which all items loaded above 0.50 on their respective dimensions without any cross-loadings [30]. Specifically, in the pilot study, the 120 teachers were individually observed and rated by a single trained rater, resulting in 120 individual observation ratings for the purposes of factor analysis. Table 1 summarizes the three CIBOS dimensions

together with sample items used in this study, which were developed based on the dimensional model derived from the taxonomies of knowledge domains in the TPACK framework and the philosophy of observational ratings within the ICALT framework [13, 29]. The reliability coefficients (alpha) obtained from the pilot study showed an acceptable level of internal consistency reliability across all dimensions (TIB: $\alpha = .89$; CIB: $\alpha = .85$; DAB: $\alpha = .83$; scale total: $\alpha = .92$). In the present study, one complete teaching session lasting for 45 minutes was observed for each teacher. Regularly scheduled classes were selected as the unit of observation rather than specially designed demonstration lessons. The participants were informed of the study being a part of a research project, but not the targeted behavioral dimensions. Two separate lessons conducted on different days, involving two different types of lessons (reading comprehension and speaking practice, depending on the teacher's schedule) were observed for each teacher at each time point and averaged into one score per dimension per teacher.

Table 1. Dimensions, Number of Items, and Sample Items of the CIBOS

Dimension	Items	Sample Item
Technology Integration Behaviors (TIB)	8	The teacher purposefully uses digital tools to present language input materials
Classroom Interaction Behaviors (CIB)	6	The teacher utilizes digital platforms to organize real-time student interaction and collaboration
Digital Assessment Behaviors (DAB)	5	The teacher employs digital tools to conduct formative classroom assessment and provide immediate feedback
Total Scale	19	—

3.4.2 Semi-structured Interview Protocol

A sample size of ten teachers was selected purposefully from the experimental group through stratified purposeful sampling considering the behavioral change after the test, where three teachers with higher behavioral change, four teachers with moderate behavioral change, and three teachers with lower behavioral change were selected for ensuring thematic saturation [31]. Three major thematic domains were considered during these interviews regarding the views of the participants regarding the training and change in their instructional behavior, and their perception about their professional development paths. Each interview was conducted for a period ranging between 30 to 45 minutes.

3.5 Data Analysis

Quantitative analyses were carried out using SPSS 26.0. Classroom observation scores were obtained independently by two trained raters per lesson, and inter-rater reliability was measured using the intraclass correlation coefficient based on the two-way random effects model, ICC(2,1). Pre- and post-difference comparisons were conducted using a 2 (group) x 2 (time) mixed ANOVA, where group served as the between-subjects factor, and time was the within-subjects factor. Whenever there was statistical significance in the interaction effect of the group x time variable, post hoc simple effects analysis was used to examine the direction of the differences between groups and within groups. Despite the nesting of teachers in three different institutions, multilevel modeling was not conducted due to the inadequacy in number of Level-2 units ($k = 3$) falling far below the minimum 20-30 units recommended for hierarchical linear modeling for accurate variance component and standard error estimation [32]. Baseline equivalence testing of demographic and performance-related variables on an individual basis was adopted as an alternative, while the quasi-experimental design

with assignment at the institutional level is clearly recognized as a study limitation. Differences in effect sizes for the three behavioral dimensions were assessed through descriptive analysis of Cohen's d values with accompanying 95% CI. All inferential analyses yielded effect size metrics (Cohen's d and partial η^2) together with p values. The total CIBOS score of teachers was calculated as the arithmetic mean of the total scores in 19 items rather than as a weighted mean score of the means of the three dimensions, resulting in slight deviations when compared against post-hoc weightings from dimension-level means. Coding was done independently by two coders following the six-stage process of thematic analysis [33], while inter-coder agreement was assessed using Cohen's Kappa coefficient.

3.6 Ethical Considerations

The project was approved by the institutional ethics committee of the affiliated university. All the participants provided their consent for participation in the study, having understood its objectives, methods, and use of data. The data collected during interviews were anonymous, using codes like T-E01 to T-E10 for the experimental group interviewees. It was made clear that all participants could terminate their involvement in the study without any conditions being applied.

4. Results

4.1 Baseline Equivalence

Table 2 contains data on demographic variables and equivalency tests for both experimental and control groups. There were no statistically significant differences among the two groups in terms of either demographic variables or pre-test variables ($p > .05$).

Table 2. Demographic Characteristics and Baseline Equivalence Between Groups

Variable	Experimental Group (n = 30)	Control Group (n = 30)	Test Statistic	p
Gender (Female / Male)	22 / 8	20 / 10	$\chi^2 = 0.32$	.574
Qualification (Master's / Doctoral)	24 / 6	22 / 8	$\chi^2 = 0.37$	.541
Teaching Experience (years, M ± SD)	8.73 ± 3.41	9.20 ± 3.68	$t = -0.52$	.607
CIBOS Pre-test Total Mean Score (M ± SD)	2.71 ± 0.39	2.68 ± 0.41	$t = 0.30$	.768

4.2 Inter-rater Reliability

The systematically trained two raters independently rated a total of 240 classroom observations (60 teachers * 2 sessions * 2 lessons per teacher). The intraclass correlation coefficient for TIB, CIB, DAB, and total scores obtained using a two-way random-effects model (ICC(2,1)) was 0.91, 0.88, 0.87, and 0.90, respectively. All of these achieved good-to-excellent reliability (ICC > 0.75) and therefore provided validity of the coded observation measures for further statistical analysis.

4.3 Effects of Training on Classroom Instructional Behaviors

Descriptive statistics and results of the two-way mixed ANOVA on CIBOS scores of the experimental group and control group at pre-test and post-test are presented in Table 3. Before conducting the mixed ANOVA, we first assessed the assumptions underlying the analysis. According to the Shapiro-Wilk tests, the scores of CIBOS at pre-test and post-test in each group did not depart from normal distribution

significantly (all $p > .05$). The Levene's tests for equality of variances also showed that the homogeneity of variance across groups could be assumed for all dimensions at pre-test and post-test (all $p > .05$). As the within-subjects factor had only two levels, the assumption of sphericity was fulfilled by default. In terms of overall scores, the interaction between group and time was significant ($F(1, 58) = 45.86$, $p < .001$, partial $\eta2 = .442$), suggesting that there was a significant difference in change pattern between the two groups regarding their overall instructional behavior. On the level of dimensions, there was a significant group-by-time interaction effect on TIB ($F(1, 58) = 38.72$, $p < .001$, partial $\eta2 = .400$), CIB ($F(1, 58) = 9.67$, $p = .003$, partial $\eta2 = .143$), and DAB ($F(1, 58) = 22.41$, $p < .001$, partial $\eta2 = .279$).

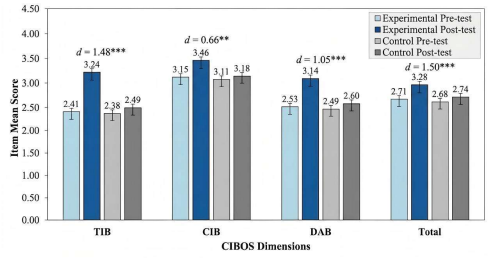
Table 3. Descriptive Statistics, Two-Way Mixed ANOVA Results, and Effect Sizes for CIBOS Dimensions

Dimension	Group	Pre-test M (SD)	Post-test M (SD)	Interaction F(1, 58)	p	Partial η^2	Within-group d (Experimental)
TIB	Experimental	2.41 (0.58)	3.24 (0.54)	38.72	< .001	.400	1.48 [0.90, 2.06]
	Control	2.38 (0.55)	2.49 (0.57)				0.20
CIB	Experimental	3.15 (0.49)	3.46 (0.45)	9.67	.003	.143	0.66 [0.13, 1.19]
	Control	3.11 (0.51)	3.18 (0.48)				0.14
DAB	Experimental	2.53 (0.61)	3.14 (0.55)	22.41	< .001	.279	1.05 [0.50, 1.60]
	Control	2.49 (0.58)	2.60 (0.60)				0.19
Total	Experimental	2.71 (0.39)	3.28 (0.37)	45.86	< .001	.442	1.50 [0.92, 2.08]
	Control	2.68 (0.41)	2.74 (0.40)				0.15

The simple effects test identified the directions of the interaction effects. Specifically, the participants in the experimental condition obtained significantly higher post-test scores compared to pre-test scores on TIB ($F(1, 58) = 42.15$, $p < .001$, $d = 1.48$ [95% CI: 0.90, 2.06]), DAB ($F(1, 58) = 24.87$, $p < .001$, $d = 1.05$ [95% CI: 0.50, 1.60]), CIB ($F(1, 58) = 10.23$, $p = .002$, $d = 0.66$ [95% CI: 0.13, 1.19]), and total scores ($F(1, 58) = 51.34$, $p < .001$, $d = 1.50$ [95% CI: 0.92, 2.08]), all at the medium-to-high effect levels ($d > 0.50$). On the other hand, there were no statistically significant increases in scores from pre-test to post-test on any dimension in the control condition (TIB: $F(1, 58) = 0.89$, $p = .349$; CIB: $F(1, 58) = 0.42$, $p = .520$; DAB: $F(1, 58) = 0.71$, $p = .403$; Total: $F(1, 58) = 0.52$, $p = .474$) with effect sizes being negligible ($d = 0.12–0.20$). This finding eliminates maturation as an alternative hypothesis. The gradient pattern along the three dimensions (TIB > DAB > CIB) is congruent with the pattern in the pre-test means (with the TIB having the smallest mean ($M = 2.41$) because of the novelty of digital technology integration and CIB having the highest mean ($M = 3.15$) because of the long-established tradition of classroom interaction as a teaching skill), indicating an effective enhancement of instructional skills. Thus, the obtained findings provide empirical support for H1. Although the directionality of

the obtained effect sizes is consistent with the predictions made in H2, it cannot be considered as a statistically proven hierarchy since the 95% confidence intervals overlap along dimensions.

As shown in Figure 3 below, there are two sets of grouped bar charts showing the comparison between the means of pre-test and post-test scores along all three dimensions for the experimental group and the control group.



Note. Error bars represent ±1 standard deviation. Effect sizes (Cohen's *d*) are displayed for the experimental group only. *** $p < .001$, ** $p < .01$.

Figure 3. Pre-test and Post-test Mean Scores of CIBOS Dimensions for Experimental and Control Groups

4.4 Qualitative Findings

Theme analysis of the interview data collected from ten teachers who were part of the experimental group resulted in three main themes and six secondary themes, with Cohen's Kappa = 0.86 intercoder agreement. All thematic categories and selected quotes are given in Table 4 below.

Table 4. Themes, Sub-themes, and Representative Quotations from Teacher Interviews

Core Theme	Sub-theme	Representative Quotation
Theme 1: Cognitive Transformation from Technology Anxiety to Pedagogical Confidence	1a. Dissolution of technology fear	T-E03: "Before the training, I was always worried that technical malfunctions during class would cause me to lose control of the classroom, so I avoided using any digital tools altogether."
	1b. Awakening of TPACK awareness	T-E08: "I began to understand that technology is not an add-on but a way of thinking that integrates with teaching content and learning objectives."
Theme 2: Restructuring of Instructional Design Thinking	2a. From content-first to goal-driven integration	T-E01: "Previously I would prepare the content first and then think about what tools to use; now I start from the learning objectives and consider how to embed technology into every stage of the lesson."
	2b. Assessment shift from summative to formative	T-E06: "Before the training I thought assessment meant final exams; now I realize I can use tools to monitor student comprehension at any point during class and adjust my teaching accordingly."
Theme 3: Peer Collaboration as a Catalyst for Behavioral	3a. Peer observation stimulating imitative	T-E07: "After seeing how a colleague used a real-time polling

Core Theme	Sub-theme	Representative Quotation
Transfer	learning	tool in class, I truly began trying it in my own teaching the following week."
	3b. Mutual evaluation promoting reflective practice	T-E05: "Peer feedback made me realize that although I had changed at the tool-use level, my interaction patterns had not actually transformed fundamentally in the way I engage with students."

Theme 1: Cognitive Shift from Technology-Related Fears to Technological Pedagogy Confidence. Multiple teachers mentioned having experienced high levels of fears and avoiding any use of technological equipment during lessons due to the fear of technology malfunction leading to classroom disturbances before the training session. After the training session was done using the systematic approach, such fears and avoidance began to dwindle down and be replaced by confidence to include technology within instructional design. As stated by one of the interviewees T-E08, "the training helped me understand that technology should not be seen as something that is added to the instructional material, but more like a mindset about the material."

Theme 2: Transformation of Instructional Design Thinking Process. Teachers unanimously agreed that the training session made them realize the need to shift their thinking processes on instruction from preparing teaching materials first, and then using technological tools to prepare materials to a more goal-oriented approach where technology should be integrated into content delivery through instructional goals. The same case applies to assessment. As mentioned by T-E06, "the training helped me understand that assessment should not only focus on the end-of-term assessments, but also continuous assessments using technological tools."

Theme 3: Peer Interaction as a Trigger for Reflection and Behavioral Change. Peer observation and collaborative feedback on teaching practices was considered by most teachers to be the most significant aspect of training that led to actual behavioral modification. Teacher T-E07 stated that witnessing the utilization of the polling application in action by a peer in a classroom environment was what prompted them to experiment with such applications themselves, indicating that the observation of peer practice reduced the mental barrier to behavioral modification through social proof. Importantly, teacher T-E05 offered even more insight in the sense that the feedback from peers indicated that although changes did happen at the surface level of using the digital tool, no significant transformation happened in terms of the interaction patterns in the classroom, which is directly reflected in the relatively smaller effect size noted for the CIB component in quantitative analysis.

While there was quite a bit of overlap between the three qualitative themes and the Interconnected Model used as a theoretical framework, there were also discrepancies. For example, Theme 1 matched quite well with the External-to-Personal pathway, indicating the connection between external input and cognitive restructuring. Theme 2 was also quite straightforwardly explained via the Personal-to-Practice pathway, illustrating how cognitive restructuring led to behavioral change. On the other hand, Theme 3 mostly described

reflection on peers' actions that served as a trigger for introspection and behavioral change. However, as discussed above, the observations provided insights that could not be accounted for in terms of the framework used, namely that the border between the Personal Domain and the Domain of Practice may be more permeable than the model predicts. Overall, triangulation between quantitative and qualitative results yielded reasonable support in that the first two themes suggested cognitive explanations behind the larger effect sizes of the TIB and DAB components, and the third theme corroborated the smaller effect size of the CIB component.

5. Discussion

5.1 Training Effectiveness and Behavioral Evidence

The principal result obtained in this study shows that training for pedagogical competency using digital methods has significant influence on EFL teachers' instructional behaviors, which can be seen from a big interaction effect of the total score of the experimental group (partial η2 = .442). These findings are compatible with several recent meta-analyses in terms of the magnitude of their findings. Thus, in one meta-analysis of 102 quantitative studies of teacher professional development (TPD), the researchers found a medium effect (Hedges' g = 0.55) [34], whereas in another synthesis of 46 experimental studies the estimated pooled impact was +0.52 standard deviations [35]. Although it would be somewhat inappropriate to directly compare within-group paired effect sizes calculated according to the Cohen's d formula and the between-groups Hedges' g effect sizes, the findings seem to be relatively similar, and thus it can be suggested that the current training program resulted in the effects equaling those typically found in the literature on TPD. In contrast to the above syntheses where mostly heterogeneous measures such as self-reported perceptions and knowledge tests were used, the main contribution of the current study is providing objective behavioral data based on structured classroom observation coding. Considering the Interconnected Model, the current quantitative findings demonstrate empirical support of the transfer process from the External Domain to the Personal and Domain of Practice domains, whereas Qualitative Themes 1 and 2 contribute to understanding of the mechanisms taking place in the Personal Domain and help uncover that training inputs are transformed into behavior changes via two cognitive mechanisms that include TPACK awareness awakening and instructional design thinking restructuring. In particular, it needs to be pointed out that the interaction effect size (partial $\eta^2 = .442$; or $f = 0.89$) is higher than the medium-to-large effect size expected in the power analysis ($f = 0.30$) and thus exceeds the anticipated value. Although this could partly depend on the thoroughness of training design based on the all five Desimone features of TPD success, it should also be mentioned that the non-blind design of classroom observations could increase the observed effects due to Hawthorne effects.

5.2 Differential Effects and Second-Order Barriers

The gradient trend in effect sizes observed between the three behavioral dimensions, namely TIB (d=1.48)>DAB (d=1.05)>CIB (d=0.66), has substantial theoretical implications but should be interpreted cautiously in light of an alternative psychometric explanation. Specifically, one possible interpretation may be that the gradient is partially due to the initial means scores for each variable. For example, TIB, which has the lowest mean (M=2.41), is in a better position to improve, whereas CIB, which has the highest mean (M=3.15), might struggle to change because of a proximity to the ceiling effect. Such a psychometric explanation is impossible to dismiss outright. That said, two pieces of evidence contradict the idea that such an explanation is the only one worth considering.

First, the qualitative evidence from Theme 3, especially the T-E05 participant, offers independent testimony to confirm that structural patterns were still resisting change, an issue unlikely to be caused by any purely psychometric factors. Second, the control group demonstrated no change whatsoever (d=0.14-0.20) despite having a similar initial baseline distribution, which suggests that being in a favorable psychometric position does not ensure improvement. Below, a theoretical interpretation will be presented where both second-order barriers and the baseline explanation will be viewed as two sides of the same coin.

This is because there is significant direct operational correspondence between the training module content (Modules 1-4) and the corresponding behavior. Previous studies have shown that professional development targeting concrete and easily transferable activities has a more significant effect on behavioral changes [22], which also explains the substantial effect of TIB. In contrast, the relatively low effect size for CIB is because second-order barriers affect classroom interaction patterns. According to the first-order/second-order barrier model, second-order barriers originate within the individual, such as deep-seated beliefs regarding teaching methods, technologies, and willingness to make changes [19]. Longitudinal research indicates that these deep-seated factors follow development paths significantly different from those of operational competencies, developing at a much slower pace despite repeated professional development programs [23]. While technical tools operation skills can be developed through professional development programs in a relatively short period of time, interaction behaviors depend on established belief structures among teachers, which will take more time and depth than 8 weeks to alter fundamentally. This was evident from interviewee T-E05 who explicitly indicated that while changes occurred concerning the tool operations, they did not affect the deep structure of the classroom interactions in Qualitative Theme 3.

5.3 Practical Implications

Based on the findings above, three evidence-based suggestions can be made regarding the design of EFL digital teaching training programs. From a content perspective, the training should go beyond mere modules about how to use the technology and include reflective and belief-questioning activities that specifically target interaction behavior changes so as to overcome the inhibitory influence of second-order barriers on interaction behavior transfer. Mechanistically, in view of the fact that collaborative activities among peers were recognized by all teachers to be extremely important in catalyzing their behaviors (Theme 3), it would be advisable that schools form communities of practice after the training is completed to enable the teachers to benefit from peer networks over an extended period of time [36]. Lastly, when it comes to the post-training stage, changes in CIB need follow-up scaffolding.

5.4 Limitations

There are four limitations that should be mentioned. First, participants were recruited from only three universities in one Chinese province, which might limit the external validity of this study, especially when the results are generalized to other areas. Eight weeks might be too little time to see significant changes in the deeply rooted attitudes towards teaching that could have an impact on teachers' interactive behaviors in the future. Learning outcomes of the students were not used as a variable of consequences domain, thus, there is no possibility to trace the impact of teachers' behavioral changes on students' learning process. Also, teachers were observed for only two lessons per time point, and thus, their everyday interactive behaviors might not be entirely captured by researchers. Lesson types were not

regulated; therefore, averaged results might conceal differences between behaviors observed during reading, speaking, listening, and writing lessons, as these lesson types require a different attitude towards digital tools usage and classroom interaction.

Moreover, the teachers in the experimental group were well aware of their inclusion in the structured training process and that their classrooms would be observed, thereby causing the Hawthorne effect where behavioral changes can be partially attributed to increased awareness of observation rather than actual internalization of the training material. While teachers in the control group were also observed twice, the imbalance between the two groups makes it impossible to separate the effects of training and the observer reactance effect. Also, grouping of the institutions into the experimental and control groups may have compromised the internal validity of the study due to uncontrolled confounding factors at an institutional level, such as institutional teaching culture, infrastructure quality, and collegiality in research conduct. The pilot validation of the CIBOS tool involved exploratory factor analysis using a sample size of 120 teachers, giving a sample-to-item ratio of 6.3:1, which is higher than the 5:1 minimum but still lower than the recommended ratio of 10:1. Moreover, there was no confirmatory factor analysis carried out on the primary study sample due to the relatively small number of subjects. Moreover, there was no blinding of the classroom observers since it was logistically difficult to blind them due to the need to observe teachers from three different institutions.

6. Conclusion

The current paper investigates the influence of 8-week training aimed at developing digital pedagogical competency of EFL teachers on their classroom instructional behaviors conducted at universities of China using quasi-experiment with structured classroom observation and semi-structured interviews. Training proved effective in increasing the levels of experimental group members' instruction behavior scores along all variables considered, whereas no changes were observed in the control group (RQ1). The effect size difference between TIB (d = 1.48), DAB (d = 1.05) and CIB (d = 0.66) was shown to be consistent with theoretical expectations in terms of its trend, yet overlapping confidence intervals failed to confirm the hypothesis that the hierarchy of effects is statistically significant (RQ2). Qualitative analysis helped identify three behavioral change mechanisms: cognitive transformation, instructional design modification, and collaborative efforts of peers being used as a stimulus to change. In this way, qualitative data contribute valuable explanations to quantitative results (RQ3). The main scientific novelty of the present study consists in collecting empirical behavioral evidence based on classroom observation and not subjective reports of participants as well as in uncovering that training effects differ depending on the behavioral dimension affected by the training, the existence of second-order barriers being responsible for inability to change the interactional dimension of behavior (RQ3).

References

[1] OECD. (2023). OECD Digital Education Outlook 2023: Towards an Effective Digital Education Ecosystem. OECD Publishing. https://doi.org/10.1787/c74f03de-en

[2] UNESCO. (2023). Global Education Monitoring Report 2023: Technology in Education — A Tool on Whose Terms? UNESCO Publishing. https://doi.org/10.54676/UZQV8501

[3] Ministry of Education of the People's Republic of China. (2022). (Teachers' Digital Literacy, JY/T 0646-2022). Beijing: Ministry of Education.

[4] Zou, B., Liviero, S., Hao, M., & Wei, C. (2023). Artificial intelligence technology for

EAP speaking skills: Student perceptions of opportunities and challenges. Technology in Society, 74, 102340. https://doi.org/10.1016/j.techsoc.2023.102340

[5] Jiang, N., Gao, J., & Yang, X. (2022). Teacher education interventions on teacher TPACK: A meta-analysis study. Sustainability, 14(18), 11791. https://doi.org/10.3390/su141811791

[6] Akram, H., Abdelrady, A. H., Al-Adwan, A. S., & Ramzan, M. (2022). Teachers' perceptions of technology integration in teaching-learning practices: A systematic review. Frontiers in Psychology, 13, 920317. https://doi.org/10.3389/fpsyg.2022.920317

[7] Müller, A. M., & Leyer, M. (2023). Understanding intention and use of digital elements in higher education teaching. Education and Information Technologies, 28, 19475–19494. https://doi.org/10.1007/s10639-023-11733-5

[8] Alanazi, M. A., & Alwadai, M. A. (2025). A systematic review on the impact of teacher professional development on digital instructional integration and teaching practices. Frontiers in Education, 10, 1541031. https://doi.org/10.3389/feduc.2025.1541031

[9] Li, S., Zhang, Y., Wang, H., & Liu, X. (2025). From challenges to chances in the digital age: Psychological transformation in EFL teachers in Chinese teacher education institutions. Frontiers in Education, 10, 1606932. https://doi.org/10.3389/feduc.2025.1606932

[10] Zhang, L., Yang, C., & Zheng, Y. (2025). Digital competence for sustainable education of pre-service teachers: A systematic literature review (2014–2024). Frontiers in Psychology, 16, 1710983. https://doi.org/10.3389/fpsyg.2025.1710983

[11] García-Delgado, M. Á., Rodríguez-Cano, S., Delgado-Benito, V., & Di Giusto-Valle, C. (2024). The digital competences necessary for the successful pedagogical practice of teachers in the digital age. Education Sciences, 14(5), 507. https://doi.org/10.3390/educsci14050507

[12] Hockly, N. (2023). Digital Literacies in Language Teaching (2nd ed.). Routledge. https://doi.org/10.4324/9781003296287

[13] Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. Teachers College Record, 108(6), 1017–1054. https://doi.org/10.1111/j.1467-9620.2006.00684.x

[14] Schmid, M., Brianza, E., & Petko, D. (2025). How valid, really? A meta-analysis of the validity evidence of TPACK self-report assessments. Learning and Instruction, 96, 102083. https://doi.org/10.1016/j.learninstruc.2025.102083

[15] Clarke, D., & Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. Teaching and Teacher Education, 18(8), 947–967. https://doi.org/10.1016/S0742-051X(02)00053-7

[16] da Ponte, J. P., Quaresma, M., Mata-Pereira, J., & Baptista, M. (2022). Teachers' learning in lesson study: Insights provided by a modified version of the interconnected model of teacher professional growth. ZDM – Mathematics Education, 54(2), 373–386. https://doi.org/10.1007/s11858-021-01317-5

[17] 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

[18] Joshi, B. M., & Khatiwada, S. P. (2024). Analyzing barriers to ICT integration in education: A systematic review. Journal of Comparative & International Higher Education, 16(5), 122–138. https://doi.org/10.32674/jcihe.v16i5

[19] Ertmer, P. A. (1999). Addressing first- and second-order barriers to change: Strategies for

technology integration. Educational Technology Research and Development, 47(4), 47–61. https://doi.org/10.1007/BF02299597

[20] Tsai, C.-C., & Chai, C. S. (2012). The "third"-order barrier for technology-integration instruction: Implications for teacher education. Australasian Journal of Educational Technology, 28(6), 1057–1060. https://doi.org/10.14742/ajet.810

[21] Zadorozhnyy, A., & Lai, C. (2025). EFL teachers' ecological barriers to integrating informal digital learning of English. TESOL Quarterly, 59(2), e3400. https://doi.org/10.1002/tesq.3400

[22] Sims, S., & Fletcher-Wood, H. (2024). The search for evidence-based features of effective teacher professional development: A critical analysis of the literature. Professional Development in Education, 50(1), 85–104. https://doi.org/10.1080/19415257.2023.2283437

[23] Xu, W., & Zammit, K. (2022). The evolution of the association between teacher technology integration and its influencing factors over time. Journal of Research on Technology in Education, 54(5), 727–747. https://doi.org/10.1080/15391523.2022.2030266

[24] Hofer, M., & Grandgenett, N. (2012). TPACK development in teacher education: A longitudinal study of preservice teachers in a secondary M.A.Ed. program. Journal of Research on Technology in Education, 45(1), 83–106. https://doi.org/10.1080/15391523.2012.10782598

[25] Creswell, J. W., & Creswell, J. D. (2023). Research Design: Qualitative, Quantitative, and Mixed Methods Approaches (6th ed.). SAGE.

[26] Creswell, J. W., & Plano Clark, V. L. (2023). Revisiting mixed methods research designs twenty years later. In C. N. Poth (Ed.), The SAGE Handbook of Mixed Methods Research Design (pp. 21–36). SAGE.

[27] 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/10.3102/0013189X09331140

[28] Antonio, R. P. (2025). Promoting technological pedagogical content knowledge (TPACK) in preservice science teacher education: A scoping review of instructional strategies, interventions, and programs. International Journal on Studies in Education, 7(1), 157–171. https://doi.org/10.46328/ijonse.302

[29] Van de Grift, W. (2007). Quality of teaching in four European countries: A review of the literature and application of an assessment instrument. Educational Research, 49(2), 127–152. https://doi.org/10.1080/00131880701369651

[30] Watkins, M. W. (2022). A Step-by-Step Guide to Exploratory Factor Analysis with R and RStudio. Routledge. https://doi.org/10.4324/9781003120001

[31] Leavy, P. (2022). Research Design: Quantitative, Qualitative, Mixed Methods, Arts-Based, and Community-Based Participatory Research Approaches (2nd ed.). Guilford Press.

[32] Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. Methodology, 1(3), 86–92. https://doi.org/10.1027/1614-2241.1.3.86

[33] Braun, V., & Clarke, V. (2022). Thematic Analysis: A Practical Guide. SAGE.

[34] Böttger-Gajewski, T., Kenner, A., & Lipowsky, F. (2025). Effects of online teacher professional development on teacher-, classroom-, and student-level outcomes: A meta-analysis. Computers & Education, 228, 105189. https://doi.org/10.1016/j.compedu.2025.105189

[35] Lynch, K., Gonzalez, K., Hill, H., & Merritt, R. (2025). A meta-analysis of the experimental evidence linking professional development interventions to teacher knowledge, classroom instruction, and student achievement. AERA Open, 11(1), 1–22.

https://doi.org/10.1177/23328584251335302

[36] Lantz-Andersson, A., Peterson, L., Hillman, T., Lundin, M., & Bergviken Rensfeldt, A. (2023). A realist review of online professional development and professional learning in digital contexts for teachers. Internet and Higher Education, 59, 100911. https://doi.org/10.1016/j.iheduc.2023.100911

[37] Kennedy, M. M. (2016). How does professional development improve teaching? Review of Educational Research, 86(4), 945–980. https://doi.org/10.3102/0034654316653907