Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Instructional Skills Workshops Improve Clinical Teachers' Self-Efficacy, but Only Some Gains Track Feedback Experience_ A Two-Cohort Study

Leyuan Xiang
Shanghai University of Traditional Chinese Medicine
Huan Mao
Shanghai University of Traditional Chinese Medicine
Jing Shu
drshujing@126.com

Shanghai University of Traditional Chinese Medicine

Research Article

Keywords: Instructional Skills Workshop, clinical teachers, teaching self-efficacy, feedback experience, faculty development, BOPPPS

Posted Date: June 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10052527/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

The Instructional Skills Workshop (ISW) develops the teaching competence of clinical teachers through repeated cycles of microteaching, peer and facilitator feedback, and re-practice. However, most studies have treated training effects as a single global outcome, rarely testing whether gains across different teaching-competence dimensions are uniform, and rarely examining which training experiences are associated with those gains. This exploratory study examined whether an ISW-based programme could enhance clinical teachers' teaching self-efficacy, and whether gains in different dimensions differed in their association with feedback experience.

Methods

A single-group pre–post design was implemented in two independent annual cohorts of core clinical teachers (2022 cohort N = 37; 2024 cohort N = 39; each trained between April and July of the respective year), using cross-cohort conceptual replication to strengthen reproducibility. Teaching self-efficacy was measured with a 12-item questionnaire adapted from the TSES, comprising three dimensions: Student Engagement (SE), Instructional Strategies (IS), and Classroom Management (CM). Feedback experience was measured with a 30-item questionnaire adapted from the AEQ, comprising five dimensions. Paired t-tests and Cohen's d assessed pre–post change, and Pearson correlations examined associations between feedback experience and self-efficacy gain scores (Δ). Given the exploratory aim and sample size, no correction for multiple comparisons was applied; cross-cohort replication served as the primary basis for credibility.

Results

All three dimensions improved significantly and substantially in both cohorts (all $p < .001$; 2022 cohort $d = 1.31–1.79$; 2024 cohort $d = 1.45–1.55$), with consistent direction and magnitude across cohorts. However, the association between gains and feedback experience showed a replicable divergence: gains in IS (and partly SE) correlated significantly with later feedback dimensions such as quality of feedback and what learners do with feedback, whereas gains in CM showed no significant association with any feedback dimension in either cohort (highest $r = .27$ in 2022; highest $r = .18$ in 2024). Given limited statistical power, the latter should be read as a failure to detect a significant association rather than evidence of no association.

Conclusions

This ISW-based programme produced large, replicable gains in teaching self-efficacy across all three dimensions. However, not all gains tracked feedback experience: improvement in classroom-management efficacy may have other sources. This divergence suggests that feedback should not be presumed a single unifying mechanism driving gains across all dimensions, a hypothesis warranting further investigation.

Background

The dual role of clinical teachers and the training dilemma

Within medical education, core clinical teachers have long carried the dual role of clinician and educator. Most enter teaching as subject-matter experts: their clinical competence is solid, yet they have rarely received systematic pedagogical training, having built their teaching ability largely through accumulated experience and self-directed trial, without a structured understanding of instructional design, classroom organisation, or learning assessment. At the same time, heavy clinical workloads leave them little time to invest in teaching development. These realities make brief, intensive, and immediately applicable faculty development programmes a better fit for this group than long-cycle, systematic training. Determining whether such short interventions genuinely improve clinical teachers' teaching, however, first requires an appropriate outcome measure. In faculty development research, teaching self-efficacy is a commonly used indicator of programme effectiveness. Defined as a teacher's belief in their capacity to organise and carry out instructional tasks effectively, it is regarded as an important psychological basis for teaching behaviour and engagement [1] and serves as an antecedent of whether teachers translate what they learn into improved practice. Among clinical teachers, it is further associated with teaching quality, greater professional fulfilment, and lower burnout[2], and it can be measured reliably through short pre–post assessment using established instruments[3], making it well suited to evaluating short interventions. We therefore adopted teaching self-efficacy as the primary outcome for evaluating the ISW. With the outcome defined, the next question is which features of a short intervention might move it, and whether there is evidence that it works.

The Instructional Skills Workshop and the evidence gap

The Instructional Skills Workshop (ISW) is precisely a short-term faculty development programme designed for this need. Originating in late-1970s Canada, the ISW uses small groups (typically 4 to 6 participants) to complete, within 24 to 30 hours, multiple cycles of "microteaching—peer and facilitator feedback—re-practice." It emphasises improving teaching through repeated mini-lessons and immediate feedback in a safe, collaborative atmosphere, and its classroom-organisation framework, BOPPPS, originated here[4]. Its core mechanism is not one-way lecturing but having participants obtain feedback on their own teaching across repeated "teach—evaluate—revise" cycles and adjust accordingly.

Prior work has reported that the ISW can change higher-education teachers' approaches to and behaviours in teaching[5], but most of this evidence comes from general higher education, and empirical evidence for clinical teachers as a group remains quite limited. The broader medical faculty development

(FD) literature likewise shows that such studies rely mainly on satisfaction and self-report, with relatively little robust behaviour-level evidence, and that single-group pre–post designs remain common in the field[6][7]. Existing ISW and similar studies have largely been one-off, small-sample satisfaction evaluations, lacking pre–post comparison, with even fewer replication tests across independent annual samples—making it hard to judge whether effects are robust or generalisable. Even within the broader field of clinical teacher training, whether training truly translates into improved teaching effectiveness remains contested[8]. More importantly, most studies have treated training effects as a whole, rarely examining whether ISW influences different dimensions of teaching competence to a consistent degree, and rarely asking which of the workshop's many design elements actually drives these gains. Answering the latter question first requires identifying a concrete target for examination. Among the many design elements of ISW, feedback is the one with the strongest rationale for being examined first: as noted above, ISW is not didactic instruction, and its "microteaching–feedback–re-practice" cycle is built around feedback, which serves as the pivot linking each teaching attempt to subsequent improvement[4]; the microteaching and faculty development literature has likewise consistently regarded feedback as the element through which reflection and instructional improvement occur. We therefore began by examining the association between feedback experience and gains across the efficacy dimensions.

Feedback experience: a mechanism to be tested, not a default engine

Among the ISW's many design elements, feedback is often seen as the most critical. This feedback takes more than one form: there is the written and oral feedback that peers and facilitators give on the spot, as well as the video-based feedback (self-observation) in which participants review recordings of their own mini-lessons against a standard. All of these are thought to help participants identify blind spots and adjust strategies, thereby raising teaching efficacy; in the clinical teaching context, how to give effective feedback has also been distilled into a set of actionable practical principles[9]. But feedback does not work automatically; its effect depends largely on the quality of the feedback and on whether learners accept and act on it[10][11]. From a peer-review perspective, feedback resembles a process actively constructed by the learner rather than passively received, and participants' interpretation and use of feedback are themselves part of learning[12]. Recent medical-education evidence similarly shows that how feedback is given and received affects its actual impact, and that feedback being perceived as effective is conditional[13][14]. There is thus a considerable distance between the intuition that feedback matters and the claim that a feedback mechanism can explain teachers' entire gains across every competence dimension.

The root of this gap is that teaching self-efficacy is itself a multidimensional construct, and its dimensions need not be driven by the same kind of experience. According to Bandura's self-efficacy theory[1], efficacy beliefs are rooted in specific task contexts, and efficacy in different competence domains may arise from different experiential sources, such as mastery experience, vicarious experience, and verbal persuasion. The widely used Teachers' Sense of Efficacy Scale (TSES) likewise divides teaching efficacy into three dimensions—student engagement, instructional strategies, and classroom management[15]—and gains in these dimensions need not all follow the same pathway.

Growth in some competencies may depend heavily on the individual's acceptance and use of feedback, whereas others may arise more from repeated practice, observational learning from peer demonstrations, or an overall boost in efficacy from the training setting, without passing through a specific feedback step. Presuming feedback to be the single engine driving all gains would obscure mechanistic differences across competence dimensions.

A more cautious approach is to treat feedback experience as an explanatory variable to be tested: first confirm whether the ISW genuinely raises teaching self-efficacy across dimensions, then examine separately whether gains in each dimension are associated with the feedback experience participants underwent. If gains in some dimensions correlate significantly with feedback experience while others do not, that divergence has theoretical value in itself: it indicates that the role of the feedback mechanism is bounded and that competence gains may have multiple sources.

The present study

On this basis, the present study took core clinical teachers selected from a Chinese university of traditional Chinese medicine and its affiliated clinical teaching institutions, used two cohorts enrolled in independent years (2022 and 2024), implemented a single-group pre–post evaluation of the same ISW training, and measured participants' feedback experience. Constrained by the ethical and organisational realities of a genuine faculty development programme, this study adopted a single-group pre–post design, using replication across independent annual samples to strengthen reproducibility; the analysis of associations between feedback experience and gains in each efficacy dimension was positioned as exploratory and hypothesis-generating (the design orientation and its limits are detailed in the Methods).

Within this framing, the study sought to answer three progressive research questions:

• **RQ1 (effect and replication):** Can the ISW robustly improve clinical teachers' teaching self-efficacy, and can that effect be replicated across two independent annual samples?

• **RQ2 (consistency across dimensions):** Are the ISW's gains in the different dimensions of teaching self-efficacy (student engagement, instructional strategies, classroom management) uniform?

• **RQ3 (association with feedback experience):** To what extent are gains in each dimension associated with the feedback experience participants underwent, and does the pattern of association diverge across dimensions?

Through a cross-year independent-sample replication design, this study aims to provide more robust evidence for the effectiveness of the ISW in clinical teacher development, and to make a cautious, exploratory test of the boundaries of feedback experience's role within it.

Methods

This study used a single-group pre–post design, implementing the same ISW training in two cohorts of core clinical teachers enrolled in independent years (2022 and 2024). The overall design and intervention structure are shown in Fig. 1. Constrained by the ethical and organisational realities of a genuine faculty development programme, no randomised control was used; instead, replication across independent annual samples was used to strengthen reproducibility. The analysis of associations between feedback experience and gains in each efficacy dimension was positioned as exploratory and hypothesis-generating (the full rationale for the design orientation is given at the end of the Background).

Participants

Participants were core clinical teachers selected from a Chinese university of traditional Chinese medicine and its affiliated clinical teaching institutions. The two cohorts attended a 2-day intensive ISW between April and July 2022 and between April and July 2024, respectively. A self-efficacy questionnaire was administered before and after the training, and a feedback-experience questionnaire was administered at the end of training. Groups were organised by hospital, placing teachers from the same affiliated hospital in the same group wherever possible, so that group members shared similar clinical departments and teaching contexts and could draw on existing collegial relationships to sharpen the relevance of feedback and the psychological safety of one another's practice sessions.

Participating teachers were selected uniformly by the university's affiliated clinical medical schools under the implementation rules of the university's "Teaching Competence Enhancement Funding Programme for Young and Mid-career Core Teachers." Selection allocated quotas to each clinical medical school, which then recommended candidates on merit, each school finalising its nominees through internal teaching review or competition (internal selection): the 2022 cohort was selected from 9 clinical medical schools (5 schools nominating 5 each and 4 schools nominating 3 each, for a formal quota of 37 in this round; an additional 2 core clinical teachers selected in the previous round, who had been unable to attend at the time because they were on assigned medical-aid missions away from the city, were carried over to attend with this round, giving an intended enrolment of 39). The 2024 cohort, building on this, added 1 medical school and was selected from 10 clinical medical schools (intended enrolment 40). Basic eligibility included: being engaged in clinical teaching; being in principle under 40 years of age; and holding a master's degree or above and an intermediate or higher professional-technical title; teaching liaison officers responsible for teaching administration were given priority among otherwise comparable candidates. All teachers enrolled voluntarily, with no mandatory participation. The effective sample sizes reported here, N = 37 (2022 cohort) and N = 39 (2024 cohort), are the numbers of valid matched cases after excluding those with missing pre- or post-test data.

Participant flow for both cohorts is summarised in Fig. 1. In the 2022 cohort, 39 teachers were nominated by 9 clinical medical schools according to quota; during training, 39 pre-test and 37 post-test questionnaires were collected. After name-by-name matching, 2 teachers had completed only the pre-test and were missing the post-test (they did not attend the post-test administration; these are not the same 2 teachers carried over for participation mentioned above) and were not included in the paired

analysis, with no other exclusions, yielding a final valid matched sample of 37 (response and matching rate 37/39 = 95%).

In the 2024 cohort, 40 teachers were nominated by 10 clinical medical schools (1 added relative to 2022) according to quota; of these, 1 teacher had completed only the pre-test and was missing the post-test (did not attend the post-test administration) and, following the same "exclude on any missing measurement" rule used for the 2022 cohort, was not included in the paired analysis, with no other exclusions, yielding a final valid matched sample of 39 (response and matching rate 39/40 = 98%). The two cohorts used identical recruitment, administration, and matching procedures; apart from those excluded for a missing post-test, no cases were actively removed for data quality or other reasons.

During the implementation of both the 2022 and 2024 programmes, the systematic faculty training that participants received in instructional design and teaching skills was organised uniformly by the university's Center for Faculty Development; they did not attend comparable systematic teaching-competence training organised by other institutions. This arrangement limited, to some degree, the interference of parallel external training on changes in self-efficacy between pre- and post-test, making the observed changes more plausibly attributable to this programme (including the ISW). However, a single-group pre–post design still cannot fully rule out systematic threats such as maturation; these limitations are addressed separately in the Discussion.

Training intervention

The two cohorts received the same training programme, delivered by the same facilitation team (both cohorts had a fixed complement of 2 identical facilitators) following the same procedure. The programme used the internationally established ISW as its basic framework, with localised adaptations made to fit the realities of clinical teachers. Given that core clinical teachers carry heavy clinical workloads and have limited time for teaching development, the programme compressed the standard procedure—originally requiring several days (about 3 days)—into a 2-day intensive format, while preserving the core training dose of the ISW. It comprised three interlinked parts.

The first part was an introductory lecture on instructional design, placed up front. Core clinical teachers mostly enter teaching as subject-matter experts and generally lack a structured understanding of instructional design; the programme therefore arranged a focused lecture on the principles of instructional design before entering the standard ISW procedure. This component is not an inherent part of the standard ISW but was added by the facilitation team based on the judgement that "instructional design is central to how teachers understand course teaching," in order first to help participants build an overall framework for understanding teaching, so that in the subsequent microteaching and feedback cycles they could examine their own teaching from an instructional-design perspective rather than remaining at fragmentary imitation of techniques.

The second part was systematic teaching and demonstration of the BOPPPS instructional framework. Building on the instructional-design introduction, the programme systematically taught the role and presentation of each of the six BOPPPS components, clarifying the dual links between BOPPPS and instructional-design principles on the one hand and the university's standard teaching requirements on the other, so that teachers could connect design concepts to their own teaching practice. A facilitator then gave a demonstration, deliberately taking "the learning psychology behind participatory learning" as the demonstration theme, to echo the up-front instructional-design introduction and let principle and procedural demonstration corroborate each other. This, too, was a localised design beyond the standard ISW.

The third part was the standard ISW microteaching cycle. In groups of 4 to 5, participating teachers completed multiple cycles of "microteaching—peer and facilitator feedback—re-practice." The training used the six BOPPPS elements—Bridge-In, Outcome, Pre-Assessment, Participatory Learning, Post-Assessment, and Summary—as a shared discourse for participants to organise microteaching and give one another feedback.

To complete the above within 2 days, the programme streamlined the elements of the standard ISW that focus on atmosphere-building and carry lower information density: the icebreaker was retained but simplified and shortened, and the team-game and drawing-summary components were removed; the number of microteaching rounds and the duration of feedback in each round were kept unchanged, to ensure that the core training dose of "practice—feedback—re-practice" was not reduced by the compression. In addition, in the video-feedback component, the programme added teaching-demonstration videos by teachers who had won national teaching competitions, for participants to observe before and after feedback. This design was intended to provide an external high-standard reference point for group feedback: when group members are similar in overall teaching experience and find it hard to point one another to higher-level improvements, an excellent demonstration can serve as a shared benchmark, raising the quality and constructiveness of feedback. Participants could also watch recordings of their own practice sessions for self-feedback, making feedback an actively constructed process.

Among these localised designs, the up-front introductory lecture on instructional design, the BOPPPS demonstration, and the award-winning teacher demonstrations in the video-feedback component together constitute a set of intervention features that distinguish this programme from the standard ISW. These features provide the necessary background for the later examination of differences in how gains in the various self-efficacy dimensions relate to feedback experience.

Measures

Teaching self-efficacy questionnaire

Teaching self-efficacy was measured with the short form of the Teachers' Sense of Efficacy Scale (TSES, 12-item short form) adapted from Tschannen-Moran and Woolfolk Hoy (2001) [15]. The original short form is a 12-item three-dimensional structure (student engagement, instructional strategies, classroom management). This study retained its three-dimensional framework and reworded the items to fit the clinical teaching context, keeping a 12-item self-report questionnaire scored on a 1–9 scale (consistent with the original TSES scoring), with items assigned to three dimensions by content, each dimension comprising 4 items:

Student Engagement (SE) comprised 4 items—"motivate students with low interest in learning," "help students believe they can do better," "help students improve self-directed learning," and "help students think critically" (items 2, 3, 4, 11); Instructional Strategies (IS) comprised 4 items—"ask good questions to guide learning," "use a variety of assessment methods," "provide an alternative explanation or example for students," and "implement different instructional strategies" (items 5, 9, 10, 12); Classroom Management (CM) comprised 4 items—"motivate students through positive feedback," "plan classroom structure and use time efficiently," "stay calm with students who do not follow the rules," and "create a collaborative learning atmosphere" (items 1, 6, 7, 8). Each dimension score is the mean of its 4 items, and the total scale score is the overall mean of the 12 items. The two cohorts used identical items and dimension assignments. The internal consistency of each dimension was assessed with Cronbach's α and corrected item-total correlations; results are reported in the "Reliability" section of the Results.

Feedback experience questionnaire

Participants' perceptions of the feedback experience during training were measured with a 30-item questionnaire adapted from the Assessment Experience Questionnaire (AEQ) of Gibbs and Simpson (2004) [16], scored on a 1–5 scale and including 13 reverse-scored items (reverse-item score = 6 − raw score). The AEQ was originally used to measure higher-education students' experience of course assessment and feedback; this study retained its scale structure and scoring, changing only the item context from "students' experience of course-assessment feedback" to "participating teachers' experience of peer and facilitator feedback in the workshop," to fit the peer-feedback setting of the ISW. The questionnaire is organised into five dimensions corresponding to the original AEQ: amount and distribution of study effort, assignments and learning, quantity and timing of feedback, quality of feedback, and what you do with the feedback. Of these, "quality of feedback" reflects the usefulness and clarity of feedback content, and "what you do with the feedback" reflects learners' acceptance of and action on feedback; treating feedback as a teaching skill that can be learned and improved through training also has empirical support in medical education[17]. Each dimension score is the mean of its constituent items (including reverse-scored items after reversal). The factor structure and applicability of this cross-context-adapted questionnaire require further validation; its internal consistency was assessed with Cronbach's α and corrected item-total correlations, with results reported in the "Reliability" section of the Results.

Data analysis

The following analyses were performed separately for each cohort. First, paired-samples t-tests compared pre- and post-test differences in the three self-efficacy dimensions and the total scale, with Cohen's d used to assess effect sizes. Next, Pearson product-moment correlations examined the associations between the five feedback-experience dimensions and the gain scores in the three self-efficacy dimensions (post minus pre, denoted ΔSE, ΔIS, ΔCM). The significance level was set at α = .05 (two-tailed). Statistical power was estimated for a one-sample correlation test at two-tailed α = .05 with N = 37–39 (G*Power 3.1). The correlation analysis between the five feedback dimensions and the three efficacy gain scores comprised 15 tests within each cohort, 30 in total across cohorts; this study treated this series of correlations as exploratory and applied no family-wise correction for multiple comparisons. Consequently, individual correlations near the significance threshold may be false positives, and interpretation relied primarily on whether a result replicated across the two cohorts rather than on the p-value of any single test.

A limitation of statistical power should also be noted: at N ≈ 37–39, the power to detect a medium correlation of r = .30 is about .56, below the commonly recommended .80. This power limitation is symmetric and must be interpreted in both directions: on the one hand, non-significant small-to-medium correlations are interpreted as a failure to detect a significant association at the current sample size rather than evidence of no association; on the other hand, the same low power means that significant positive correlations (e.g., r ≈ .39 within a single cohort) are themselves unstable estimates with wide confidence intervals, and a single significant result may overestimate the true effect. This study therefore does not rely on the significance of any single test; for both non-significant null results and significant positive results, replication across the two cohorts is the primary basis for credibility, and the Results and Discussion are framed by this symmetric standard throughout.

Results

Reliability

The adapted self-efficacy questionnaire showed acceptable-to-good internal consistency (Cronbach's α) in both cohorts: for the three dimensions (4 items each), α at pre-test/post-test was SE .78/.86, IS .68/.83, and CM .74/.80 in the 2022 cohort, and SE .82/.85, IS .75/.87, and CM .77/.82 in the 2024 cohort; corrected item-total correlations for each dimension's items were not below .30 in either cohort at either time point. The α for the IS dimension at 2022 pre-test (.68) was slightly below the conventional .70 threshold, attributable to the limited score variance of a short four-item dimension at pre-test; after training, the dimension's reliability rose to .83, with all other dimensions in the acceptable-to-good range.

The adapted feedback-experience questionnaire (AEQ) also showed good internal consistency in both cohorts: the full-scale (30-item) α was .94 in the 2022 cohort and .94 in the 2024 cohort; the α for the five dimensions in the two cohorts was amount and distribution of study effort .76 / .75, assignments and learning .76 / .70, quantity and timing of feedback .84 / .86, quality of feedback .79 / .80, and what you do with the feedback .85 / .87 (2022 / 2024, respectively); corrected item-total correlations for each

dimension's items were not below .30 in either cohort. The reliability coefficients of both questionnaires were highly consistent across the two independent annual samples, in keeping with this study's cross-year replication orientation.

Pre–post change in self-efficacy: consistent gains in both cohorts

Both independent cohorts showed significant gains in all three self-efficacy dimensions and the total scale, with effect sizes reaching the large-effect level (Cohen's d > 0.8) and consistent direction and magnitude across the two samples (Table 1).

Table 1. Paired pre–post comparison of self-efficacy in the two cohorts

Cohort	Dimension	Pre M (SD)	Post M (SD)	Difference M (SD)	t (df)	p	Cohen's d
2022 (N=37)	Student Engagement SE	5.78 (0.95)	7.14 (0.84)	+1.36 (0.99)	8.37 (36)	<.001	1.38
2022 (N=37)	Instructional Strategies IS	5.72 (0.79)	7.30 (0.79)	+1.58 (0.88)	10.91 (36)	<.001	1.79
2022 (N=37)	Classroom Management CM	5.99 (0.82)	7.20 (0.76)	+1.22 (0.93)	7.95 (36)	<.001	1.31
2024 (N=39)	Student Engagement SE	5.42 (1.19)	7.13 (0.92)	+1.71 (1.16)	9.21 (38)	<.001	1.47
2024 (N=39)	Instructional Strategies IS	5.42 (1.26)	7.20 (0.93)	+1.78 (1.23)	9.04 (38)	<.001	1.45
2024 (N=39)	Classroom Management CM	5.29 (1.28)	7.13 (0.94)	+1.84 (1.19)	9.66 (38)	<.001	1.55

In the 2022 cohort, gains in all three dimensions were highly significant (all p < .001), with effect sizes ranging from 1.31 to 1.79; the 2024 cohort likewise showed highly significant gains in all three dimensions (all p < .001), with effect sizes ranging from 1.45 to 1.55. Notably, the classroom-management dimension showed large-effect gains in both cohorts (2022 d = 1.31; 2024 d = 1.55), no weaker than the other two dimensions, and in the 2024 cohort its effect size (d = 1.55) was in fact the highest of the three dimensions. In other words, judged by the magnitude of pre–post gains, the ISW promoted all three self-efficacy dimensions broadly and evenly, with comparable gains across the three. The distribution of effect sizes across the three dimensions in both cohorts is shown in Fig. 2.

Association between feedback experience and self-efficacy gains: divergence across dimensions

Although the three dimensions showed comparable gains, their patterns of association with feedback experience diverged clearly and replicably (Table 2; Fig. 3).

Table 2. Pearson correlations between the five feedback-experience dimensions and self-efficacy gain scores

Feedback dimension	ΔSE (2022)	ΔIS (2022)	ΔCM (2022)	ΔSE (2024)	ΔIS (2024)	ΔCM (2024)
Amount and distribution of study effort	.18 (.286)	.12 (.479)	.06 (.725)	.12 (.467)	.18 (.273)	.06 (.717)
Assignments and learning	.22 (.190)	.19 (.260)	.14 (.408)	.15 (.362)	.22 (.178)	.09 (.586)
Quantity and timing of feedback	.28 (.093)	.33* (.046)	.21 (.212)	.24 (.141)	.31 (.055)	.16 (.330)
Quality of feedback	.31 (.062)	.36* (.029)	.24 (.152)	.19 (.247)	.26 (.110)	.11 (.505)
What you do with the feedback	.34* (.040)	.39* (.017)	.27 (.106)	.27 (.096)	.35* (.029)	.18 (.273)

Note. Cells show Pearson correlation coefficients r (with exact two-tailed p values in parentheses). *p < .05 (two-tailed). The 2022 cohort N = 37 (df = 35; critical value ≈ r = .325); the 2024 cohort N = 39 (df = 37; critical value ≈ r = .316). Exact p values were computed as $t = r \cdot \sqrt{(df)}/\sqrt{(1-r^2)}$. Across the 30 correlation tests in this series, no family-wise correction was applied; at α = .05, approximately 1.5 (30 × .05) false positives are expected by chance alone, so individual significant correlations should be interpreted conservatively, with inference relying primarily on replication across the two cohorts (see Data analysis and Discussion).

Gains in instructional-strategies efficacy (ΔIS) correlated significantly with "what you do with the feedback" in both cohorts (2022 r = .39; 2024 r = .35; both p < .05); for "quantity and timing of feedback," ΔIS correlated significantly in the 2022 cohort (r = .33, p < .05), while the 2024 cohort showed the same direction without reaching significance (r = .31, p = .055); in the 2022 cohort, ΔIS also correlated significantly with "quality of feedback" (r = .36, p < .05). Gains in student-engagement efficacy (ΔSE) correlated significantly with "what you do with the feedback" in the 2022 cohort (r = .34, p < .05), with the same direction in the 2024 cohort but not reaching significance (r = .27).

By contrast, gains in classroom-management efficacy (ΔCM) showed no significant correlation with any of the five feedback dimensions in either cohort: the highest in the 2022 cohort was r = .27 (with "what you do with the feedback"), and the highest in the 2024 cohort was r = .18. Following the statistical-power framing above, this should be stated as a failure to detect a significant association at the current sample size, rather than as classroom-management gains being entirely unrelated to feedback experience.

Within feedback experience there was also a consistent gradient: the dimensions more strongly associated with efficacy gains were mostly quality of feedback, what you do with the feedback, and

quantity of feedback, whereas the amount-and-distribution-of-study-effort and assignments-and-learning dimensions correlated lower overall and did not reach significance. This pattern was likewise visible in both cohorts, indicating that learners' acceptance of and response to feedback are more closely associated with changes in self-efficacy than the input into and amount of feedback given.

Summary

Combining the results of the two independent cohorts yields two robust findings. First, the ISW training significantly and substantially improved clinical teachers' teaching self-efficacy across all three dimensions, and this broad gain was replicated in two independent annual samples. Second, these gains were unequally associated with feedback experience: gains in instructional-strategies efficacy (and partly student-engagement efficacy) correlated significantly with later feedback dimensions such as quality of feedback and what you do with the feedback, whereas gains in classroom-management efficacy showed no detectable significant association with any feedback dimension at the current sample size. In other words, not all self-efficacy gains can be traced to the feedback experience an individual underwent.

The strength of replication warrants careful statement. With 15 correlations within each cohort and 30 in total, and no correction for multiple comparisons, the significance of any single test should be viewed conservatively, with replication across the two cohorts as the criterion. In terms of specific replication: the 2022 cohort had 4 correlations reaching $p < .05$ (ΔIS × what you do with the feedback, ΔIS × quality of feedback, ΔIS × quantity and timing of feedback, ΔSE × what you do with the feedback), of which only 1 (ΔIS × what you do with the feedback) likewise reached $p < .05$ in the 2024 cohort; the others, while consistent in direction, did not reach significance in the 2024 cohort. Thus the only significant association robustly replicated across cohorts was the positive correlation between instructional-strategies efficacy gains and what learners do with feedback; the remaining IS/SE correlations should be regarded as direction-consistent but weakly replicated, suggestive results. It must be stressed that this conservative stance is symmetric for positive and null results: under the low-power conditions of $N \approx 37–39$, a significant positive correlation within a single cohort (e.g., $r \approx .39$) is also an unstable estimate and should not be treated as an established association; it is more credible only because it replicated across cohorts for ΔIS × what you do with the feedback, not because any single test reached $p < .05$. What truly replicated robustly across the two cohorts were the two more macro-level patterns: large gains across all three dimensions, and the absence of any significant association between ΔCM and all feedback dimensions. Possible explanations for this divergence are developed in the Discussion.

Discussion

Why classroom-management efficacy gains are unrelated to feedback experience: an explanation from the training design

The more noteworthy finding of this study is not that the programme broadly improved clinical teachers' teaching self-efficacy, but that this improvement diverged clearly and replicably across dimensions in its association with feedback experience. Gains in instructional-strategies efficacy and student-engagement efficacy related mainly to later feedback dimensions such as quality of feedback and teachers' use of feedback; gains in classroom-management efficacy were different—they reached the large-effect level in both cohorts, no weaker in magnitude than the other dimensions, yet showed no detectable significant association with any feedback dimension (highest r = .27 in 2022; highest r = .18 in 2024). Even reading this conservatively as a failure to detect an association rather than confirmation of no association, one thing still holds: classroom-management efficacy gains did not co-vary with teachers' acceptance and use of feedback the way the other two dimensions did, suggesting they may have another source and need not be realised through the feedback cycle. This is consistent with a multi-source view of self-efficacy: Bandura noted that efficacy beliefs are rooted in specific task contexts, and efficacy in different competence domains can arise from different sources—enactive mastery, vicarious experience, verbal persuasion[1]—so gains in different teaching dimensions need not all travel the same feedback pathway. It should first be noted that the following explanation of this divergence is exploratory and hypothesis-generating; moreover, there is a measurement-level alternative explanation that cannot be distinguished from the design explanation (see the end of this section), so the mechanism proposed below should be regarded as a hypothesis to be tested rather than an established conclusion.

The design explanation below fits this study's training programme. What this study implemented was not the standard ISW but an arrangement that, beyond its feedback cycle, added an up-front introductory lecture on instructional design and deliberately took the learning psychology behind participatory learning as the main thread of the BOPPPS demonstration. These components convey, through principle-based teaching and demonstration, the overall framework for organising a lesson, and do not depend on participants first obtaining feedback on their own microteaching performance before adjusting. What the feedback cycle most directly acts on are those teaching behaviours that can be specifically pointed out by peers and facilitators in successive microteaching sessions and immediately revised—such as how to ask guiding questions, how to use varied assessment methods, and how to organise participatory activities. This chain—acquiring a specific skill, believing one can use it, and thereby raising efficacy—is consistent with the mechanism by which skill-based interventions raise teacher self-efficacy[18], and also accords with a core view in feedback research: whether feedback translates into improvement depends on the learner's acceptance and active use of it, rather than on the giving of feedback itself[10][11][12]. These fall precisely within the scope of instructional strategies and student engagement, which is why gains in these two dimensions correlated significantly with teachers' "use of feedback." Classroom-management efficacy, by contrast, is more about an overall grasp of classroom structure—planning the pace of a lesson, using time efficiently, fostering a collaborative yet orderly atmosphere; this kind of competence is more like a design awareness running throughout, hard to slice into specific actions that can be fed back item by item within a single microteaching session, and thus more likely to be established as a whole in the up-front instructional-design lecture and reinforced through demonstration, rather than accumulated step by step through the feedback cycle. Accordingly, an

explanation consistent with the training design is that classroom-management efficacy gains were driven more by this design, which directly imparts the overall teaching framework, than by feedback on individual performance; if that hypothesis holds, it would explain why these gains were large yet lacked association with feedback experience.

This, however, is only one of several possible explanations. An equally plausible alternative is at the measurement level: classroom management is itself a more holistic, diffuse competence, and prior factor-structure research indicates that, over and above general teaching efficacy in the TSES, it remains a distinct dimension with significant independence[19], capturing teachers' overall grasp of classroom structure and order. The items of such holistic competence (lesson pace, time use, order and atmosphere) may inherently be hard to move with the specific feedback an individual receives in a single microteaching session, so that even without the up-front lecture component, ΔCM and feedback experience might still show a weak association. In other words, attributing classroom-management efficacy gains to the up-front lecture (the design explanation) and holding that its items are inherently insensitive to individual feedback (the measurement explanation) can both produce the present data pattern, and this study's design cannot distinguish the two—precisely the reason, stated above, that the design explanation should be regarded as a hypothesis to be tested. Although the two explanations differ in mechanism, both point to the same judgement, consistent with this study's position in the Background: feedback should not be presumed the single engine driving all gains. Teachers' different teaching competencies may follow different developmental pathways—some more dependent on the individual's acceptance and use of feedback, others arising more from direct acquisition of an overall teaching framework and observation of demonstrations. This study's training programme happened to contain both kinds of mechanism, and the divergence in association patterns across dimensions may be precisely the result of each operating on different competencies, warranting examination in future research with more direct designs.

This divergence has implications for the design of clinical faculty development programmes. It suggests that short-term workshops aimed at improving teachers' different teaching competencies should perhaps not rely solely on the feedback-cycle pathway, but should deliberately incorporate, beyond feedback, direct teaching and demonstration of instructional-design and classroom-organisation principles, so as to address the holistic competencies that are harder to acquire through successive feedback.

Threats to internal validity and their mitigation

As a single-group pre–post study, this study's pre–post gains must be carefully distinguished from the true training effect against several alternative explanations.

First, regression to the mean. The participants were core clinical teachers produced through merit-based recommendation by each clinical medical school on the basis of teaching performance and competitive internal selection (teaching "PK"); since the selection explicitly involved evaluation of teaching

performance, it cannot be ruled out that some selectees happened to be at a temporary low point in teaching self-rating, with part of their post-test rebound attributable to statistical regression[20]. This threat cannot be wholly eliminated in a single-group pre–post design, but the fact that both independent annual cohorts showed gains of consistent direction and magnitude, with effect sizes far larger than regression to the mean can typically explain, lowers the persuasiveness of this alternative to some degree. To test this more directly, we examined the correlation between each dimension's pre-test score and its gain (Δ): all three dimensions showed significant moderate-to-strong negative correlations in both cohorts (2022 cohort: ΔSE, ΔIS, ΔCM with pre-test $r = -.42, -.38, -.45$, respectively, all $p < .05$; 2024 cohort: $-.51, -.48, -.53$, respectively, all $p \leq .002$), that is, those with lower pre-test scores gained more. This pattern is consistent with regression to the mean, suggesting that part of the gain may contain a statistical-regression component to be taken into account when interpreting Δ scores. At the same time, it clearly shows that the data have no ceiling effect (a ceiling would manifest as limited gains among high scorers, not larger gains among low scorers), indicating that the large gains across the three dimensions were not artificially compressed or inflated by a measurement ceiling among high-baseline participants. Notably, the baseline–gain negative correlations were of similar strength across the three dimensions and did not show classroom management to be more susceptible to regression than the other two; regression to the mean therefore cannot explain the later divergence in the association pattern between ΔCM and feedback experience.

Second, maturation and history. The two cohorts were two years apart, and if the institution undertook other faculty-development initiatives in the interim, their contribution to self-efficacy gains cannot be wholly excluded. As described in the Methods, during both programmes the systematic teaching-competence training participants received was organised uniformly by the university's Center for Faculty Development, and they did not attend comparable training organized by other institutions, which limited the interference of parallel external training to some degree; moreover, training was conducted immediately after teachers were selected into the core clinical teacher programme, with the pre-test administered before the workshop and the post-test at its end, so the interval between pre- and post-test was merely the 2-day intensive training, on which time scale the influence of maturation is extremely limited.

Third, social desirability and the Hawthorne effect. Participants knew they were undergoing training, and post-test self-ratings may have been inflated by expectation and being observed; this is a common limitation of self-report efficacy measures, and a single-group pre–post design cannot fully separate it from the true training effect. Future studies could corroborate findings with more objective indicators such as behavioural observation or third-party ratings.

Replication across independent annual samples mitigates the above threats only to a limited and conditional extent. If the two cohorts were broadly comparable in selection criteria, institutional environment, and parallel training, then the stable re-emergence of the effect in two independent samples can to some degree support that it is not due to a single year's chance factor; but precisely because the two cohorts shared the same training programme and implementation team, any

systematic bias present in the same direction in both cohorts (such as social desirability, the Hawthorne effect, or a common regression tendency) cannot be ruled out by replication itself. The replication evidence in this study should therefore be understood as support for the reproducibility of the effect, not as ruling out threats to internal validity.

Limitations and future directions

First, this study used a single-group pre–post design, and the up-front instructional-design lecture and participatory-learning demonstration were implemented bundled with the standard ISW procedure rather than manipulated or controlled as independent experimental conditions; the above explanations are therefore inferences based on design features and association patterns, not causal evidence. Single-group pre–post designs are common in the evaluation of educational and training interventions and have pragmatic value under ethical and organisational constraints, but their conclusions are susceptible to biases such as maturation and regression to the mean (these threats are detailed above in "Threats to internal validity and their mitigation") and require cautious interpretation[21][22]; confirming a causal relationship between the training and teaching self-efficacy gains still awaits randomised controlled trials or more rigorous quasi-experimental designs (such as waitlist control, non-equivalent control groups, or regression discontinuity). This study used Δ scores as gain measures correlated with feedback experience; difference scores have known limitations of correlation with baseline, and future studies could consider baseline adjustment (e.g., ANCOVA) to mitigate this[19]. Second, limited by sample size and statistical power, the classroom-management dimension showed no detectable significant association, which cannot be equated with confirmed independence; larger samples are needed for future testing. Third, this study measured feedback experience rather than the experience of the up-front lecture component; to directly verify the above explanation, subsequent studies should simultaneously measure teachers' perceptions of the instructional-design lecture and principle demonstration and examine their association with classroom-management efficacy gains. Even so, the consistent divergence pattern replicated across two independent years already provides fairly robust preliminary support for the judgement that gains in different competence dimensions arise from different mechanisms, and points the way to testing this with more rigorous designs.

Conclusions

This study took two cohorts of core clinical teachers enrolled in independent years, implemented a single-group pre–post evaluation of the same Instructional Skills Workshop training, and measured participants' feedback experience. The results show that this ISW-based, localized training programme significantly and substantially improved clinical teachers' teaching self-efficacy across the three dimensions of student engagement, instructional strategies, and classroom management, with all three effect sizes reaching the large-effect level and consistent direction and magnitude across two independent annual samples. This cross-year independent-sample replication provides relatively robust evidence of reproducibility for the effectiveness of this ISW-based localized training programme in

clinical teachers, a population for which prior evidence has been quite limited. However, these gains were unequally associated with feedback experience: gains in instructional-strategies efficacy (and partly student-engagement efficacy) correlated significantly with dimensions such as quality of feedback and what learners do with feedback, whereas gains in classroom-management efficacy showed no detectable significant association with any feedback dimension in either cohort. This replicable divergence suggests that feedback need not be the single mechanism driving gains across all competence dimensions, and that improvement in classroom-management efficacy may have another source.

This was an exploratory single-group pre–post study; constrained by its design, sample size, and the fact that the up-front lecture component was not independently controlled, the above interpretation of mechanism remains an inference based on association patterns rather than causal evidence, with the relevant limitations detailed in the Discussion. Within these limits, the consistent results replicated across years offer preliminary support that this ISW-based programme helps improve clinical teachers' teaching self-efficacy; clarifying "which training experiences drive which competence gains" remains key to refining this training pathway and testing it with more rigorous designs in future.

Declarations

Ethics approval and consent to participate

This study was an educational evaluation of a routine faculty-development programme for university clinical teachers. It used anonymous self-report questionnaires and involved no patients, clinical data, or any invasive/biomedical intervention, and no identifiable sensitive personal information was collected. In accordance with Article 32 of the Measures for Ethical Review of Life-Science and Medical Research Involving Humans (People's Republic of China, 2023), which exempts research conducted using anonymous information that causes no harm to participants and involves no sensitive personal information or commercial interest, and consistent with the principles of the Declaration of Helsinki for minimal-risk educational evaluation, this study qualified for exemption from formal ethics review. This determination was reviewed and confirmed as eligible for ethics exemption by the Office of Academic Affairs, Shanghai University of Traditional Chinese Medicine. All participants were enrolled voluntarily and were informed before completing the questionnaires that their anonymous, aggregated responses would be used for programme-quality evaluation and educational research; voluntary completion of the questionnaire was taken as informed consent. All procedures adhered to the relevant principles of the Declaration of Helsinki.

Consent for publication

Not applicable. This article contains no identifiable individual data, images, or case information.

Availability of data and materials

The de-identified datasets supporting the conclusions of this study are available from the corresponding author upon reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This study received no specific external funding.

Authors' contributions

Leyuan Xiang was responsible for manuscript writing and statistical data analysis; Huan Mao was responsible for workshop implementation and data collection; Jing Shu provided guidance on the overall writing approach of the manuscript. All authors have read and approved the final manuscript.

Acknowledgements

The authors thank all participating clinical teachers, and the affiliated clinical medical schools for their support in teacher selection and organisational coordination.

References

1. Bandura A. Self-efficacy: the exercise of control. New York: W. H. Freeman; 1997.
2. Tipwong A, Hall NC, Snell L, Chamnan P, Moreno M, Harley JM. Clinical teaching self-efficacy positively predicts professional fulfillment and negatively predicts burnout amongst Thai physicians: a cross-sectional survey. BMC Med Educ. 2024;24:361. https://doi.org/10.1186/s12909-024-05325-1.
3. Wu PL; Tseng YC; Chen LC; Tseng SM; Pai HC. Development and validation of the Clinical Nursing Teacher Self-Efficacy Scale. Asian Nursing Research. 2022;16:125–33.
4. Feng X, Wu W, Bi Q. Reform of teaching and practice of the integrated teaching method BOPPPS-PBL in the course “clinical haematological test technique.” BMC Med Educ. 2024;24:773. https://doi.org/10.1186/s12909-024-05765-9.
5. Dawson D; Borin P; Meadows K; Britnell J; Olsen K; McIntyre G. The impact of the Instructional Skills Workshop on faculty approaches to teaching. Toronto: Higher Education Quality Council of Ontario; 2014.

6. Steinert Y, Mann K, Anderson B, Barnett BM, Centeno A, Naismith L, et al. A systematic review of faculty development initiatives designed to enhance teaching effectiveness: A 10-year update: BEME Guide No. 40. Medical Teacher. 2016;38:769–86. https://doi.org/10.1080/0142159X.2016.1181851.
7. Steinert Y, Mann K, Centeno A, Dolmans D, Spencer J, Gelula M, et al. A systematic review of faculty development initiatives designed to improve teaching effectiveness in medical education: BEME Guide No. 8. Medical Teacher. 2006;28:497–526. https://doi.org/10.1080/01421590600902976.
8. Breckwoldt J, Svensson J, Lingemann C, Gruber H. Does clinical teacher training always improve teaching effectiveness as opposed to no teacher training? A randomized controlled study. BMC Med Educ. 2014;14:6. https://doi.org/10.1186/1472-6920-14-6.
9. Burgess A, Van Diggele C, Roberts C, Mellis C. Feedback in the clinical setting. BMC Med Educ. 2020;20:460. https://doi.org/10.1186/s12909-020-02280-5.
10. Hattie J, Timperley H. The Power of Feedback. Review of Educational Research. 2007;77:81–112. https://doi.org/10.3102/003465430298487.
11. Boud D, Molloy E. Rethinking models of feedback for learning: the challenge of design. Assessment & Evaluation in Higher Education. 2013;38:698–712. https://doi.org/10.1080/02602938.2012.691462.
12. Nicol D, Thomson A, Breslin C. Rethinking feedback practices in higher education: a peer review perspective. Assessment & Evaluation in Higher Education. 2014;39:102–22. https://doi.org/10.1080/02602938.2013.795518.
13. Hardavella G, Aamli-Gaagnat A, Saad N, Rousalova I, Sreter KB. How to give and receive feedback effectively. Breathe. 2017;13:327–33. https://doi.org/10.1183/20734735.009917.
14. Daou D, Sabra R, Zgheib NK. Factors That Determine the Perceived Effectiveness of Peer Feedback in Collaborative Learning: a Mixed Methods Design. MedSciEduc. 2020;30:1145–56. https://doi.org/10.1007/s40670-020-00980-7.
15. Tschannen-Moran M; Woolfolk Hoy A. Teacher efficacy: capturing an elusive construct. Teaching and Teacher Education. 2001;17:783–805.
16. Gibbs G; Simpson C. Conditions under which assessment supports students' learning. Learning and Teaching in Higher Education. 2004;1:3–31.
17. Van Blankenstein FM, O'Sullivan JF, Saab N, Steendijk P. The effect of peer modelling and discussing modelled feedback principles on medical students' feedback skills: a quasi-experimental study. BMC Med Educ. 2021;21:332. https://doi.org/10.1186/s12909-021-02755-z.
18. Wang H-T, Chien C-F, Chen C-Y. Enhancing skill practice and feedback in a higher education teacher training course using the online flipped teaching in Taiwan. Teaching and Teacher Education. 2025;162:105073. https://doi.org/10.1016/j.tate.2025.105073.
19. Perera HN, Calkins C, Part R. Teacher self-efficacy profiles: Determinants, outcomes, and generalizability across teaching level. Contemporary Educational Psychology. 2019;58:186–203. https://doi.org/10.1016/j.cedpsych.2019.02.006.

20. Barnett AG, van der Pols JC, Dobson AJ. Regression to the mean: what it is and how to deal with it. International Journal of Epidemiology. 2005;34:215–20. https://doi.org/10.1093/ije/dyh299.
21. Miller CJ, Smith SN, Pugatch M. Experimental and quasi-experimental designs in implementation research. Psychiatry Research. 2020;283:112452. https://doi.org/10.1016/j.psychres.2019.06.027.
22. Vickers AJ, Altman DG. Analysing controlled trials with baseline and follow up measurements. BMJ. 2001;323:1123–4. https://doi.org/10.1136/bmj.323.7321.1123.

Figures

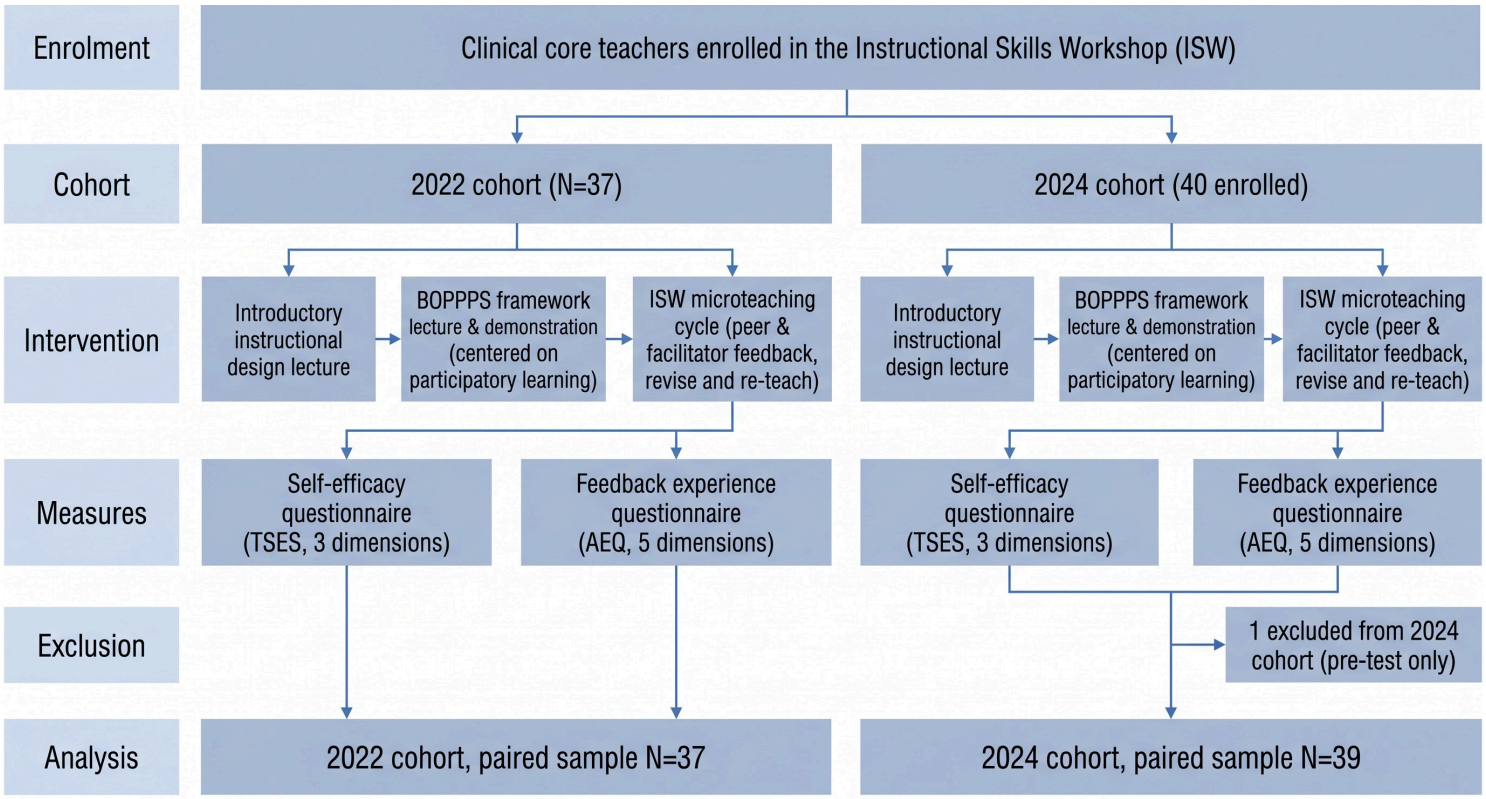


Fig. 1 Flow diagram of the study

Figure 1

Study design and intervention structure

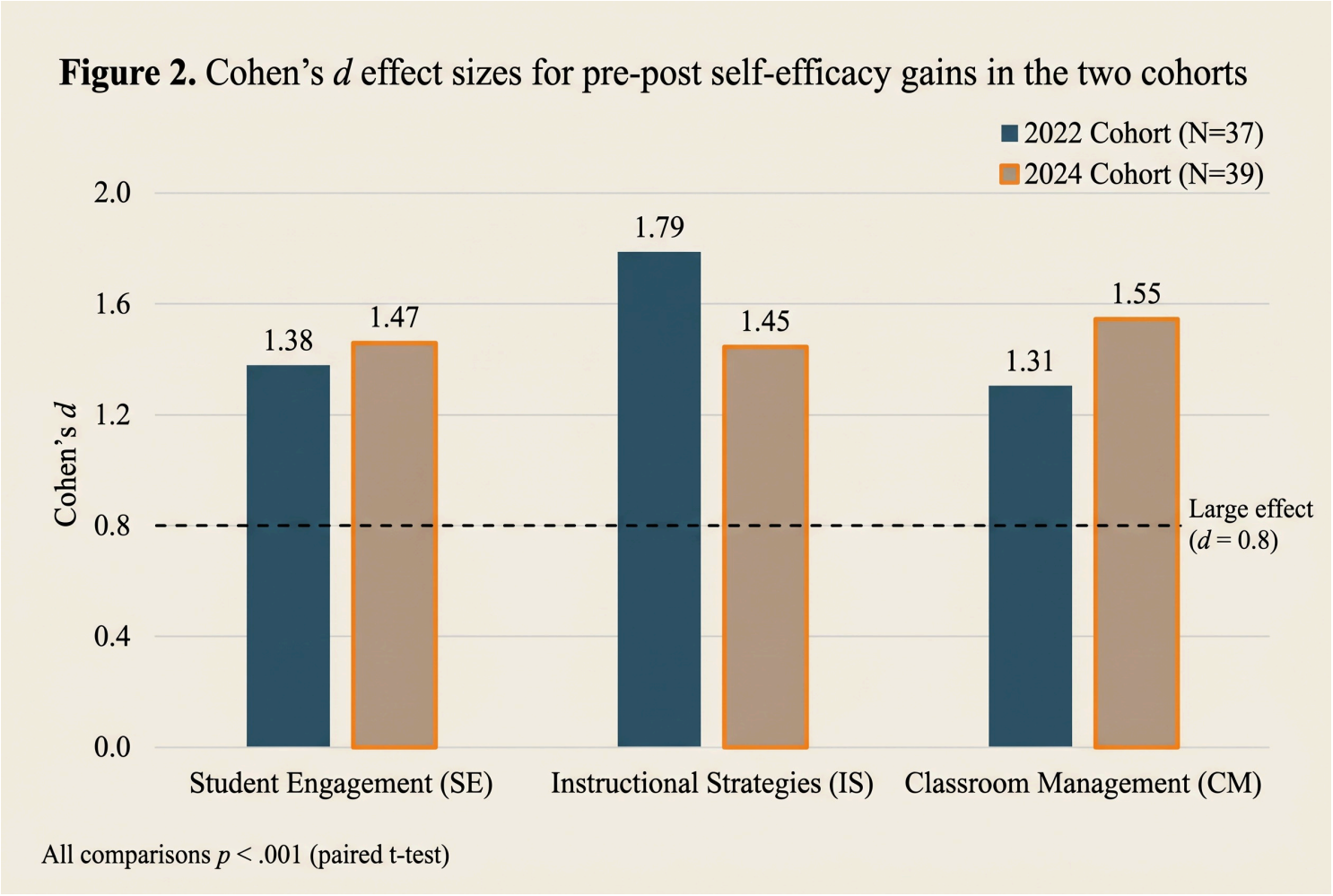


Figure 2

Cohen's d effect sizes for pre–post self-efficacy gains in the two cohorts

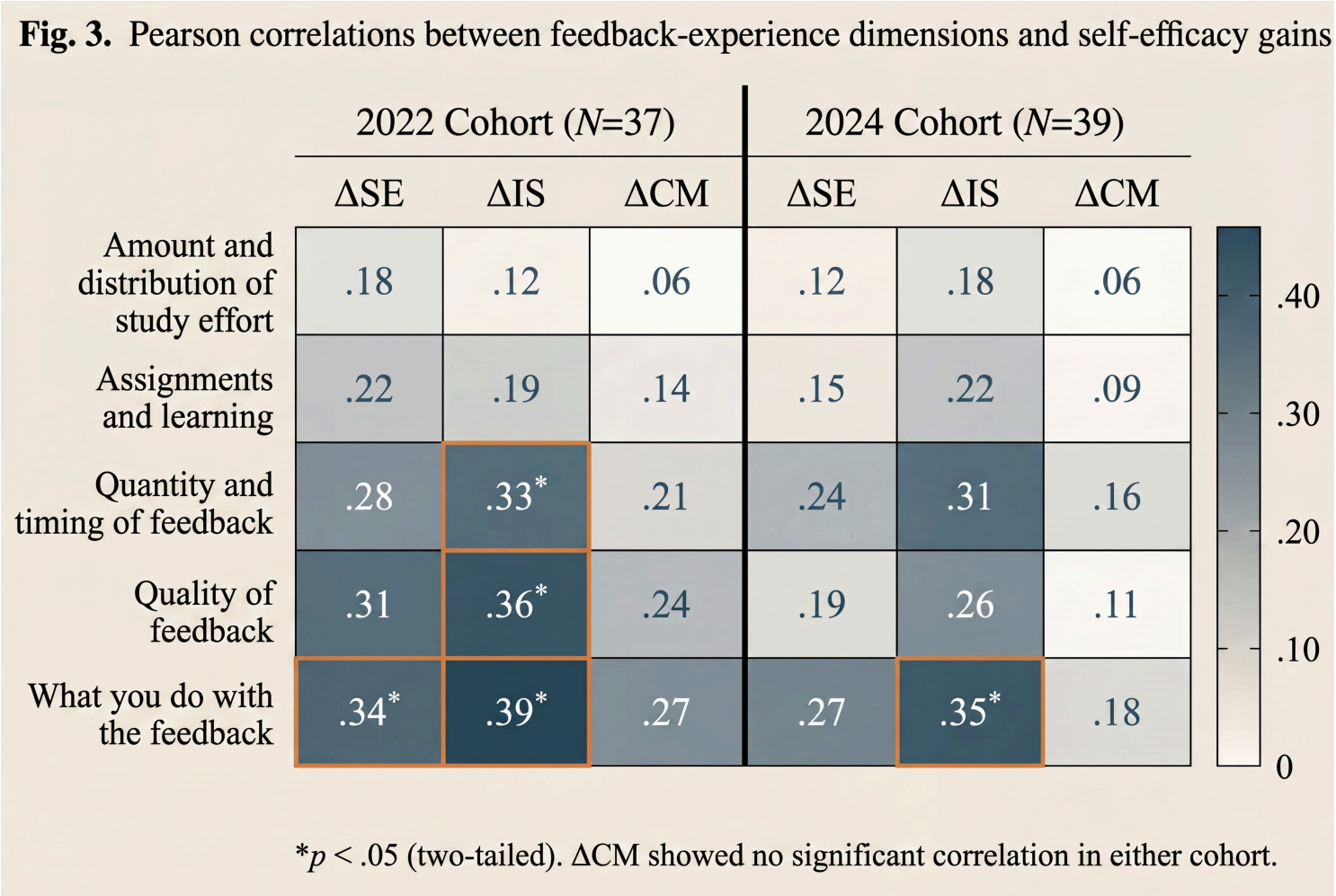


Figure 3

Feedback-experience correlation heatmap

Supplementary Files

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

- AdditionalfileQuestionnaires.docx