

Beyond Peer and Supervisor: How a Four-Source Feedback Configuration Cultivates Reflective Independence in Pre-Service Music Teachers

Chen Liu*

Wuhan Conservatory of Music, Jiefang Road, Wuhan, China

**Corresponding author: Chen Liu.*

Abstract

Third-year pre-service music teachers in Chinese conservatoires approach the school-placement year with limited confidence in judging and improving their own lesson designs once the daily presence of a supervising professor recedes. The present study investigates the mechanisms through which a four-source feedback configuration in microteaching, combining self, peer, graduate-mentor, and faculty-advisor assessment, cultivates the reflective capacities pre-service music teachers will need to draw on independently after they leave the university. Grounded in Schön's theory of reflective practice and Korthagen and Kessels' ALACT model, the study adopts an exploratory mixed-methods case study design with eight third-year music education students at a Chinese conservatoire. Data comprise paired pre- and post-intervention questionnaires ($n = 8$), individual retrospective interviews with all eight participants, and structured reflective journals completed after each microteaching cycle (with four selected for in-depth analysis). The findings suggest that the four-source configuration supports reflective independence through two complementary mechanisms: cultivating evaluative standards when participants assess peers, and surfacing self-perceived blind spots when they encounter discrepant evaluations of their own designs. Participants converged on a tiered model of source differentiation in which the graduate-mentor tier, largely absent from existing multi-source assessment research, emerges as a distinctive scaffold between peer and faculty advisor.

Keywords

reflective practice, pre-service music teachers, multi-source feedback, microteaching, graduate mentor

1. Introduction

By the third year of undergraduate study, music education students in Chinese conservatoires stand at the threshold of school placement. Within months they will design real lessons for real children and confront questions that until now have been answered for them by faculty advisors: Was this lesson well designed? Did it achieve what I intended? What should I do differently next time? In school teaching, no university supervisor stands beside the novice teacher each lesson. The novice must address these questions independently, and doing so is fundamentally reflective work: looking back on what has happened, identifying what was significant, and constructing alternatives for what comes next [1]. Whether the pre-service teacher has developed the capacity to do this work independently is therefore the most consequential question pre-service preparation can address, and it is the question on which third-year students in our setting express the deepest

uncertainty. Microteaching, the principal pre-placement pedagogical site in Chinese conservatoires, often delivers post-session feedback as one-way evaluation, leaving the cycle from action to reflection incomplete. The structural problem is not the absence of evaluation in microteaching but the absence of conditions under which evaluation becomes reflection [2,3]. The microteaching programme examined in the present study addresses this problem through a four-source feedback configuration that incorporates self, peer, and faculty-advisor assessment and, distinctively, graduate-mentor assessment by advanced students who have themselves taught in schools. The remainder of this article reviews the literature on reflective practice and multi-source feedback that situates this configuration as a candidate scaffold for reflective independence, and then reports an empirical study of how the configuration prepares pre-service music teachers for the reflective work they will need to undertake independently after school placement. Drawing on paired pre- and post-intervention data, retrospective interviews, comparison-reflection journals, and a post-session recording from third-year students at a Chinese conservatoire, the study is guided by two research questions: RQ1. How do the four sources of feedback in microteaching, self, peer, graduate mentor, and faculty advisor, each elicit distinct forms of reflection among pre-service music teachers? RQ2. Through which mechanisms, operating on the evaluator and on the recipient, does the four-source configuration support the reflective independence that pre-service music teachers will need in the school workplace?

2. Literature Review

2.1 Pre-Service Music Teacher Preparation

Music teacher education has long grappled with the problem of preparing graduates not only to know music but to teach it under the contingent conditions of real classrooms. Ballantyne [4] found that early-career music teachers identified limited preparation in pedagogical content and professional knowledge as the principal source of classroom difficulty; programmes treating general education and music knowledge as separate domains left graduates without the integrative judgement teaching demands. Shaw [5], drawing on conversations with conservatoire alumni about their entry into the music education workforce, reached a parallel conclusion. The most consequential gap in pre-service preparation lay not in technical or musical content but in the capacities that allow graduates to keep learning from their own classroom experience after the university has receded into the background. Both studies converge on a single point. What pre-service music teachers most need, both studies suggest, is not more content but the capacity for reflective independence, the ability to judge and improve one's own lesson designs without ongoing external direction [4,5]. The 2022 Chinese Curriculum Standards for Compulsory Education in the Arts [6] intensify this requirement, presupposing a teacher capable of reflecting independently on her own work in relation to the four core competencies.

2.2 Reflective Practice as the Foundational Framework

Schön's [1] *The Reflective Practitioner* remains the foundational text for understanding how professionals develop expertise through reflection on their own action. Schön argued that professional knowledge is generated through a continuing conversation between practitioner and situation, not through the application of general rules to particular cases. Reflection-in-action occurs when a practitioner, in the midst of doing, becomes aware that the situation is not unfolding as expected and adjusts on the spot. Reflection-on-action occurs when, after the event, the practitioner steps back to examine what happened and what might have been done differently. For teachers, Schön [7] further argued that reflective capacities develop most fully under the guidance of a coach who creates conditions in which the novice can both act and look back on action without the inhibiting weight of formal evaluation. Translating Schön's framework into a workable pedagogy proved difficult. Korthagen and Kessels [8] addressed this by proposing the ALACT model, which articulates the reflective cycle into five phases, Action, Looking back, Awareness of essential aspects, Creating alternatives, and Trial, and identifies the third phase, Awareness, as the critical point at which reflection moves from description to insight. Subsequent research has confirmed that without sustained attention to the Awareness phase, pre-service teachers tend to settle for surface-level reflection that does not reorganise their understanding [9]. Within music teacher education, Gardiner [10] made an additional contribution that is essential for the present argument. In a study of student-teacher reflections, she showed that the content of these reflections is structured by the evaluative ideologies the student teachers have already internalised. What

they notice, and what they overlook, depends on the standards they bring to noticing. Reflection thus requires alternative ways of seeing to make one's own ways of seeing visible.

2.3 Multi-Source Feedback in Microteaching: Two Complementary Mechanisms

Multi-source feedback designs in pre-service teacher education typically combine self, peer, and supervisor assessment, with the educational rationale resting on the differing vantage points each source offers [2]. Each source has well-documented limitations when relied on in isolation. Self-assessment is constrained by novices' limited evaluative experience: because pre-service teachers have not yet fully internalised the criteria of good lesson design, their judgements of their own work tend to be inaccurate, and self-assessment accuracy improves only as task-relevant expertise develops [11]. Peer assessment broadens the range of perspectives, but it can be inhibited by interpersonal considerations, such as reluctance to criticise classmates, and by peers' shared inexperience, both of which constrain the reliability and depth of the feedback peers can provide [12]. Supervisor assessment, though diagnostically the most authoritative, risks producing dependency rather than reflective autonomy when it is delivered as a one-way verdict that students receive passively rather than actively interpret and act upon [3,13]. The argument for combining sources rests on two distinct mechanisms that prior literature has rarely articulated together. The first is the cultivation of evaluative standards in the assessor. When pre-service teachers are asked to formulate feedback on peers' lessons, they must mobilise and apply their own criteria of good lesson design, and the repeated exercise of these criteria strengthens them as internal cognitive resources for the assessor's own future work. From the perspective of pre-service preparation, this mechanism addresses precisely the difficulty motivating the present study: how to develop in pre-service teachers the internal standards by which they will eventually judge their own work independently in the school workplace. The second mechanism operates on the recipient. When a pre-service teacher's own lesson is evaluated from multiple distinct vantage points in close succession, the discrepancies among those vantage points create the surprises that, for Schön [1], provoke reflection. Gardiner's [10] observation about the structuring role of internalised standards becomes especially significant here. Because internalised standards determine what one notices, encountering alternative readings of the same lesson makes those standards visible. Multi-source feedback thus functions less as evaluation than as structured exposure to alternative ways of seeing. Both mechanisms can be read as a process of internalisation in Vygotsky's [14] sense, in which a capacity first appears in social interaction and is then reconstructed on the psychological plane. The evaluative standards a pre-service teacher exercises aloud when assessing peers, and the alternative readings she encounters when her own lesson is judged, are in the first instance social and external. Reflective independence is the form they take once internalised as a teacher's own inner resources. Empirical examination of this dual mechanism in pre-service music teacher education remains limited. Yıldız and Yıldız [2] studied the use of self, peer, and teacher feedback in microteaching but did not analytically separate the mechanisms operating on assessor and recipient, and Tilahun [3] documented inadequacies of feedback delivery but did not examine the configuration of feedback sources.

2.4 The Graduate Mentor: A Tier Largely Absent from Existing Research

A further feature of the microteaching configuration examined in the present study has been almost entirely overlooked in prior multi-source feedback research, the routine presence of a graduate-mentor tier of feedback contributed by advanced students who have themselves taught in primary or secondary schools. From the perspective of reflective practice theory, this tier occupies a position of particular interest. Schön [7] argued that the development of reflective practice in pre-service preparation depends on the presence of a coach whose role is neither that of an authoritative judge nor that of a fellow novice, but that of a more experienced practitioner who can articulate practice in ways the novice can follow. Shaw [5] demonstrated that conservatoire alumni functioning as near-peer mentors contribute distinctive value to graduating students' professional preparation precisely because they have walked recently along the same path. The graduate-mentor role examined in the present study is structurally similar to Shaw's alumni in providing classroom-experienced, non-authoritative guidance, but differs in being routinely embedded in each cycle of microteaching evaluation rather than entering as occasional visitors. A targeted search of the Web of Science, Scopus, and ERIC databases (combining the keywords graduate mentor, near-peer mentor, multi-source feedback, peer feedback, microteaching, reflective practice, and pre-service teacher, with results screened for studies that treated feedback source as an analytic variable) returned no study that analytically distinguished the graduate-mentor tier from peer and faculty tiers in microteaching, examined the kind of reflective activity

it specifically elicits, or considered how it contributes to the transition from pre-service preparation to in-service reflective independence. Existing work has examined peer, self, and supervisor feedback separately [2,3] and near-peer mentoring in other contexts [5], but not this particular tier as a distinct source. Taken together, the literature points to a gap with both theoretical and practical urgency: the assessor-side and recipient-side mechanisms of multi-source feedback have rarely been examined simultaneously in pre-service music education, and the potential role of a graduate-mentor tier, implied by Schön's [7] account of coaching and Gardiner's [10] analysis of internalised evaluative standards, remains empirically untested. The present study addresses this gap by investigating how a four-source configuration incorporating the graduate mentor operates in microteaching.

3. Methodology

This study adopts a qualitative-dominant mixed-methods design [15]. Quantitative data provide a group-level indication of change across the intervention; qualitative data carry the principal analytic weight, with quantitative and qualitative findings integrated to support convergent inferences. The sections below describe the setting, participants, data sources, and analytical procedures.

3.1 Settings and Participants

The study was conducted within a required foundational teacher-preparation course for third-year undergraduates at a Chinese conservatoire, scheduled at four credit hours per week across one semester. The track is a professional curriculum whose graduates enter school music-teaching positions after placement. By this point, students have completed two years of foundational coursework in the professional skills required of music teachers. Each microteaching cycle involves a 30-minute lesson based on a topic from the People's Music Publishing House textbook series, followed by a structured post-session evaluation discussion. The discussion follows a fixed four-source sequence: self-assessment, unprompted peer feedback from the other seven students, graduate-mentor feedback (from two graduate students with classroom teaching experience together with the first author), and the supervising faculty advisor's concluding evaluation. Following each post-cycle discussion, all participants complete a structured Reflective Comparison Journal integrating scores and feedback from the four feedback tiers. The eight participants comprise the entire cohort of third-year students in the track during the year of data collection (six women, two men, reflecting the broader gender distribution in Chinese music education programmes). The sample is best characterised as a complete cohort case. Pseudonyms are used throughout. All eight completed the pre-intervention questionnaire, the full microteaching programme, the post-intervention questionnaire, and an individual retrospective interview.

3.2 Data Sources

The Reflective Practice in Lesson Design Scale (RPLDS), developed for this study from Schön [1] and Korthagen and Kessels [8], comprises thirteen 5-point Likert items in five conceptual clusters: reflection-in-action (Items 1–2), retrospective reflection-on-action (Items 3–5), prospective reflection in independent lesson design (Items 6–9), differential awareness of feedback sources (Item 10), and reflective independence (Items 11–13). Two additional post-only items (14–15) capture retrospective appraisal of the intervention. The pre-intervention version was administered at the start of the semester and the post-intervention version at its conclusion. All eight participants completed both administrations, yielding eight matched pairs. Content validity was supported through expert review by the supervising professor, a senior music teacher educator with over two decades of experience, who examined each item's alignment with Schön's [1] and Korthagen and Kessels' [8] constructs. The five-cluster structure was derived deductively from theory rather than empirically from the data. Given the small sample ($n = 8$), internal consistency coefficients are not reported, as alpha estimates from such small samples are unstable and potentially misleading [16]. The RPLDS functions in this study not as a validated psychometric instrument but as a structured descriptive tool whose primary role is to provide a systematic, item-level map of change to guide qualitative inquiry. Each participant completed an individual semi-structured retrospective interview of approximately 30 minutes following five open-ended core questions: (Q1) criteria for independently judging another teacher's lesson plan, (Q2) a recent instance of self-monitoring during independent design, (Q3) the structure of self-diagnosis when facing difficulty, (Q4) specific feedback that resurfaced in independent design work, and (Q5) the most consequential change in thinking after the intervention. Interviews were conducted in Mandarin Chinese. Every participant completed

a Reflective Comparison Journal after each of their own microteaching cycles. The journal comprises four sections: (A) descriptive record and pre-feedback confidence, (B) a four-source score matrix on five lesson-design dimensions, (C) a structured differential analysis prompting participants to identify maximum-discrepancy dimensions, attribute reasons for self-external rating gaps, characterise the distinctive contributions of peer / graduate-mentor / faculty-advisor sources, and specify concrete revisions, and (D) a reflexive self-monitoring section. Section C constitutes the principal mechanism converting four-source feedback into reflective awareness. From the full set of journals collected, four were selected as typical cases for in-depth analysis in this paper, corresponding to the teaching cycles of P1, P3, P5, and P6, on the basis of completeness and the diversity of lesson topics they represented.

3.3 Analytical Framework and Procedure

Qualitative data were analysed using reflexive thematic analysis [17,18], following a hybrid procedure that combined deductive coding, guided by Schön's [1,7] reflective practice theory as the primary lens. Two distinctions guided coding, reflection-in-action versus reflection-on-action, and descriptive reflection versus reflection achieving what Korthagen and Kessels [8] term Awareness of essential aspects. The study anchors analysis in Schön and Korthagen, reflecting the concern with reflective independence as a domain-specific professional capacity. Analysis proceeded in six iterative phases following Braun and Clarke [17,18]. Three thematic levels resulted, feedback-source differentiation, reflective mechanisms (assessor-side and recipient-side), and reflective independence as internalised dialogue. Quantitative data were analysed descriptively at total, cluster, and item levels. Given the small matched sample ($n = 8$), statistical significance is interpreted with caution. The Wilcoxon signed-rank test serves as a descriptive benchmark for the qualitative findings rather than as a confirmatory test. Trustworthiness was supported through triangulation across four data sources, audit-trail documentation, thick description, and ongoing reflexive memos [19,20].

3.4 Researcher Positionality and Ethics

The first author occupied a dual role as researcher and as one of the graduate-mentor evaluators. Four procedural separations mitigated the resulting risks. First, faculty-advisor feedback was contributed independently by the supervising professor, who was not involved in coding or analysis. Second, peer feedback was elicited without prompting. Third, interviews used open-ended, non-leading questions. Fourth, emergent themes were reviewed with the supervising professor at multiple checkpoints. Ethical approval was obtained through the institutional research ethics process. All eight participants provided written informed consent for audio-recording, journal content, and anonymised quotation.

4. Findings

The findings are presented in four sections. The first establishes the group-level pattern of change in reflective practice capacity. The second develops two qualitative themes characterising the forms of reflection elicited by the four-source configuration (RQ1). The third articulates the two complementary mechanisms through which the configuration supported reflective independence (RQ2). The fourth presents the cross-source differentiation pattern that emerged across the four selected reflective journals.

4.1 Outcome Pattern (Quantitative): Change in Reflection Capacity

All eight participants exhibited positive change in overall reflective practice capacity, with no participant showing a decline in total score (Table 1). Group-level scores rose from a pre-intervention mean of 46.1 ($SD = 6.36$) to a post-intervention mean of 57.6 ($SD = 4.27$), corresponding to +11.5 points on a 65-point scale. A Wilcoxon signed-rank test yielded $T = 0$ with all eight signed differences positive ($W+ = 36$, $W- = 0$). The effect size $r = .63$ represents a large effect [21]. This result is interpreted as a descriptive signal rather than a confirmatory inference, given the small sample size and the absence of a comparison group.

Cluster-level analysis revealed that change was concentrated on the clusters most directly corresponding to the study's theoretical focus, reflection-in-action ($\Delta = +1.25$, $M = 3.06 \rightarrow 4.31$) and reflective independence ($\Delta = +1.38$, $M = .88 \rightarrow 4.25$). The remaining clusters showed moderate change (retrospective reflection-on-action $\Delta = +0.67$, prospective reflection in independent lesson design $\Delta = +0.66$) or approached ceiling (differential source awareness $\Delta = +0.25$, $M = 4.62$ at pre-test). At the item level, two items within these clusters

showed the largest individual gains and anchor the qualitative findings that follow. Item 2, indexing within-process reflection-in-action, shifted from $M = 2.88$ to 4.50 ($\Delta = +1.62$), and Item 12 (“I have a fairly accurate judgement of the quality of my own lesson design”) showed the single largest pre-post change of any item ($M = 2.50 \rightarrow 4.25$, $\Delta = +1.75$). Two post-only items provided convergent corroboration. Item 14 (multi-source feedback exposed blind spots invisible to solo reflection) yielded $M = 5.00$, with all eight participants selecting “strongly agree.” Item 15 (confidence in independent reflection during the future school placement) yielded $M = 4.75$. The qualitative analysis traces the interactional processes through which these gains were produced.

Table 1: Individual and Group Change in RPLDS Mean Score (N = 8).

Participant	Pre M	Post M	Δ	Direction
P1	3.54	4.31	+0.77	↑
P2	4.15	4.69	+0.54	↑
P3	3.38	4.85	+1.46	↑
P4	3.54	4.77	+1.23	↑
P5	3.08	3.85	+0.77	↑
P6	4.15	4.31	+0.15	↑
P7	2.77	4.31	+1.54	↑
P8	3.77	4.38	+0.62	↑
Group (SD)	3.55 (0.48)	4.43 (0.33)	+0.88	All ↑

Note. Wilcoxon: $W+ = 36$, $W- = 0$, $T = 0$. $r = .63$.

4.2 Emergent Reflective Patterns (Qualitative Themes)

Reflexive thematic analysis produced two principal themes: criteria internalisation (Theme 1) and the delayed activation of multi-source feedback as an internal dialogue (Theme 2). The first characterises the evaluative standards participants now hold and the routes through which those standards became internal. The second characterises how earlier four-source feedback resurfaces in their later independent work. Together they describe a coherent process by which external four-source feedback became internalised as a self-sustaining reflective practice.

4.2.1 Theme 1: Criteria internalisation

Across all eight interviews, participants articulated an organised set of criteria they reported applying when independently judging a music lesson plan, marked by sequenced articulation (“first I would look at..., then I would examine...”). P4 offered the most theoretically integrated formulation:

Taking up someone else’s music lesson plan, the very first thing I look at is whether the framework is structurally complete, whether it clearly contains the learning objectives, the key and difficult points, the instructional process with episodes and timing, the board design, and the post-lesson reflection. This is my first threshold of judgement, because the integrity of the framework reflects whether the designer possesses an integrative planning capacity. If the basic blocks are missing, or if the blocks are disconnected from one another, no amount of fine detail in any one section can find its footing.

(P4, Interview Q1)

P4 then proceeded through five further evaluative dimensions: curriculum-standard alignment, student-experience coherence, methodological appropriateness, assessment integration, and adaptive planning. The breadth and organisation of this repertoire, echoed across all eight interviews, indicates that participants now hold an internalised set of evaluative standards they can activate without external prompting.

The journals document two routes through which these standards became internal. The first route is the recipient’s confrontation with discrepant evaluations of her own lesson. Across the four selected reflective journals, the same pattern appears: a substantial gap between self-assessment scores and external evaluator scores, followed by an explicit reckoning with the standards producing that discrepancy. P1, reflecting on her Xingfu Ge lesson, recorded a self-assessment mean of 2.6 against a faculty-advisor mean of 1.0, and isolated the largest single discrepancy on D2 (instructional sequence and structure), where her self-rating of 3 met a faculty-advisor rating of 1:

I gave myself a 3 on D2. I thought the lesson was quite rich, with singing, rhythmic movement, lyric creation, and I talked about concepts like mode, chenci, and folk song definition. Every section got students involved,

and it felt exploratory. But the supervisor gave me a 1. He said, “You threw a pile of elements on the table. You talked about so many things, but how exactly does the character of Jingchu folk song manifest in this Xingfu Ge? Take the three-note scale, that’s the most typical concept for capturing Jingchu folk song. You should have dug into that and taught it thoroughly, instead of laying every piece of knowledge in front of the students.” I’m only now realising our standards were different. I used to think a music lesson meant telling students every bit of knowledge a song contains, as long as I spread the content out, it was a good lesson. I mistook listing concepts for depth. The supervisor’s standard was different: pick one core point, go deep, and really ground it in the music itself. Two completely different levels: I was at “how much was thrown out,” he was at “how much was actually picked up.”

(P1, Reflective Journal, Section C)

The discrepancy is not a numerical disagreement but a marker of a different evaluative standard. The act of comparison, which the journal’s structured prompts oblige the participant to undertake, converts the numerical discrepancy into a surfaced realisation about a previously invisible criterion. The same pattern recurs across the other three selected journals (P3, P5, P6), each with its own discrepancy dimensions and its own re-articulation of standards.

The second route is the assessor’s repeated exercise of evaluating peers. When participants were asked what they would now look for in evaluating another teacher’s plan, what emerged were criteria actively rehearsed in their own role as peer evaluator. P6 wrote:

If I were now to evaluate another classmate’s lesson design, I would attend to two standards first. First, whether the whole lesson has a single clear main thread, whether every episode is linked to the next and the cues from the introduction are actually picked up and resolved later, not ‘many episodes that stand independently.’ Second, whether the cultural understanding and the musical knowledge actually land on the musical body of the work, whether the lesson makes clear the ‘why’ rather than treating cultural background as a label stuck on the outside of the music. These two standards come directly from the supervisor’s criticism of ‘fractured episodes, surface cultural exposition’ and from the graduate mentor’s criticism of ‘rhythm teaching that flows on the surface.’ I hope that in the next round my self-assessment can move closer to the external evaluators’ judgement.

(P6, Reflective Journal, Section D)

P6 is articulating a standard she has appropriated, traced to its specific origin in the four-source feedback experience, and made her own. The two routes are complementary: confronting discrepancy reveals what one had not seen, while assessing others establishes the standards by which one can begin to see for oneself. This capacity underpins independent judgement of participants’ own future lesson designs, and the mechanisms that account for it are taken up in the Discussion.

4.2.2 Theme 2: Delayed activation

The second theme, perhaps the most distinctive finding of the study, concerns the way specific utterances from earlier four-source feedback episodes resurfaced spontaneously in participants’ independent lesson-design work weeks or months later. P8 captured the phenomenon with the greatest analytic clarity:

Independent lesson planning is no longer a closed-door, solitary process. It has become a kind of cross-temporal dialogue, between my supervisor and me, between my classmates and me, even between my earlier self and me. When I face a design problem, voices from those four-source discussions rise up in my mind and offer different angles on what I am about to do.

(P8, Interview Q4)

P4 supplied a particularly fine-grained instance. Preparing a lesson on Li Huanzhi’s Spring Festival Overture and reaching a sticking point in her opening design, she reported:

At precisely the moment I was hesitating, Professor Sun’s sentence suddenly emerged from memory: ‘In music teaching you must enter from the musical elements themselves, and let students understand from those elements why a work has this particular musical feeling.’ I had heard him say this at the time of the post-session discussion and had thought it made sense, but it was only at this moment in my own planning, when I had reached a real design impasse, that I truly understood its operational meaning... That sense of

‘suddenly remembering’ is, I think, the moment when an internalised professional judgement standard is reactivated at exactly the right time.

(P4, Interview Q4)

P3 described the same phenomenon in language that maps almost directly onto Schön’s [1] original formulation:

What the four-source evaluation gave me is no longer a particular technique, those gradually fade, but a habit of second-order thinking. Above my own teaching design I now suspend an observing self, who keeps asking: What is the rationale of this episode? What is its disciplinary basis? What does it actually mean for the growth of students’ aesthetic experience? This inner dialogue is no longer a deliberate effort at reflection, it has become as natural in my planning as breathing.

(P3, Interview Q5)

The four-source feedback configuration did not produce reflective independence by training participants to follow a procedure. It produced it by leaving in their minds a chorus of internalised voices, peer, graduate mentor, faculty advisor, and their own evolving self-evaluating voice, that could be recruited later when participants confronted novel design problems.

4.3 Cross-Source Differentiation Pattern: The Tier Model

The most striking finding in the qualitative data is the convergent characterisation that emerged across the four selected reflective journals’ Section C, in which each participant was asked to specify what the peer, graduate mentor, and faculty advisor each attended to. Despite no shared template language, the four journals produced strikingly congruent descriptions of a tiered model (Table 2).

Table 2: Convergent Three-Tier Differentiation of Feedback Sources (from Four Selected Reflective Journals).

Tier	Focus of Reflection	Representative Wording (Synthesised Across P1, P3, P5, P6)
Peer	Immediate experiential and operational accuracy	“What did this lesson sound and look like to me sitting in the back?”. “Which sentence was said wrong, which step felt awkward?” Surface impressions that do not explain why the design failed at its root.
Graduate Mentor	Classroom feasibility and instructional logic	“If you put this lesson in a real classroom, how would students react and where would they lose their footing?” Focuses on landability, logical breakdowns, sequencing errors, and absence of cognitive closure in actual school teaching.
Faculty Advisor	Disciplinary reasoning and professional standards	“What is the design thinking, the disciplinary grounding, and the rigour of scholarship behind this lesson?” Interrogates foundational concepts, professional documentation standards, and scholarly accuracy.

Note. Characterisation emerged inductively from four reflective journals selected for in-depth analysis (P1, P3, P5, P6), produced without coordination among participants.

P5 formulated the model with the most analytic explicitness:

Looking back over the three sources’ feedback, I find that they attended to angles that are visibly different. Peers attended to ‘how did this lesson sound and look,’ the immediate experience layer. The graduate mentors attended to ‘if this lesson were taken into a real classroom, how would it land,’ the practical feasibility layer. The supervisor attended to ‘what is the design thinking behind this lesson,’ the disciplinary reasoning layer. The three angles helped me in different ways. Peers told me where I said something wrong or did something wrong. Mentors told me what would go wrong if this were placed in a real classroom. The supervisor told me what was missing in my design thinking and disciplinary understanding. If I had only one of them, my understanding would have been incomplete in a definite way.

(P5, Reflective Journal, Section C)

That four participants writing about four different lessons without coordination converged on essentially the same model is a strong qualitative signal that the differentiation is a feature of the four-source configuration itself rather than an artefact of any particular lesson. Within this characterisation, the graduate-mentor tier appears as a distinctively positioned source, sitting between peers, who share participants’ lack of classroom experience, and the faculty advisor, whose authoritative voice may produce dependency rather than reflective autonomy. Its distinctive contribution is feedback that is classroom-grounded yet non-authoritative, addressing

“how will this land in a real classroom?” from the position of someone who has herself recently faced that question.

5. Discussion

Read together, the three sets of findings point to a single theoretical claim: for these pre-service teachers, reflective independence is not a solitary trait exercised in isolation but an internalised, multi-voiced dialogue. The quantitative gains, the spontaneous re-emergence of earlier feedback during independent planning, and the convergent tier model all describe the same underlying phenomenon, the relocation of an external, social evaluative conversation onto the psychological plane of the individual planner. On this reading, the contribution of multi-source feedback is not that it delivers more comprehensive evaluative information at the moment of microteaching, but that it installs a set of distinguishable internal voices, traceable to Schön's [1,7] coaching relationship and to Vygotsky's [14] account of internalisation, that the teacher can later recruit when no supervisor is present. The clearest empirical signature of this relocation is the delayed-activation pattern: participants spontaneously recalled specific utterances from earlier four-source episodes weeks or months later. P8's formulation of independent planning as “a kind of cross-temporal dialogue” identifies independent reflection not as a solitary act but as a sustained inner conversation among multiple internalised standpoints, each tracing back to a specific moment in the feedback history; P3's description of a “second-order observing self” is recognisable as the internalised form of what was, in the post-session discussion, the faculty advisor's externally posed question. For multi-source feedback to function this way, its different voices must remain distinguishable to the participant, so that they can later be internalised as separate inner voices rather than fused into undifferentiated evaluative noise, a requirement the tiered model that emerged in the journals meets. This phenomenon is produced by two mechanisms that the findings document on opposite sides of the feedback exchange. On the recipient side, evaluating a single lesson from several vantage points in close succession generates discrepancies, and a discrepancy is precisely the kind of surprise that, for Schön [1], provokes reflection. Because internalised standards govern what a novice is able to notice [10], a divergent reading of her own lesson does more than correct a score. It renders her own prior standard visible to her, as P1's reinterpretation of “structure” illustrates. The discrepancy must be worked on, not merely recorded, which is why the structured comparison prompts rather than the raw scores carry the reflective load. On the assessor side, being required to formulate feedback on a peer's lesson obliges the novice to mobilise and articulate her own criteria of good design, and the repeated exercise of those criteria consolidates them as durable internal resources, as P6's appropriation of the supervisor's and mentor's standards shows. The two mechanisms are complementary: judging others builds the standards, and being judged from multiple angles reveals where those standards are still missing. Both are moments in which an evaluative capacity first rehearsed in social interaction is reconstructed as an individual one. The graduate-mentor tier merits separate attention because it functions as a key structural variable rather than as an additional voice. Peers and the participant share the same inexperience, so peer feedback cannot reach the question of how a lesson will behave in a real classroom, whereas the faculty advisor can reach that question but speaks with an authority that tends to convert feedback into a verdict and to foster dependency. The graduate mentor occupies the developmental position between them: classroom-experienced enough to speak credibly to feasibility, yet close enough to the participants' situation to remain non-authoritative. In Vygotskian [14] terms, the mentor is the interlocutor whose competence sits just beyond the participants' current reach, which is why the four journal authors independently identified the mentor's “how will this land in a real classroom?” as a distinct layer. This account specifies what near-peer mentoring contributes [5] and extends it by showing that the function can be embedded routinely within each feedback cycle, rather than arriving as an occasional alumni encounter, and that it is empirically separable from peer and faculty-advisor feedback.

The study therefore makes a threefold contribution. Theoretically, it reframes multi-source feedback in pre-service education from a device for delivering information to a scaffold that instills recruitable internal voices, and it distinguishes the assessor-side and recipient-side mechanisms that prior work has rarely separated. Methodologically and substantively, it identifies the graduate-mentor tier as an analytically distinct and developmentally consequential source. For practice, three implications follow. First, microteaching evaluation should ensure that several distinct evaluative voices contribute in differentiated ways, with the differentiation itself made visible to participants, and the reflective comparison journal used here is one practical means. Second, the graduate-mentor role should be cultivated deliberately, recruiting graduate students who have at

least preliminary classroom experience and who can articulate their classroom judgement in terms accessible to pre-service teachers. Third, evaluation cycles should build in explicit comparison episodes that require participants to confront discrepancies between self-assessment and external evaluation, since discrepancies that remain merely numerical do not, on their own, surface the standards that produced them. These implications speak to a core demand of China's arts curriculum standards [6], which presuppose teachers capable of reflecting independently on their own work against the four core competencies. Consistent with this, participants' confidence in reflecting independently during the forthcoming placement rose alongside their reflective independence (Item 15, $M = 4.75$), although the present study was not designed to test that relationship directly. Three limitations warrant acknowledgement. First, the sample is small, comprising eight participants in a single Chinese conservatoire. While the complete-cohort character of the case supports internal coherence, the small N constrains the statistical interpretation of the pre-post data and the generalisability of the qualitative themes; replication with larger cohorts and ideally with comparison conditions would substantially strengthen the empirical case. Second, the first author's dual role as researcher and graduate-mentor evaluator inevitably shaped both access to the setting and the interpretive lens applied to the data, despite the four procedural separations described. Third, the post-intervention measurements were administered within weeks of the conclusion of the microteaching cycles, while participants were still situated within the supporting institutional context, so the persistence of these capacities into school placement and early-career teaching remains an open question. Future research could incorporate a validated teacher self-efficacy instrument administered before and after the intervention, and longitudinal follow-up could test whether the cross-temporal dialogue these participants described continues to function once they are removed from the setting in which it was cultivated.

6. Conclusion

Pre-service music teachers in Chinese conservatoires enter their fourth-year school placement at a moment when the supervisory structures of the university recede and the daily responsibility for judging and improving their own lesson designs becomes theirs alone. The present study has examined how a four-source feedback configuration in microteaching prepares them for this transition. The study has identified two complementary mechanisms through which the configuration supports reflective independence, cultivating evaluative standards when participants assess peers, and surfacing self-perceived blind spots when they encounter discrepant evaluations of their own designs. The four sources differentiated, in participants' own characterisations, into three functionally distinct tiers, peer, graduate mentor, and faculty advisor, with the graduate-mentor tier emerging as a particularly consequential scaffold occupying the developmental space between peer and faculty advisor. The study reframes multi-source feedback as a mechanism for installing internalised voices that pre-service teachers can recruit in independent practice, and identifies design features through which microteaching evaluation can serve the cultivation of reflective independence.

References

- [1] Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. Basic Books.
- [2] Yıldız, P. K., & Yıldız, Y. (2021). Reflective practices in micro teaching from the perspective of preservice teachers: Teacher feedback, peer feedback and self-reflection. *Reflective Practice*, 22(6), 822–837.
- [3] Tilahun, S. (2023). Teacher educators' use of feedback to facilitate reflective practice among pre-service teachers during microteaching. *Cogent Education*, 10(2), 2257121.
- [4] Ballantyne, J. (2006). Reconceptualising preservice teacher education courses for music teachers: The importance of pedagogical content knowledge and skills and professional knowledge and skills. *Research Studies in Music Education*, 26(1), 37–50.
- [5] Shaw, R. D. (2024). Preparing conservatoire students for the music education workforce: Conversations with alumni. *Music Education Research*, 26(3), 1–14.
- [6] Ministry of Education of the People's Republic of China. (2022). *Curriculum standards for compulsory education in the arts* (2022 ed.). Beijing Normal University Press.

- [7] Schön, D. A. (1987). *Educating the reflective practitioner*. Jossey-Bass.
- [8] Korthagen, F. A. J., & Kessels, J. P. A. M. (1999). Linking theory and practice: Changing the pedagogy of teacher education. *Educational Researcher*, 28(4), 4–17.
- [9] Korthagen, F., Loughran, J., & Russell, T. (2006). Developing fundamental principles for teacher education programs and practices. *Teaching and Teacher Education*, 22(8), 1020–1041.
- [10] Gardiner, R. (2020). To see ourselves as others see us: Student teacher reflections as articulations of ideology. *Music Education Research*, 22(1), 15–28.
- [11] Panadero, E., Brown, G. T. L., & Strijbos, J.-W. (2016). The future of student self-assessment: A review of known unknowns and potential directions. *Educational Psychology Review*, 28(4), 803–830.
- [12] Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68(3), 249–276.
- [13] Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, 43(8), 1315–1325.
- [14] Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- [15] Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE.
- [16] Cronbach, L. J., & Shavelson, R. J. (2004). My current thoughts on coefficient alpha and successor procedures. *Educational and Psychological Measurement*, 64(3), 391–418.
- [17] Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101.
- [18] Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, 11(4), 589–597.
- [19] Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. SAGE.
- [20] Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis: Striving to meet the trustworthiness criteria. *International Journal of Qualitative Methods*, 16(1), 1–13.
- [21] Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Funding

This research was supported by the Wuhan Conservatory of Music 2026 Postgraduate Yan Tu Zhu Meng Program Project (Grant No. YTZM202618).

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.

Copyrights

Copyright for this article is retained by the author (s), with first publication rights granted to the journal. This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).