



MAKING PERFORMANCE TEACHABLE: CURRICULAR ALIGNMENT AND MICROTEACHING IN GROUP PIANO TEACHER EDUCATION

Min Xu, Pan Li, Baiyun Zhou, Xuan Yi

Jiangxi Normal University, P. R. China

E-mail: xumin@jxnu.edu.cn, 004937@jxnu.edu.cn,
005845@jxnu.edu.cn, 005856@jxnu.edu.cn

Abstract

Many pre-service music teachers can play a passage before they can explain, model, or teach it to someone else. This exploratory mixed-methods pilot examines that gap in a required group-piano course at a public normal university in mainland China. The course had been redesigned one year earlier, leaving an archive of checkpoint sheets, SMART practice logs, learning diaries, classroom interaction notes, microteaching plans, and end-of-course assessment forms. In Stage 1, two intact second-year cohorts of non-piano-major undergraduates ($n = 20$) worked through either a redesigned syllabus, organised around a Learning Outcomes-Activities-Evidence matrix and a weekly Listen-Read-Play-Teach cycle, or a conventional demonstration-and-practice routine. Stage 1 is reported descriptively because the archive did not retain medians, interquartile ranges, Mann-Whitney U values, or individual-level distributions, and the per-group sample was too small for responsible inference. Stage 2 re-read process materials from a wider explanatory sample ($n = 40$), using a codebook focused on listening, reading, playing, peer teaching, and difficulty turning performance knowledge into teachable language. The redesigned class cleared more threshold-approved checkpoint tasks, kept more target-linked practice logs, and wrote microteaching plans that named cues, broke demonstrations into steps, and anticipated likely beginner errors. The comparison class more often recognised problems but stopped before teachable explanation. The study offers a modest example of how visible criteria, recurring teach-back, and a revisitable evidence trail can help make practical performance knowledge more teachable.

Keywords: curricular alignment, microteaching, mixed-methods pilot, music teacher education, pre-service teachers

Introduction

Skill-based teacher education often treats competent performance as a bridge to competent teaching. The bridge is not automatic. Pre-service teachers may perform a routine fluently and still struggle to explain why it works, model it for a beginner, or anticipate the first mistake another learner is likely to make (Darling-Hammond, 2006; Dragnić-Cindrić & Anderson, 2025).

This difficulty is not unique to music. Laboratory courses, studio work, physical education, and other practice-heavy routes also ask candidates to convert action into instruction. Recent teacher-education studies suggest that this conversion is stronger when assignments require candidates to use professional knowledge in teaching activity, not only to describe that knowledge in written coursework (Carss et al., 2023; Crawford et al., 2023; Jones & Clarke, 2025; Pale et al., 2023; Spooner-Lane et al., 2023; Trent, 2023). Pale et al. (2023) show early-years candidates connecting academic knowledge with classroom action when course tasks made that connection necessary. Carss et al. (2023) report a related effect when assessment tasks were sequenced across a programme. Studies of emotional preparedness, digital literacy,

critical thinking, and inclusive-schooling readiness point to the same practical condition: the capability has to be visible while candidates still have time to rehearse it (Brígido Mero et al., 2024; Demir Evcimen et al., 2025; Durante, 2025; Protacio, 2025; Vančíková et al., 2025).

Group piano is a useful place to examine this question. The course typically serves non-piano-major undergraduates who will later need keyboard skills for accompaniment, classroom demonstration, and rehearsal support in school music settings (Blackwell & Matherne, 2024; Kong, 2025; Shaw, 2024; Young, 2013).

Keyboard fluency alone does not prepare a student to explain touch, sequence practice for a peer, or diagnose a repeated mistake. For that reason, recent music teacher-education scholarship has pushed against the older assumption that pedagogical knowledge can be left for a later methods course (Tucker et al., 2025; Weidner et al., 2025). The Chinese setting makes this problem sharper. Student-centred reform and formative-assessment discourse have entered school music education, but teacher authority, demonstration-led traditions, and large classes remain familiar classroom conditions (Zhang & Leung, 2023, 2025; Zhang et al., 2025; Zhang et al., 2026). For pre-service music teachers in this context, performance anxiety and teaching anxiety may also interact, making the move from doing to teaching still harder (Cui et al., 2024).

In such a setting, that move rarely happens by accident; it has to be designed for.

Alignment and Visible Criteria in Practical Courses

Constructive alignment and backward design begin from a practical premise: name the outcomes, organise teaching around them, and gather evidence that students have reached them (Biggs, 1996; Biggs & Tang, 2011; Harden, 2007). In teacher education, the value of that premise lies less in tidy syllabus design than in what it makes visible during learning. Carss et al. (2023) show how linked assessment tasks can support candidates' engagement with teaching as inquiry, while Crawford et al. (2023) argue that experiential assessment places pre-service teachers in more authentic evaluative situations. Kerimoğlu and Altun (2024) add that backward design matters most when it changes what candidates can see and do across a course, not only how the course is described on paper.

Practical courses especially need visible criteria. When the standard remains only in the instructor's judgement, students may know that something is wrong without knowing what failed or what to try next. Sadler's (1989) point remains useful here: formative assessment works only when learners understand quality, compare their current attempt with that standard, and act on the difference. Hattie and Timperley (2007), and later Brooks et al. (2019), frame the same problem through feedback questions about direction, current performance, and next steps. Brookhart (2024) extends the argument to teachers' own assessment knowledge, treating it as contextual and action-oriented.

Alignment is used here in that limited sense. Newby and Cornelissen (2025) warn that constructive alignment can become reductive when treated as a universal logic, and that warning is relevant to practical teacher education. The present study uses alignment more narrowly: as a tool for connecting five course outcomes with weekly tasks and with evidence students and the instructor could revisit.

Pedagogical Transformation in Practical Teacher Education

Shulman (1986) treated teaching as a problem of transformation. Knowing subject matter is not enough; the teacher must turn it into explanations, demonstrations, analogies, and judgements that another learner can use. Later work on theory-praxis relations, teaching performance assessment, and school-based judgements of teaching ability reaches a similar point from different directions (Chen & Tu, 2025; Pale et al., 2023; Spooner-Lane et al., 2023).

Music teacher education makes the problem sharper. Musical understanding often sits in the hand and ear before it is available as language. A student may feel the right touch for a phrase, yet still be unable to name the principle, plan the teaching sequence, or predict a beginner's likely mistake (Kong, 2025; Shaw, 2024).

In this study, pedagogical explanation and demonstration refers to four linked moves: explaining why a passage is played in a particular way, modelling that way for another learner, diagnosing likely errors, and adapting the explanation so a peer can act on it. The construct is close to current music teacher-education accounts of pedagogical knowledge, but it is not the same as carrying a performance technique from one piece to another.

Microteaching, Peer Feedback, and Self-Regulated Learning

Microteaching was deliberately brief in this course. A typical slot lasted no more than five minutes and sat inside the weekly lesson, rather than appearing as a separate assessment event. That placement matters. A recent review confirms that microteaching remains a common and useful way for pre-service teachers to rehearse pedagogy before practicum demands increase (Iliasova et al., 2025). The literature also warns against treating it as automatically educative. Trent (2023) shows that candidates may experience microteaching as emotionally demanding when instructor support is weak; Gunawardena (2023) shows that reflective structures can shift the activity away from display and toward reciprocity.

Feedback raises the same issue from another side. Peer comments improve pedagogical judgement only when students know the criteria, trust the exchange, and have strategies for responding to what they hear (Cano García, 2024; Cavanagh & Tran, 2025; Franke et al., 2024; Okumu et al., 2024). Self-regulated learning studies make a comparable demand: learners need visible standards, occasions to monitor their work, and a next action that is concrete enough to try (Nicol & Macfarlane-Dick, 2006; Panadero, 2017). In music education, feedback has often remained teacher-led and task-centred. Recent studies by Blackwell et al. (2023), Blackwell and Matherne (2024), and McPherson and Blackwell (2025) show, however, that structured practice in giving and receiving feedback can help pre-service music teachers build judgements they can use in later teaching.

Teach-back routines therefore serve a design function in practical courses. They require students to put tacit knowledge into words, test whether another learner can act on those words, and then revise both the playing and the explanation.

Research Aim and Questions

Group piano is treated here as a teacher-education problem, not only as a course-reform problem. The study reports a two-stage exploratory mixed-methods pilot of an aligned group-piano design for non-piano-major pre-service music teachers in mainland China. Three questions guided the analysis.

RQ1. In an exploratory intact-class comparison, what descriptive pattern of checkpoint attainment, reported practice time, and end-of-course outcomes is associated with the aligned design relative to a conventional condition?

RQ2. How was pedagogical explanation and demonstration evidenced in day-to-day course work within the aligned design?

RQ3. Which features of the aligned design appear to support the move from performance toward pedagogical explanation and demonstration?

Research Methodology

Design

The materials analysed here came from a redesign of a required group-piano course at a public normal university in mainland China. The redesign had taken place one academic year before the present study. After reviewing the archive, we organised the analysis as a two-stage exploratory mixed-methods pilot. Stage 1 compared two intact classes under ordinary teaching conditions, looking for classroom-level patterns that distinguished the aligned design from a conventional approach. Stage 2 extended the sample within the same course sequence and re-read process materials – learning logs, interaction records, and microteaching plans – to clarify how pedagogical explanation and demonstration developed.

The two stages had different roles. Stage 1 located a pattern in a very small classroom sample, while Stage 2 carried the explanatory work. Random redistribution across classes was not possible because departmental timetabling and intact cohort assignments had already been fixed. The archive therefore offered a practical basis for linking classroom outcomes to learning processes, but not a basis for treating the redesign as a causal experiment.

Context and Participants

The course was a required group-piano class for second-year undergraduates in a music teacher-education programme. All participants were non-piano majors studying keyboard skills for accompaniment, classroom demonstration, rehearsal support, and school-based music teaching. The archived project records show that students with extensive prior piano study had been excluded from the analytic sample; the exclusion threshold was more than ten years of formal training or a Grade-10 piano examination background. Because the analysis relied on archived course records, participant descriptions are limited to variables preserved across both stages: year level, programme status, non-piano-major status, and the prior-piano screening criteria. Age and gender distributions were not preserved in the archived analytic files.

Stage 1 included 20 students drawn from two intact classes (aligned group $n = 10$; comparison group $n = 10$). Stage 2 included 40 students (aligned group $n = 20$; comparison group $n = 20$), comprising the Stage 1 students plus additional students from parallel classes in the same year level and course sequence. The parallel classes followed the same course sequence, entry criteria, assignment requirements, assessment regime, checkpoint thresholds, repertoire bank, and semester schedule. That uniformity supports their use as an explanatory extension rather than as a separate cohort.

The course ran across two semesters for a total of 32 class hours, with one scheduled class period per week. Across both stages, the same course instructor taught all sections, the same repertoire bank and comparable repertoire difficulty were used, and overall time allocation plus core technical content were held constant. This arrangement reduced between-section variability, although course-instructor effects cannot be fully separated from design effects.

Table 1
Archived Baseline Diagnostic Procedures and Comparability Summary

Baseline Domain	Assessment Approach	Archived Between-Group Result
Touch and Tone Control	Diagnostic task rated by four independent raters using a four-level rubric	Archived comparability records indicated no baseline imbalance; full group descriptive statistics were not preserved.
Rhythmic Stability	Diagnostic task rated by four independent raters using a four-level rubric	Archived comparability records indicated no baseline imbalance; full group descriptive statistics were not preserved.
Sight Reading	Diagnostic task rated by four independent raters using a four-level rubric	Archived comparability records indicated no baseline imbalance; full group descriptive statistics were not preserved.

Note. The archived records preserved each diagnostic procedure and a general comparability judgement, but not the underlying group descriptive statistics. The baseline stage is used only to describe initial classroom comparability and not to support inferential claims.

Course Design and Comparison Condition

The comparison condition followed a familiar group-piano pattern: teacher demonstration, student imitation, after-class practice, periodic repertoire checks, with assessment focused on end products. Students received feedback, but process records were sparse and the link between weekly tasks and final outcomes remained largely implicit.

The aligned condition used a deliberately structured design. Five learning outcomes were specified in advance: touch and tone control; rhythmic stability and ensemble coordination; sight reading and self-monitoring; style and musical expression; and pedagogical explanation and demonstration. A Learning Outcomes-Activities-Evidence matrix linked weekly tasks, evidence sources, and attainment thresholds for each outcome.

The aligned lessons used the Listen-Read-Play-Teach sequence each week. First, students listened to modelled contrasts. They then read scores and short technique or expression scripts before moving into segmented rehearsal and peer review. The final step was a brief microteaching task in which a student taught a short excerpt or strategy to classmates. Weekly task sheets kept the criteria visible, while logs, checkpoint records, peer comments, and microteaching plans left a recurring process record across the semester. In practice, the sequence moved from goal prompt to modelled listening, from script preparation to rehearsal, and from peer teaching to evidence-based feedback. The comparison group worked with similar repertoire difficulty and comparable overall time on task. What differed was the explicit linkage among outcomes, weekly tasks, and evidence. Figure 1 summarises the overall study design and the aligned learning cycle.

Table 2
Two-Stage Study Design and Data Sources

Stage	Sample	Purpose	Main Sources	Analysis
Stage 1: Exploratory Classroom Comparison	20 students (10 aligned, 10 comparison)	To identify directional classroom-pattern differences between an aligned design and a conventional group-piano approach	Baseline diagnostics; classroom checkpoint records; weekly practice logs; end-of-course assessment	Descriptive comparison only (means and SDs); no inferential testing
Stage 2: Process Analysis	40 students (20 aligned, 20 comparison)	To explain how pedagogical explanation and demonstration developed in practice	Learning logs; classroom interaction texts; microteaching plans	Theory-informed coding; node comparison; per- student averages; illustrative excerpts

Note. Stage 2 included the Stage 1 participants plus additional students from parallel classes in the same course sequence. Integration occurred at interpretation, with Stage 2 used to explain the patterns observed in Stage 1.

Figure 1
Two-stage study design and aligned learning cycle.

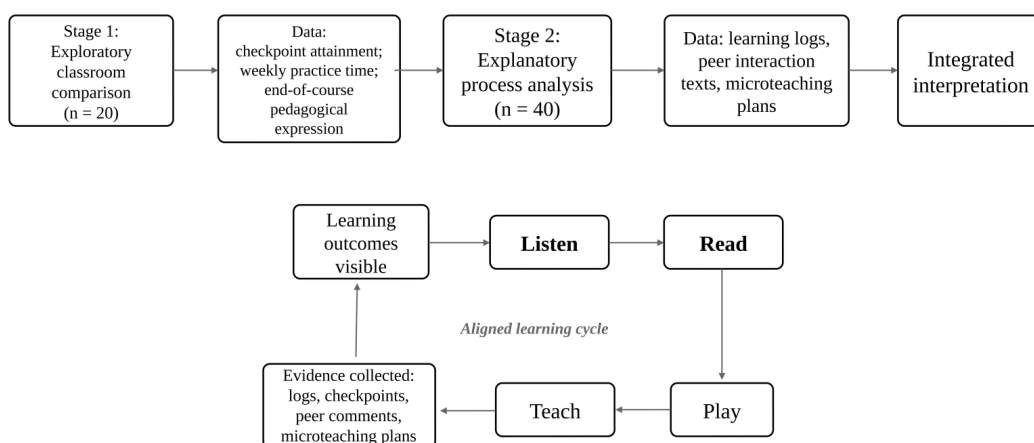


Table 3
Condensed LO-A-E Matrix for the Aligned Group-Piano Design

Learning Outcome	Key Activities	Evidence	Attainment Threshold
LO1. Touch and Tone Control	Contrastive listening; technique script cards; slow, segmented, and synthesis practice	Rubric ratings; 32-bar error sheet	Rubric average of at least 3 on a 4-point scale and error rate no higher than 5%
LO2. Rhythmic Stability and Ensemble Coordination	Rhythmic cueing; duet rehearsal by part; peer review	Collaboration rating; one-pass record	Collaboration score of at least 80/100 and one-pass rate of at least 90%
LO3. Sight Reading and Self-Monitoring	Score-reading prompts; SMART practice logs; self- and peer feedback	Sight-reading error record; weekly logs	Error rate no higher than 5%; at least five sessions and 180 minutes per week
LO4. Style and Musical Expression	Style listening; verbal explanation; expressive script cards	Performance rubric; style quiz	Rubric average of at least 3 on a 4-point scale and quiz score of at least 80/100
LO5. Pedagogical Explanation and Demonstration	Microteaching; decomposed modelling; peer teaching	Microteaching rubric; lesson plan; peer comments	Microteaching score of at least 80/100 and application pass rate of at least 85%

Note. In the archived materials, the fifth dimension was originally labelled teaching expression; the more explicit term pedagogical explanation and demonstration is used here. The microteaching rubric focused on alignment of goals and tasks, demonstration and decomposition, correction and feedback, teaching language and organisation, and transfer-task design.

Data Sources and Measures

Stage 1 used four archived sources. Baseline diagnostics covered touch and tone control, rhythmic stability, and sight reading; four independent raters scored each diagnostic task. The touch-and-tone rubric considered action stability, touch control, tone and dynamic layering, and accuracy. Rhythmic stability and ensemble work were judged through beat stability, entry handling, real-time adjustment, rehearsal engagement, and collaboration. Sight-reading records preserved note errors, rhythm errors, and completion time.

Classroom checkpoint records formed the second source. These records showed how many checkpoint tasks each student completed and passed during the two-semester course. A task counted only when three conditions were met: the student completed an item from the shared course bank, met the threshold for that checkpoint, and received teacher confirmation. Different excerpts from the same piece were not counted separately. The checkpoint totals should therefore be read as cumulative threshold-approved tasks from a shared instructional bank, rather than as counts of full-length pieces. The third source was the weekly SMART practice log, which summarised daily practice time. In the aligned group, students had to enter a target, an indicator, and a method before recording minutes. The fourth source was the performance-based end-of-course assessment. It judged tone control, rhythmic stability, performance presentation, and pedagogical explanation and demonstration; the final dimension required students to explain and model key musical points, not merely perform them. The archived records reported an interrater reliability coefficient of $ICC = .86$ for that assessment.

Stage 2 turned to the process archive: learning logs, classroom interaction texts, and microteaching plans. The logs recorded weekly practice content, difficulty points, strategies, and self-feedback. Interaction texts preserved teacher questions, peer discussion, and immediate

feedback notes. Microteaching plans showed how students organised a short teaching sequence around objectives, demonstration steps, common errors, and corrective cues. The original project did not use video. As a result, plans, observation notes, and rating sheets became the main record of microteaching. These materials capture documented pedagogical planning and explanation more directly than the full quality of live demonstration in interaction. The microteaching rubric asked raters to look for four features in each plan. Goals and tasks had to align. Demonstrations had to break the difficulty into manageable parts. Feedback needed to be actionable and verifiable through a second attempt. The transfer task had to be usable beyond the immediate exercise. Coding units were complete meaning-bearing sentences or standalone lesson-plan items. Generic comments such as “I did not practise well today” were excluded unless they also included a specific error analysis, strategy, explanation, or feedback move. The original qualitative materials were written in Chinese; the authors translated the excerpts into English and checked them against the source wording during drafting.

Coding paid particular attention to pedagogical specificity. A statement such as “practise the left hand more” was not coded as a pedagogical explanation because it named effort without naming a cue, target, error, or teaching move. By contrast, a statement that identified the likely source of instability, proposed a decomposed step, or indicated how a peer would verify improvement was treated as evidence of explanation that could be used in teaching. The distinction kept routine self-report language from being overread as pedagogical development.

Data Analysis

Stage 1 is reported descriptively only. The archived summary files preserved group means and standard deviations, along with attainment counts on the end-of-course rubric. They did not preserve the underlying Mann-Whitney U values, medians, interquartile ranges, or individual-level distributions. Those values cannot be responsibly reconstructed from what remains. Given the *n* of ten per group, the missing distributional information, and the intact-class design under one course instructor, inferential testing would not be appropriate. Continuous indicators are therefore reported as means and standard deviations, and end-of-course outcomes as attainment counts. Z statistics, *p* values, and effect-size estimates from earlier outcome tables have been removed. The numerical findings function as classroom-level indicators that motivate the explanatory work in Stage 2, not as treatment-effect estimates.

The qualitative analysis moved between pre-set categories and new patterns in the archive. The first version of the codebook came from the aligned task chain: listening as input, reading as scripted retention, playing as output, and teaching as peer explanation. A further node was added for difficulty translating embodied performance knowledge into teachable language. About one fifth of the materials were double-coded. When coders disagreed, the discussion returned to the code definitions, and the node boundaries were revised before coding the full dataset. The archived project notes reported Cohen’s $\kappa = .83$.

Triangulation was built into the analysis rather than added after coding. We compared logs with interaction texts and plans, checked coder agreement on the shared subset, and integrated the two phases only at interpretation. Reference counts were treated as traces of process, not competence scores. After reconciliation, the final node definitions were applied to the full archive; excerpts were kept only when the same pattern appeared in at least two source types.

Trustworthiness, Ethics, and Transparency

The course instructor (the second author) was involved in the original redesign of the course. That involvement gave us contextual insight but also raised the possibility of interpretive

bias. Because the same course instructor designed and taught all sections, design features and teacher enactment remain intertwined in the archive. We tried to reduce over-interpretation by relying on independent rating for rubric-based measures, by keeping instructional conditions stable across cohorts, and by triangulating across multiple process sources rather than a single reflective account. Node frequencies in the qualitative phase were interpreted against archived excerpts, the wording of the instruments, and the structure of the learning tasks themselves.

Institutional ethics approval was obtained before data collection. Identifying details of the approving body and the approval reference are withheld for blind review and can be provided to the editor on request. All participants gave informed consent, and archived materials were anonymised before analysis. The full dataset is not publicly available because the archive contains potentially identifiable educational data. De-identified excerpts, coding definitions, and summary tables may be shared on reasonable request and subject to institutional ethics requirements.

AI-assisted language tools were used only after study design, data collection, analysis, and interpretation had been completed. In the earlier copy-editing stage, Grammarly Pro was limited to spelling, grammar, punctuation, awkward-phrasing alerts, and basic readability or formatting prompts of the kind a copy-editor might flag. Grammarly Pro's generative writing functions were not used, and outputs from Generate, Rewrite, Paraphrase, Ideas, Shorten, or comparable generative rephrasing functions were not accepted or retained. No AI tool was used to create research ideas, generate findings, code data, perform statistical operations, interpret results, select excerpts, or replace authorial judgment. The authors checked the revised wording against the underlying archive, made the final decisions, and take full responsibility for the manuscript.

Research Results

Stage 1: Exploratory Classroom Comparison

The archived baseline diagnostics indicated no obvious imbalance between the two intact classes in touch and tone control, rhythmic stability, or sight reading at the start of the course. This judgement does not rule out selection effects or other pre-existing differences. It supports only a limited claim: the two classes were not visibly mismatched at baseline. The baseline summaries are therefore used to describe initial classroom comparability, not to support inferential claims.

Across two semesters, the aligned group completed and passed more checkpoint tasks and reported longer daily practice time than the comparison group. Table 4 reports the descriptive summaries. The aligned class moved more often from task attempt to documented threshold attainment within the shared task bank, and its members logged more minutes of daily practice. These figures are read as classroom-level indicators that motivate the process analysis, not as evidence of an effect.

Earlier versions of the archived output included a large effect-size estimate for the checkpoint measure. That estimate is no longer reported. With ten students per group, and without preserved medians, interquartile ranges, Mann-Whitney U values, or individual-level distributions, we cannot inspect whether outliers, skew, or other distributional features shaped the estimate. Table 4 therefore stays with the descriptive summaries that the archive preserved. The difference between group means is used only as a classroom-level pattern for Stage 2 to explain; it is not treated as an effect-size estimate or as evidence of causality.

Table 4
Stage 1 Continuous Outcomes from Archived Summary Records (Descriptive Only)

Measure	Aligned Group M (SD), <i>n</i> = 10	Comparison Group M (SD), <i>n</i> = 10	Descriptive Difference (<i>M</i> ₁ – <i>M</i> ₂)
Checkpoint tasks passed from the shared course bank	28.67 (3.21)	11.58 (2.86)	17.09
Daily practice time (minutes)	59.37 (8.65)	50.68 (9.12)	8.69

Note. Values reproduce the descriptive summaries preserved in the archived Stage 1 output. The checkpoint measure refers to threshold-approved tasks from the shared course bank, not counts of full-length pieces. Inferential statistics are not reported because the archived summary did not retain Mann-Whitney U values, medians, or interquartile ranges, and inferential testing was judged inappropriate given the sample size and intact-class design under a single instructor. The classroom-level contrast is reported as an exploratory directional indicator that motivates Stage 2.

Attainment counts at the end of the course pointed in the same direction. On each rubric dimension, the aligned group contained more students who met the benchmark than the comparison group. The strongest contrast concerned pedagogical explanation and demonstration: eight of ten aligned students reached the benchmark, compared with three of ten in the comparison group. No formal significance test is reported for these counts. For the reasons stated in Data Analysis, they are used only as a classroom-level pattern to be interpreted alongside the process archive.

Table 5
Stage 1 Attainment Counts on the End-of-Course Assessment (Descriptive Only)

Dimension	Aligned Group Attained (<i>n</i> = 10)	Comparison Group Attained (<i>n</i> = 10)
Tone Control	9	7
Rhythmic Stability	8	6
Performance Presentation	7	6
Pedagogical Explanation and Demonstration	8	3

Note. The final pedagogical-explanation dimension corresponds to the archived rubric label teaching expression. Counts are reported as contextual pilot evidence from an intact-class comparison and should not be read as causal. Inferential statistics are not reported (see Data Analysis).

Stage 2: Process Evidence in Three Themes

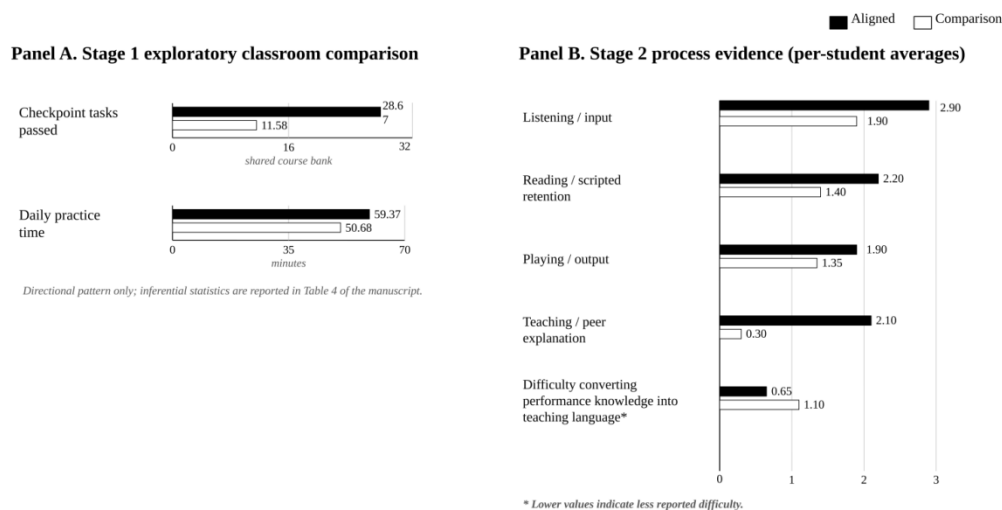
Table 6 reports the Stage 2 process markers for the two groups. Figure 2 places the descriptive Stage 1 patterns from Table 4 beside the Stage 2 markers from Table 6, giving readers a visual route into the integration. The following three themes then show how the coded traces appeared in ordinary course work and how they help make sense of the small-sample classroom pattern.

Table 6
Stage 2 Coded Process Markers

Process Node	Aligned References	Comparison References	Per-Student Average (A/C)	Interpretive Meaning
Listening / Input	58	38	2.90 / 1.90	Students identified modelled targets and performance contrasts
Reading / Scripted Retention	44	28	2.20 / 1.40	Students annotated scores and turned musical problems into usable scripts
Playing / Output	38	27	1.90 / 1.35	Students performed, monitored, and corrected their work
Teaching / Peer Explanation	42	6	2.10 / 0.30	Students explained, demonstrated, and guided peers
Difficulty turning performance knowledge into teachable language	13	22	0.65 / 1.10	Students struggled to convert performance experience into usable teaching language

Note. Reference counts come from the archived coding summaries and function as process markers. They do not serve as direct measures of ability. A/C = aligned/comparison.

Figure 2
Summary of Stage 1 outcome patterns and Stage 2 process evidence.



Making Tacit Performance Knowledge Explicit

Pedagogical explanation produced the clearest qualitative contrast. The aligned group generated 42 references to peer explanation; the comparison group generated six. The aligned group also showed fewer references to difficulty translating performance knowledge into teachable language, 13 against 22. The excerpts illustrate the difference. One comparison-group

student wrote: “My accompaniment keeps getting stuck. I know my fingers are not independent enough, but if I had to teach this later, I would not know how to tell elementary pupils, in words, where the effort should come from.” In an aligned-group microteaching plan, a student wrote: “During the demonstration, remind the peer to imagine the wrist as a spring-loaded shock absorber and to release immediately after the key is depressed.”

The two excerpts point to different stages of pedagogical work. The first recognises a technical difficulty but stops at recognition. The second converts an embodied action into an image, a demonstration point, and a correction strategy.

In the aligned materials, private sensation more often became public explanation. Students named the problem, connected it to a target, and planned a demonstration that a peer could try. Many entries also included a second attempt or a check on whether the cue had worked. Pedagogical explanation here is therefore not a matter of fluent speech alone. It requires choices: what to foreground, where to divide the movement or sound, which error is likely to occur, and what cue can support self-correction. The comparison materials tended to stop earlier, at problem recognition rather than preparation for another learner.

Visible Criteria Reshaped Self-Regulated Practice

Listening- and reading-related nodes were also more frequent in the aligned group: 58 versus 38 for listening, and 44 versus 28 for reading. These numbers suggest that modelled contrasts, score annotation, and script cards became part of the working method rather than peripheral preparation. Aligned-group logs usually followed a target-strategy-evidence-next-step structure. Comparison-group entries more often named errors, or noted a general need for more practice, without linking those observations to a criterion.

Longer practice time is not, on its own, a reliable sign of better learning. Students can repeat an unstable pattern for an hour and improve little. What the aligned logs suggested was narrower: practice time was more often tied to a target, a method, and a basis for judging progress. The weekly form supported that pattern by requiring target, indicator, and method entries before minutes could be recorded. These prompts do not prove learning by themselves, but they help explain why aligned-group records contained more criterion-linked self-monitoring and more specific statements about what would be checked next.

Reading and listening operated, in the aligned design, as organising actions rather than warm-ups. Students were asked to listen for contrasts, mark score features, prepare short verbal scripts, and then use those scripts during later playing and teaching. That sequence mattered because it gave students language before full performance stability had been reached. Rather than waiting until a passage felt automatic, the aligned class moved repeatedly between hearing, labelling, playing, and explaining.

Peer Teaching Created Pedagogical Accountability

Microteaching also brought accountability into the work. Under the aligned criteria, a student had to set a short objective, divide the demonstration into manageable steps, identify likely errors, and provide feedback that could be tested in a follow-up task. The plans and peer-feedback sheets show this shift. Aligned-group students more often described how they would isolate a rhythmic problem, simplify an accompaniment pattern, check whether a peer could reproduce the correction, or adapt the cue for later classroom use.

The rubric made that shift visible. In LO5, raters checked the fit between goal and task, the decomposition of the demonstration, the actionability of feedback, and the reusability of the transfer task. These criteria made teaching quality part of the threshold. Accurate playing alone was not enough for a student to meet the requirement.

The comparison group, by contrast, more often stopped at performance completion. Students recognised that a passage was difficult or that a coordination problem remained, but they less often documented how they would teach through that difficulty for someone else. The issue was not an absence of effort. It was the absence of routine occasions for making that effort pedagogically accountable.

Integrated Interpretation

Table 7 aligns the principal Stage 1 patterns with the process evidence that helps explain them. The integration does not treat the qualitative phase as proof of the quantitative phase. It shows, more modestly, how the process archive renders the small-sample classroom pattern more interpretable.

More checkpoint attainment lined up with visible thresholds and repeated evidence use. Longer reported practice lined up with target-strategy-evidence-next-step logs. The more favourable pedagogical-explanation pattern lined up with decomposed modelling, peer explanation, and fewer translation barriers. The contribution of the aligned design therefore appears to lie less in increasing activity than in reorganising how activity was interpreted, recorded, and recycled into later learning.

Table 7
Joint Display of Stage 1 Patterns, Stage 2 Process Evidence, and Integrated Inferences

Stage 1 Pattern	Stage 2 Process Evidence	Integrated Inference
More checkpoint tasks were attained in the aligned group.	Outcomes reappeared in logs, checkpoint sheets, and microteaching criteria; students more often moved from target to evidence to revision.	Visible thresholds made task completion legible and supported repeated movement from attempt to documented attainment.
Longer reported daily practice time appeared in the aligned group.	Aligned logs more often followed a target-strategy-evidence-next-step structure; listening and score reading were used as organising actions rather than preliminaries.	Practice time appears to have become more purposeful because time was attached to a visible target and a method of checking progress, rather than simply longer.
A more favourable end-of-course pattern appeared on pedagogical explanation and demonstration.	The aligned archive contained more peer explanation, more decomposed modelling, and fewer instances of difficulty turning performance knowledge into teachable language.	Embedded teach-back routines appeared to help students put tacit performance knowledge into words a peer could use.

Note. The joint display does not treat the qualitative phase as proof of the quantitative phase. Its function is explanatory: to show how the process archive makes the small-sample classroom pattern more interpretable.

Discussion

Alignment as a Practical Organiser

Alignment mattered in this course less as a slogan than as a set of small constraints on weekly work. In the comparison condition, practice could still mean sitting longer at the keyboard, repeating the assigned bars, and waiting for the course instructor’s next correction. The aligned group was asked to do something narrower and more visible: name the target,

attach it to a criterion, leave evidence of the attempt, and then try to teach the point to a peer. That change helps explain why their records contained more target-linked practice and more usable teaching language. This interpretation still has to remain cautious. One course instructor taught all sections, and the second author enacted the redesigned course; design effects, cohort effects, and instructor enactment therefore cannot be separated cleanly.

Visible criteria also shaped the work before assessment occurred. They directed attention and gave revision a point of reference. Students could see where effort should go, rather than waiting for a retrospective judgement (Brookhart, 2024; Sadler, 1989).

The process findings follow that logic. In the aligned group, logs more often named a target and connected that target to a next step. Score annotation and listening were used as preparation for later action, not as brief preliminaries. The LO-A-E matrix gave students a vocabulary for practice that was less available in the comparison condition. This reading is consistent with backward-design work when alignment is embedded in instructional activity rather than left at the planning stage (Kerimoğlu & Altun, 2024), and with programme-level studies of assessment mapping (Carss et al., 2023; Crawford et al., 2023). Newby and Cornelissen's (2025) warning still matters: constructive alignment should not become a totalising philosophy. The aim here is narrower. Expectations must be visible enough for students to practise with direction and for teachers to respond with evidence.

A Rival Hypothesis: Documentation Volume or Pedagogical Capability?

One rival explanation needs to be considered. The aligned group may look stronger because its students were required to leave more written traces, not because their pedagogical capability developed further. If documentation volume alone explains the pattern, the main claim of the study becomes much weaker.

The archive does not fit a simple volume explanation. If the issue were merely more writing, higher reference counts should appear across all five process nodes in roughly similar proportions. They do not. The largest gap appears on the teaching/peer-explanation node, where the per-student ratio is 2.10 to 0.30. That is about a seven-fold difference. On the playing/output node, where both groups had similar reasons to write, the difference is only 0.55 per student. The aligned-group surplus is concentrated where the LO-A-E matrix required pedagogical reasoning, rather than being spread evenly across the writing demands of the course.

The entries themselves also point beyond length. The comparison group did write about playing; students recorded errors, frustration, and awareness that something was wrong. What was often missing was the move from a recognised problem to a teachable explanation. The wrist-as-shock-absorber example works differently from the comparison student's statement about "where the effort should come from". It anticipates a peer's likely misunderstanding and sets up both the cue and the demonstration. More words alone would not account for that difference.

The LO5 rubric provides another check. To meet the aligned-group threshold, students had to show that the goal and task fitted together, that the demonstration decomposed the difficulty, that feedback could be verified through a second attempt, and that the transfer task could be reused. A long but generic plan would not have met those requirements. The aligned design did increase documentation, but it increased a particular kind of documentation: records had to display pedagogical reasoning before they counted toward the threshold.

The limitation remains. The archive cannot fully separate documentation volume from pedagogical capability. A cleaner test would require a comparison condition that demands equivalent documentation without the LO-A-E scaffolding. Until such work is done, this rival explanation should remain visible rather than dismissed.

Why Embedding Teaching within Skill Learning Mattered

The strongest difference between the two conditions was probably the decision to embed teaching inside skill learning, rather than postponing it to practicum or a later methods course. Students in the aligned group were repeatedly asked to explain fingering logic, model a technique, diagnose a problem, or offer a cue another learner could use. Shulman (1986) identified this kind of work as central to teaching: the transformation of private competence into publicly usable explanation.

The wrist-release plan makes the transformation concrete: a student framed the wrist as a spring-loaded hinge that releases after key contact. This was not simply a colourful phrase. It gave the peer a bodily image, a demonstration focus, and a correction to test. Similar moves appeared more often in the aligned group. That pattern sits well with microteaching research that emphasises low-stakes practice in explanation, feedback, and reflection before practicum (Iliasova et al., 2025). It also needs the cautions in the literature. Trent (2023) shows that microteaching can carry emotional pressure; Gunawardena (2023) shows that reflective structure helps keep the activity educative. In this course, the slots were brief, frequent, and criterion-supported, which may have kept teach-back developmental rather than merely evaluative.

Bringing teaching into the piano course also changed the standard of success. If the course rewards only accurate or expressive playing, students can remain in the role of individual performers. Once they must explain and demonstrate for a peer, they have to select the first feature to show, anticipate misunderstanding, judge which error matters most, and decide how improvement will be checked. Their own execution still matters, but it is no longer the whole task. Another learner's uptake becomes part of the work.

In Chinese music classrooms, teacher demonstration has long been a central pedagogical resource. Student-centred reform has not removed the visibility of modelling or teacher authority in music lessons (Zhang & Leung, 2023, 2025). The aligned course relocated part of that resource. Demonstration became something students had to prepare, decompose, and justify for peers, rather than something they mainly received from the instructor. In this limited sense, the design redistributed a familiar teacher-led practice for the preparation of future teachers.

Implications for Practical Teacher Education

The value of the LO-A-E design is not confined to music. Its broader relevance lies in a recurring problem across practical teacher education: candidates can often perform more than they can explain. Recent work on pre-service competencies in critical thinking, digital literacy, inclusive-schooling readiness, action research, and emotional preparation has reached a comparable conclusion, namely that professional capacities need to be made visible during preparation rather than assumed to appear later (Brígido Mero et al., 2024; Demir Evcimen et al., 2025; Durante, 2025; Kuzu & Erdoğan Kayabaşı, 2025; Protacio, 2025; Vančíková et al., 2025).

The design worked through several modest components rather than through a single intervention. Outcomes returned throughout the semester in logs, checkpoint sheets, and microteaching criteria, so students repeatedly encountered what counted as progress. The Listen-Read-Play-Teach cycle turned that return into a routine. The evidence trail then gave the instructor and students a record of practice, explanation, and revision that could be used in later feedback. Teach-back did not need to become a long weekly mini-lesson. Its value lay in being brief, regular, and tied to criteria.

Shulman's framework clarifies the extension made by this course. Practical teacher-education students often possess tacit knowledge before they can represent it for another learner. Courses therefore need tasks that convert that knowledge into language, cues, and

demonstrations. The same logic could travel to science methods, physical education, or laboratory-based teacher education, provided each practical task includes a short cycle of explaining, modelling, and checking.

Limitations and Future Research

Several limitations shape how this study should be read. The exploratory classroom comparison relied on intact classes from a single institution, one course instructor, and a very small sample. Quantitative patterns must therefore be read as contextual rather than causal evidence; cohort differences and instructor effects cannot be ruled out. The archived records preserved only limited participant information, with fuller demographic details unavailable. The archived quantitative summaries also did not preserve every statistic desirable for a full nonparametric report — medians, interquartile ranges, and Mann-Whitney U values were not retained, which is one reason Stage 1 is reported descriptively in the present version. No classroom video was available, so some aspects of embodied demonstration and peer response were inevitably lost. Because the qualitative materials were translated from Chinese archival sources, an additional interpretive layer is involved, even though translations were checked against the original wording. Stage 2 is therefore best understood as an explanatory extension within the same instructional context, not as an independent replication.

The rival-hypothesis concern returns here. Future studies should include a comparison condition that demands equivalent documentation without the LO-A-E scaffolding, so that the contribution of pedagogical content and the contribution of writing volume can be separated more cleanly than is possible in the present archive.

These limitations lead to a focused research agenda. Multi-site work could test whether similar alignment structures support pedagogical explanation beyond one course and could separate design from instructor effects more convincingly. Video, stimulated recall, and classroom interaction analysis would make the movement from self-regulation to explanation easier to trace. Follow-up into practicum would show whether explanations rehearsed in a skill course survive in school classrooms. Another useful line would examine cognitive load: staging listening, reading, playing, and teaching across time may help candidates manage the combined demands of performing and instructing. In music education, future work should also connect course-level design with Chinese assessment reforms and with how schools recognise teacher candidates' teaching ability (Chen & Tu, 2025; Zhang et al., 2025; Zhang et al., 2026).

Conclusions and Implications

The two-stage exploration points to a modest design claim. In the aligned group, practice became easier to see and review: targets remained visible, expectations were repeated, and competence had to be demonstrated for someone else. The Stage 1 figures — more checkpoint attainment, longer reported practice, and more students meeting the pedagogical-explanation benchmark — are therefore treated as a classroom-level indicator for RQ1, not as evidence of an effect. Stage 2 explains where that indicator had substance. Pedagogical explanation and demonstration appeared through repeated movement among hearing, labelling, playing, and teach-back, with logs, score annotation, decomposed modelling, and peer explanation supplying the main traces (RQ2).

Across the two stages, three features carried most of the design work. Visible criteria gave students a working language. Recurring teach-back gave tacit performance knowledge a public form. A light evidence trail made feedback and revision possible across weeks (RQ3). For practical teacher education, the implication is direct: a course that asks students to perform and to teach must build the passage from performance to explanation into ordinary coursework.

The contribution lies in the combination, not in a general claim that microteaching helps. Used together, visible criteria, teach-back, and revisitable evidence can make performance knowledge more teachable inside an ordinary practical course.

These conclusions should be read as design-oriented propositions for future research, not as evidence of treatment efficacy. The next test should use cleaner conditions: multiple instructors, multiple sites, and a documentation-equivalent comparison condition.

Acknowledgments and Language-Editing Disclosure

The authors thank the School of Music, Jiangxi Normal University, for institutional and logistical support related to the instructional implementation and the organization of archived course records. Grammarly Pro was used only for limited English-language editing to improve clarity and grammar after study design, data collection, analysis, and interpretation had been completed. Its use was restricted to linguistic refinement. All substantive content, intellectual contributions, analytic decisions, and final wording were generated, reviewed, and approved by the human authors, who take full responsibility for the manuscript.

Ethics

Institutional ethics approval was obtained prior to data collection from the School of Music, Jiangxi Normal University (Approval No. IRP-JXNU-SOM-2025047), and all participants provided informed consent. The ethics approval form is supplied as a separate submission file.

Declaration of Interest

The authors declare no competing interest.

References

- Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, 32(3), 347–364. <https://doi.org/10.1007/BF00138871>
- Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university* (4th ed.). Open University Press.
- Blackwell, J., & Matherne, N. (2024). Developing feedback skills in preservice music teachers. *International Journal of Music Education*, 42(2), 243–256. <https://doi.org/10.1177/02557614231163237>
- Blackwell, J., Matherne, N., & McPherson, G. E. (2023). A PRISMA review of research on feedback in music education and music psychology. *Psychology of Music*, 51(3), 716–729. <https://doi.org/10.1177/03057356221109328>
- Brígido Mero, M., Rodríguez de Castro, A., Lejárraga García, A., Laviña Pérez, I., & Nistal Anta, V. (2024). Emotional training for prospective teachers in science education. *Problems of Education in the 21st Century*, 82(2), 236–251. <https://doi.org/10.33225/pec/24.82.236>
- Brooks, C., Carroll, A., Gillies, R. M., & Hattie, J. (2019). A matrix of feedback for learning. *Australian Journal of Teacher Education*, 44(4), 14–32. <https://doi.org/10.14221/ajte.2018v44n4.2>
- Brookhart, S. M. (2024). Educational assessment knowledge and skills for teachers revisited. *Education Sciences*, 14(7), Article 751. <https://doi.org/10.3390/educsci14070751>
- Cano García, E. (2024). Peer feedback for teaching professional development: Conditions for it to take effect. *Cogent Education*, 11(1), Article 2391577. <https://doi.org/10.1080/2331186X.2024.2391577>
- Carss, W., Cowie, B., & Cook, S. (2023). Mapping assessment tasks and lecturer considerations: A case study of a NZ primary teacher education programme. *Australian Journal of Teacher Education*, 48(9), 100–114. <https://doi.org/10.14221/1835-517X.6624>
- Cavanagh, M., & Tran, D. (2025). Pre-service teachers' feedback on their peers' lesson plans. *Asia-Pacific Journal of Teacher Education*, 53(4), 464–478. <https://doi.org/10.1080/1359866X.2025.2539240>

- Chen, L., & Tu, C. (2025). Chinese secondary school teachers' perceptions of teacher candidates' teaching ability and performance. *Australian Journal of Teacher Education*, 50(3), 35–53. <https://doi.org/10.14221/1835-517X.6413>
- Crawford, R., Jenkins, L. E., & Wan, L. (2023). Experiential learning projects as assessment in initial teacher education. *Australian Journal of Teacher Education*, 48(1), 37–54. <https://doi.org/10.14221/1835-517X.5787>
- Cui, C., Xie, X., & Yin, Y. (2024). Exploring the relationships among music performance anxiety, teaching anxiety, and self-efficacy of Chinese preservice music teachers. *Frontiers in Psychology*, 15, Article 1373454. <https://doi.org/10.3389/fpsyg.2024.1373454>
- Darling-Hammond, L. (2006). Constructing 21st-century teacher education. *Journal of Teacher Education*, 57(3), 300–314. <https://doi.org/10.1177/0022487105285962>
- Demir Evcimen, F. B., Ulukaya Öteleş, Ü., & Koçoğlu, E. (2025). The effect of critical thinking practices on digital footprint. *Problems of Education in the 21st Century*, 83(3), 306–322. <https://doi.org/10.33225/pec/25.83.306>
- Dragnić-Cindrić, D., & Anderson, J. L. (2025). Developing pre-service teachers' pedagogical content knowledge: Lessons from a science methods class. *Education Sciences*, 15(7), Article 860. <https://doi.org/10.3390/educsci15070860>
- Durante, P. G. C. (2025). Advancing critical literacy through critical literature pedagogy-based principles of literary education. *Problems of Education in the 21st Century*, 83(2), 196–217. <https://doi.org/10.33225/pec/25.83.196>
- Franke, U., Backfisch, I., Scherzinger, L., Lachner, A., Scheiter, K., & Stürmer, K. (2024). Do prompts and strategy instruction contribute to pre-service teachers' peer-feedback on technology integration? *Educational Technology Research and Development*, 72, 3117–3138. <https://doi.org/10.1007/s11423-024-10403-8>
- Gunawardena, M. (2023). Mentoring pre-service teachers: The THIINK4 reflective cycle. *Australian Journal of Teacher Education*, 48(4), 19–37. <https://doi.org/10.14221/1835-517X.5901>
- Harden, R. M. (2007). Outcome-based education: The future is today. *Medical Teacher*, 29(7), 625–629. <https://doi.org/10.1080/01421590701729930>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Iliasova, L., Nekrasova, I., Mena, J., & Estrada-Molina, O. (2025). Microteaching on pre-service teachers' education: Literature review. *Frontiers in Education*, 10, Article 1562975. <https://doi.org/10.3389/feduc.2025.1562975>
- Jones, S. E., & Clarke, D. (2025). Enhancing pre-service teachers' preparedness for professional experience placement and employment: Authentic school-university partnership opportunities. *Australian Journal of Teacher Education*, 50(3), 121–145. <https://doi.org/10.14221/1835-517X.6974>
- Kerimoğlu, E., & Altun, S. (2024). Backward design in pre-service teacher education to enhance curriculum knowledge. *Journal of Teaching and Learning*, 18(2), 128–149. <https://doi.org/10.22329/jtl.v18i2.8625>
- Kong, S.-H. (2025). Music education training for kindergarten teachers: A workshop integrating Shulman's pedagogical content knowledge and pedagogical reasoning and action model. *Journal of Early Childhood Teacher Education*, 46(1), 85–102. <https://doi.org/10.1080/10901027.2024.2410399>
- Kuzu, Ç. İ., & Erdoğan Kayabaşı, K. (2025). Determining the digital literacy levels of mathematics pre-service teachers. *Problems of Education in the 21st Century*, 83(4), 563–578. <https://doi.org/10.33225/pec/25.83.563>
- McPherson, G. E., & Blackwell, J. (2025). Teacher feedback in collegiate instrumental music lessons. *Journal of Research in Music Education*, 73(2), 217–236. <https://doi.org/10.1177/00224294241265014>
- Newby, L., & Cornelissen, L. (2025). A neutral toolkit? For a fundamental critique of constructive alignment. *Teaching in Higher Education*, 30(7), 1687–1702. <https://doi.org/10.1080/13562517.2025.2465987>
- Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199–218. <https://doi.org/10.1080/03075070600572090>
- Okumu, M. O., Lammert, C., & Hargura, H. M. (2024). Teacher candidates' perceptions of peer and instructor feedback through video annotations: Whose advice do they prefer? *Education Sciences*, 14(12), Article 1361. <https://doi.org/10.3390/educsci14121361>

- Pale, M., Ohi, S., & Kee, L. (2023). Towards pre-service teachers' theory-praxis nexus in early years English and literacy education: A pilot study. *Australian Journal of Teacher Education*, 48(1), 55–73. <https://doi.org/10.14221/1835-517X.5880>
- Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*, 8, Article 422. <https://doi.org/10.3389/fpsyg.2017.00422>
- Protacio, J. V. (2025). Predictors of concept formation and educational contextualization competencies in geometry of pre-service mathematics teachers. *Problems of Education in the 21st Century*, 83(2), 218–236. <https://doi.org/10.33225/pec/25.83.218>
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119–144. <https://doi.org/10.1007/BF00117714>
- Shaw, L. (2024). Conservatoire students' perspectives on instrumental music teacher education: Developing pedagogical knowledge for the future music education workforce. *Journal of Education for Teaching*, 50(1), 155–171. <https://doi.org/10.1080/02607476.2023.2228233>
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14. <https://doi.org/10.3102/0013189X015002004>
- Spooner-Lane, R., Broadley, T., Curtis, E., & Grainger, P. (2023). Development of a teaching performance assessment in Australia: What did we learn? *Australian Journal of Teacher Education*, 48(1), 1–16. <https://doi.org/10.14221/1835-517X.5692>
- Trent, J. (2023). How are preservice teachers discursively positioned during microteaching? The views of student teachers in Hong Kong. *Australian Journal of Teacher Education*, 48(6), 80–97. <https://doi.org/10.14221/1835-517X.6124>
- Tucker, O. G., Nicholson, G., & Jordan, R. C. (2025). Navigating competing demands within music teacher education: Biesta's three domains of educational purpose. *International Journal of Education & the Arts*, 26(25). <https://doi.org/10.26209/ijea26n25>
- Vančíková, K., Čiliaková, R., & Hušlová Orságová, M. (2025). Conceptualizations of inclusive schooling by pre-service teachers in Slovakia. *Problems of Education in the 21st Century*, 83(5), 689–704. <https://doi.org/10.33225/pec/25.83.689>
- Weidner, B. N., Wilson, J., & Royston, N. S. (2025). Self-study of faculty perspectives of music teacher education curricular reform. *International Journal of Education & the Arts*, 26(23). <https://doi.org/10.26209/ijea26n23>
- Young, M. M. (2013). University-level group piano instruction and professional musicians. *Music Education Research*, 15(1), 59–73. <https://doi.org/10.1080/14613808.2012.737773>
- Zhang, L.-X., & Leung, B.-W. (2023). Context matters: Adaptation of student-centred education in China school music classrooms. *Music Education Research*, 25(4), 418–434. <https://doi.org/10.1080/14613808.2023.2230587>
- Zhang, L.-X., & Leung, B.-W. (2025). Defining music demonstration lessons: A unique performance-based lesson type improving teachers' instructional skills in Chinese mainland education. *British Journal of Music Education*, 42(1), 65–79. <https://doi.org/10.1017/S0265051724000214>
- Zhang, L.-X., Yan, Z., & Wang, X. (2025). Understanding music teachers' formative assessment intention and implementation: A Chinese mainland context. *International Journal of Music Education*, 43(1), 44–60. <https://doi.org/10.1177/02557614251315461>
- Zhang, L.-X., Qiu, Y., Zou, Y., & Yan, Z. (2026). From policy to practice: A longitudinal interview study of experienced music teachers on implementing formative assessment in China. *Music Education Research*. Advance online publication. <https://doi.org/10.1080/14613808.2026.2615436>

Received: March 18, 2026

Revised: April 28, 2026

Accepted: June 18, 2026

Cite as: Xu, M., Li, P., Zhou, B., & Yi, X. (2026). Making performance teachable: Curricular alignment and microteaching in group piano teacher education. *Problems of Education in the 21st Century*, 84(3), 505–524. <https://doi.org/10.33225/pec/26.84.505>

Min Xu (Corresponding author)	PhD, Senior Lecturer, School of Music, Jiangxi Normal University, Ziyang Avenue 99, Nanchang, Jiangxi 330022, P. R. China E-mail: xumin@jxnu.edu.cn ORCID: https://orcid.org/0009-0003-5183-3635
Pan Li	PhD, Associate Professor, School of Music, Jiangxi Normal University, Ziyang Avenue 99, Nanchang, Jiangxi 330022, P. R. China E-mail: 004937@jxnu.edu.cn ORCID: https://orcid.org/0009-0009-2514-2893
Baiyun Zhou	PhD, Senior Lecturer, School of Music, Jiangxi Normal University, Ziyang Avenue 99, Nanchang, Jiangxi 330022, P. R. China E-mail: 005845@jxnu.edu.cn ORCID: https://orcid.org/0009-0007-4715-9392
Xuan Yi	PhD, Senior Lecturer, School of Music, Jiangxi Normal University, Ziyang Avenue 99, Nanchang, Jiangxi 330022, P. R. China E-mail: 005856@jxnu.edu.cn ORCID: https://orcid.org/0009-0007-9387-2308