

Evaluation of the Bachelor of Music Education Degree Programme (PISMP) at Institut Pendidikan Guru Malaysia (IPGM): A CIPP Model Analysis

Asyira Ismail¹, Mohd Isha Awang²

¹School of Education and Modern Languages, College of Arts and Sciences (CAS), Universiti Utara Malaysia, Sintok, Kedah, Malaysia

²School of Education and Modern Languages, College of Arts and Sciences (CAS), Universiti Utara Malaysia, Sintok, Kedah, Malaysia

HOW TO CITE:

Asyira Ismail, Mohd Isha Awang. Evaluation of the Bachelor of Music Education Degree Programme (PISMP) at Institut Pendidikan Guru Malaysia (IPGM): A CIPP Model Analysis
International Journal of Special Education, 41(20s), 1770-1785

ABSTRACT

This study evaluates the implementation of the Bachelor of Teaching (PISMP) Music Education programme at Institut Pendidikan Guru Malaysia (IPGM) using Stufflebeam and Shinkfield's (2007) Context, Input, Process, Product (CIPP) model. Although IPGM plays a strategic role in producing specialist primary school music teachers, systematic, model-based evaluation of the Music Education curriculum has remained markedly limited relative to core academic programmes. An explanatory sequential mixed-methods design was employed, drawing on a census sample of 84 trainee teachers (from a population of 90) across five IPGM campuses, supplemented by semi-structured interviews with three purposively selected trainees. Quantitative data were analysed descriptively and inferentially using Pearson correlation and independent-samples t-tests, while qualitative data were analysed thematically following Braun and Clarke (2006). Findings indicate consistently favourable perceptions across all four CIPP dimensions, with the strongest ratings recorded for curriculum content suitability ($M = 4.35$) and comparatively lower, though still acceptable, ratings for process-level implementation ($M = 4.05$), particularly music equipment and individual practice-room adequacy. Curriculum suitability correlated significantly and moderately with trainee quality ($r = .551, p < .01$), as did the adequacy of learning support ($r = .554, p < .01$), and these quantitative patterns were corroborated by the qualitative interview evidence. The discussion situates these findings within the broader literature on teacher preparation and music pedagogy, arguing that the interdependence among the four CIPP dimensions offers a more complete explanatory account than any single dimension considered alone. The study contributes an empirically grounded, model-based evaluation that is currently scarce in Malaysian music teacher education and proposes targeted, evidence-based recommendations for continuous quality improvement (CQI) of the PISMP Music Education curriculum.

Keywords: programme evaluation; CIPP model; music teacher education; curriculum evaluation; Institut Pendidikan Guru Malaysia

COPYRIGHT STATEMENT:

Copyright: © 2026 Authors.
Open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Teacher education occupies a foundational position within any national education system, since the quality of classroom instruction that a country is ultimately able to deliver is bounded by the quality of the teachers it prepares (Darling-Hammond, 2006). This premise has been repeatedly reaffirmed in comparative education research: nations that have achieved sustained gains in student learning outcomes are, almost without exception, those that have invested deliberately and continuously in the structure, content, and delivery of their teacher preparation programmes (Darling-Hammond, 2017). In Malaysia, this responsibility is discharged principally through Institut Pendidikan Guru Malaysia (IPGM), the national teacher training institution mandated to produce primary and secondary school teachers in line with the aspirations set out in the Malaysia Education Blueprint 2013–2025. IPGM operationalises this mandate through the Bachelor of Teaching (Program Ijazah Sarjana Muda Perguruan, PISMP), a structured, MQF-aligned undergraduate degree offered across specialised subject options, one of which is Music Education.

Music education, as a field of teacher preparation, presents a distinctive set of pedagogical demands that distinguish it from core academic subjects. It is simultaneously cognitive, psychomotor, and affective in its scope, requiring trainee teachers to develop theoretical mastery of musical concepts, practical instrumental and vocal competence, and the pedagogical sensitivity needed to nurture creativity and expression in young learners (Hallam, 2010; Hallam et al., 2017). Music education is also widely understood to serve broader developmental functions beyond the subject itself, contributing to learners' cognitive development, social integration, and personal confidence (Hallam, 2010). Preparing teachers who are capable of realising these outcomes in the classroom is therefore a demanding undertaking, one that depends not merely on the transmission of musical knowledge but on an integrated curriculum that connects theory, practical skill, professional pedagogy, and reflective practice (Bautista et al., 2017a; Bautista et al., 2017b).

Within the Malaysian policy landscape specifically, this demand has become more pronounced rather than less. National curricular reform under the Malaysia Education Blueprint has repeatedly emphasised the cultivation of twenty-first-century competencies, creativity, critical thinking, and digital literacy chief among them, across all subject areas, including the arts (Kementerian Pendidikan Malaysia, 2013). For a subject such as Music Education, this policy emphasis translates into an expectation that trainee teachers not only master musical content but also become confident, technologically literate, and reflective practitioners capable of adapting their teaching to increasingly diverse primary school contexts. Meeting this expectation places a correspondingly heavier burden on the design and delivery of the PISMP Music Education curriculum, since any shortfall in curriculum content, learning resources, or pedagogical delivery is liable to be felt directly in trainees' eventual classroom readiness. It is precisely this chain, from curriculum design through resourcing and delivery to eventual trainee readiness, that a systematic, model-based evaluation is best positioned to examine, and it is this chain that the present study sets out to trace empirically.

Despite this complexity, curriculum evaluation efforts within Malaysian teacher education have historically concentrated on core academic subjects, such as Mathematics, Science, and the national and international languages, where student outcomes are more directly and visibly linked to national examination performance. Specialist and co-curricular programmes, including Music Education, have by comparison received considerably less systematic evaluative attention. This is a consequential gap. Where curriculum evaluation is absent or fragmented, decisions about curriculum revision, resource allocation, and staff development tend to be driven by anecdotal impression rather than by a defensible evidence base, and opportunities for continuous quality improvement (CQI) are correspondingly diminished (Fitzpatrick, Sanders, & Worthen, 2011). The absence of a holistic, model-based evaluation of the PISMP Music Education programme at IPGM is therefore not

simply an academic oversight; it represents a genuine constraint on the institution's capacity to make informed, defensible decisions about how the programme should evolve.

This study addresses that gap directly. It undertakes a systematic evaluation of the PISMP Music Education programme at IPGM using the CIPP evaluation model developed by Stufflebeam and refined in Stufflebeam and Shinkfield (2007), a framework selected specifically for its capacity to examine a programme not merely in terms of its outcomes, but across the full arc of its rationale, resourcing, implementation, and results. Framed through this model, the study is organised around four specific research problems. First, it examines whether the content of the Music Education curriculum is suitably aligned with the demands of primary school music teaching and with national curricular policy, and whether this alignment is associated with the quality of trainee teachers who complete the programme. Second, it examines the adequacy of the learning environment that supports the curriculum, including physical facilities, digital infrastructure, and reference resources, and whether this adequacy, in turn, is related to trainee outcomes. Third, it examines how the curriculum is actually delivered in practice, including the extent to which teaching and learning incorporate twenty-first-century pedagogy, continuous assessment, and professional practice standards. Fourth, it examines the resulting product of the programme: the level of theoretical, practical, and pedagogical mastery achieved by trainee teachers, and their overall satisfaction with the learning experience the programme has provided.

In pursuing these four lines of enquiry, the study makes several distinct contributions. Empirically, it provides one of the first systematic, model-based evaluations of a specialist music teacher education programme within the Malaysian IPGM system, addressing a gap that has persisted despite the growing international literature on music teacher preparation (Bautista et al., 2017b; Christophersen et al., 2023). Methodologically, it demonstrates the value of an explanatory sequential mixed-methods design for evaluating a practice-based, creative teacher education curriculum, in which quantitative patterns are

elaborated and contextualised by qualitative interview evidence rather than treated as sufficient in isolation. Practically, the study is oriented explicitly towards continuous quality improvement: rather than producing a purely descriptive account of the programme, it is designed to surface specific, evidence-based points of intervention that IPGM and the Ministry of Education can act upon. The remainder of this article proceeds as follows. The next section situates the study within the broader literature on programme evaluation and music teacher education, justifies the selection of the CIPP model over competing evaluation frameworks, and sets out the conceptual framework guiding the analysis. This is followed by a description of the study's research design, sample, instrumentation, and analytical procedures. The findings are then presented and discussed in turn, before the article concludes with a synthesis of the study's contributions, its practical and policy implications, its limitations, and directions for future research.

2. Literature Review and Framework of Analysis

The evaluation of educational programmes has a long and contested history, and the framework chosen to guide any given evaluation carries substantive implications for what that evaluation is capable of revealing. One of the earliest and most influential contributions to this field is Tyler's (1949) objectives-based model, which conceives of evaluation primarily as the process of determining the extent to which pre-specified behavioural objectives have been achieved. Tyler's model established evaluation as a legitimate and systematic component of curriculum work, and its emphasis on clearly defined objectives continues to inform curriculum design practice today. However, precisely because it is anchored so tightly to pre-specified outcomes, the Tylerian model is comparatively insensitive to the context in which a programme operates and to the process through which those outcomes are, or are not, achieved. For a creative and practice-based discipline such as music education, where much of what is valuable about a trainee teacher's development, professional confidence, pedagogical adaptability, artistic sensibility, is not

easily reducible to a discrete behavioural objective, this limitation is especially consequential.

Subsequent evaluation models sought, in different ways, to address this limitation. Kirkpatrick's (1998) four-level model, comprising reaction, learning, behaviour, and results, was developed originally within corporate training contexts and has since been widely adapted for educational evaluation. Its principal strength lies in its clarity and its attention to the transfer of learning into subsequent behaviour, but it was designed with comparatively short, discrete training interventions in mind, and it offers limited conceptual purchase on the structure and process of an extended, multi-year academic degree programme such as the PISMP. Stake's (1967) Countenance model, by contrast, explicitly broadens the evaluative gaze beyond outcomes to include the antecedents and transactions, that is, the context and process, that shape a programme's results, and in doing so anticipates much of the logic that would later be formalised in the CIPP model. Its responsive, judgement-based orientation, however, depends heavily on the evaluator's qualitative interpretation of programme description and standards, which, while valuable for context-rich case evaluation, constrains its capacity to support the systematic, comparable, decision-oriented evaluation that an institution such as IPGM requires across multiple campuses and cohorts.

Other frameworks address still narrower evaluative concerns. The Modified Budgeting System (MBS), for instance, foregrounds questions of financial efficiency and resource allocation, but has little to say about pedagogical quality or trainee development, the very dimensions most central to teacher education. Bell's systems model and the CIRO framework (Context, Input, Reaction, Outcome), meanwhile, were developed principally within industrial and organisational training contexts, and their emphasis on participant reaction and immediate organisational outcomes sits uneasily with the longer-term, developmental character of initial teacher preparation. Taken together, this comparative reading of the evaluation literature points towards a consistent conclusion: models that isolate a single evaluative concern, whether

objectives, reaction, efficiency, or outcome, are structurally unable to capture the multidimensional character of a teacher education curriculum, in which the suitability of what is taught, the adequacy of what supports its teaching, the quality of how it is actually delivered, and the results it eventually produces are all analytically distinct but empirically interdependent.

It is against this backdrop that the CIPP model, developed by Stufflebeam and most comprehensively articulated in Stufflebeam and Shinkfield (2007), was selected as the guiding framework for the present study. Unlike outcome-only or reaction-only models, CIPP evaluates a programme across four interlocking dimensions: Context, which examines the rationale for the programme and the needs it is intended to address; Input, which examines the resources, strategies, and plans mobilised in service of the programme; Process, which examines the actual implementation of the programme, including the fidelity and quality of its delivery; and Product, which examines the outcomes the programme produces, both intended and unintended. This structure is deliberately decision-oriented rather than merely descriptive: Stufflebeam's explicit purpose in developing the model was to support ongoing programme improvement, not simply to render a summative verdict on programme success or failure (Fitzpatrick, Sanders, & Worthen, 2011, as reflected more broadly in Stufflebeam's CIPP tradition). For an evaluation whose ultimate purpose is to inform continuous quality improvement of the PISMP Music Education curriculum, rather than to produce a one-off judgement of adequacy, this decision-oriented logic makes CIPP substantially more fit for purpose than any of the alternative models considered above.

The suitability of the CIPP model for the present study is reinforced by its demonstrated applicability to teacher education and creative-discipline curricula in prior research. Studies applying CIPP to teacher training programmes have consistently found that context-level curriculum design and input-level resourcing exert measurable influence on process-level

implementation, which in turn shapes product-level trainee outcomes, a sequential logic that aligns closely with the research problems articulated in the introduction to this article. At the same time, the literature on music teacher education specifically underscores why each of the four CIPP dimensions warrants close attention in this particular context. On the Context dimension, curriculum alignment with national policy and with the practical demands of primary classroom teaching has been identified as a persistent challenge in specialist teacher training, given the tension between broad national curricular frameworks and the highly specific pedagogical skills that music teaching requires (Schipper, 2010; Schippers & Grant, 2016). On the Input dimension, the adequacy of practice facilities, instruments, and increasingly, digital and ICT infrastructure, has been repeatedly identified as a determinant of trainee confidence and skill development in music education (Partti & Karlsen, 2010), a finding with direct relevance given the growing expectation that trainee teachers graduate with competence in music technology and digitally mediated pedagogy (Gaunt & Westerlund, 2013).

On the Process dimension, the literature on twenty-first-century pedagogy in teacher education emphasises the importance of moving beyond didactic transmission towards reflective, practice-embedded learning, in which trainee teachers are supported to develop what Schön (1983) termed reflection-in-action and what Shulman (1987) characterised as pedagogical content knowledge, the integrated capacity to translate subject expertise into teachable form. In music education specifically, this process-level concern extends to the integration of continuous assessment, ensemble-based collaborative learning, and professional practicum experience, each of which has been linked in prior research to stronger trainee outcomes (Karlsen, 2010; Hargreaves et al., 2011; Holdhus et al., 2021). On the Product dimension, finally, the literature converges on two related outcome constructs: the trainee's demonstrated mastery of theoretical, practical, and pedagogical knowledge, and the trainee's subjective satisfaction with the learning experience, both of which have been shown to

correlate with subsequent professional confidence and retention in the teaching profession (Sadler, 2010; Boud & Dawson, 2021).

Notwithstanding this substantial body of literature on the individual components of music teacher education, systematic, model-based evaluation that examines these four dimensions together, as an interdependent system rather than as discrete concerns, remains conspicuously underdeveloped within the Malaysian context specifically. Existing Malaysian studies of teacher training evaluation have tended either to concentrate on core academic subject programmes or to examine isolated aspects of music pedagogy, such as creativity or assessment practice, without situating these within a comprehensive evaluative framework. This constitutes the specific research gap that the present study is designed to address: the absence, within Malaysian music teacher education, of a systematic, empirically grounded evaluation that examines curriculum content, learning resources, implementation quality, and trainee outcomes as an integrated whole. Addressing this gap is not merely of academic interest. Given IPGM's explicit institutional commitment to continuous quality improvement of its teacher training curricula (IPGM, 2017), a rigorous, model-based evaluation of this kind is a necessary precondition for curriculum decisions that are genuinely evidence-based rather than impressionistic.

This gap is thrown into sharper relief by developments in the international music teacher education literature over the past decade, which has increasingly called for exactly the kind of integrated, system-level evaluation that the Malaysian literature currently lacks. Christophersen et al. (2023), reflecting on the changes needed in music teacher education internationally, argue that meaningful reform depends on understanding how curriculum, resourcing, and delivery interact as a connected system rather than as separate domains to be addressed in isolation. Bautista et al. (2017b), reviewing the literature on music teacher professional development, reach a related conclusion, noting that the evidence base on individual professional-development components is rarely synthesised into a coherent account

capable of guiding curriculum reform. The present study is designed explicitly to close this integration gap for the Malaysian case: rather than treating curriculum content, learning resources, delivery quality, and trainee outcomes as separate lines of enquiry, it examines them jointly, within a single evaluative framework, in a manner that is intended to be of direct practical use to IPGM's own continuous quality improvement processes.

The conceptual framework guiding the present study, presented in Figure 1, operationalises the four CIPP dimensions specifically for the PISMP Music Education context, translating the general

logic of the model into the specific research problems set out in the introduction. The Context dimension is operationalised as the suitability and alignment of curriculum content with national policy and classroom demands; the Input dimension as the adequacy of physical facilities, digital infrastructure, and reference resources; the Process dimension as the quality of teaching and learning delivery, including the integration of twenty-first-century pedagogy and continuous assessment; and the Product dimension as trainees' resulting mastery of knowledge and skill and their overall satisfaction with the programme.

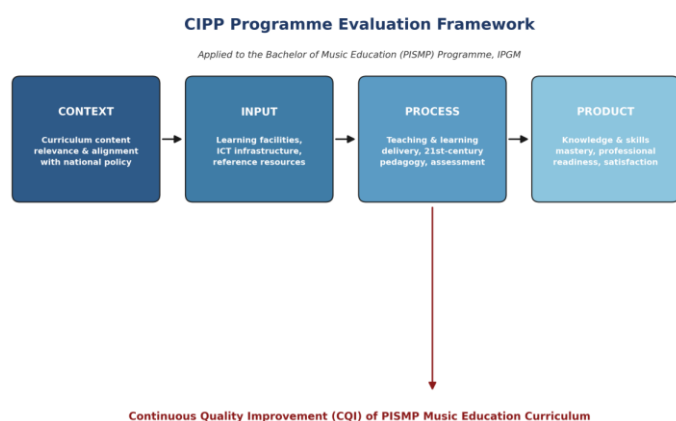


Figure 1. Conceptual Framework: CIPP Evaluation Model Applied to the PISMP Music Education Programme

Source: Adapted from Stufflebeam and Shinkfield (2007).

As Figure 1 illustrates, the four dimensions are conceived not as isolated evaluative targets but as sequentially and functionally interdependent: the suitability of the curriculum's Context shapes the demands placed on Input provision, the adequacy of Input in turn conditions the quality of Process-level implementation, and the cumulative effect of Context, Input, and Process is ultimately realised in Product-level outcomes. This interdependence carries a direct methodological implication that is reflected throughout the remainder of this article: the four dimensions are examined individually in order to establish their respective empirical status, but they are also, and just as importantly, examined in relation to one another, so that a comparatively weaker finding in one dimension can be interpreted in light of its plausible

downstream or upstream connection to the others. The feedback loop returning to curriculum improvement in Figure 1 further reflects the explicitly formative, decision-oriented purpose of the CIPP model: the findings generated across all four dimensions are intended, ultimately, to feed directly into continuous quality improvement of the programme, consistent with the broader quality assurance obligations that apply to Malaysian higher education programmes under the Malaysian Qualifications Framework.

3. Methodology of the Study

In order to examine the four CIPP dimensions in the interdependent manner set out above, this study adopts an explanatory sequential mixed-methods design, following the typology described by Creswell and Poth (2018) and Ivankova et al. (2006). This design was selected deliberately over either a purely quantitative or purely qualitative

alternative. A purely quantitative survey design would have been capable of establishing the statistical patterns of trainees' perceptions across the four CIPP dimensions, but would have offered little insight into why those patterns take the shape they do, an explanatory gap that is especially consequential for a creative, practice-based discipline such as music education, where numerical ratings alone can obscure important variation in trainees' lived experience. Conversely, a purely qualitative design would have provided richer contextual understanding but would have lacked the breadth and generalisability that a census-based quantitative survey affords across the full population of trainees. The explanatory sequential design addresses both limitations directly: quantitative data collection and analysis are conducted first, in order to establish the overall pattern of trainees' perceptions across the CIPP dimensions, and this is followed by a subsequent qualitative phase, in which semi-structured interviews are used to explain, elaborate, and contextualise the quantitative patterns that emerge. This sequencing is not incidental; rather, it directly enables the triangulated interpretation that the conceptual framework in Section 2 requires, since a purely numerical finding, such as a comparatively lower mean score for Process-level implementation, gains genuine explanatory depth only once it can be connected to the specific experiences trainees describe in their own words.

The population for this study comprised all trainee teachers enrolled in the PISMP Music Education programme across the IPGM campuses at which the specialisation is offered, totalling 90 individuals. Given the small, clearly bounded, and institutionally accessible nature of

this population, a census approach was adopted rather than a probability sampling strategy, consistent with Creswell and Poth's (2018) observation that census sampling is appropriate where a population is limited and directly relevant to the evaluative focus of the study. Of the 90 trainees in the population, 84 returned complete and usable questionnaires, a response rate that comfortably exceeds the minimum sample size of 73 recommended by Krejcie and Morgan (1971) for a population of this size at a 95 per cent confidence level and a five per cent margin of error, and which therefore provides a statistically defensible basis for the descriptive and inferential analyses reported in Section 4.

As shown in Figure 2, the achieved sample was predominantly female ($n = 64$, 76 per cent) relative to male trainees ($n = 20$, 24 per cent), a gender composition broadly consistent with teacher training programmes in Malaysia more generally, and one that mirrors the international pattern of female-dominated enrolment in primary teacher education (UNESCO, 2022). Respondents were drawn from five campuses at which the PISMP Music Education specialisation is offered, with the largest concentrations recorded at IPG Kampus Perlis ($n = 37$, 44 per cent) and IPG Kampus Perempuan Melayu Melaka ($n = 35$, 41 per cent), and considerably smaller cohorts at IPG Kampus Tuanku Bainun, IPG Kampus Ipoh, and IPG Kampus Ilmu Khas ($n = 4$ each, 5 per cent). This uneven distribution reflects the genuine, uneven national offering of the Music Education specialisation across IPGM campuses, rather than any artefact of the sampling procedure, and it is treated explicitly as a boundary condition when the generalisability of the study's findings is discussed in Section 5.

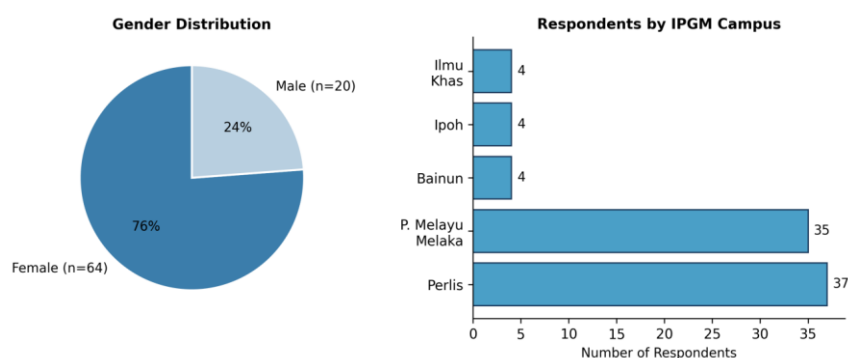


Figure 2. Respondent Profile by Gender and IPGM Campus (N = 84)

Source: Fieldwork data collected for this study (N = 84).

Two research instruments were employed to generate the quantitative and qualitative data respectively. The primary quantitative instrument was a structured questionnaire comprising items distributed across the four CIPP dimensions and rated on a five-point Likert scale, adapted and refined through expert content validation prior to administration. This was complemented by a semi-structured interview protocol administered to three purposively selected trainee teachers, chosen specifically for their active, sustained engagement with the curriculum across their period of study. Although three informants is a comparatively small number in absolute terms, this figure is consistent with established qualitative sampling guidance that emphasises information power and thematic saturation over

numerical size (Patton, 1990; Guest, Bunce, & Johnson, 2006), particularly where, as here, the informant group is deliberately homogeneous in relevant experience and the analytic purpose is explanatory elaboration of an existing quantitative pattern rather than the generation of new, freestanding theory.

The internal consistency of the questionnaire was assessed prior to the main study using Cronbach's alpha, computed separately for each of the four CIPP dimensions. As reported in Table 1, all four coefficients exceeded the widely accepted threshold of .70 (Nunnally & Bernstein, 1994), ranging from .884 for the Context dimension to .975 for the Product dimension, indicating strong internal consistency across the instrument as a whole and providing an adequate empirical basis for proceeding to the main data collection.

Table 1. Internal Consistency (Cronbach's Alpha) of the Study Instrument by CIPP Dimension

Dimension	Construct(s)	Items	Cronbach's α
Context (K)	Curriculum content suitability	A01–A06	.884
Input (I)	Facilities, learning support, resources	B01–B38	.954
Process (Pr)	Teaching quality, pedagogy, assessment	C01–C35	.973
Product (Pd)	Skills mastery and trainee satisfaction	D01–D28	.975

Source: Reliability analysis conducted for this study using SPSS.

Several procedural safeguards were built into the design to strengthen the trustworthiness of the resulting data. Formal permission to conduct the study was obtained from the relevant education planning and research authorities and from each participating IPGM campus prior to fieldwork, and informed consent was obtained individually from all questionnaire respondents and interview informants. The questionnaire itself was subject to expert content validation prior to the pilot stage, in which item wording, response format, and dimension coverage were reviewed and refined based on expert feedback before reliability testing was conducted, consistent with established practice in instrument development (Nunnally & Bernstein, 1994). For the qualitative strand, trustworthiness was supported through methodological triangulation between the questionnaire and interview data, member-checking of interview transcripts with informants

prior to analysis, and an audit trail of the coding process, in line with recommended practice for thematic analysis (Braun & Clarke, 2006). Taken together, these procedures were intended to ensure that the quantitative and qualitative strands of the study could be integrated with confidence in Section 4, where their combined evidentiary weight, rather than either strand considered alone, forms the basis for the study's central claims.

Once collected, quantitative data were analysed in SPSS, drawing on descriptive statistics, means, standard deviations, and frequency distributions, to characterise trainees' perceptions across all four dimensions, and on inferential statistics, specifically Pearson product-moment correlation and independent-samples t-tests, to examine relationships between dimension-level constructs and overall trainee quality, following the interpretive benchmarks for correlational effect size proposed by Cohen (1988). For each of the

four CIPP dimensions, an overall dimension-level mean was computed as the unweighted average of the item-level means within that dimension's constituent construct(s), consistent with standard practice in Likert-scale survey reporting; these composite means are reported in Section 4.1 and visualised in Figure 3, while the underlying item-level means and standard deviations are reported in the corresponding tables (e.g., Table 2 for the Context dimension). Qualitative interview data, meanwhile, were transcribed verbatim and analysed thematically following the six-phase inductive procedure described by Braun and Clarke (2006), in which codes are generated from the data itself rather than imposed from a pre-existing framework, before being clustered into broader themes and interpreted in relation to the corresponding CIPP dimension. This analytic pairing, descriptive and inferential statistics on one hand, inductive thematic analysis on the other, is what ultimately enables the triangulated, explanatory account of programme evaluation that the study's mixed-methods design is intended

to produce, and it is this account that is presented and discussed in the following section.

4. Findings and Discussions

4.1 Findings of the Study

Figure 3 presents an overview of trainees' mean perception scores across the four CIPP dimensions, each computed as the unweighted average of its constituent item-level means (see Section 3 for the computation procedure). All four means fell within the high-interpretation band conventionally applied to five-point Likert data, yet a discernible gradient is evident across dimensions: Context recorded the highest mean ($M = 4.35$), followed by Product ($M = 4.22$), Input ($M = 4.10$), and Process ($M = 4.05$). This overall pattern, curriculum design rated somewhat more favourably than its day-to-day implementation, recurs at the item level across each of the four dimensions and forms the organising thread for the remainder of this subsection.

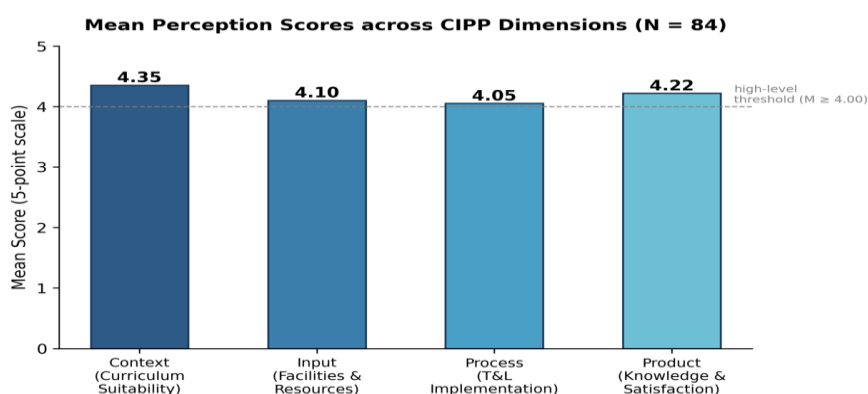


Figure 3. Mean Perception Scores across the Four CIPP Dimensions (N = 84)

Source: Fieldwork data collected for this study (N = 84).

At the Context level, item-level means ranged from 4.13 to 4.44, with the curriculum's delivery of theoretical knowledge ($M = 4.44$, $SD = 0.61$) and its development of practical problem-solving skill ($M = 4.44$, $SD = 0.63$) rated most favourably, as detailed in Table 2. The lowest-rated Context item, entrepreneurial-skills development ($M = 4.13$, $SD = 0.80$), remained within the high-

interpretation band but exhibited the greatest response variability of any item in the dimension, suggesting uneven trainee exposure to entrepreneurship-related content across campuses and cohorts. Curriculum suitability, taken as a composite construct, correlated significantly and moderately with trainee quality ($r = .551$, $p < .01$), a relationship whose strength is consistent with Cohen's (1988) benchmark for a moderate effect.

Table 2. Item-Level Mean Scores for Curriculum Content Suitability (Context Dimension, N =

84)

Curriculum Item	M	SD
Delivery of theoretical knowledge and principles	4.44	0.608
Development of practical problem-solving skill	4.44	0.628
Integration of digital and communication skills	4.32	0.662
Interpersonal skills and professional ethics	4.39	0.621
Entrepreneurial-skills development	4.13	0.803
Professionalism, leadership, and teamwork	4.40	0.608

Source: Fieldwork data collected for this study (N = 84).

At the Input level, structural provisions were generally well rated, including room-scheduling arrangements (M = 4.19, SD = 0.84) and classroom size (M = 4.21, SD = 0.89). The lowest-rated items within the dimension concerned the adequacy of music equipment for trainee needs (M = 3.55, SD = 1.31) and the provision of individual practice rooms (M = 3.57, SD = 1.24), with high-speed internet access in lecture rooms rated only slightly higher (M = 3.75, SD = 1.15) and also carrying a comparatively large standard deviation. This pattern indicates that equipment and practice-space adequacy, rather than digital infrastructure specifically, constitute the most consistently under-resourced aspect of the Input dimension, with ICT provision

representing a secondary, though still notable, area of variability across campuses. Adequacy of learning support overall correlated significantly with trainee quality ($r = .554$, $p < .01$), a relationship of comparable strength to that observed for the Context dimension.

At the Process level, the overall dimension mean of 4.05 was the lowest of the four CIPP dimensions. Inferential analysis nonetheless confirmed statistically significant relationships between trainee quality and each of curriculum implementation, learning facilities, continuous assessment practice, and overall process quality, as detailed in Table 3, indicating that, notwithstanding its comparatively modest absolute rating, process-level implementation nonetheless exerts a measurable and statistically significant influence on trainee outcomes.

Table 3. Inferential Relationships Between Process-Level Implementation and Trainee Quality (N = 84)

Relationship with Trainee Quality	t	df	p
Curriculum implementation	2.12	82	.037
Learning facilities	3.02	82	.003
Continuous assessment practice	3.92	82	< .001
Overall process quality	4.22	82	< .001

Source: Independent-samples t-test results reported for this study (N = 84).

As shown in Table 3, all four relationships reached statistical significance at the $p < .05$ level, with continuous assessment practice and overall process quality showing the strongest associations with trainee quality ($p < .001$ in both cases). This pattern indicates that even the most weakly rated component of the Process dimension, curriculum implementation, still exerts a statistically meaningful influence on trainee outcomes, reinforcing the interpretation that process-level

delivery is a substantive rather than peripheral determinant of programme quality.

At the Product level, the dimension mean recovered to 4.22, the second-highest of the four dimensions. Thematic analysis of the interview data, conducted following Braun and Clarke's (2006) inductive procedure, identified three recurring themes among the three trainee informants, summarised in Table 4. Across all three informants, trainees consistently described feeling adequately prepared for classroom practice, a qualitative pattern that corroborates,

and adds explanatory texture to, the dimension mean reported above. comparatively strong quantitative Product-

Table 4. Qualitative Themes Identified in Product-Dimension Interview Data (N = 3 Informants)

Theme	Description
Satisfaction with learning experience and perceived programme quality	Informants described an overall positive learning experience and a favourable perception of the programme's quality across their period of study.
Confidence in professional knowledge and pedagogical skill	Informants reported growing confidence in their theoretical grounding and pedagogical competence, linking this directly to their sense of readiness for classroom teaching.
Perceived relevance of curriculum content to the local primary school context	Informants viewed the curriculum content as directly applicable to the specific demands of teaching music in local primary school classrooms.

Source: Thematic analysis of interview data conducted for this study, following Braun and Clarke's (2006) inductive procedure.

As Table 4 shows, the three themes trace a coherent arc from general satisfaction, through growing professional and pedagogical confidence, to a concrete sense of the curriculum's applicability in the primary classroom. Read alongside the quantitative Product-dimension mean, this qualitative pattern reinforces the interpretation that trainees leave the programme not only rating it favourably in the abstract, but with a specific, articulable sense of what the programme has equipped them to do.

4.2 Discussion of the Findings

Read together, the findings reported in Section 4.1 support a coherent interpretation that is consistent with, and indeed anticipated by, the conceptual framework set out in Section 2. The PISMP Music Education curriculum appears to be well designed at the Context level: trainees rate its theoretical and practical content highly, and this rating is significantly associated with trainee quality, echoing Darling-Hammond's (2017) broader observation that well-structured curriculum content is a necessary, though not sufficient, condition for effective teacher preparation. However, the full translation of this well-designed Context into Process-level delivery appears to be constrained by uneven Input-level provision, most conspicuously in respect of music equipment and individual practice-space adequacy. This pattern is consistent with international research on music teacher education,

which has repeatedly identified inadequate practice facilities and instrumental resources as a persistent constraint on the realisation of otherwise well-designed teacher preparation curricula (Partti & Karlsen, 2010; Gaunt & Westerlund, 2013), and it extends this line of literature to the specific, and previously under-examined, context of Malaysian primary music teacher education.

The comparatively lower Process-dimension mean observed in this study warrants particular attention, not because it falls outside the high-interpretation band, but because of what it plausibly signals about the sequential relationship between Input and Process posited in the conceptual framework. Where music equipment and practice-space provision are constrained, as the Input-level findings suggest they are for at least a subset of trainees, their downstream effect on Process-level delivery, for instance, on the feasibility of hands-on instrumental practice or on the logistics of continuous assessment, offers a plausible explanatory mechanism for the pattern observed. This interpretation is reinforced by the inferential findings reported in Section 4.1: even though the Process dimension recorded the lowest overall mean, its constituent relationships with trainee quality were nonetheless all statistically significant, indicating that process-level implementation is not a peripheral concern but a substantive determinant of programme outcomes, consistent with Shulman's (1987) argument that pedagogical content knowledge, the capacity to translate subject expertise into teachable classroom practice, is central rather than

incidental to teacher effectiveness.

At the same time, the Product-dimension findings suggest that this constraint has not yet substantially eroded trainee outcomes. Trainees continue to report strong theoretical and practical knowledge mastery and high overall satisfaction, and the qualitative interview data corroborate this pattern in trainees' own descriptions of feeling prepared for classroom practice. One plausible explanation, consistent with the sequential logic of the conceptual framework, is that the comparatively strong Context-level foundation, evident in the high curriculum-suitability ratings reported in Section 4.1, is currently sufficient to offset the more moderate Process-level conditions. This should not, however, be read as grounds for complacency. The moderate correlations reported for both the Context dimension ($r = .551$) and the Input dimension ($r = .554$) indicate that curriculum design and learning-support adequacy each account for a meaningful, though partial, share of variance in trainee quality; sustained Input-level constraints, if left unaddressed, could plausibly erode Product-level outcomes over time even where they have not yet done so, a risk consistent with longitudinal findings elsewhere in the teacher education literature regarding the cumulative effects of infrastructural under-provision (Sadler, 2010).

These findings also carry more specific implications when interpreted against the demographic and institutional profile of the sample. Because the achieved sample was drawn from five IPGM campuses of markedly unequal size, three of which contributed only four respondents each, the aggregate Input-dimension findings likely mask meaningful between-campus variation that the present sample size is insufficiently powered to test statistically. It is nonetheless a reasonable, literature-consistent inference that campuses with smaller Music Education cohorts may be more exposed to infrastructural strain than the two larger campuses, since institutional investment in specialist facilities is often driven by enrolment scale. This constitutes a specific, evidence-based target for future institutional attention, one considerably more precise than a generic call for infrastructure investment, and it is a

recommendation to which the study returns in the concluding section.

It is also worth situating these findings within the broader tradition of CIPP-based evaluation research from which the study's conceptual framework is drawn. A recurring finding across CIPP evaluations of teacher training programmes internationally is that Context and Product dimensions tend to be rated more favourably than Input and Process, precisely the asymmetry observed in this study, and this recurring pattern has been interpreted in the literature as evidence that curriculum designers are generally more successful at articulating what a programme should achieve than institutions are at consistently resourcing and delivering it in practice (Fitzpatrick et al., 2011, drawing on the broader CIPP evaluation tradition). Read in this light, the findings reported here are not an isolated institutional weakness specific to IPGM, but an instance of a pattern that appears to recur across teacher training evaluation more generally, which in turn suggests that the practical remedy is unlikely to lie in redesigning the curriculum's Context dimension, already comparatively strong, but rather in strengthening the institutional mechanisms, budgetary, logistical, and supervisory, that translate a well-designed curriculum into consistently well-resourced and well-delivered classroom practice. This distinction matters for how IPGM prioritises its continuous quality improvement efforts: it argues for investment directed at Input and Process, rather than further Context-level curriculum revision, as the more efficient lever for improving Product-level outcomes going forward.

Finally, the entrepreneurial-skills item, the single lowest-rated item across the entire questionnaire, merits discussion in its own right. Entrepreneurship has become an increasingly prominent expectation within Malaysian higher education curricula generally, reflecting national policy emphasis on graduate employability and self-reliance, yet its integration into a specialist, arts-based curriculum such as Music Education is conceptually less straightforward than its integration into, for instance, a business or technical curriculum. The comparatively low and highly variable rating recorded for this item

suggests that its current curricular treatment may not yet be well matched to the specific professional pathway that Music Education trainees are being prepared for, namely, classroom teaching rather than independent entrepreneurial practice. This finding, considered alongside the broader Context-dimension results, suggests that curriculum refinement efforts might usefully focus on reframing entrepreneurial content in terms more directly relevant to a music teacher's actual professional context, for example, initiating and sustaining school-based musical activities and community engagement, rather than treating it as a generic, subject-independent competency.

5. Conclusion

This study set out to address a specific and consequential gap in Malaysian teacher education research: the absence of a systematic, model-based evaluation of the PISMP Music Education programme at Institut Pendidikan Guru Malaysia, a specialist curriculum that has historically received considerably less evaluative attention than core academic programmes. Using Stufflebeam and Shinkfield's (2007) CIPP model as its guiding framework, and an explanatory sequential mixed-methods design as its methodological vehicle, the study has generated an empirically grounded account of the programme's Context, Input, Process, and Product dimensions, and, importantly, of the relationships among them.

The study's central finding is that trainees' perceptions of the programme are consistently positive but not uniform across dimensions: curriculum content suitability is rated most favourably, music equipment and individual practice-space adequacy constitute a specific and identifiable point of strain within an otherwise adequately resourced learning environment, process-level implementation, while rated comparatively lower, nonetheless exerts a statistically significant influence on trainee quality, and product-level outcomes, knowledge mastery and trainee satisfaction, remain strong despite these implementation-level constraints. Taken together, these findings support the broader theoretical proposition, embedded in the study's conceptual framework, that the four CIPP dimensions function as an interdependent system

rather than as four separate, independently interpretable evaluative concerns; a full account of programme quality requires attention to the relationships between dimensions, not merely to their individual scores.

These findings carry direct and, in several respects, actionable implications for both policy and institutional practice. At the policy level, the specific, campus-linked nature of the Input-dimension findings suggests that IPGM and the Ministry of Education would be better served by targeted investment in music equipment and individual practice-room provision, informed by campus-level need, rather than by uniform, institution-wide resource allocation that risks under-serving the smaller Music Education cohorts identified in this study. At the level of curriculum design, the entrepreneurial-skills finding points towards a specific, tractable revision: reframing this component of the curriculum around music-teaching-relevant applications of entrepreneurship, such as school-based musical programming and community music engagement, rather than treating it as a generic transferable skill disconnected from trainees' actual professional trajectory. At the level of programme delivery, the significant relationships identified between trainee quality and each of curriculum implementation, facilities, and continuous assessment practice indicate that process-level quality assurance, classroom observation, peer review of practicum supervision, and structured feedback on assessment practice, warrants sustained institutional attention even where, as here, it does not yet register as the programme's weakest dimension in absolute terms.

Methodologically, the study demonstrates the value of an explanatory sequential mixed-methods CIPP evaluation for specialist, practice-based teacher education programmes more broadly, illustrating how quantitative dimension-level patterns and qualitative thematic evidence can be combined to produce an interpretation that neither data source could sustain on its own. This methodological demonstration may be of use to other IPGM campuses, and indeed to other specialist teacher training programmes within the Malaysian system, seeking to conduct comparable

evaluations of their own curricula.

The study is nonetheless subject to several limitations that should temper the generalisability of its findings. First, its cross-sectional design captures trainees' perceptions at a single point in time and cannot speak to how these perceptions, or the underlying programme conditions that shape them, may change over the course of trainees' study or after their graduation into full-time teaching. Second, the study is restricted to currently enrolled trainees and therefore excludes any assessment of graduates' subsequent, in-service performance in actual primary school classrooms, a product-level outcome of arguably even greater ultimate significance than trainees' self-reported mastery and satisfaction while still enrolled. Third, although the census sampling strategy achieved a strong response rate relative to the accessible population, the marked imbalance in campus-level representation, with three of five campuses contributing only four respondents each, limits the statistical power available for any disaggregated, campus-level analysis, even though such analysis would be of considerable practical value given the Input-dimension findings discussed above.

Future research should accordingly pursue several extensions of the present study. A logical next step would be to broaden data collection across a wider and more evenly distributed range of IPGM campuses, nationally, in order to test directly the campus-level infrastructural variation that the present study can only infer. A longitudinal design that follows trainees from enrolment through to their first years of in-service teaching would allow researchers to examine whether the Product-level outcomes documented here, knowledge mastery and satisfaction while still in training, translate into sustained classroom effectiveness once trainees enter full-time practice. Finally, future studies might usefully examine mediating variables not directly addressed here, such as trainee self-efficacy, mentor and cooperating-teacher support during practicum, and institutional leadership

commitment to specialist curricula, each of which plausibly moderates the Input–Process–Product relationships identified in this study and would deepen understanding of precisely how curriculum design is, or is not, translated into effective classroom teaching. In sum, while this study establishes that the PISMP Music Education programme at IPGM is, on the evidence available, functioning soundly across all four CIPP dimensions, it simultaneously identifies specific, addressable points of strain, principally in music equipment and practice-space provision and in the curricular framing of entrepreneurial skill, that offer IPGM a concrete, evidence-based agenda for the continuous quality improvement of Malaysian music teacher education going forward.

More broadly, this study contributes to a growing recognition, evident across the international literature reviewed in Section 2, that specialist and arts-based teacher education programmes require the same rigour of systematic evaluation traditionally reserved for core academic subjects, not merely as a matter of institutional parity, but because the developmental and pedagogical complexity of a field such as music education arguably demands more, rather than less, evaluative attention. As Malaysia continues to pursue the twenty-first-century competency agenda articulated in its national education policy, the capacity of institutions such as IPGM to evaluate, and thereby continuously improve, the specialist curricula that sit alongside core academic programmes will be an increasingly consequential determinant of the quality of teaching that Malaysian classrooms, in all subjects, ultimately receive. It is in this spirit that the present study offers its findings, not as a final verdict on the PISMP Music Education programme, but as one deliberately evidence-based contribution to an evaluative conversation that IPGM and the wider Malaysian teacher education community are encouraged to continue.

References

The following references are cited in this article and were independently verified against their original publication records.

Bautista, A., Toh, G. Z., Mancenido, Z., & Wong, J. (2017a). Curriculum implementation in music education: A

-
- review. *Music Education Research*, 19(3), 245–260.
- Bautista, A., Yau, X., & Wong, J. (2017b). High-quality music teacher professional development: A review of the literature. *Music Education Research*, 19(4), 455–469.
- Boud, D., & Dawson, P. (2021). What feedback literate teachers do: An empirically-derived competency framework. *Assessment & Evaluation in Higher Education*, 46(2), 158–171.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101.
- Christophersen, C., Aróstegui, J.-L., Holdhus, K., Kenny, A., Knudsen, J. S., Lindgren, M., Väkevä, L., & Grieg Viig, T. (2023). Music teacher education for the future: Reflections on change. *Action, Criticism & Theory for Music Education*, 22(3), 7–40.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). SAGE Publications.
- Darling-Hammond, L. (2006). Constructing 21st century teacher education. *Journal of Teacher Education*, 57(3), 300–314.
- Darling-Hammond, L. (2017). Teacher education around the world: What can we learn from international practice? *European Journal of Teacher Education*, 40(3), 291–309.
- Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2011). *Program evaluation: Alternative approaches and practical guidelines* (4th ed.). Pearson.
- Gaunt, H., & Westerlund, H. (Eds.). (2013). *Collaborative learning in higher music education*. Routledge.
- Guest, G., Bunce, A., & Johnson, L. (2006). How many interviews are enough? An experiment with data saturation and variability. *Field Methods*, 18(1), 59–82.
- Hallam, S. (2010). The power of music: Its impact on the intellectual, social and personal development of children and young people. *International Journal of Music Education*, 28(3), 269–289.
- Hallam, S., Creech, A., & Varvarigou, M. (2017). Well-being and music leisure activity through the lifespan: A psychological perspective. In R. Mantie & G. D. Smith (Eds.), *Oxford handbook of music making and leisure* (pp. 31–60). Oxford University Press.
- Hargreaves, D., Miell, D., & MacDonald, R. (Eds.). (2011). *Musical imaginations: Multidisciplinary perspectives on creativity, performance, and perception*. Oxford University Press.
- Holdhus, K., Murphy, R., & Espeland, M. I. (Eds.). (2021). *Music education as craft: Reframing theories and practices*. Springer.
- Institut Pendidikan Guru Malaysia. (2017). *Pelan strategik transformasi IPG 2016–2025*. IPGM.
- Ivankova, N. V., Creswell, J. W., & Stick, S. L. (2006). Using mixed-methods sequential explanatory design: From theory to practice. *Field Methods*, 18(1), 3–20.
- Karlsen, S. (2010). BoomTown Music Education and the need for authenticity: Informal learning put into practice in Swedish post-compulsory music education. *British Journal of Music Education*, 27(1), 35–46.
- Kementerian Pendidikan Malaysia. (2013). *Malaysia Education Blueprint 2013–2025 (Preschool to post-secondary education)*. Ministry of Education Malaysia.
- Kirkpatrick, D. L. (1998). *Evaluating training programs: The four levels* (2nd ed.). Berrett-Koehler Publishers.
- Krejcie, R. V., & Morgan, D. W. (1971). Determining sample size for research activities. *Educational and Psychological Measurement*, 30(3), 607–610.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.
- Partti, H., & Karlsen, S. (2010). Reconceptualising musical learning: New media, identity and community in music education. *Music Education Research*, 12(4), 369–382.
- Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed.). SAGE Publications.
-

-
- Sadler, D. R. (2010). Beyond feedback: Developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education*, 35(5), 535–550.
- Schippers, H. (2010). *Facing the music: Shaping music education from a global perspective*. Oxford University Press.
- Schippers, H., & Grant, C. (Eds.). (2016). *Sustainable futures for music cultures: An ecological perspective*. Oxford University Press.
- Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. Basic Books.
- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57(1), 1–23.
- Stake, R. E. (1967). The countenance of educational evaluation. *Teachers College Record*, 68(7), 523–540.
- Stufflebeam, D. L., & Shinkfield, A. J. (2007). *Evaluation theory, models, and applications*. Jossey-Bass.
- Tyler, R. W. (1949). *Basic principles of curriculum and instruction*. University of Chicago Press.
- UNESCO. (2022). *Gender equality in and through the teaching profession*. UNESCO.
- .