

A CONCEPTUAL MODEL OF AUTHENTIC ASSESSMENT IN MALAYSIAN PRIMARY MUSIC EDUCATION

Joshua Raj Pitchemuthu ¹

Mohd Hassan Abdullah ²

¹ Faculty of Music and Performing Arts, Universiti Pendidikan Sultan Idris, 35900 Tanjong Malim, Perak, Malaysia (E-mail: p20231000033@siswa.upsi.edu.my)

² Faculty of Music and Performing Arts, Universiti Pendidikan Sultan Idris, 35900 Tanjong Malim, Perak, Malaysia (Corresponding Author E-mail: mohd@upsi.edu.my)

Article history

Received date : 15-8-2026

Revised date : 16-8-2026

Accepted date : 15-10-2026

Published date : 20-10-2026

To cite this document:

Pitchemuthu, J. R., & Abdullah, M. H. (2026). A conceptual model of authentic assessment in Malaysian primary music education. *Journal of Islamic, Social, Economics and Development (JISED)*, 11 (87), 186 – 205.

Abstract: *Authentic assessment in primary music education requires more than the inclusion of performance tasks in classroom assessment. It requires a coherent assessment architecture that connects what pupils are expected to demonstrate, how evidence is generated and captured, what constitutes quality, and how teachers make defensible judgements. This article presents and critically examines a conceptual model of authentic assessment for Malaysian primary music education, grounded in evidence generated through a Design and Development Research (DDR) Type 2 study. The study comprised three phases: needs analysis involving the Year 1–6 Music Education Curriculum and Assessment Standard Documents (DSKP), six daily lesson plans, structured interviews with 10 teachers and a 30-teacher pilot; design and development involving 130 teachers, comparative curriculum analysis across Ireland, Germany, New Zealand, Canada, Australia and Japan, and expert review; expert review and content validation involving eight experts; and Phase III usability evaluation involving six experts using the Fuzzy Delphi Method (FDM). The analysis identified four interdependent assessment components—performance or assessment strategy, assessment criteria, assessment tools, and rubrics—mapped across four primary Music Education modules: Music Language, Musical Skills, Music Creativity and Music Appreciation. Document analysis recorded relative frequencies of 16 for assessment strategy, 18 for criteria, 13 for tools and 4 explicit references to rubrics, indicating an uneven specification of how assessment evidence and quality are operationalised. Content validation yielded an S-CVI/Ave of 0.985, while all 32 framework items met the predefined FDM acceptance criteria. The 11-item usability assessment recorded an overall mean of 4.65 (SD = 0.31). The proposed model conceptualises authentic assessment as an evidence-alignment system rather than a collection of assessment instruments. Its contribution lies in connecting curriculum intentions with authentic musical performance, explicit criteria, fit-for-purpose evidence tools and rubric-supported professional judgement, while incorporating formative feedback, reflection and teacher professional judgement. The model provides a context-sensitive conceptual basis for strengthening authentic assessment practice in Malaysian primary Music Education.*

Keywords: *authentic assessment; performance assessment; music education; assessment criteria; assessment tools; rubrics; Malaysia*

Introduction

Assessment in primary music education presents a distinctive challenge because musical learning involves cognitive, psychomotor, affective, social and expressive dimensions. Pupils demonstrate musical understanding through singing, instrumental performance, movement, listening, musical creation, collaboration and communication of musical ideas. Assessment approaches that rely predominantly on written or decontextualised tasks may therefore provide only a partial representation of pupils' musical knowledge, skills and development. Authentic assessment is particularly relevant because it emphasises meaningful tasks through which learners apply knowledge and skills, demonstrate performance and generate evidence of learning within relevant contexts (Koh, 2017; Villarroel et al., 2018).

In music education, however, authentic assessment should not be equated simply with performance. A musical performance becomes meaningful assessment evidence when the intended learning, task demands, assessment criteria, evidence collection and judgement processes are appropriately aligned. Contemporary perspectives on authenticity also emphasise learner participation, agency and meaning-making, as well as the relationship between assessment and subsequent learning (Ajjawi et al., 2024; McArthur, 2023). This is particularly important in music because the quality of learning may involve observable technical features alongside less easily standardised dimensions such as creativity, interpretation and expression. Assessment therefore requires sufficient structure to support transparent judgement while retaining space for musical individuality.

The relationship among assessment criteria, assessment tools and rubrics is central to this process. Criteria establish the qualities expected in a performance or learning outcome, while assessment tools capture and document relevant evidence. Rubrics organise criteria into descriptions of performance quality and can support transparency and consistency in professional judgement. These components are interdependent rather than interchangeable. Weakly specified criteria may lead to inconsistent judgements, inappropriate tools may fail to capture important aspects of musical learning, and overly restrictive rubrics may reduce complex musical performances to mechanical indicators (Dawson, 2017; Villarroel et al., 2018). Authentic assessment consequently requires coherence among curriculum intentions, musical performance, criteria, evidence and judgement.

This issue is particularly relevant to Malaysian primary Music Education, which is implemented through the Primary School Standard Curriculum (Revised 2017) within a broader classroom-based assessment system that emphasises continuous evidence collection and teacher professional judgement (Kementerian Pendidikan Malaysia, 2019). The curriculum encompasses four areas (Music Language, Musical Skills, Music Creativity and Music Appreciation), each requiring assessment approaches suited to different forms of musical knowledge, performance and expression. However, curriculum expectations alone do not fully specify how musical activities should be systematically connected with assessment criteria, evidence-collection tools and rubric-supported judgement.

Evidence from the present development study indicates that assessment elements are represented in the curriculum documents examined, although their explicit specification varies. The document analysis recorded frequencies of 16 for assessment strategy, 18 for assessment criteria, 13 for assessment tools and 4 for explicit references to rubrics. These figures describe the coded references identified in the materials analysed and should not be interpreted as direct measures of teachers' classroom practices. Together with teacher evidence indicating

substantial reliance on continuous observation and practical performance, the findings suggest a need to examine how assessment activities can be connected more systematically to the generation, documentation and interpretation of evidence of musical learning. The issue is therefore not simply whether assessment takes place, but whether its different elements function coherently as an assessment system.

Research on authentic assessment and musical creativity highlights the educational importance of formative feedback, reflection and flexible criteria that support meaningful musical learning without imposing unnecessarily narrow definitions of quality (Bolden & DeLuca, 2022; Bolden et al., 2020, 2023). These perspectives reinforce the need to balance transparent assessment expectations with the flexibility required to accommodate pupils' different musical responses, creative processes and levels of development.

Despite these contributions, the functional relationship among authentic performance, assessment criteria, evidence-collection tools and rubrics remains insufficiently articulated for Malaysian primary Music Education. Existing scholarship offers important principles for authentic assessment, formative assessment, creativity assessment and rubric design, but these principles do not, by themselves, establish how the elements can be integrated into a coherent, curriculum-responsive model for this context. In particular, a clearer account is needed of how curriculum intentions can guide the selection of meaningful musical tasks, the identification of relevant criteria, the collection of appropriate evidence and the interpretation of performance through rubric-supported professional judgement.

Accordingly, this article proposes a conceptual model of authentic assessment for Malaysian primary Music Education, developed through a Design and Development Research (DDR) Type 2 study. The model integrates four interrelated components—performance or assessment strategy, assessment criteria, assessment tools and rubrics—across the four curriculum modules: Music Language, Musical Skills, Music Creativity and Music Appreciation. Its central contribution is to establish a curriculum-responsive structure for organising authentic assessment within Malaysian primary Music Education.

Literature Review

Authentic Assessment in Primary Music Education

Authentic assessment emphasises meaningful tasks through which learners apply knowledge and skills, demonstrate understanding and engage in reflection and judgement. Koh (2017) highlights meaningful performance and higher-order thinking, while Villarroel et al. (2018) emphasise realism, cognitive challenge and evaluative judgement in assessment design. More recent perspectives extend authenticity beyond the format of an assessment task to include learner participation, agency and meaningful engagement with learning (Ajjawi et al., 2024; McArthur, 2023).

In primary Music Education, authentic assessment may involve singing, instrumental performance, ensemble work, musical creation, listening and interpretation. However, a performance task is not automatically authentic merely because pupils demonstrate musical skills. Its value depends on whether the activity reflects the intended learning outcomes and generates relevant evidence for assessment. Teachers therefore need to connect curriculum expectations with appropriate tasks, criteria and evidence of learning. This is particularly important in music, where technical accuracy interacts with creativity, expression and

individual interpretation. Authentic assessment consequently requires a balance between meaningful musical activity and transparent judgement.

Formative Assessment and Musical Creativity

Formative assessment is important in music learning because pupils develop through practice, experimentation, feedback and revision. Bolden and DeLuca (2022) highlight the role of explicit criteria, context-sensitive teaching and self-assessment in supporting musical creativity. Similarly, Bolden et al. (2020) identify criteria, reflection and self-assessment as important considerations in creativity assessment, while Bolden et al. (2023) emphasise flexible success criteria and supportive feedback. Booth (2022) also illustrates how formative assessment can support pupils' progression in group composition.

These studies suggest that assessment should support learning during musical activity rather than merely determine achievement at the end of a task. For example, feedback during a composition activity may help pupils reconsider rhythmic organisation, coordination or expressive intent before presenting their work. Nevertheless, feedback is useful only when it relates to the intended learning and provides opportunities for improvement. Assessment criteria must therefore be sufficiently clear to guide pupils while remaining flexible enough to accommodate different creative responses. This balance is essential in primary music classrooms, where pupils may demonstrate learning through varied musical ideas and levels of experience.

Assessment Evidence and the Selection of Tools

Authentic assessment requires appropriate evidence to support meaningful and defensible judgements. In Music Education, evidence may include live performance, classroom observation, audio or video recordings, portfolios, project work, self-assessment and reflective accounts. Each source serves a different purpose. Observation can capture pupils' responses during musical activity, recordings preserve performances for subsequent review, and portfolios can document development over time. Reflective accounts may also provide insight into pupils' musical decisions and understanding.

Schultz et al. (2022) discuss the relationship between authentic assessment and intended learning outcomes, while Ketonen et al. (2025) examine authentic assessment as a support for reflection among student teachers. Although these studies are not specifically situated in Malaysian primary Music Education, they offer relevant perspectives on aligning assessment purposes, evidence and reflection. A checklist, for instance, may document specified behaviours but may not adequately capture musical expression or creativity. Similarly, a recording provides evidence of a performance but does not independently establish its quality.

The selection of tools should therefore be guided by the evidence required, rather than by the availability of particular instruments. Although multiple sources may provide a more comprehensive account of learning, excessive documentation can increase teachers' workload without necessarily improving assessment quality. Effective assessment requires sufficient, relevant and manageable evidence that can be interpreted in relation to the intended learning outcomes.

Assessment Criteria and Rubrics in Music Education

Assessment criteria identify the qualities against which evidence of learning is interpreted. Depending on the task, these may include pitch accuracy, rhythmic control, tempo, articulation,

tone production, ensemble coordination, musical expression, originality and interpretation. Criteria need to reflect the purpose of the activity rather than assume that every dimension of musical learning is equally important in every task.

Research on creativity assessment highlights the difficulty of establishing criteria that are both explicit and flexible. Broad criteria may provide insufficient guidance for consistent judgement, whereas excessively narrow indicators may restrict creative expression (Bolden et al., 2020; Bolden et al., 2023). For example, a rhythmic composition may require coherent organisation and purposeful musical choices while allowing pupils to experiment with different patterns. The assessment should recognise the quality of pupils' musical decisions rather than require identical responses.

Rubrics can support this process by describing levels or dimensions of performance quality. Dawson (2017) emphasises the importance of clarity in rubric design and the need to consider how rubrics are used in practice. In music, overly fragmented descriptors may fail to represent the relationship between technical, expressive and creative qualities, while vague descriptions may provide insufficient guidance. Rubrics should therefore clarify expectations and support teacher professional judgement rather than replace it. Their usefulness depends on the alignment of criteria, performance evidence and descriptions of quality.

Feedback, Reflection and Musical Individuality

Authentic assessment should connect judgement with feedback, reflection and subsequent learning. Wake et al. (2024) indicate that authentic assessment can be perceived as relevant and engaging, but may also create uncertainty when learners are unfamiliar with the approach. Clear expectations, modelling and appropriate feedback can help learners understand assessment requirements and respond constructively.

In music classrooms, pupils may need repeated opportunities to practise, listen, evaluate and revise their work. Self-assessment and peer feedback can encourage them to consider musical quality and explain their decisions, provided that these activities are guided by relevant criteria. Assessment should also recognise that consistency does not require every pupil to produce the same musical response. Instead, criteria should establish common expectations while allowing appropriate variation in creativity, interpretation and expression.

This creates a continuing design challenge: assessment must be sufficiently structured to support transparent judgement without reducing musical learning to mechanical scoring. Professional judgement remains important because the qualities of musical performance may not be fully represented by fixed indicators. The alignment of criteria, evidence, rubrics and feedback can help teachers make more defensible judgements while supporting pupils' individual musical development.

Synthesis of the Literature and Research Gap

The literature highlights the importance of authentic assessment, formative feedback, musical creativity, explicit assessment criteria and rubric-supported professional judgement in Music Education (Bolden & DeLuca, 2022; Bolden et al., 2020, 2023; Dawson, 2017; Villarroel et al., 2018). However, these elements are often discussed separately, and their functional relationships within a coherent framework for Malaysian primary Music Education remain insufficiently articulated. The specific research gap concerns the need for a context-specific conceptual framework that systematically integrates four essential components: performance or

assessment strategy, assessment criteria, assessment tools and rubrics. Without such integration, the relationship between curriculum expectations, the generation of musical evidence, the interpretation of performance quality and the provision of feedback may remain unclear. This need is particularly relevant across the four Music Education modules, namely Music Language, Musical Skills, Music Creativity and Music Appreciation, which involve different forms of musical knowledge, performance and expression. Therefore, the present study addresses this gap by developing a conceptual model that integrates these four components within an iterative process of professional judgement, feedback, reflection and improvement to support coherent and evidence-informed authentic assessment in Malaysian primary Music Education.

Conceptual Foundation of the Model

The model is based on the alignment of curriculum intentions, musical activities and the evidence required to evaluate pupils' learning. It was developed through Design and Development Research (DDR) Type 2, drawing on curriculum document analysis, teacher evidence, comparative curriculum analysis and expert evaluation. The model was developed through Design and Development Research (DDR) Type 2, drawing on curriculum document analysis, teacher evidence, comparative curriculum analysis and expert evaluation.

The model comprises four interrelated components: performance or assessment strategy, assessment criteria, assessment tools and rubrics. These components are applied across the four Malaysian primary Music Education modules: Music Language, Musical Skills, Music Creativity and Music Appreciation. The following sections explain the specific function of each component in translating curriculum expectations into assessable musical learning.

Curriculum Intent and Musical Learning Domains

Curriculum intent establishes the purpose and direction of assessment by identifying the knowledge, skills and understanding pupils are expected to demonstrate. The four curriculum modules represent different but related dimensions of musical learning. Music Language focuses on musical knowledge, notation and literacy. Musical Skills emphasises performance, technique and musical execution. Music Creativity involves musical exploration, experimentation, composition and the development of musical ideas. Music Appreciation focuses on listening, interpretation, musical response and reflection.

These modules require different forms of assessment evidence. For example, pupils may demonstrate Music Language through identifying musical symbols or applying notation, while Musical Skills may be assessed through singing, instrumental performance or ensemble activities. Music Creativity may require evidence of both the creative process and the resulting musical product, whereas Music Appreciation may involve listening responses, interpretation and reflection. Consequently, the model does not assume that a single assessment strategy or tool is equally suitable for every module. Instead, curriculum intent guides the selection of appropriate musical tasks, assessment criteria and evidence collection methods.

Performance or Assessment Strategy

Performance or assessment strategy provides opportunities for pupils to demonstrate their learning through meaningful musical activities. The model incorporates approaches such as practical performance, continuous observation, formative assessment, summative assessment, portfolio assessment, self-assessment, peer assessment and project-based assessment. These approaches provide different opportunities to observe musical development and collect

evidence of learning. However, performance alone does not guarantee authentic assessment. Each activity must be aligned with the intended curriculum outcomes and supported by appropriate assessment criteria and evidence collection procedures. Within the model, performance or assessment strategy therefore serves as the means through which intended musical learning becomes observable and can be evaluated systematically.

Assessment Criteria and Dimensions of Musical Quality

Assessment criteria identify the qualities used to interpret pupils' musical learning. Depending on the intended outcome, relevant criteria may include pitch accuracy, rhythmic control, tempo, articulation, tone production, ensemble coordination, musical expression, creativity, notation, interpretation and reflection. The selection of criteria should be guided by the curriculum module and the purpose of the musical activity.

The model positions assessment criteria as a link between curriculum expectations and the interpretation of evidence. Criteria should be sufficiently clear to guide teacher judgement and communicate expectations to pupils. At the same time, they should allow appropriate flexibility for creativity, interpretation and individual musical expression. For example, a composition task may require rhythmic coherence and purposeful musical organisation without requiring every pupil to produce the same musical response. This approach supports transparent assessment while recognising the varied ways pupils may demonstrate musical understanding.

Assessment Tools and Evidence Collection

Assessment tools provide the means to capture, document and organise evidence generated through musical activities. The model includes observation records, checklists, anecdotal records, performance records, audio or video recordings, portfolios, worksheets, project documentation and reflective records. The choice of tool depends on the type of evidence required to assess the intended learning.

Different tools serve different purposes. Observation records can document pupils' participation and responses during musical activities, while recordings preserve performances for subsequent review. Portfolios can demonstrate development over time, and reflective records can provide information about pupils' musical decisions and learning experiences. However, the availability of a tool does not automatically ensure meaningful assessment. The evidence collected must be relevant to the assessment criteria and sufficient to support professional judgement. The model therefore treats assessment tools as evidence collection mechanisms rather than complete assessment solutions.

Rubrics and Professional Judgement

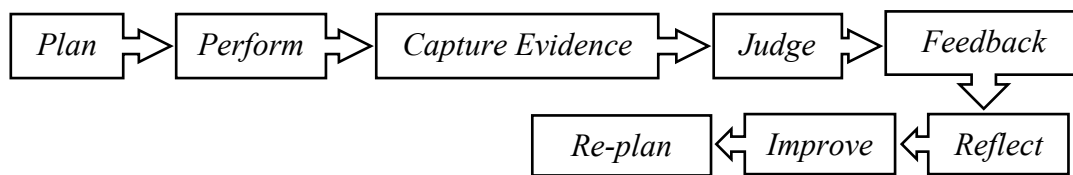
Rubrics provide a structured basis for interpreting evidence against assessment criteria. They may include performance descriptors, developmental levels and descriptions of expected quality. Their purpose is to make assessment expectations clearer and support more transparent and consistent judgements. In the proposed model, rubrics are distinguished from assessment tools because tools capture or document evidence, whereas rubrics guide the interpretation of that evidence.

Professional judgement remains important because musical learning involves technical, expressive, creative and interpretive dimensions that may interact within a single performance. Rubrics should therefore support teachers in evaluating evidence rather than replace their professional interpretation. Descriptors need to be sufficiently clear to guide judgement without

reducing musical quality to mechanical scoring or restricting legitimate creative responses. The relationship between assessment criteria, evidence and rubric descriptors is consequently essential to the coherence of the model.

Iterative Assessment, Feedback and Reflection

The four assessment components operate within an iterative process that connects assessment with continuing musical development. The process can be represented as follows:



During planning, curriculum intentions guide the selection of musical activities and assessment approaches. Pupils then demonstrate their learning through performance or other meaningful musical tasks. Assessment tools capture relevant evidence, which is interpreted against the criteria and supported by rubric-based professional judgement. Feedback communicates strengths and areas for improvement, while reflection helps pupils consider their learning and musical decisions. The resulting information informs further practice, revision and subsequent assessment planning.

This iterative process distinguishes the model from an approach that focuses only on final performance or scores. Evidence can be accumulated throughout the learning process, allowing teachers to consider pupils' progress rather than relying exclusively on a single assessment event. Feedback and reflection also provide opportunities for pupils to improve their work and develop greater awareness of their musical learning.

Overall Structure of the Conceptual Model

The overall structure of the model is designed to maintain alignment between curriculum expectations and the evidence used to evaluate musical learning. Its components operate within an iterative process in which assessment judgements inform feedback, reflection and subsequent teaching and learning activities.

The model provides a curriculum-responsive structure for applying authentic assessment across Music Language, Musical Skills, Music Creativity and Music Appreciation, while retaining professional judgement and flexibility for musical expression. Its application across Music Language, Musical Skills, Music Creativity and Music Appreciation provides a curriculum-responsive basis for assessment in Malaysian primary Music Education. The model is intended to support the systematic collection and interpretation of evidence while preserving teacher professional judgement and the flexibility required to recognise musical creativity, expression and individual development.

Methodology

Research Design

This study employed a Design and Development Research (DDR) Type 2 approach to develop and evaluate an authentic assessment framework for Malaysian primary Music Education. The research comprised three phases: Needs Analysis, Design and Development, and Usability Evaluation. Evidence from each phase informed the subsequent phase, connecting identified assessment needs with framework development, expert consensus and usability evaluation.

Phase I: Needs Analysis

Phase I involved a literature review, analysis of the Music Education Curriculum and Assessment Standard Documents (*DSKP-Dokumen Standard Kurikulum dan Pentaksiran*) for Years 1–6, analysis of six Daily Lesson Plans (*RPH-Rancangan Pengajaran Harian*), structured interviews with 10 primary music teachers, and a pilot study involving 30 teachers. The document analysis examined the representation of assessment strategies, criteria, tools and rubrics. The interviews explored teachers' assessment practices, challenges and perceived needs. The pilot study was used to examine questionnaire clarity and assess the internal consistency of the instrument using Cronbach's alpha. Findings from these sources informed the initial requirements for the framework's four assessment components.

Phase II: Design and Development

Phase II involved a questionnaire administered to 130 primary school music teachers. The responses were analysed using SPSS and descriptive statistics, including frequencies, percentages, means and standard deviations, as appropriate to the data. The results provided evidence about teachers' assessment practices and needs and contributed to the framework's design.

An international curriculum comparison examined assessment practices in Ireland, Australia, New Zealand, Canada, Japan and Germany. The countries were selected purposively based on the availability of relevant literature, curriculum documents and information on music education and assessment, as well as their use in previous comparative research. The comparison considered assessment strategies, criteria, tools and rubrics. International practices were critically examined as sources for contextualised framework development rather than adopted directly.

The findings from the questionnaire, curriculum analysis, teacher interviews and international comparison were synthesised to develop the draft framework. Eight experts then evaluated the draft using expert assessment procedures, including the Fuzzy Delphi Method (FDM) and Content Validity Index (CVI). The FDM assessment examined 32 framework items, while CVI provided evidence concerning content relevance. The draft was refined in response to the quantitative results and expert feedback.

Phase III: Usability Evaluation

Phase III evaluated the usability of the refined framework through a separate panel of six experts. The evaluation instrument comprised 11 items covering the main framework, core modules and assessment components. Experts provided ratings using a five-point Likert scale. Responses were analysed using the Fuzzy Delphi Method (FDM), including Triangular Fuzzy Numbers, threshold values ((d)), percentage of expert agreement and defuzzification scores.

The acceptance criteria were a threshold value of $d \leq 0.20$, expert agreement of at least 75%; and a fuzzy score (A) of at least 0.50. These criteria were applied together to determine whether an item met the predefined consensus requirements. All 11 usability items met the study's FDM acceptance criteria. The overall usability mean was 4.65 ((SD = 0.31)), indicating very high perceived usability among the participating experts.

Content Validity and Data Analysis

The Content Validity Index (CVI) was used to assess the content relevance of the expert evaluation instrument. At the draft-framework validation stage, 8 experts assessed 8 elements representing the 4 Music Education modules. The item-level CVI (I-CVI) ranged from 0.88 to 1.00, and the average scale-level CVI (S-CVI/Ave) was 0.985. The thesis applies an I-CVI acceptance value of at least 0.78 and an S-CVI/Ave value of at least 0.90 as indicators of acceptable content validity.

CVI and FDM served different purposes. CVI provided evidence of content relevance, whereas FDM established expert consensus on the proposed framework items and, in Phase III, the usability items. The Phase II FDM analysis of 32 framework items yielded threshold values ranging from 0.067 to 0.172, expert agreement from 75% to 100%, and fuzzy scores from 0.700 to 0.775. All 32 items met the predefined acceptance criteria. These results are distinct from the Phase III usability results for the 11-item instrument.

Qualitative evidence from the curriculum documents, lesson plans and teacher interviews was examined through content and thematic analysis. Questionnaire data were analysed descriptively using SPSS. Expert comments were considered alongside the CVI and FDM results to inform refinement of the framework.

Model Development and Integration of Evidence

The model was developed through an iterative synthesis of findings across the 3 phases. Phase I identified assessment needs and gaps through literature review, curriculum and lesson plan analysis, teacher interviews and a pilot study. Phase II integrated teacher questionnaire findings, international curriculum comparison and expert review to develop and refine the draft framework. Phase III provided further evidence through content validation, expert consensus and usability evaluation.

The document analysis recorded 16 references to assessment strategies, 18 to assessment criteria, 13 to assessment tools and 4 explicit references to rubrics. These frequencies represent coded references in the materials analysed rather than direct measures of classroom practice. The findings highlighted the need for a more structured relationship among assessment strategies, criteria, tools and rubrics.

The synthesis of curriculum analysis, teacher responses, international comparisons and expert feedback informed the integration of the four assessment components across the four Malaysian primary Music Education modules. The model was subsequently refined through content validation and expert evaluation. It therefore represents an evidence-informed and expert-evaluated conceptual framework. However, its effectiveness in routine classroom implementation and its contribution to pupil learning require further empirical investigation. The model therefore represents an evidence-informed and expert-evaluated conceptual framework. However, its effectiveness in routine classroom implementation and its

contribution to pupil learning require further empirical investigation. Table 1 summarises the evidence base and development process.

Table 1. Evidence Base and Development of the Model

Phase	Evidence sources	Analytical purpose	Contribution to the model
Phase I: Needs Analysis	Literature review; DSKP Years 1–6; six Daily Lesson Plans; interviews with 10 teachers; 30-teacher pilot	Identify assessment practices, needs, gaps and recurring assessment elements	Establish initial requirements for performance, criteria, tools and rubrics
Phase II: Design and Development	Questionnaire involving 130 teachers; comparative analysis of six countries; review of the draft framework by eight experts	Examine teacher evidence, identify relevant international practices, and evaluate and refine the draft framework	Structure and contextualise the four assessment components
Phase III: Content Validation and Usability Evaluation	Six experts; FDM; 11-item usability assessment	Assess expert consensus and perceived usability	Provide further validation and usability evidence for the refined model

Source: Compiled by the author based on the doctoral research, “*Pembangunan Kerangka Pentaksiran Autentik dalam Mata Pelajaran Pendidikan Muzik Sekolah Rendah di Malaysia.*”

Findings and Model Development

Assessment Gaps in the Malaysian Music Education Curriculum

The document analysis revealed that assessment-related elements were represented unevenly across the materials examined. The recorded frequencies were 16 for assessment strategies, 18 for assessment criteria, 13 for assessment tools and four explicit references to rubrics. The relatively low frequency of explicit rubric references should not be interpreted as evidence that assessment standards are absent from the curriculum. Rather, it highlights a potential gap between describing levels of achievement and providing structured descriptors that support consistent judgements of musical performance.

Although the *Dokumen Standard Kurikulum dan Pentaksiran (DSKP)* specifies learning standards and performance levels, teachers may require more operational guidance to translate these expectations into observable evidence of achievement. For example, assessing singing accuracy may involve considering pitch, rhythm, tempo, pronunciation, breathing, tone and expression. The study identified musical criteria including articulation, posture, rhythmic accuracy, dynamics, ensemble coordination, creativity, notation, musical symbols and values. The proposed model organises these criteria into task-relevant dimensions that can be adapted to different musical activities rather than treating every characteristic as an independent scoring item.

Assessment Strategies and Authentic Musical Performance

The second finding concerns the importance of practical performance and continuous observation in primary Music Education. Teacher questionnaire findings indicated that

continuous observation received the highest mean rating ($M = 4.82$), followed by practical performance ($M = 4.79$). The comparative analysis of six international curriculum systems in Ireland, Australia, New Zealand, Canada, Japan and Germany also identified recurring assessment approaches, including performance, observation, formative assessment, portfolios, projects, reflection, and self- and peer assessment.

These findings support an assessment approach that considers both the musical product and the learning process. A single performance may demonstrate achievement at a particular moment, but observation across activities, supported by formative feedback and records of progress, can provide a broader picture of pupils' developing musical competence. Accordingly, the model positions performance within a continuing assessment process rather than treating it solely as a final evaluative event.

Assessment Tools and the Collection of Evidence

The third finding concerns the relationship between assessment tools and the evidence they are intended to capture. The study identified tools and records including observation checklists, anecdotal notes, performance records, audio and video recordings, portfolios, worksheets and reflective documentation. The model organises these tools according to their evidentiary functions. Observation records capture pupils' participation and responses during musical activities; performance records document demonstrations of musical skills; recordings preserve performances for subsequent review; portfolios illustrate development over time; and reflective documentation provides evidence of pupils' interpretations of their learning.

This functional approach emphasises that selecting an assessment tool is not sufficient, in itself, to make an assessment authentic. The tool must be appropriate to the task, the intended learning evidence and the criteria used to interpret that evidence. The model therefore connects assessment strategies, criteria and tools within a coherent structure.

Rubrics Between Criteria and Judgement

The fourth finding highlights the role of rubrics in translating assessment criteria into judgements about the quality of musical performance. The proposed model accommodates analytic rubrics, which distinguish performance dimensions, and holistic judgements, which consider musical performance as an integrated whole. The choice depends on the purpose of the task and the qualities being assessed.

Expert feedback also highlighted the need for clearer terminology, more operationalised elements, stronger attention to musical experience and practical implementation guidance. These considerations indicate that rubric development involves more than formatting assessment information into a table. Rubric descriptors need to communicate expectations clearly, align with the task and support reasoned judgements while allowing appropriate professional judgement and musical expression.

Content Validation and Usability Evaluation

The fifth finding concerns the expert evaluation of the proposed model. The study reported a Scale-Level Content Validity Index based on the average method (S-CVI/Ave) of 0.985. In the Phase II evaluation, all 32 retained items met the study's Fuzzy Delphi Method (FDM) acceptance criteria. The threshold values ranged from 0.067 to 0.172, expert agreement ranged from 75% to 100%, and fuzzy scores ranged from 0.700 to 0.775.

The subsequent Phase III usability evaluation involved six experts and an 11-item assessment. The overall usability mean was 4.65 (SD = 0.31). Framework suitability received a mean of 4.81 (SD = 0.22), content relevance 4.60 (SD = 0.35), and perceived effectiveness 4.54 (SD = 0.39). These findings indicate favourable expert evaluations of the model's suitability, relevance and perceived effectiveness.

Nevertheless, expert consensus, content validity and perceived usability should not be equated with demonstrated improvements in pupils' musical learning. The findings support the model as a content-validated and expert-evaluated framework. Further classroom-based research is needed to examine its implementation, feasibility across different teaching contexts and contribution to pupil learning outcomes.

Conceptualisation of the Model

The findings were synthesised into a conceptual model comprising four interrelated components: performance or assessment strategy, assessment criteria, assessment tools and rubrics. These components are intended to operate across the four Malaysian primary Music Education modules: Music Language, Musical Skills, Music Creativity and Music Appreciation. Curriculum expectations provide the foundation for selecting appropriate tasks, while feedback, reflection and improvement connect the components through an iterative assessment process.

Table 2 illustrates how the four components can be applied across the curriculum modules. The examples are intended to demonstrate the model's structure rather than prescribe a fixed set of assessment tasks or criteria for every classroom situation.

Table 2. Conceptualisation of the Authentic Assessment Model

Curriculum module	Performance evidence	Criteria	Tools	Rubric
Music Language	Reading notation, identifying symbols and explaining musical concepts	Notation accuracy, terminology, application and aural recognition	Worksheets, oral responses, observation checklists and short digital quizzes	Accuracy, understanding and application
Musical Skills	Singing, recorder playing, percussion, movement and ensemble performance	Pitch, rhythm, tempo, tone, technique, posture, coordination and expression	Performance records, checklists, audio or video recordings and anecdotal notes	Technical control, musicality and ensemble interaction
Music Creativity	Improvisation, short composition, arranging and creative response	Originality, coherence, musical decision-making, experimentation and communication	Project portfolios, scores or sketches, recordings and reflection	Creative process, product quality and justification
Music Appreciation	Listening responses, interpretation,	Aural awareness, musical vocabulary,	Listening logs, oral responses, portfolios and	Evidence-based responses,

comparison, discussion and reflection	interpretation, and cultural and contextual understanding	reflective records	interpretation and reflection
---	--	-----------------------	----------------------------------

Source: Compiled by the author based on *the DSKP Pendidikan Muzik Tahun 1–6* and findings from the doctoral research, “*Pembangunan Kerangka Pentaksiran Autentik dalam Mata Pelajaran Pendidikan Muzik Sekolah Rendah di Malaysia.*”

Table 3 clarifies the operational relationship among the four components. Each component addresses a distinct assessment question, but none is sufficient in isolation. Their alignment is central to the model's intended application.

Table 3. Operational Relationship Among Performance, Criteria, Tools and Rubrics

Component	Key question	Operational function	Risk if isolated	Model response
Performance or assessment strategy	How will learning be demonstrated?	Creates authentic opportunities to generate observable evidence	Performance may become a one-off event with limited diagnostic value	Integrate performance with continuous observation, formative feedback and projects
Assessment criteria	What qualities matter?	Defines the dimensions of quality to be assessed	Criteria may be too general or excessively fragmented	Use clear, task-specific and developmentally appropriate criteria
Assessment tools	How will evidence be captured?	Documents performance and learning over time	Tools may become paperwork disconnected from actual evidence	Select tools according to their evidentiary purpose
Rubrics	How will quality be interpreted?	Structures teacher judgement and communicates expectations	Rigid descriptors may reduce musical expression to mechanical scoring	Use meaningful developmental descriptors while retaining professional judgement

Source: Compiled by the author based on *the DSKP Pendidikan Muzik Tahun 1–6* and findings from the doctoral research, “*Pembangunan Kerangka Pentaksiran Autentik dalam Mata Pelajaran Pendidikan Muzik Sekolah Rendah di Malaysia.*”

Table 4 summarises the principal validation and usability results. The Phase II FDM findings relate to the 32 framework items, whereas the usability results were obtained separately during Phase III.

Table 4. Validation and Usability Evidence for the Model

Evidence indicator	Result	Interpretation
Content validity (S-CVI/Ave)	0.985	Strong content-validity evidence for the proposed framework elements
FDM threshold across 32 items (Phase II)	0.067–0.172	All items met the study's threshold criterion of $d \leq 0.20$
Expert agreement across 32 items (Phase II)	75%–100%	All items met the minimum agreement criterion of 75%
Fuzzy score, A (Phase II)	0.700–0.775	All items met the minimum fuzzy-score criterion of $A \geq 0.50$
Framework suitability (Phase III)	M = 4.81, SD = 0.22	Favourable expert evaluation of suitability
Content relevance (Phase III)	M = 4.60, SD = 0.35	Favourable expert evaluation of content relevance
Perceived effectiveness (Phase III)	M = 4.54, SD = 0.39	Favourable expert evaluation of perceived effectiveness
Overall usability (Phase III)	M = 4.65, SD = 0.31	Favourable overall expert evaluation of usability

Source: Author's compilation and analysis of expert evaluation data from the doctoral research, "*Pembangunan Kerangka Pentaksiran Autentik dalam Mata Pelajaran Pendidikan Muzik Sekolah Rendah di Malaysia.*"

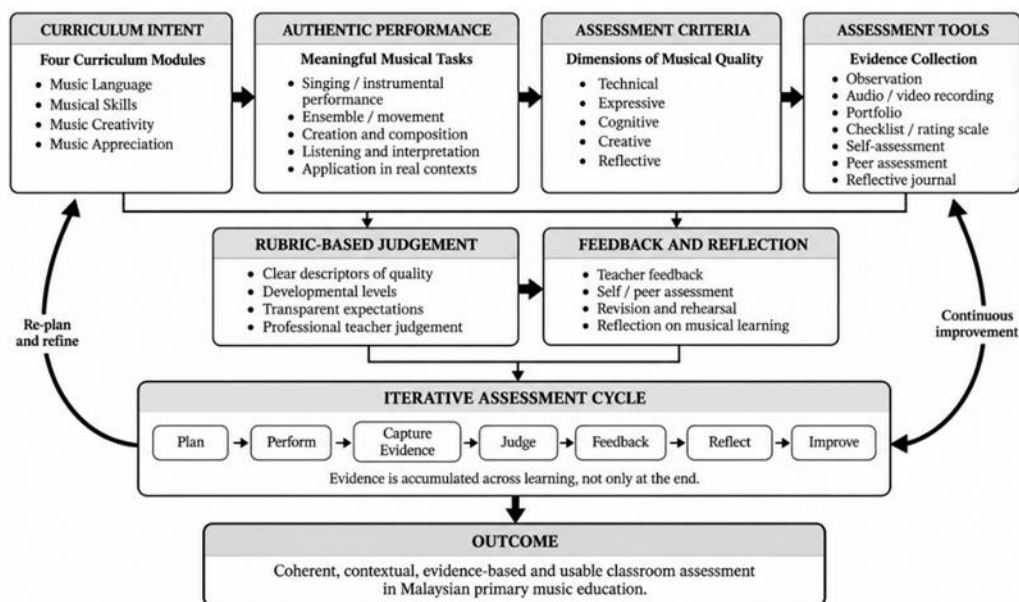


Figure 1. Conceptual Model of Authentic Assessment Integrating Performance, Criteria, Assessment Tools and Rubrics

Source: Developed by the author based on the findings of the doctoral research, "*Pembangunan Kerangka Pentaksiran Autentik dalam Mata Pelajaran Pendidikan Muzik Sekolah Rendah di Malaysia,*" and expert evaluation of the proposed model.

Analysis and Discussion

The principal contribution of the model lies in addressing the imbalance identified in the curriculum document analysis, particularly the limited explicit representation of rubrics compared with assessment strategies, criteria and tools. This finding highlights the need to examine how assessment expectations can be translated into clearer descriptors of musical

performance quality, while recognising that the coded frequencies reflect the documents analysed rather than teachers' classroom practices.

First, the findings highlight the importance of alignment between assessment criteria and rubrics. The curriculum analysis recorded 18 references to assessment criteria but only four explicit references to rubrics. This imbalance suggests that criteria were more visible than structured descriptors for interpreting performance quality within the materials analysed. However, these frequencies represent coded references rather than direct measures of classroom practice. Rubrics are not inherently superior to other assessment approaches; their value depends on their purpose, design and application (Dawson, 2017). Within the proposed model, rubrics support more transparent judgements by clarifying expectations and describing different levels of musical quality.

The findings suggest that practical performance and observation need to be supported by evidence that is relevant to the intended learning outcomes. For example, a singing task may combine observation notes, a recording and pupil reflection to provide complementary evidence of performance and learning progress. The model addresses this issue by connecting performance or assessment strategy with explicit criteria and appropriate assessment tools. For example, a singing task may be assessed through observation notes, audio or video recordings, a checklist and pupil reflection. These sources can provide complementary evidence of performance quality and learning progress. This approach is consistent with Villarroel et al. (2018), who emphasise the importance of realism, cognitive challenge and evaluative judgement in authentic assessment.

Third, the model recognises the need to assess musical creativity without reducing it to conformity. Although explicit criteria are necessary for making judgements, overly prescriptive descriptors may restrict pupils' expressive choices. Bolden et al. (2020) examined the role of criteria in creativity assessment, while Bolden et al. (2023) highlighted the importance of flexible success criteria. Accordingly, assessment criteria for creative musical tasks may address coherence, intentionality, experimentation, musical communication and pupils' explanations of their decisions. These criteria provide a basis for evaluating the quality of musical work while allowing different forms of creative expression.

The findings support collecting evidence across multiple learning activities rather than relying exclusively on a single performance. Observation records, successive recordings and reflective statements can help teachers examine pupils' musical development over time. The assessment tools component therefore supports the collection of evidence appropriate to the task and learning outcome. This approach also reflects practices identified in the international curriculum comparison, including observation, performance recording, portfolio development and reflection. The model consequently encourages teachers to consider a connected set of evidence rather than relying exclusively on an isolated assessment event.

The model positions rubrics as aids to professional judgement rather than substitutes for teachers' interpretation of musical performance. Teachers must consider the relevance and context of evidence, particularly when evaluating creativity and expression (Ajjawi et al., 2024). Teachers therefore need to interpret descriptors in relation to pupils' learning stages, musical experiences and classroom contexts. Discussion of performance exemplars and collaborative interpretation of criteria may help strengthen consistency while preserving appropriate professional discretion.

Sixth, the model reflects the importance of contextual adaptation. The comparison of curriculum and assessment practices in Ireland, Germany, New Zealand, Canada, Australia and Japan provided an international reference for identifying relevant approaches. However, these practices cannot simply be transferred to Malaysian primary schools without considering differences in curriculum requirements, resources, teacher preparation and educational contexts. The proposed model was developed through the synthesis of international practices with Malaysian curriculum documents, teacher perspectives and expert input. Its relevance to Malaysia therefore rests on this contextual integration rather than the direct adoption of a model from another education system.

The model also addresses the relationship between consistency and musical individuality. Shared criteria can support clearer expectations and more transparent assessment decisions, but excessive prescription may limit creativity and expression. A balanced approach distinguishes essential indicators of musical quality from the artistic choices pupils can make independently. For example, a composition task may require rhythmic coherence and purposeful organisation while allowing pupils to select their own sounds, patterns and expressive approaches. Rubrics should therefore communicate expectations without requiring every pupil to produce an identical musical outcome.

Finally, the findings must be interpreted within the scope of the study. The content validity result ($S-CVI/Ave = 0.985$) and overall usability rating ($M = 4.65$, $SD = 0.31$) provide evidence of expert agreement on content validity and perceived usability. They do not establish that the model improves pupils' musical achievement, enhances creativity or reduces teachers' workload. Furthermore, the study's data collection in Selangor limits the extent to which its findings can be generalised to all Malaysian primary schools without further investigation. Implementation may also be affected by class size, teaching time, access to recording equipment and the demands of documenting individual pupil progress.

Future research should therefore examine the model through classroom-based implementation across diverse school settings. Such research should investigate its practicality, teacher workload, pupils' assessment experiences, consistency of professional judgement and potential contribution to musical learning. Overall, the model offers a structured yet flexible framework for connecting performance or assessment strategy, assessment criteria, assessment tools and rubrics within an iterative process of feedback, reflection and improvement. Its contribution lies in making the relationships among these components more explicit while recognising the contextual and creative nature of primary Music Education.

Limitations and Future Research

The first limitation is geographical and contextual. The teacher survey was situated in Selangor, and the model reflects the Malaysian primary school context represented by the study. It should not be assumed that all states, school types or classroom conditions will implement the model identically. Replication across states and school contexts is required.

Second, the model was validated through expert consensus and usability assessment rather than a longitudinal implementation study. The high usability mean and FDM acceptance provide important evidence, but they do not establish causal effects on pupil learning. Future research should use classroom-based implementation designs to examine changes in musical performance, creativity, reflection and teacher assessment consistency.

Third, authentic assessment may increase documentation demands if every performance is recorded or every activity receives a detailed rubric. Future implementation research should therefore examine the minimum evidence needed for defensible judgement. A sustainable model should improve assessment quality without creating disproportionate administrative burden.

Fourth, the model requires further work on inclusive assessment. Pupils differ in musical background, language, confidence, physical access and opportunities to practise outside school. Future studies should test how criteria and evidence tools can maintain common learning expectations while allowing appropriate pathways for participation.

Fifth, digital and AI-supported assessment warrants further investigation. Emerging technologies can assist with recording, transcription, feedback and evidence management, but they also introduce questions of reliability, privacy, bias and the authenticity of pupil-generated work. Research should investigate where technology genuinely improves assessment and where teacher observation remains indispensable.

Conclusion

This study developed and evaluated a conceptual model of authentic assessment for Malaysian primary Music Education through a three-phase Design and Development Research (DDR) Type 2 process. The findings revealed uneven explicit representation of assessment components in the curriculum materials analysed, particularly the limited references to rubrics compared with assessment strategies, criteria and tools. In response, the proposed model integrates four interrelated components—performance or assessment strategy, assessment criteria, assessment tools and rubrics—across the four curriculum modules: Music Language, Musical Skills, Music Creativity and Music Appreciation.

The model's principal contribution lies in establishing a curriculum-responsive structure for planning and evaluating authentic assessment. It supports teachers in aligning learning outcomes with appropriate musical activities, identifying task-specific criteria, selecting relevant evidence-collection tools and using rubrics to inform professional judgement. Its application across the four modules allows assessment practices to reflect different dimensions of musical learning, including musical knowledge, technical performance, creativity, interpretation and appreciation. The integration of formative feedback, self-assessment, peer assessment and reflection further positions assessment as an ongoing process that supports pupils' musical development rather than merely a means of determining achievement at the end of a learning activity.

The validation findings provide evidence of the model's content validity and perceived usability within the scope of the study. The S-CVI/Ave of 0.985 and overall usability mean of 4.65 (SD = 0.31) indicate favourable expert evaluations of the proposed framework. Nevertheless, these findings should be interpreted in relation to the study's design. Expert consensus and usability ratings do not establish that the model improves pupils' musical achievement, creativity or learning experiences, nor do they demonstrate its effectiveness across all Malaysian primary school contexts. Furthermore, the teacher data were collected in Selangor, which limits the extent to which the findings can be generalised without further investigation.

Future research should examine the model through classroom-based implementation across diverse school settings in Malaysia. Particular attention should be given to its practicality,

teacher workload, consistency of professional judgement, inclusive assessment practices and contribution to pupils' musical learning. Such investigations would help determine how the model can be implemented sustainably while maintaining appropriate assessment expectations and flexibility for individual musical expression. Overall, the model provides an evidence-informed and expert-evaluated foundation for strengthening authentic assessment in Malaysian primary Music Education, with its longer-term educational contribution to be established through further empirical research.

Acknowledgements

The authors express their sincere appreciation to the Faculty of Music and Performing Arts, Sultan Idris Education University (UPSI), for their institutional support and academic guidance. The authors also extend their gratitude to the experts and primary school music teachers who contributed to this study through their valuable insights and participation.

References

- Abdul Rahman, F. B., Mustafa, Z., & Kharuddin, A. F. (2021). Employing fuzzy Delphi technique to validate Multiple Intelligence based instructional teaching module for preschool children. *Southeast Asia Early Childhood Journal*, 10(1), 62–71.
- Ajjawi, R., Tai, J., Dollinger, M., Dawson, P., Boud, D., & Bearman, M. (2024). From authentic assessment to authenticity in assessment: Broadening perspectives. *Assessment & Evaluation in Higher Education*, 49(4), 499–510.
- Ariffin, A., Mohd Yusof, N. A. A., Siraj, S., & Md Nor, M. (2018). Fuzzy Delphi Method (FDM): Determining phase for multicultural-based model of peace education curriculum for preschool children. *Journal of Research, Policy & Practice of Teachers and Teacher Education*, 8(1), 5–17.
- Boini, B., Asnawi, A., Nugraha, M. T., & Moh. Ainin, M. (2023). Pengembangan instrumen penilaian autentik pada pembelajaran. *At Turots: Jurnal Pendidikan Islam*, 5(1), 799–806.
- Bolden, B., & DeLuca, C. (2022). Nurturing student creativity through assessment for learning in music classrooms. *Research Studies in Music Education*, 44(1), 273–289.
- Bolden, B., DeLuca, C., Kukkonen, T., Roy, S., & Wearing, J. (2020). Assessment of creativity in K–12 education: A scoping review. *Review of Education*, 8(2), 343–376.
- Bolden, B., Rickey, N., & DeLuca, C. (2023). Nurturing musical creativity through assessment for learning. *Music Educators Journal*, 109(4), 18–26.
- Booth, N. (2022). The significance of formative assessment for pupils' spiral progression in English lower secondary school group composing. *British Journal of Music Education*, 39(1), 120–124.
- Chua, C., Mohd Kosnin, A., & Yeo, K. J. (2024). Fuzzy Delphi method for A-level mathematics technological pedagogical and content knowledge module. *International Journal of Evaluation and Research in Education*, 13(1), 441–453.
- Dawson, P. (2017). Assessment rubrics: Towards clearer and more replicable design, research and practice. *Assessment & Evaluation in Higher Education*, 42(3), 347–360.
- Gedera, D. (2023). A holistic approach to authentic assessment. *Asian Journal of Assessment in Teaching and Learning*, 13(2), 23–34.
- Kementerian Pendidikan Malaysia. (2019a). *Panduan pelaksanaan Pentaksiran Bilik Darjah (PBD)* (Edisi ke-2). Bahagian Pembangunan Kurikulum.
- Kementerian Pendidikan Malaysia. (2019b). *Surat Pekeliling Ikhtisas Bilangan 5 Tahun 2019: Pelaksanaan Kurikulum Standard Sekolah Rendah (KSSR) (Semakan 2017) Pendidikan Seni Visual dan KSSR (Semakan 2017) Pendidikan Muzik bagi menggantikan mata pelajaran Pendidikan Kesenian mulai tahun 2020.*

- Kementerian Pendidikan Malaysia. (2025). *Panduan Pengurusan Pentaksiran Berasaskan Sekolah (PBS) Edisi 1 Tahun 2025*.
- Ketonen, L., Körkkö, M., & Pöysä, S. (2025). Authentic assessment as a support for student teachers' reflection. *European Journal of Teacher Education*, 48(3), 467–488.
- Koh, K. H. (2017). Authentic assessment. In K. Hytten (Ed.), *Oxford Research Encyclopedia of Education*. Oxford University Press.
- McArthur, J. (2023). Rethinking authentic assessment: Work, well-being, and society. *Higher Education*, 85, 85–101.
- Pitchemuthu, J. R., & Abdullah, M. H. (2026). Designing authentic performance assessment in primary music education: A framework for evaluating musical creativity and performance. *International Journal of Research and Scientific Innovation*, 13(3), 2313–2332.
- Shaw, B. P. (2024). Artificial intelligence and assessment: Three implications for music educators. *Music Educators Journal*, 111(2), 19–25.
- Schultz, M., Young, K., Gunning, T. K., & Harvey, M. L. (2022). Defining and measuring authentic assessment: A case study in the context of tertiary science. *Assessment & Evaluation in Higher Education*, 47(1), 77–94.
- Villarroel, V., Bloxham, S., Bruna, D., Bruna, C., & Herrera-Seda, C. (2018). Authentic assessment: Creating a blueprint for course design. *Assessment & Evaluation in Higher Education*, 43(5), 840–854.
- Wake, S., Pownall, M., Harris, R., & Birtill, P. (2024). Balancing pedagogical innovation with psychological safety? Student perceptions of authentic assessment. *Assessment & Evaluation in Higher Education*, 49(4), 511–522.