

Authentic Assessment in Primary Music Education: A Comparative Analysis of Seven Curriculum Systems and Implications for Malaysia

Joshua Raj Pitchemuthu, Mohd Hassan Abdullah

Sultan Idris Education University, Tanjong Malim, Malaysia

DOI: <https://doi.org/10.51244/IJRSI.2026.1308000247>

Received: 5 September 2026; Accepted: 10 September 2026; Published: 18 September 2026

ABSTRACT

Authentic assessment has gained increasing importance in competency-oriented education because it enables learners to demonstrate knowledge, skills, and competencies through meaningful and contextually relevant tasks. This approach is particularly relevant to primary music education, where learning involves performance, singing, instrumental playing, movement, composition, improvisation, listening, collaboration, and reflection. However, comparative evidence on how authentic assessment is represented in primary music curriculum documents across different education systems remains limited. This study therefore aimed to compare authentic assessment provisions in seven selected curriculum systems: Malaysia, Japan, Western Australia, Ontario (Canada), North Rhine–Westphalia (Germany), Ireland, and New Zealand. A qualitative comparative document analysis was employed using official curriculum documents and assessment guidance from the selected systems. The analysis focused on four dimensions: assessment strategies, assessment criteria, assessment tools, and assessment rubrics. The findings revealed five common principles across the seven curriculum systems: performance-based musical learning, continuous formative assessment, multidimensional musical competence, multiple sources of evidence, and criterion-referenced judgement. The comparison also identified differences in the specificity of curriculum guidance, explicitness of performance descriptors, learner participation, and the range of assessment tools identified for classroom use. Malaysia demonstrates a substantial policy foundation through the Primary School Standard Curriculum (KSSR), Curriculum and Assessment Standard Document (DSKP), and Classroom-Based Assessment (PBD). Nevertheless, the findings indicate opportunities to strengthen the operational connection between curriculum expectations, authentic musical tasks, assessment evidence, music-specific criteria, and performance descriptors. In particular, music-specific analytic descriptors, classroom exemplars, diversified evidence sources, and strengthened teacher assessment literacy could complement the existing TP1–TP6 structure. The study contributes a comparative perspective on authentic assessment in primary music education and provides evidence-informed directions for strengthening assessment practices within the Malaysian context.

Keywords: authentic assessment; primary music education; comparative curriculum analysis; music assessment; classroom-based assessment; competency-based education; Malaysia

INTRODUCTION

Background and Rationale

Educational assessment has increasingly moved beyond the measurement of factual recall towards the demonstration and development of competencies. Contemporary curriculum reforms emphasise creativity, collaboration, communication, critical thinking, self-regulation, and the capacity to apply knowledge in meaningful situations. The OECD Learning Compass 2030 positions learner agency and the mobilisation of knowledge, skills, attitudes, and values as central to future-oriented education, while UNESCO (2021) calls for education systems that support holistic, creative, reflective, and socially responsible learners. These developments have strengthened the role of assessment as part of learning rather than as a separate activity conducted only at the end of instruction (OECD, 2010; UNESCO, 2021).

Authentic assessment is particularly compatible with this shift because it asks learners to demonstrate competence through meaningful performance, application, creation, reflection, and problem-solving. In contrast

to assessment approaches that depend mainly on written responses, authentic assessment seeks evidence that is closely connected to the nature of the learning being assessed. Formative assessment reinforces this orientation by using evidence of learning to guide feedback, adaptation of instruction, and subsequent learner action (Panadero et al., 2019).

Music education provides a particularly strong context for examining authentic assessment. Musical learning is inherently experiential and multidimensional. Primary pupils learn through performing, listening, creating, moving, composing, improvising, interpreting, and working with others. These activities require the integration of cognitive understanding, psychomotor coordination, affective engagement, creativity, communication, and social participation. Assessment that focuses exclusively on written knowledge therefore risks overlooking important dimensions of musical competence.

Authentic Assessment in Primary Music Education

In music education, authenticity is closely related to the extent to which assessment allows learners to demonstrate musical understanding through actual musical activity. Performance tasks can reveal technical control and expressive interpretation; composition and improvisation can reveal creative decision-making; ensemble activities can reveal listening, communication, and collaboration; and reflective activities can reveal learners' understanding of their own musical decisions. Consequently, authentic assessment is not synonymous with performance assessment alone. It is better understood as an integrated approach in which the assessment task, evidence, criteria, and judgement are aligned with meaningful musical learning.

Contemporary music education research has increasingly highlighted performance-based assessment, formative feedback, portfolios, learner reflection, composition, and improvisation as ways of capturing the breadth of musical learning (Biasutti & Concina, 2021; Burnard, 2013; Kaschub & Smith, 2022). The challenge is to ensure that such assessment remains educationally useful and sufficiently transparent for teachers and learners. This requires clear criteria, appropriate evidence sources, and assessment descriptors that support professional judgement without reducing musical learning to a rigid scoring exercise.

Problem Statement

Although authentic assessment is increasingly advocated internationally, its representation within primary music curriculum documents varies considerably. Some curriculum systems provide detailed achievement standards and performance descriptors, whereas others provide broader expectations and place greater responsibility on teachers to interpret evidence. This variation is important because curriculum documents shape not only what students are expected to learn but also what counts as evidence of achievement and how teachers are expected to make assessment judgements.

In Malaysia, the KSSR, DSKP, and PBD provide an established policy basis for continuous classroom assessment and practical learning. However, the existence of these policies does not necessarily mean that teachers receive equally explicit guidance for assessing complex musical performances, creativity, collaboration, reflection, and musical interpretation. Malaysian classroom assessment research has also identified issues relating to teacher readiness, implementation, workload, and assessment guidance (Chin, Thien, & Chew, 2019; Tan et al., 2022). A systematic comparison with other curriculum systems can therefore help distinguish between what is already established in Malaysian policy and what may require stronger operational support.

Research Gap

Existing literature on authentic assessment and music education has generally examined particular assessment approaches, classroom implementation, teacher perceptions, assessment literacy, or reforms within individual educational settings. Comparative curriculum research, meanwhile, has frequently concentrated on curriculum content, pedagogy, and educational policy, with assessment receiving less attention as a multidimensional object of comparison. There is consequently limited synthesis of how primary music curricula translate authentic assessment into assessment strategies, criteria, tools, and rubrics across different education systems.

This study addresses the gap by placing authentic assessment at the centre of a qualitative comparative analysis of seven selected primary music curriculum systems. The study does not attempt to rank the systems or infer classroom implementation directly from curriculum policy. Instead, it identifies patterns in documentary evidence and considers how those patterns may inform assessment development in Malaysia.

Research Objectives

To compare authentic assessment provisions embedded within the primary music curricula of Malaysia, Japan, Western Australia, Ontario (Canada), North Rhine–Westphalia (Germany), Ireland, and New Zealand.

To identify common principles and distinctive characteristics of authentic assessment across the seven selected curriculum systems using four analytical dimensions: assessment strategies, assessment criteria, assessment tools, and assessment rubrics.

To derive evidence-informed implications for strengthening authentic assessment in Malaysian primary music education.

Research Questions

How are authentic assessment provisions embedded within the primary music curricula of Malaysia, Japan, Western Australia, Ontario (Canada), North Rhine–Westphalia (Germany), Ireland, and New Zealand?

What common principles and distinctive characteristics emerge across the seven selected curriculum systems when authentic assessment is examined through assessment strategies, assessment criteria, assessment tools, and assessment rubrics?

What implications can be derived from the comparative analysis for strengthening authentic assessment in Malaysian primary music education?

LITERATURE REVIEW

Authentic Assessment and Competency-Based Education

Authentic assessment has developed alongside broader changes in the purposes of schooling, particularly the movement from knowledge transmission towards the development and demonstration of competencies. In competency-oriented education, achievement is not defined solely by what learners know but also by their ability to mobilise knowledge, skills, attitudes, and values in meaningful contexts. Contemporary educational frameworks therefore increasingly emphasise learner agency, creativity, collaboration, communication, reflection, and the application of learning in context (OECD, 2020; UNESCO, 2021).

Authentic assessment is compatible with this orientation because it requires learners to demonstrate understanding through tasks that are meaningful and connected to the nature of the learning being assessed. Rather than relying exclusively on decontextualised written responses, authentic assessment can involve performance, creation, application, problem-solving, reflection, and other forms of observable activity. The emphasis is therefore placed on generating meaningful evidence of learning and judging that evidence against explicit expectations.

The relationship between authentic assessment and formative assessment is particularly important. Formative assessment uses evidence generated during learning to provide feedback and inform subsequent teaching and learning decisions. Panadero et al., 2019; Yan et al., 2021 highlight the importance of evaluative judgement, including learners' capacity to judge their own work and the work of others, in supporting self- and co-regulated learning. This perspective strengthens the role of assessment as part of learning rather than as an activity conducted only after instruction has been completed.

Assessment can therefore serve several interconnected purposes: documenting achievement, identifying learning needs, providing feedback, supporting reflection, and guiding subsequent learning. Within an authentic assessment approach, these purposes are not treated as separate activities. Instead, assessment tasks, evidence, criteria, judgement, and feedback are connected to the intended learning outcomes.

For the present study, authentic assessment is operationally understood as assessment that requires learners to demonstrate musical knowledge and competence through meaningful musical activity and that uses evidence, criteria, and feedback to support continued learning. This definition provides a practical basis for examining how authentic assessment is represented in primary music curriculum documents.

Assessment in Music Education

Music education provides a particularly appropriate context for examining authentic assessment because musical learning is inherently experiential, practical, and multidimensional. Primary pupils develop musical understanding through activities such as singing, instrumental performance, listening, movement, composition, improvisation, interpretation, and ensemble participation. These activities require learners to integrate cognitive understanding with psychomotor coordination, expressive decision-making, creativity, communication, and social interaction.

Consequently, assessment in music cannot be fully represented through written tests alone. Performance tasks can provide evidence of technical control, accuracy, interpretation, and expression; composition can reveal creative decision-making; improvisation can demonstrate the ability to generate and transform musical ideas; ensemble activities can provide evidence of listening and collaboration; and reflection can reveal learners' awareness of their musical decisions and learning processes.

The literature on music assessment consequently supports a broader conception of evidence in which performance and creative activity are considered alongside other forms of assessment evidence. Authentic assessment in music should not, however, be equated simply with performance assessment. A performance may constitute an authentic task, but authentic assessment also requires attention to the alignment among the task, evidence, criteria, judgement, and feedback.

This distinction is important because musical competence extends beyond technical accuracy. A pupil may demonstrate accurate pitch and rhythm while showing limited expressive interpretation, creativity, collaboration, or reflective understanding. Conversely, a creative musical response may demonstrate significant musical understanding even when technical execution is still developing. Effective assessment therefore requires criteria that reflect the multidimensional nature of musical learning.

Assessment in music also involves professional judgement. Teachers must interpret evidence relating to qualities such as expression, creativity, interpretation, participation, and collaboration. These qualities are observable, but they cannot always be reduced to a single objectively correct answer. Assessment approaches therefore need to provide sufficient transparency and consistency while retaining appropriate space for teacher professional judgement.

Research on assessment in general music education similarly indicates that assessment practices are influenced by teachers' beliefs, curriculum requirements, teacher autonomy, and educational policy. A review of assessment in general music education highlights the importance of developing appropriate and authentic assessment approaches and strengthening teachers' capacity to use assessment evidence to support instruction and learning.

For primary music education, this suggests that an effective assessment system should provide multiple opportunities for pupils to demonstrate musical learning through authentic activity. It should also provide teachers with practical means of collecting, interpreting, and communicating evidence of that learning.

Teacher Assessment Literacy and Professional Judgement

The implementation of authentic assessment depends substantially on teachers' assessment literacy. Teacher

assessment literacy involves more than knowledge of assessment terminology or the ability to administer assessment instruments. It includes the capacity to understand assessment purposes, design appropriate assessment tasks, collect and interpret evidence, apply criteria, provide feedback, and use assessment information to support teaching and learning.

DeLuca, LaPointe-McEwan, and Luhanga (2016) identify assessment literacy as an important professional requirement and examine how assessment standards across different educational contexts have increasingly incorporated assessment for learning and broader dimensions of classroom assessment. Their work demonstrates that contemporary assessment literacy extends beyond technical procedures towards the meaningful use of assessment within teaching and learning.

This issue is particularly relevant to authentic assessment in music. Music teachers frequently make judgements about performance quality, expression, creativity, interpretation, collaboration, and musical understanding. Such judgements require professional knowledge and cannot always be determined through mechanical scoring. Clear criteria and descriptors can support teachers in making more transparent and defensible judgements, but they should not eliminate professional interpretation.

Rubrics can play an important supporting role in this process. A well-designed rubric can communicate what quality performance looks like, clarify expectations for learners, support consistency in judgement, and provide a basis for formative feedback. However, overly prescriptive rubrics may reduce complex musical qualities to isolated numerical scores. The challenge is therefore to develop assessment descriptors that provide sufficient clarity while preserving the qualitative and contextual dimensions of musical performance.

Teacher assessment literacy is also relevant to the selection of assessment tools. Different learning outcomes require different forms of evidence. For example, a performance task may be appropriate for assessing expressive singing, an observation record may be useful for ensemble participation, a composition product may provide evidence of creativity, and a reflective response may provide evidence of learners' awareness of their musical decisions. The quality of assessment therefore depends not on the number of tools used but on the appropriateness of the evidence collected in relation to the intended learning.

In the Malaysian context, this consideration is particularly important because Classroom-Based Assessment (PBD) requires teachers to make professional judgements about learners' progress. Strengthening teacher assessment literacy can therefore support the more effective use of existing curriculum and assessment structures rather than requiring the replacement of the current system.

Comparative Curriculum Perspective

Comparative curriculum research provides a means of examining how different education systems respond to shared educational purposes while maintaining distinctive cultural, institutional, and policy contexts. Curriculum comparison can identify patterns of convergence and divergence without assuming that one system is inherently superior to another. Adamson and Morris (2014) emphasise the importance of carefully defining the basis and unit of comparison when examining curricula across contexts.

In the present study, the unit of comparison is the curriculum system represented by the selected official documents. The study therefore compares Malaysia, Japan, Western Australia, Ontario, North Rhine–Westphalia, Ireland, and New Zealand at the level of documented curriculum expectations and assessment provisions. Western Australia, Ontario, and North Rhine–Westphalia are treated specifically as jurisdictional cases rather than as complete representations of Australia, Canada, and Germany.

This distinction is methodologically important because curriculum systems differ in terminology, organisational structure, degree of prescription, assessment traditions, and the allocation of responsibility between curriculum authorities and teachers. A concept such as formative assessment, for example, may be explicitly named in one curriculum while being represented through descriptions of teaching, observation, progression, or feedback in another.

Comparative analysis is therefore most useful when attention is given to both explicit and implicit curriculum provisions. In this study, four dimensions provide a common analytical structure: assessment strategies, assessment criteria, assessment tools, and assessment rubrics. These dimensions enable the analysis to move from how learning is assessed, to what qualities are judged, to what evidence is collected, and finally to how judgements about quality are structured.

The comparative perspective is particularly valuable for the Malaysian context. Rather than treating international systems as models to be copied, comparison can identify practices and principles that may provide useful reference points for strengthening existing Malaysian policy. This approach is consistent with the purpose of the study, which is not to rank education systems but to identify common principles, distinctive characteristics, and context-sensitive implications for Malaysian primary music education.

Conceptual Orientation of the Study

The literature reviewed indicates that authentic assessment in primary music education can be understood as an integrated relationship among curriculum expectations, authentic musical activity, evidence of learning, assessment criteria, judgement, and feedback. The present study therefore conceptualises authentic assessment through four interconnected analytical dimensions: assessment strategies, assessment criteria, assessment tools, and assessment rubrics.

Assessment strategies describe how musical learning is elicited, observed, and supported. In primary music education, these may include performance, creative musical tasks, observation, continuous assessment, reflection, and formative assessment. Assessment criteria identify the qualities or competencies used to judge musical achievement, including performance, pitch, rhythm, tempo, expression, creativity, collaboration, music literacy, participation, and reflection.

Assessment tools represent the sources through which evidence is collected and documented. These may include performance tasks, observation records, portfolios, journals, self-assessment, peer assessment, and audio or video documentation. Assessment rubrics and performance descriptors provide structures for interpreting the quality or level of demonstrated achievement.

The four dimensions are considered interconnected rather than isolated. Curriculum expectations establish what learners are expected to demonstrate; assessment strategies determine how that learning is elicited; assessment tools provide evidence of learning; criteria identify the qualities against which evidence is interpreted; and rubrics or descriptors support judgement about the quality and level of achievement. Feedback then connects assessment evidence to subsequent learning and progression.

This conceptual orientation provides the analytical foundation for the comparative examination of the seven curriculum systems. It also provides a basis for interpreting the Malaysian findings without assuming that international practices can be transferred directly to the Malaysian context. The ultimate purpose is to identify evidence-informed principles that can complement existing Malaysian curriculum and classroom-based assessment arrangements.

The conceptual orientation connects four assessment dimensions with the nature of musical competence and with the final purpose of strengthening Malaysian assessment. The relationship is presented in Figure 1.

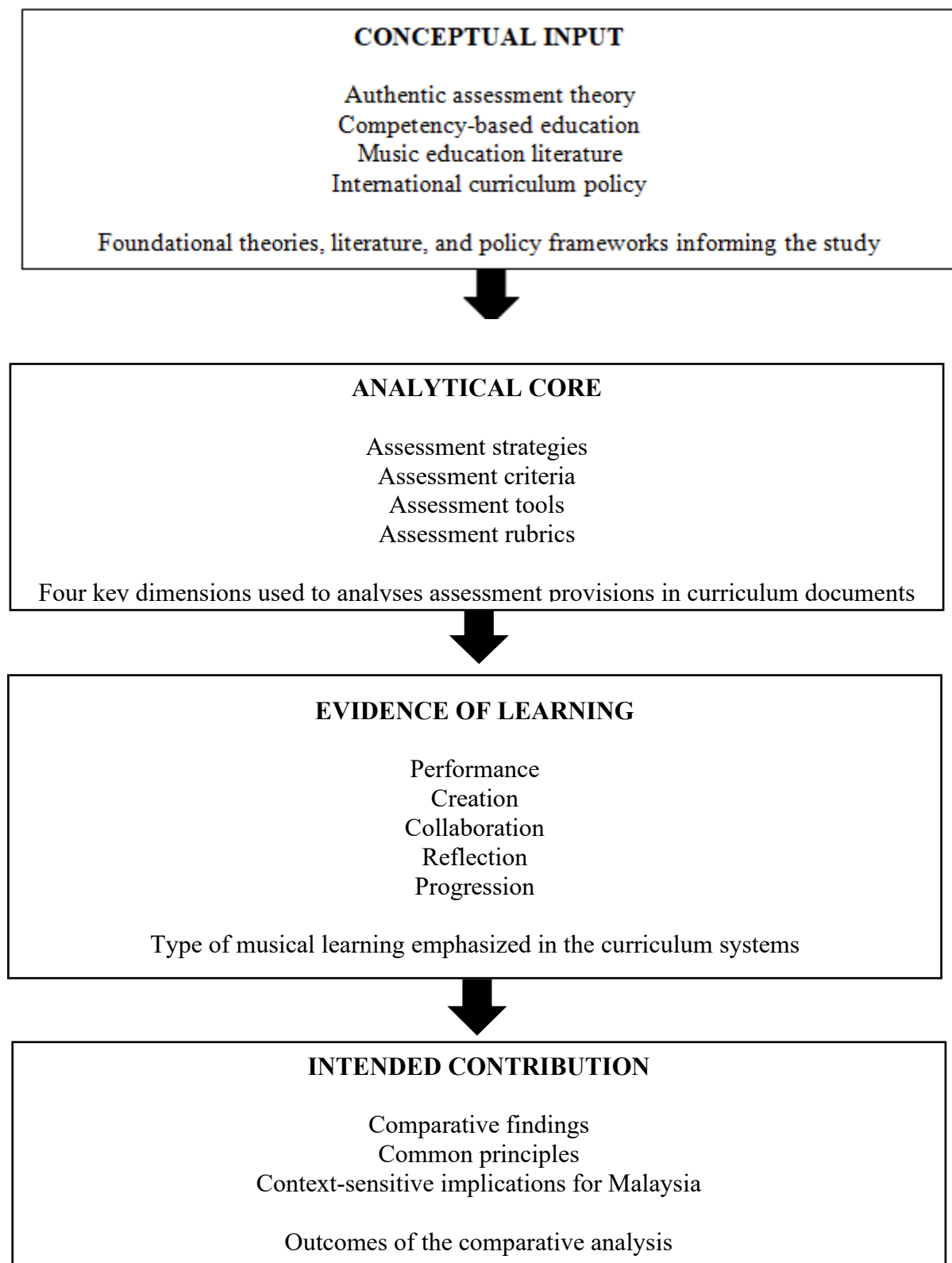


Figure 1. Conceptual orientation for the comparative analysis of authentic assessment in primary music education.

METHODOLOGY

Research Design

This study employed a qualitative comparative document analysis design to examine how authentic assessment is represented within selected primary music curriculum systems. Document analysis is appropriate for

examining the meaning, structure, and implications of policy and curriculum texts (Bowen, 2009). A comparative approach was adopted to identify similarities, differences, common principles, and distinctive characteristics in the assessment provisions contained within the selected curriculum documents.

The study focused specifically on curriculum-level evidence. Therefore, the analysis examined what the selected curriculum documents explicitly state or imply about authentic assessment rather than attempting to determine how consistently the documented provisions are implemented by teachers in actual classrooms. This distinction is important because curriculum expectations and classroom implementation represent different levels of educational practice.

The comparative analysis was organised around four dimensions: assessment strategies, assessment criteria, assessment tools, and assessment rubrics. These dimensions were selected because together they provide a systematic way of examining how curriculum expectations are translated into assessment practices, evidence, and judgements of musical learning.

Curriculum Systems and Documentary Corpus

Seven curriculum systems were selected for comparison: Malaysia, Japan, Western Australia, Ontario (Canada), North Rhine–Westphalia (Germany), Ireland, and New Zealand. The selected documents represent official or formally endorsed curriculum documentation relating to primary or elementary music education.

Table 1. Curriculum Systems and Documentary Corpus

Case	Curriculum/document	Curriculum authority	Year
Malaysia	Dokumen Standard Kurikulum dan Pentaksiran (DSKP): Pendidikan Muzik Tahun 1–6 (KSSR Semakan)	Ministry of Education Malaysia	2017
Japan	Course of Study for Elementary Schools: Music (Grades 1–6)	Ministry of Education, Culture, Sports, Science and Technology (MEXT)	2017
Western Australia	Western Australian Curriculum: The Arts – Music	School Curriculum and Standards Authority (SCSA)	2018
Ontario	The Ontario Curriculum, Grades 1–8: The Arts	Ontario Ministry of Education	2009, revised
North Rhine–Westphalia	Music Curriculum for Primary Schools in North Rhine–Westphalia	Ministry of School and Education, NRW	2021
Ireland	Primary School Curriculum: Music	National Council for Curriculum and Assessment (NCCA)	1999
New Zealand	The Arts in the New Zealand Curriculum	Ministry of Education New Zealand	2007

Note. Western Australia, Ontario, and North Rhine–Westphalia are treated as jurisdictional cases. The comparison therefore does not claim to represent every curriculum authority within Australia, Canada, or Germany.

Table 1 establishes the documentary corpus used for the comparative analysis. The cases were selected to provide variation in geographical location, curriculum structure, and approaches to music education while maintaining a common focus on primary or elementary music. The use of official curriculum documents also provides a consistent basis for examining assessment expectations at the policy and curriculum level.

Inclusion Criteria

Documents were included in the study when they met the following criteria:

1. The document was an official or formally endorsed curriculum document.
2. The document addressed primary or elementary music education.
3. The document contained curriculum expectations, learning outcomes, assessment guidance, or performance expectations relevant to music.
4. The document contained sufficient information to permit analysis through the four analytical dimensions used in the study.
5. The document represented the curriculum version selected for the study.

These criteria were applied to establish a documentary corpus that was sufficiently relevant and comparable for the purposes of the study. The analysis did not seek to include every available policy document or supplementary resource from each jurisdiction. Instead, it focused on the principal curriculum documents that could provide evidence relevant to authentic assessment in primary music education.

- Official or formally endorsed curriculum documentation was used.
- The document addressed primary or elementary music education.
- The document contained curriculum expectations, learning outcomes, assessment guidance, or performance expectations relevant to music.
- The document was sufficiently detailed to permit analysis through the four study dimensions.
- The document represented the curriculum version selected for the study.

Analytical Framework

The study used four analytical dimensions to examine authentic assessment provisions consistently across the seven curriculum systems. The dimensions were selected because they establish a logical connection between how learning is assessed, what is judged, what evidence is collected, and how judgements are made.

Table 2. Analytical Framework for the Comparative Analysis

Dimension	Definition in this study	Indicative codes
Assessment strategies	Ways in which musical learning is elicited, observed, and supported.	Formative; performance; continuous assessment; observation; creative tasks; project-based learning; reflection
Assessment criteria	Qualities or competencies used to judge musical achievement.	Performance; pitch; rhythm; tempo; expression; creativity; collaboration; literacy; reflection; participation
Assessment tools	Instruments or evidence sources used to document learning.	Performance task; observation record; portfolio; journal; self-assessment; peer assessment; audio/video evidence
Assessment rubrics	Descriptors or structures used to interpret levels and quality of performance.	Achievement levels; competency descriptors; analytic descriptors; holistic judgement; observable indicators

Table 2 forms the analytical core of the study. Assessment strategies address how learning is elicited and supported; assessment criteria identify what aspects of musical achievement are judged; assessment tools

identify the sources through which evidence is collected; and assessment rubrics describe how the quality or level of achievement is interpreted.

The four dimensions were applied to all seven curriculum systems. This common analytical structure enabled the study to compare the curriculum documents systematically while recognising differences in terminology and curriculum organisation across jurisdictions.

Research Procedure

The research procedure consisted of three broad phases: data preparation, comparative analysis, and synthesis.

Table 3. Research Procedure

Phase	Procedure	Purpose	Output
I. Data preparation	Identify and select official curriculum documents	Establish an authoritative and comparable documentary corpus	Seven curriculum cases
I. Data preparation	Familiarise with each document	Understand curriculum terminology, structure, and assessment orientation	Document profiles
II. Comparative analysis	Apply four-dimensional coding	Extract assessment-related evidence systematically	Coded evidence matrix
II. Comparative analysis	Analyse each curriculum independently	Avoid decontextualised comparison	Case assessment profiles
II. Comparative analysis	Compare cases across dimensions	Identify similarities, differences, and distinctive practices	Cross-case matrix
III. Synthesis	Conduct thematic synthesis	Identify recurring principles across systems	Five common principles
III. Synthesis	Derive Malaysian implications	Translate findings without copying any single system	Evidence-informed recommendations

Table 3 shows the sequence followed in the analysis. The first phase established and familiarised the documentary corpus. Each curriculum was initially considered independently to reduce the risk of interpreting provisions without regard to their original curricular context.

In the second phase, the four analytical dimensions were applied systematically to each curriculum. Relevant provisions concerning assessment strategies, criteria, tools, and rubrics were identified and organised into case profiles. The individual case profiles were then compared through a cross-case matrix to identify areas of convergence, divergence, and distinctive practice.

The final phase involved thematic synthesis. Patterns recurring across the seven curriculum systems were identified and subsequently interpreted in relation to Malaysian primary music education. The Malaysian implications were derived from the comparative findings rather than by directly adopting the assessment arrangements of any single international curriculum system.

Comparative Coding and Interpretation

The coding process focused on identifying explicit and, where appropriate, sufficiently clear implicit representations of authentic assessment within the curriculum documents. Evidence was organised according to the four analytical dimensions.

For comparative reporting, the study used two categories to indicate the degree of representation:

✓ = explicitly represented in the curriculum documentation.

△ = present but less explicit, more implicit, or substantially dependent on teacher interpretation.

The symbols were used as a means of organising the comparative findings rather than as numerical scores or measures of curriculum quality. The purpose was to make differences in the explicitness of curriculum provisions visible while avoiding the ranking of curriculum systems.

The analysis therefore does not claim that a curriculum marked with a △ is necessarily weaker than one marked with a ✓. Instead, the distinction indicates the extent to which the relevant assessment provision is explicitly identifiable within the selected documentary corpus.

Trustworthiness

Trustworthiness was addressed through the principles of credibility, dependability, confirmability, and transferability (Lincoln & Guba, 1985).

Credibility

Credibility was supported through the use of official curriculum sources and repeated familiarisation with each document. The analysis remained focused on the documentary evidence and avoided making claims about classroom implementation that could not be established from the curriculum documents.

Dependability

Dependability was strengthened by applying the same four analytical dimensions to all seven curriculum cases. The use of a consistent analytical structure provided a common basis for comparison while allowing the specific terminology and organisation of each curriculum system to be retained.

Confirmability

Confirmability was supported by grounding interpretations in the content of the selected curriculum documents. The analysis focused on documented curriculum expectations and assessment provisions rather than relying solely on assumptions about how music assessment should operate.

Transferability

Transferability was addressed through clear descriptions of the selected curriculum systems, documentary corpus, inclusion criteria, analytical dimensions, and boundaries of interpretation. In particular, Western Australia, Ontario, and North Rhine–Westphalia were treated as jurisdictional cases rather than as complete representations of Australia, Canada, and Germany.

Methodological Boundaries

The methodology deliberately distinguishes curriculum provision from classroom implementation. The findings therefore describe what is represented in the selected curriculum documents and do not establish whether teachers consistently implement these provisions in practice.

Similarly, the study does not seek to rank the seven curriculum systems. The purpose of the comparison is to identify common principles, distinctive characteristics, and potentially useful implications for Malaysian primary music education.

The proposed Malaysian implications are consequently interpreted as evidence-informed directions rather than validated prescriptions. Any proposed framework or assessment enhancement arising from the comparative findings would require further empirical investigation, including expert validation and classroom-based research.

FINDINGS

Cross-System Overview

The comparative analysis identified a clear pattern of convergence in the broad orientation of primary music assessment across the seven selected curriculum systems: Malaysia, Japan, Western Australia, Ontario (Canada), North Rhine–Westphalia (Germany), Ireland, and New Zealand. Across the systems, assessment is closely associated with practical musical activity, continuous observation, performance, creativity, and learner development. The main differences occur in the degree of curricular prescription, the range of explicitly identified assessment tools, the visibility of learner participation, and the specificity of performance descriptors.

The findings are organised according to the four analytical dimensions established in the methodology: assessment strategies, assessment criteria, assessment tools, and assessment rubrics. The comparison is intended to identify patterns across the selected curriculum systems rather than to rank them.

The comparative analysis identified a clear pattern of convergence in the broad orientation of primary music assessment. Across the seven curriculum systems, assessment is closely associated with practical musical activity, continuous observation, performance, creativity, and learner development. Differences occur mainly in the degree of curricular prescription, the range of explicitly named assessment tools, the visibility of learner participation, and the specificity of performance descriptors.

Assessment Strategies

Performance-based assessment and formative assessment emerged as the most consistently represented assessment strategies across the seven curriculum systems. Musical learning is assessed through activities such as singing, instrumental performance, creative musical tasks, listening, movement, composition, improvisation, and ensemble participation. This pattern demonstrates that musical activity itself constitutes an important source of evidence of learning.

All seven curriculum systems incorporate continuous assessment and classroom observation. However, the explicitness of particular strategies varies. Western Australia, Ontario, and New Zealand provide comparatively explicit guidance concerning formative processes, achievement expectations, and learner reflection. Japan and North Rhine–Westphalia provide greater space for teacher interpretation, while Malaysia combines structured curriculum standards with Classroom-Based Assessment (PBD).

Table 4. Comparison of Assessment Strategies Across the Seven Curriculum Systems

Assessment strategy	Malaysia	Japan	Western Australia	Ontario	NRW	Ireland	New Zealand
Formative assessment	✓	✓	✓	✓	✓	✓	✓
Performance assessment	✓	✓	✓	✓	✓	✓	✓
Continuous assessment	✓	✓	✓	✓	✓	✓	✓
Classroom observation	✓	✓	✓	✓	✓	✓	✓
Creative musical tasks	✓	✓	✓	✓	✓	✓	✓
Project-based learning	△	△	✓	✓	△	✓	✓
Reflective learning	△	△	✓	✓	△	✓	✓

Student self-assessment	△	△	✓	✓	△	△	✓
Peer assessment	△	△	✓	✓	△	△	✓

Note. ✓ = explicitly represented; △ = present but less explicit or substantially dependent on teacher interpretation.

Table 4 demonstrates strong convergence in the use of performance, formative assessment, continuous assessment, classroom observation, and creative musical activities. These strategies appear consistently across all seven curriculum systems. The greatest differences are observed in strategies that require more explicit learner participation, such as self-assessment, peer assessment, and reflective learning.

The pattern indicates that the core orientation of primary music assessment across the selected systems is strongly connected to doing music rather than relying exclusively on written assessment. At the same time, some systems make learner participation and reflection more explicit than others.

For Malaysia, the findings suggest that the existing curriculum provides a substantial basis for performance-oriented and continuous assessment. The comparatively less explicit representation of self-assessment, peer assessment, reflective learning, and project-based approaches indicates areas where additional guidance could potentially strengthen the range of evidence used in classroom assessment.

Assessment Criteria

The second analytical dimension concerns the qualities or competencies used to judge musical achievement. Across the seven curriculum systems, the criteria reflect a multidimensional understanding of musical competence.

Musical performance, pitch accuracy, rhythm and tempo, expressive interpretation, creativity, collaboration, music literacy, and participation are recurrent areas of assessment. Improvisation and reflective evaluation are also represented, although their degree of explicitness varies across the selected curriculum systems.

Table 5. Comparison of Assessment Criteria Across the Seven Curriculum Systems

Assessment criterion	Malaysia	Japan	Western Australia	Ontario	NRW	Ireland	New Zealand
Musical performance	✓	✓	✓	✓	✓	✓	✓
Pitch accuracy	✓	✓	✓	✓	✓	✓	✓
Rhythm and tempo	✓	✓	✓	✓	✓	✓	✓
Expressive interpretation	✓	✓	✓	✓	✓	✓	✓
Creativity and composition	✓	✓	✓	✓	✓	✓	✓
Improvisation	✓	✓	✓	✓	△	✓	✓
Ensemble collaboration	✓	✓	✓	✓	✓	✓	✓
Music literacy/notation	✓	✓	✓	✓	✓	✓	✓
Reflection and evaluation	△	△	✓	✓	△	✓	✓
Values and participation	✓	✓	✓	✓	✓	✓	✓

Note. ✓ = explicitly represented; △ = present but more implicit or dependent on teacher interpretation.

Table 5 shows that all seven curriculum systems recognise musical learning as multidimensional. Technical aspects such as pitch, rhythm, and tempo are combined with expressive, creative, collaborative, and participatory dimensions.

Malaysia is particularly notable for the specificity of several technical criteria identified in its DSKP, including pitch, rhythm, tempo, articulation, posture, breathing, recorder technique, notation, dynamics, tonal quality, and ensemble participation. This indicates that Malaysian primary music curriculum documentation already contains substantial subject-specific assessment content.

The comparative issue for Malaysia is therefore not primarily the absence of assessment criteria. Rather, the finding concerns the extent to which these criteria are consolidated into clear and coherent performance descriptors that can support teachers in interpreting different levels of musical achievement.

Assessment Tools

The third analytical dimension examines the tools and evidence sources used to document musical learning. Classroom observation and performance tasks are the most consistently represented tools across the seven curriculum systems. These tools are particularly appropriate to music because they enable teachers to collect evidence while pupils are engaged in actual musical activity.

Differences become more visible in the use of portfolios, reflective journals, self-assessment, peer assessment, and audio/video documentation. These approaches provide opportunities to capture learning processes and development over time in addition to evidence from individual performances.

Table 6. Comparison of Assessment Tools Across the Seven Curriculum Systems

Assessment tool	Malaysia	Japan	Western Australia	Ontario	NRW	Ireland	New Zealand
Classroom observation	✓	✓	✓	✓	✓	✓	✓
Observation record/checklist	✓	✓	✓	✓	✓	✓	✓
Performance task	✓	✓	✓	✓	✓	✓	✓
Portfolio	△	△	✓	✓	△	✓	✓
Reflective journal	△	△	✓	✓	△	✓	✓
Self-assessment	△	△	✓	✓	△	△	✓
Peer assessment	△	△	✓	✓	△	△	✓
Audio/video documentation	△	△	✓	✓	△	△	✓
Teacher anecdotal records	✓	✓	✓	✓	✓	✓	✓

Note. ✓ = explicitly incorporated; △ = present implicitly or mainly through teacher professional judgement.

Table 6 indicates that observation and performance remain the dominant sources of assessment evidence across the seven curriculum systems. These tools provide direct evidence of pupils' musical activity and therefore correspond closely with the practical nature of music learning.

The differences are more apparent in the use of additional evidence sources. Western Australia, Ontario, and New Zealand show comparatively stronger explicit representation of portfolios, reflective records, self-assessment, peer assessment, and audio/video documentation. Malaysia has a strong foundation for direct

evidence collection through observation and performance but comparatively fewer explicit curriculum references to these additional evidence sources.

The finding does not suggest that Malaysian teachers should simply increase the number of assessment tools used. Rather, the evidence suggests that assessment tools could be selected according to the nature of the learning outcome. For example, performance evidence may be appropriate for musical performance, observation records for ensemble participation, creative products for composition, and reflective evidence for learners' awareness of musical decisions.

Assessment Rubrics and Performance Descriptors

The fourth analytical dimension concerns the structures used to interpret levels and quality of musical performance. This dimension produced one of the clearest differences among the seven curriculum systems.

All seven systems provide some form of performance expectation, achievement level, or competency description. However, the degree of specificity varies. Western Australia, Ontario, and New Zealand provide comparatively explicit descriptors that can support analytic interpretation of performance, whereas Malaysia, Japan, North Rhine–Westphalia, and Ireland rely more strongly on broader expectations and teacher professional judgement.

Table 7. Comparison of Assessment Rubrics and Performance Descriptors Across the Seven Curriculum Systems

Rubric component	Malaysia	Japan	Western Australia	Ontario	NRW	Ireland	New Zealand
Performance levels	✓	✓	✓	✓	✓	✓	✓
Achievement descriptors	△	△	✓	✓	△	△	✓
Analytic descriptors	△	△	✓	✓	△	△	✓
Holistic judgement	✓	✓	✓	✓	✓	✓	✓
Competency descriptors	✓	✓	✓	✓	✓	✓	✓
Observable performance indicators	△	△	✓	✓	△	△	✓
Creativity indicators	△	✓	✓	✓	△	✓	✓
Reflection criteria	△	△	✓	✓	△	✓	✓
Student-friendly expectations	△	△	✓	✓	△	✓	✓

Note. ✓ = explicitly represented; △ = present generally or interpreted through teacher judgement.

Table 7 demonstrates that the principal comparative difference concerns the specificity of assessment descriptors rather than the existence of assessment expectations themselves. Malaysia uses TP1–TP6 within PBD to represent progression, providing a national structure for describing levels of achievement.

However, TP1–TP6 is broader than a music-specific analytic rubric that differentiates qualities such as pitch accuracy, rhythm, expression, creativity, collaboration, and interpretation at successive levels. Consequently, the findings suggest that the existing Malaysian structure could be complemented by more detailed music-specific descriptors without replacing the national TP1–TP6 framework.

Locally developed lesson-level rubrics can provide such additional detail, but these remain primarily teacher-generated rather than systematically embedded in the national music curriculum. The comparative evidence

therefore points towards strengthening the connection between existing achievement levels and observable, music-specific performance indicators.

Five Common Principles

The cross-system synthesis identified five principles that recur across the seven selected curriculum systems. These principles represent the central comparative finding of the study and demonstrate areas of convergence despite differences in curriculum structures and terminology.

Table 8. Five Common Principles of Authentic Assessment in Primary Music Education

Principle	Evidence across the four dimensions	Meaning for primary music assessment
Performance-based musical learning	Performance and creative activity are central strategies; musical performance and expression recur as criteria; performance tasks and observation are common tools; achievement expectations describe performance quality.	Students demonstrate learning through actual musical activity rather than relying exclusively on written testing.
Continuous formative assessment	Formative and continuous assessment are widely represented; progress and development are assessed through observation; feedback is connected to learning; descriptors support progression.	Assessment operates as part of instruction and supports improvement over time.
Multidimensional musical competence	Criteria include technical skill, expression, creativity, collaboration, literacy, participation, and reflection; tools capture different forms of evidence.	Musical achievement is understood as more than technical accuracy.
Multiple sources of evidence	Observation and performance are universal or near-universal, while portfolios, reflection, self/peer assessment, and documentation appear more strongly in some systems.	A combination of evidence can provide a fuller representation of musical development.
Criterion-referenced judgement	All systems connect assessment to curriculum expectations, performance levels, achievement standards, or competency descriptions.	Judgement focuses on demonstrated achievement against expectations rather than relative ranking.

Table 8 synthesises the findings from Tables 4–7. The five principles are not presented as five independent assessment techniques. Instead, they represent interconnected characteristics of authentic assessment that recur across the selected curriculum systems.

The first principle, performance-based musical learning, reflects the disciplinary nature of music: pupils demonstrate learning through actual musical activity. The second principle, continuous formative assessment, emphasises the use of assessment throughout learning rather than only at the end of instruction.

The third principle, multidimensional musical competence, demonstrates that musical achievement includes technical, expressive, creative, collaborative, literacy, participatory, and reflective dimensions. The fourth principle, multiple sources of evidence, highlights the value of combining different forms of evidence when they serve a clear pedagogical purpose. Finally, criterion-referenced judgement indicates that assessment is generally connected to curriculum expectations, achievement standards, or competency descriptions rather than solely to comparisons among pupils.

Integrated Comparative Model

The findings demonstrate that the four analytical dimensions operate as an interconnected assessment system rather than as isolated categories. Curriculum expectations establish what pupils are expected to learn;

assessment strategies determine how that learning is elicited and demonstrated; assessment tools provide evidence of learning; and criteria and rubrics support the interpretation and judgement of that evidence.

Table 9. Integrated Logic of Authentic Assessment Derived from the Comparative Findings

Assessment component	Guiding question	Function
Curriculum expectation	What should pupils learn?	Establishes intended musical learning and competencies.
Assessment strategy	How will learning be demonstrated?	Identifies meaningful ways for pupils to demonstrate musical learning.
Evidence / tool	What evidence will be collected?	Provides observable and documentable evidence of learning.
Criterion / rubric	How will quality be judged?	Establishes the qualities, levels, or indicators used to interpret achievement.
Feedback & progression	How will evidence inform the next learning step?	Connects assessment evidence with feedback, improvement, and subsequent learning.

Table 9 represents the integrated logic emerging from the cross-system comparison. The model demonstrates that authentic assessment is most meaningful when the different assessment components are aligned. A curriculum expectation should lead to an appropriate assessment strategy, the strategy should generate relevant evidence, and the evidence should be interpreted using appropriate criteria or descriptors. Feedback then connects the judgement to subsequent learning.

This integrated relationship is particularly relevant to primary music education because musical learning is demonstrated through activity and develops progressively over time.

Malaysia in Comparative Perspective

The comparative findings do not support the conclusion that Malaysian primary music assessment is fundamentally behind the other selected systems. On the contrary, Malaysia already incorporates many principles associated with contemporary authentic assessment through KSSR, DSKP, and PBD.

The principal difference identified through the comparison is more accurately located at the level of operational detail. Malaysia has substantial curriculum content addressing musical performance and technical competence, while some of the international cases provide clearer connections among learning expectations, evidence sources, performance descriptors, and learner participation.

The findings therefore suggest that the next stage of development in Malaysia should focus on strengthening implementation support rather than introducing an entirely new assessment philosophy.

In particular, the comparative evidence points towards four areas for further strengthening:

1. clearer connections between curriculum expectations and authentic musical tasks;
2. more explicit music-specific assessment indicators and performance descriptors;
3. appropriate diversification of evidence sources; and
4. strengthened support for teacher professional judgement and assessment literacy.

These findings provide the basis for the implications and evidence-informed framework presented in the subsequent section of the study.

DISCUSSION

Overview of the Findings

The findings indicate that authentic assessment is broadly represented within the curriculum documents of the seven selected primary music education systems, although the degree of explicitness and operational detail varies. Across Malaysia, Japan, Western Australia, Ontario, North Rhine–Westphalia, Ireland, and New Zealand, the selected curriculum documents generally articulate assessment in relation to practical musical activity, formative processes, continuous observation, and the development of multidimensional musical competence. These findings reflect documented curriculum provisions and should not be interpreted as evidence of how authentic assessment is implemented in actual classroom practice.

The comparison of the selected curriculum documents also identified five common principles: performance-based musical learning, continuous formative assessment, multidimensional musical competence, multiple sources of evidence, and criterion-referenced judgement. These principles suggest that authentic assessment in primary music education is not defined by a single assessment instrument. Rather, it is characterised by the alignment of curriculum expectations, meaningful musical activity, evidence of learning, criteria for judgement, and feedback.

The discussion below interprets these findings in relation to the literature and considers their implications for Malaysian primary music education.

Performance-Based Musical Learning as a Central Feature

One of the clearest findings is the consistent emphasis on performance and practical musical activity across the seven curriculum systems. This finding is consistent with the nature of music as a practical and experiential discipline. Pupils demonstrate musical understanding not only through verbal or written responses but also through singing, instrumental performance, movement, composition, improvisation, listening, and ensemble participation.

The prominence of performance-based assessment supports the broader literature on authentic assessment, which emphasises the demonstration and application of learning through meaningful tasks. In music education, the authenticity of an assessment task is particularly important because the discipline itself requires learners to perform, create, interpret, respond to, and communicate through music.

However, the findings also indicate that performance assessment should not be treated as synonymous with authentic assessment. A performance task becomes more meaningful when it is connected to clear curriculum expectations, appropriate criteria, relevant evidence, and feedback. Consequently, the authenticity of assessment depends not simply on whether pupils perform but on what they are asked to demonstrate and how their performance is interpreted.

For Malaysian primary music education, this finding is significant because the existing curriculum already provides a strong foundation for performance-oriented assessment. The implication is therefore not that performance assessment needs to be introduced as an entirely new practice. Rather, existing performance activities could be more explicitly connected to assessment criteria and descriptors that communicate what constitutes developing, proficient, or more advanced musical achievement.

Continuous Formative Assessment and Learning Progression

The findings also show strong convergence in the representation of formative and continuous assessment. Across the selected curriculum systems, assessment is generally positioned as an ongoing process rather than an activity restricted to the end of a learning sequence.

This finding is consistent with the literature on formative assessment and evaluative judgement. Panadero et al., (2019), for example, emphasise the importance of assessment processes that support learners' ability to evaluate

their own learning and participate in self- and co-regulated learning. The findings of the present study extend this perspective to primary music education by demonstrating that formative principles can be incorporated into practical musical activity through observation, performance feedback, reflection, and progressive development.

In music education, continuous assessment is particularly valuable because musical skills develop progressively. A single performance may not provide sufficient evidence of a pupil's development. Repeated observation and performance opportunities allow teachers to identify changes in technique, musical understanding, expression, creativity, participation, and confidence over time.

For Malaysia, the findings suggest that PBD already provides an important structural basis for continuous assessment. The opportunity for further development lies in strengthening the connection between ongoing evidence and music-specific progression. For Malaysia, the findings suggest that PBD provides an important policy structure for continuous assessment. The comparative analysis further indicates that clearer music-specific progression descriptors could provide additional guidance for interpreting documented expectations of musical development. However, whether such descriptors are currently used by teachers, or how they might function in classroom practice, cannot be established from the documentary evidence examined in this study.

Multidimensional Musical Competence

Another important finding is the multidimensional nature of the assessment criteria identified across the seven curriculum systems. Musical competence extends beyond technical accuracy to include expression, creativity, collaboration, music literacy, participation, and reflection.

This finding has important implications for the design of authentic assessment. If assessment focuses predominantly on technical accuracy, it may provide only a partial representation of musical learning. For example, pitch and rhythm accuracy may demonstrate aspects of technical competence, but they do not necessarily capture expressive interpretation, creative thinking, ensemble interaction, or reflective understanding.

The multidimensional character of music learning also supports the argument that assessment should be aligned with the breadth of curriculum expectations. Where curriculum documents identify multiple dimensions of musical learning, assessment should provide opportunities for pupils to demonstrate those dimensions.

For Malaysia, this is particularly relevant because the DSKP contains substantial subject-specific expectations concerning areas such as pitch, rhythm, tempo, articulation, posture, breathing, recorder technique, notation, dynamics, tonal quality, and ensemble participation. The finding therefore suggests that the issue is not a fundamental lack of musical assessment content. Instead, greater coherence may be achieved by connecting these expectations more explicitly to assessment evidence and performance descriptors.

Multiple Sources of Evidence

The comparison identified observation and performance as dominant sources of evidence across the selected curriculum systems, while portfolios, reflective records, self-assessment, peer assessment, and audio/video documentation were more explicitly represented in some systems.

This finding supports the principle that no single assessment tool can necessarily capture the full range of musical learning. Different tools provide different forms of evidence. Performance may provide direct evidence of musical execution and interpretation, observation may document participation and interaction, creative products may provide evidence of composition or improvisation, and reflection may provide evidence of learners' awareness of their musical decisions.

However, the findings should not be interpreted as suggesting that the number of assessment tools should simply be increased. Excessive use of assessment instruments may create additional workload without necessarily improving the quality of evidence. The more important consideration is the alignment between the learning outcome and the evidence source.

For Malaysian primary music education, this suggests a selective rather than expansive approach to assessment tools. Teachers could use different evidence sources when they provide meaningful information about particular learning outcomes. Such an approach would strengthen the authenticity of assessment while remaining compatible with existing classroom-based assessment practices.

Assessment Criteria, Rubrics, and Teacher Professional Judgement

One of the most significant comparative findings concerns the specificity of assessment rubrics and performance descriptors. The selected systems generally provide expectations concerning achievement, but the extent to which these expectations are translated into detailed and observable performance descriptors varies.

This finding is closely related to teacher assessment literacy. DeLuca et al. (2016) emphasise that teacher assessment literacy involves more than technical assessment procedures and includes the capacity to understand assessment purposes, collect and interpret evidence, and use assessment to support learning. This is particularly important in music, where teachers frequently make professional judgements about qualities such as expression, creativity, interpretation, collaboration, and musicality.

The findings suggest that rubrics can support this professional judgement when they provide clear indicators without reducing complex musical qualities to overly mechanical scoring procedures. A useful music rubric should therefore balance clarity and professional judgement.

For Malaysia, the existing TP1–TP6 structure provides a national framework for describing levels of achievement. The comparative findings suggest that this structure could potentially be complemented by more detailed music-specific descriptors. Such descriptors could clarify what different levels of achievement look like in relation to pitch, rhythm, expression, creativity, collaboration, music literacy, and other relevant dimensions.

This would not require replacing TP1–TP6. Instead, the national structure could remain the overarching reporting framework while music-specific indicators provide additional support for classroom-level judgement.

The Role of Learner Participation

The findings reveal another important difference among the curriculum systems: the degree to which learner participation in assessment is explicitly identified. Self-assessment, peer assessment, reflection, and learner-friendly expectations are more explicitly represented in some of the selected systems than in others.

This finding is relevant to contemporary conceptions of assessment in which learners are not viewed solely as recipients of teacher judgement. Learners can also participate in evaluating their own work, responding to feedback, identifying areas for improvement, and developing awareness of quality.

In primary music education, learner participation can be incorporated naturally through activities such as listening back to a recorded performance, reflecting on an ensemble activity, identifying strengths and areas for improvement, or responding to peer performances using simple criteria.

For Malaysia, this represents a potential area for strengthening authentic assessment. Nevertheless, the findings do not justify the conclusion that learner participation is absent from Malaysian practice. Rather, the comparison indicates that it is less explicitly represented at the curriculum-document level than in some of the other cases. Further research would therefore be required to determine the extent to which such practices are already implemented in Malaysian classrooms.

Comparative Implications for Malaysia

The findings provide an important basis for interpreting Malaysia's position within the comparative analysis. The evidence does not support the conclusion that Malaysian primary music assessment is fundamentally underdeveloped or lacks an authentic assessment orientation. Malaysia already possesses a substantial policy and curriculum foundation through KSSR, DSKP, and PBD.

The more significant issue concerns the degree of operational specificity connecting curriculum expectations to authentic assessment practices. The comparison suggests that Malaysia could strengthen the existing system by making certain connections more explicit.

Table 10. Comparative Implications for Strengthening Malaysian Primary Music Assessment

Comparative finding	Comparative finding	Comparative finding
Performance-based musical learning	Performance and practical musical activities are embedded in the music curriculum.	Connect musical activities more explicitly to assessment criteria and evidence.
Continuous formative assessment	PBD provides a structure for ongoing assessment and progression.	Strengthen music-specific progression indicators and feedback processes.
Multidimensional musical competence	DSKP includes technical, expressive, creative, literacy, and participation-related expectations.	Consolidate these dimensions into coherent assessment criteria.
Multiple sources of evidence	Observation and performance provide important classroom evidence.	Provide guidance on selecting additional evidence sources according to learning outcomes.
Criterion-referenced judgement	TP1–TP6 provides a national achievement-level structure.	Complement TP1–TP6 with music-specific performance descriptors.
Learner participation	Learner development and classroom assessment are incorporated within PBD.	Make self-assessment, peer assessment, and reflection more explicit where pedagogically appropriate.
Teacher professional judgement	Teachers make classroom-based judgements using curriculum expectations.	Strengthen assessment literacy and support for music-specific judgement.

Table 10 summarises the main implications of the comparative findings for Malaysia. The table does not propose replacing the existing Malaysian assessment system. Instead, it identifies areas where existing structures could be strengthened through greater clarity, coherence, and subject-specific support.

This distinction is important because curriculum reform should be context-sensitive. Practices developed in other education systems cannot necessarily be transferred directly to Malaysia because curriculum structures, teacher roles, assessment expectations, and institutional conditions differ. The international cases therefore function as sources of comparative insight rather than models for direct adoption.

Implications for Authentic Assessment Framework Development

The findings provide a conceptual basis for strengthening authentic assessment in Malaysian primary music education. The four analytical dimensions—assessment strategies, assessment criteria, assessment tools, and assessment rubrics—can be understood as interconnected components of an assessment process.

The implications are particularly relevant to the development of a framework that seeks to connect curriculum expectations with authentic musical activity and evidence-based judgement.

Table 11. Implications for an Evidence-Informed Authentic Assessment Framework

Framework component	Framework component
Curriculum expectations	Identify clearly what musical knowledge, skills, competencies, and dispositions pupils are expected to demonstrate.
Authentic assessment strategies	Provide meaningful opportunities for pupils to demonstrate learning through performance, creation, listening, movement, collaboration, and other musical activities.
Assessment criteria	Use music-specific criteria that represent technical, expressive, creative, collaborative, literacy, and reflective dimensions where relevant.
Evidence / assessment tools	Select evidence sources according to the learning outcome rather than relying on a single assessment method.
Rubrics / performance descriptors	Provide observable and developmentally appropriate descriptors that support consistent professional judgement.
Learner participation	Incorporate self-assessment, peer assessment, and reflection when appropriate to pupils' developmental level and learning purpose.
Feedback and progression	Use assessment evidence to identify strengths, areas for improvement, and subsequent learning steps.
Teacher assessment literacy	Support teachers in interpreting musical evidence, applying criteria, providing feedback, and making defensible judgements.

Table 11 translates the comparative findings into framework-level implications. It should be interpreted as an evidence-informed conceptual orientation, not as an empirically validated framework.

The findings therefore provide a foundation for subsequent development and validation. Expert review, teacher consultation, and classroom-based investigation would be necessary before the proposed framework could be regarded as validated for wider implementation.

Contribution to the Literature

The study contributes to the literature in three main ways. First, it provides a comparative examination of authentic assessment in primary music education across seven curriculum systems. This contributes to an area in which assessment is often discussed within individual national or jurisdictional contexts.

Second, the study identifies five cross-system principles that provide a common conceptual basis for understanding authentic assessment in primary music education: performance-based musical learning, continuous formative assessment, multidimensional musical competence, multiple sources of evidence, and criterion-referenced judgement.

Third, the study provides a context-sensitive interpretation of these findings for Malaysia. Rather than positioning Malaysia as deficient in comparison with other systems, the study identifies opportunities to strengthen the operational connection between existing curriculum expectations, authentic musical tasks, assessment evidence, criteria, rubrics, and feedback.

Caution in Interpreting the Findings

The findings should be interpreted within the methodological boundaries of the study. The analysis examined curriculum documents and therefore identifies documented assessment provisions, not necessarily classroom

practices.

A curriculum document may provide broad expectations while leaving considerable room for teacher interpretation. Conversely, teachers may employ assessment practices that are not explicitly described in national curriculum documents. The comparative findings therefore should not be interpreted as definitive evidence that one education system implements authentic assessment more effectively than another.

Similarly, the ✓ and △ classifications indicate differences in the explicitness of documentary representation rather than differences in educational quality. A △ does not necessarily indicate an absence of the practice; it indicates that the practice is less explicitly identifiable within the selected documentary corpus.

These distinctions are important for maintaining an appropriately cautious interpretation of the comparative evidence.

Overall Discussion

Overall, the findings indicate that authentic assessment in primary music education is characterised less by a particular assessment instrument than by the alignment of musical learning, authentic activity, evidence, criteria, judgement, and feedback.

The seven curriculum systems share a broad orientation towards practical musical learning and continuous assessment, while differing in the extent to which learner participation, assessment tools, and performance descriptors are explicitly specified. For Malaysia, the findings suggest that the existing curriculum and PBD structures provide a strong foundation on which more detailed music-specific assessment guidance could be developed.

The most appropriate direction is therefore not wholesale replacement of the existing Malaysian assessment system but incremental strengthening of its operational detail. Greater clarity in music-specific criteria and performance descriptors, purposeful use of multiple evidence sources, opportunities for learner reflection, and continued development of teacher assessment literacy could enhance the implementation of authentic assessment while remaining compatible with the existing curriculum structure.

These conclusions provide the basis for the study's proposed implications and conceptual framework while recognising that further empirical research is required to establish its validity, usability, and effectiveness in Malaysian primary music classrooms.

Implications For Malaysian Primary Music Education

The comparative findings have several implications for strengthening authentic assessment in Malaysian primary music education. The findings do not suggest that Malaysia requires a complete replacement of its existing assessment system. Rather, they indicate that the existing KSSR, DSKP, and Classroom-Based Assessment (PBD) structures provide an important foundation that could be strengthened through clearer connections among curriculum expectations, authentic musical activities, assessment evidence, criteria, performance descriptors, and feedback (Ministry of Education Malaysia, 2017; Ministry of Education Malaysia, 2019).

The comparison also indicates that practices identified in other curriculum systems should not be transferred directly to Malaysia. Differences in curriculum structures, teacher responsibilities, assessment expectations, and educational contexts need to be considered. The international cases are therefore best understood as sources of comparative insight for contextualised improvement, rather than as models for direct adoption.

Strengthening the Existing Malaysian Assessment Structure

The first implication concerns the relationship between curriculum expectations and authentic assessment activities. The Malaysian primary music curriculum already contains practical musical learning and subject-specific expectations (Ministry of Education Malaysia, 2017). The opportunity is to make the connection between these expectations and assessment tasks more explicit.

For example, learning activities involving singing, instrumental performance, composition, movement, listening, or ensemble work can be deliberately designed to generate evidence related to specified learning expectations. This would help teachers identify not only whether a pupil has participated in an activity but also what musical knowledge, skill, or competency is demonstrated through that activity.

Table 12. Evidence-Informed Implications for Strengthening Malaysian Primary Music Assessment

Comparative finding	Existing Malaysian foundation	Evidence-informed implication
Performance-based musical learning	DSKP and PBD provide a basis for practical musical activities and performance.	Provide classroom exemplars showing how authentic musical tasks can be linked to learning standards, assessment evidence, and achievement expectations.
Multidimensional musical competence	DSKP contains technical and musical expectations, including pitch, rhythm, tempo, expression, notation, performance, and participation.	Develop clearer and more observable assessment indicators for relevant musical competencies.
Multiple sources of evidence	Observation and performance provide important sources of classroom evidence.	Encourage purposeful selection of additional evidence sources, such as selected recordings, portfolios, reflective responses, self-assessment, or peer feedback, when appropriate to the learning outcome.
Rubric specificity	TP1–TP6 provides an established progression structure for classroom-based assessment.	Complement the existing structure with music-specific performance descriptors and exemplars rather than replacing TP1–TP6.
Teacher professional judgement	PBD involves teacher interpretation of evidence and judgement of pupil achievement.	Strengthen teacher assessment literacy through professional learning, moderation, exemplars, and guidance for music-specific assessment.
Learner participation	PBD provides opportunities to monitor pupil development and learning progress.	Make self-assessment, peer feedback, and reflection more explicit where developmentally and pedagogically appropriate.
Feedback and progression	PBD supports ongoing assessment of pupil achievement.	Strengthen the connection between assessment evidence, feedback, identified learning needs, and subsequent musical learning.

Note. The implications presented in Table 10 are derived from the comparative documentary analysis and should be interpreted as evidence-informed directions rather than prescriptions for direct adoption.

Table 12 summarises the principal implications arising from the comparative analysis. The table is intentionally framed around strengthening existing Malaysian provision rather than introducing a separate assessment system. The findings indicate that Malaysia already possesses several important components of authentic assessment, particularly practical musical learning, continuous assessment, curriculum-based expectations, and classroom-based professional judgement.

The first area for strengthening is the alignment between curriculum expectations and authentic musical tasks. Clearer classroom exemplars could demonstrate how a particular learning standard may be assessed through an authentic musical activity and what evidence teachers could collect from that activity. Such exemplars would provide practical support without requiring teachers to abandon existing curriculum and assessment arrangements.

The second area concerns music-specific assessment indicators. The DSKP contains substantial musical content, including technical and performance-related expectations. The implication is therefore not to add substantially more curriculum content, but to organise relevant expectations into assessment indicators that are sufficiently clear and observable to support classroom judgement.

Purposeful Use of Multiple Evidence Sources

The comparative findings also suggest that Malaysian primary music assessment could benefit from a more purposeful consideration of different evidence sources. Observation and performance are important because they provide direct evidence of pupils' engagement in musical activity. However, additional evidence may be useful when it provides information that cannot be adequately captured through a single performance or observation.

For example, a selected recording could provide evidence of performance development over time, a portfolio could document creative work, and a short reflective response could provide evidence of pupils' understanding of their own musical decisions. Self-assessment and peer feedback could also be used selectively to encourage pupils to participate in evaluating musical learning.

This implication should not be interpreted as a recommendation to increase the amount of assessment documentation. Evidence diversification should not become assessment accumulation. The selection of an assessment tool should be determined by the learning outcome and the type of evidence required.

Music-Specific Performance Descriptors

A further implication concerns the specificity of performance descriptors. The comparative analysis indicates that some curriculum systems provide more detailed descriptions of achievement than others. In Malaysia, TP1–TP6 provides an established structure for describing levels of pupil achievement within PBD (Ministry of Education Malaysia, 2019).

The findings suggest that this existing structure could potentially be complemented by music-specific performance descriptors and exemplars. Such descriptors could clarify what different levels of musical achievement may look like in areas such as pitch, rhythm, expression, creativity, collaboration, music literacy, and performance.

Importantly, this does not require replacing TP1–TP6 with a new standardised music assessment system. Instead, music-specific descriptors could function as supporting assessment resources that help teachers interpret musical evidence while retaining professional judgement and contextual flexibility.

Strengthening Teacher Assessment Literacy

Teacher professional judgement is central to authentic music assessment because many aspects of musical achievement are qualitative and context-dependent. Qualities such as expression, interpretation, creativity, musical interaction, and ensemble participation cannot always be represented adequately through highly mechanical scoring procedures.

The implication for Malaysia is therefore not simply to provide additional assessment forms. Teachers also require support in interpreting evidence and applying criteria appropriately. Teacher professional development could therefore provide practical, structured support for primary generalist teachers who may have limited specialised music training. Rather than introducing additional assessment procedures, professional learning modules could use short, music-specific examples to demonstrate how curriculum expectations can be translated into observable assessment criteria. Teachers could be guided through sample singing, instrumental, rhythmic, creative, and ensemble tasks, followed by annotated examples showing how different levels of musical performance might be interpreted using the proposed descriptors. Moderated scoring activities could then allow teachers to compare their judgements, discuss differences in interpretation, and develop greater confidence in applying the criteria. Short guidance materials, worked rubric examples, and video or audio exemplars could further support teachers in recognising musical evidence and applying the descriptors consistently while

retaining appropriate professional judgement. Such professional development would complement the existing PBD structure rather than introduce a separate assessment system.

Table 13. Priority Areas for Teacher Assessment Support

Area	Possible support
Authentic assessment design	Examples of authentic musical tasks aligned with learning standards
Assessment criteria	Music-specific observable indicators
Performance judgement	Annotated performance exemplars
Rubric use	Guidance on applying descriptors without over-standardising musical judgement
Evidence collection	Examples of appropriate evidence for different learning outcomes
Formative feedback	Strategies for giving specific and actionable musical feedback
Learner participation	Age-appropriate approaches to self-assessment, peer feedback, and reflection
Moderation	Opportunities for teachers to discuss and compare interpretations of musical evidence

Table 13 translates the broader implications into areas of practical teacher support. The purpose is to improve clarity and confidence in assessment judgement, rather than to increase administrative requirements.

Learner Participation in Assessment

The comparison also suggests potential for strengthening learner participation in assessment. Self-assessment, peer feedback, and reflection are more explicitly represented in some of the selected curriculum systems than in others.

For Malaysian primary music education, such practices could be introduced selectively and according to pupils' developmental levels. For example, pupils could listen to a recording of their own performance, identify one strength and one area for improvement, or use simple criteria to provide constructive feedback to a peer.

Nevertheless, the comparative evidence does not establish that such practices are absent from Malaysian classrooms. The finding concerns their explicit representation within the selected curriculum documentation. Further empirical research would therefore be required to determine how extensively these practices are already used by Malaysian music teachers.

Developing an Evidence-Informed Malaysian Framework

The comparative findings provide a basis for organising authentic assessment around an integrated sequence:

Curriculum expectation → Authentic musical task → Evidence/tool → Criterion/rubric → Feedback and progression

This sequence emphasises alignment among what pupils are expected to learn, how they demonstrate learning, what evidence is collected, how achievement is judged, and how assessment information informs subsequent learning.

Table 14. Proposed Structure of an Evidence-Informed Authentic Assessment Framework

Component	Guiding question	Malaysian application
Curriculum expectation	What should pupils learn?	Identify relevant musical knowledge, skills, and competencies from the curriculum.
Authentic musical task	How will learning be demonstrated?	Use meaningful activities such as singing, instrumental performance, composition, improvisation, listening, movement, and ensemble work.
Evidence / tool	What evidence will be collected?	Select performance, observation, creative products, recordings, portfolios, or reflection according to the learning outcome.
Criterion / rubric	How will quality be judged?	Apply relevant music-specific criteria and performance descriptors alongside the existing PBD structure.
Feedback and progression	How will evidence inform the next learning step?	Use assessment evidence to identify strengths, areas for improvement, and subsequent musical learning.

Table 14 illustrates the proposed integrated structure. It is important to emphasise that this framework is conceptual and evidence-informed rather than empirically validated. It represents a synthesis of the patterns identified through the comparative documentary analysis and should therefore be subjected to further expert review and classroom-based investigation before any claim of wider validity or effectiveness is made.

Implications for Future Research

The findings provide several directions for future research. First, studies could investigate how authentic assessment is currently implemented by Malaysian primary music teachers and identify the practical opportunities and constraints associated with its use.

Second, future research could examine the relevance and clarity of music-specific assessment indicators and performance descriptors through expert validation. Such work could determine whether the proposed assessment components are appropriate for different year levels and musical learning outcomes.

Third, classroom-based research could investigate the usability of the proposed framework and examine how its implementation influences assessment practice, feedback, learner participation, and the documentation of musical progress.

These areas require further empirical investigation because the present study examines documented curriculum provisions rather than classroom implementation. The findings therefore provide an evidence-informed basis for future development but do not establish that the proposed approaches are already effective in Malaysian classrooms.

Overall Implication for Malaysia

Overall, the comparative findings indicate that Malaysian primary music education already has a substantial foundation for authentic assessment through KSSR, DSKP, and PBD. The principal opportunity lies not in replacing the existing assessment philosophy but in strengthening its operational specificity and subject relevance.

The most appropriate direction is therefore to develop clearer connections among curriculum expectations, authentic musical tasks, assessment evidence, music-specific criteria, performance descriptors, feedback, and progression. Such development should preserve the flexibility required for professional teacher judgement while providing sufficient guidance to support clarity and consistency.

The international curriculum systems examined in this study consequently serve as comparative sources of insight rather than templates for direct adoption. Any proposed enhancement should be adapted to the Malaysian curriculum, assessment structure, teacher context, and needs of primary music learners.

The proposed framework should therefore be regarded as a conceptual, evidence-informed direction for future development, with expert validation and classroom-based research required before claims can be made regarding its validity, usability, reliability, or effectiveness.

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

- Adamson, B., & Morris, P. (2014). Comparing curricula. In M. Bray, B. Adamson, & M. Mason (Eds.), *Comparative education research: Approaches and methods* (2nd ed., pp. 309–332). Springer. https://doi.org/10.1007/978-3-319-05594-7_11
- Biasutti, M., & Concina, E. (2021). Online composition: Strategies and processes during collaborative electroacoustic composition. *British Journal of Music Education*, 38(1), 58–73. <https://doi.org/10.1017/S0265051720000157>
- Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, 9(2), 27–40. <https://doi.org/10.3316/QRJ0902027>
- Burnard, P., & Power, A. (2013). *Teaching and assessing creativity in music education*. Routledge.
- Chin, H., Thien, L. M., & Chew, C. M. (2019). The reforms of national assessments in the Malaysian education system. *Journal of Nusantara Studies*, 4(1), 93–111. <https://doi.org/10.24200/jonus.vol4iss1pp93-111>
- DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016). Teacher assessment literacy: A review of international research and future directions. *Assessment in Education: Principles, Policy & Practice*.
- Flick, U. (2022). *An introduction to qualitative research* (7th ed.). SAGE.
- Kaschub, M., & Smith, J. (2022). Creativity, composition, and authentic assessment in music education. *Music Educators Journal*.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. SAGE.
- Ministry of Education Malaysia. (2013). *Malaysia Education Blueprint 2013–2025 (Preschool to Post-Secondary Education)*. Ministry of Education Malaysia.
- Ministry of Education Malaysia. (2017). *Dokumen Standard Kurikulum dan Pentaksiran (DSKP): Pendidikan Muzik Tahun 1–6*. Bahagian Pembangunan Kurikulum, Ministry of Education Malaysia.
- Ministry of Education Malaysia. (2019). *Panduan pelaksanaan pentaksiran bilik darjah* (Edisi ke-2). Bahagian Pembangunan Kurikulum, Kementerian Pendidikan Malaysia.
- OECD. (2019). *OECD Learning Compass 2030: A series of concept notes*. OECD Publishing.
- Panadero, E., Broadbent, J., Boud, D., & Lodge, J. M. (2019). Using formative assessment to influence self- and co-regulated learning: The role of evaluative judgement. *European Journal of Psychology of Education*, 34(3), 535–557. <https://doi.org/10.1007/s10212-018-0407-8>
- Tan, T. H., Ahmad Fauzi Mohd Ayub, Manisah Mohd Shah, & Ajurun Begum Ahamed. (2022). Implementation of classroom-based assessment in Malaysia. *International Journal of Academic Research in Progressive Education and Development*, 11(2).
- UNESCO. (2021). *Reimagining our futures together: A new social contract for education*. UNESCO. <https://doi.org/10.54675/ASRB4722>
- Westbury, I., Hopmann, S., & Riquarts, K. (Eds.). (2000). *Teaching as a reflective practice: The German Didaktik tradition*. Routledge.
- Yan, Z., Li, Z., Yang, M., & Lao, H. (2021). Assessment for learning in the digital age: A systematic review. *Educational Research Review*, 34, 100426.