INSTRUMENTAL MUSIC AS A MEDIATED MULTIMODAL PROMPT IN UNIVERSITY ENGLISH LANGUAGE TEACHING: A CONCEPTUAL FRAMEWORK

Maristela G. Binongo-Sy, DALC
University of Science and Technology of Southern Philippines, Cagayan de Oro City, Philippines
Corresponding author: maristela.sy@ustp.edu.ph
Date of Publication: July 31, 2026

ABSTRACT: *Instrumental music is often used in language classrooms, but studies do not always distinguish passive background listening from long-term musical training or teacher-guided language work. This conceptual paper argues that a short, lyric-free excerpt can support university English language teaching only when it is treated as a text to be interpreted rather than as ambient sound. The framework brings together relational authenticity, social-semiotic multimodality, research on music–language processing, and critical-thinking theory. Its central sequence links an audible feature to guided attention, a purposeful English response, and an explicit assessment criterion. Purposeful expression, interpretative reasoning, and critical evaluation are defined as related but separate outcomes. Attention, affective engagement, and willingness to communicate are proposed as possible causal mediators, whereas proficiency, musical familiarity, preference, cultural knowledge, task complexity, listening conditions, and teacher facilitation are treated as moderators. The paper contributes a bounded and testable account of how instrumental music may become a language-learning prompt. It does not claim that music itself improves learning; instead, it specifies the instructional conditions and evidence needed to evaluate that claim.*

Keywords: Critical thinking, English language teaching, Instrumental music, Interpretative skills, Multimodality, Purposeful expression.

1. INTRODUCTION

University English language teaching (ELT) is expected to develop learners who can do more than reproduce grammatically acceptable sentences. Students must select language for an audience and purpose, interpret meanings that are not stated directly, compare alternative viewpoints, and justify conclusions with relevant evidence. These demands are especially visible in academic writing, oral presentation, media analysis, and workplace communication, where meaning is distributed across words, images, sounds, gestures, and cultural conventions. Nevertheless, classroom practice often continues to privilege verbal texts with predetermined meanings and answers. Such materials are useful for linguistic control, but they provide limited opportunities for learners to explain how they reached an interpretation or to defend one plausible judgment among several alternatives.

Authentic materials respond partly to this limitation because they originate in communicative and cultural settings beyond the language classroom. Their educational value, however, is not inherent in their source. Gilmore argued that authenticity involves relationships among the material, the task, the learner, and the social purpose of communication [1]. A film score may therefore be an authentic cultural artifact, but playing it during a lesson does not automatically create authentic learning. Authenticity emerges when learners use English to construct meanings that matter to them, communicate those meanings to others, and respond to the interpretations of an actual audience.

Instrumental music is a potentially valuable but under-specified authentic material. It communicates without lexical content through tempo, rhythm, dynamics, timbre, texture, repetition, contrast, and tension. These features shape perceived emotion and expressive intention [2]. Listeners also construct stories from instrumental music, but the similarity of their narratives varies across cultural groups, demonstrating that musical interpretation is both socially patterned and open to variation [3]. This combination of perceptible evidence and interpretative openness can support description, inference, comparison, narrative construction, and evaluation. Because instrumental music supplies no lyrics for learners to copy, it may also require more self-authored language than a song-based fill-in-the-blank activity.

The central research problem is that the phrase music in language learning combines exposures that are theoretically and pedagogically different. Musical aptitude, years of formal training, active singing or rhythm exercises, teacher-guided interpretation of a short excerpt, preparatory listening, and continuous background music do not involve the same processes. Their outcomes are equally diverse: studies have examined speech perception, pronunciation, vocabulary, anxiety, enjoyment, academic achievement, creativity, interpretation, and critical thinking. When these exposures and outcomes are treated as a single phenomenon, the proposition that "music improves language learning" becomes too broad to confirm, reject, or apply responsibly.

Evidence from music cognition illustrates the problem. Music and language both organize auditory events across time and into structured sequences, and some processing resources appear to overlap [4, 5]. Yet shared processing is a theoretical basis for possible interaction, not proof that brief listening transfers to language performance. Patel's OPERA hypothesis specifies demanding conditions under which sustained musical training may influence speech processing, including neural overlap, precision, emotion, repetition, and attention [6]. A systematic review and meta-analysis likewise found positive effects of music training on some auditory and linguistic outcomes, but the evidence concerned repeated training rather than a single classroom listening episode [7]. More recently, a meta-analysis reported a positive relationship between musical ability and second-language acquisition [8]. These findings cannot be used without qualification to predict the effect of an instrumental prompt delivered for several minutes in an ELT lesson.

Classroom findings are promising but similarly heterogeneous. University foreign-language learners have

reported favorable perceptions of music, and more frequent music listening has shown a modest association with lower language anxiety [9]. A music-mediated EFL intervention reported gains in academic achievement, creative thinking, and self-esteem [10], while another intervention highlighted learners' awareness of socio-emotional skills [11]. These studies justify further investigation, but they do not show that every kind of music benefits every language task. Systematic and experimental evidence on background music shows inconsistent, null, or detrimental effects depending on task demands, musical characteristics, and listener differences [12–15]. In particular, lyrics may compete with the verbal processing required for reading or writing [13, 14]. The relevant pedagogical question is therefore not whether music is beneficial in general, but whether a teacher can convert specific audible features into prompts that elicit assessable language and reasoning.

The literature contains another gap. Research has examined music-related outcomes and, separately, multimodal language learning, but it has not sufficiently specified the instructional chain connecting an instrumental cue to a language response and then to an assessment criterion. Without this chain, positive outcomes may be attributed to music when they actually result from novelty, teacher enthusiasm, collaborative interaction, extra time on task, or reduced anxiety. Negative outcomes may likewise reflect excessive volume, unfamiliarity, divided attention, or an overly difficult task rather than a lack of pedagogical potential. A conceptual framework is needed before an empirical test because the intervention, proposed mechanism, outcome, moderators, and rival explanations must first be distinguished.

This paper addresses that need by conceptualizing instrumental music as a mediated multimodal prompt, not as passive background sound. It focuses on university ELT and uses Ennio Morricone's lyric-free title theme for The Good, the Bad and the Ugly as an illustrative artifact rather than as empirical data [16]. The work contains pronounced contrasts in timbre, dynamics, texture, and tension that can make the proposed instructional relationships concrete. The paper asks:

1. Which theories explain how a lyric-free musical artifact may support purposeful expression, interpretative skills, and critical thinking in university ELT?
2. How can audible musical features be converted into a progressive sequence of observable English-language performances?
3. Which propositions, moderators, boundary conditions, and rival explanations should guide future empirical tests of the framework?

The contribution is conceptual rather than causal. The paper defines constructs that have often been used loosely, separates evidence that concerns different forms of musical engagement, and develops propositions that can guide intervention design, measurement, and comparison conditions. The literature was selected purposively rather than systematically. Priority was given to foundational theoretical works, systematic reviews and meta-analyses, and recent empirical studies that directly addressed music–language relations, background music, music-mediated language learning, multimodality, or critical thinking. The review therefore supports theory building but does not estimate an overall effect or claim exhaustive coverage.

2. THEORETICAL FOUNDATION

2.1 Authenticity as a relationship among artifact, task, learner, and purpose

Authenticity theory explains why the origin of a material is insufficient to determine its educational value. Gilmore distinguished several meanings of authenticity and emphasized that authentic language learning depends on more than exposure to an unmodified real-world text [1]. Applied to instrumental music, this position yields two levels of authenticity. The composition has material authenticity because it was created for artistic and cultural purposes rather than for ELT. The activity acquires communicative authenticity only when the learner makes meaningful choices, produces language for a recognizable audience, and has a reason to communicate beyond satisfying a predetermined answer key.

This distinction prevents a common conceptual error. A commercially produced soundtrack remains an authentic artifact even when used in a classroom, but a task that asks students merely to list isolated adjectives may still be pedagogically thin. By contrast, asking learners to propose a scene, explain how a change in timbre supports their interpretation, and respond to a competing account creates a more authentic communicative demand. The learner is not retrieving the teacher's hidden meaning; the learner is constructing and defending a situated meaning.

2.2 Social-semiotic multimodality and transduction

Social-semiotic multimodality provides the principal explanation for how nonverbal sound can generate language activity. Kress proposed that meaning is made through multiple modes, each with different affordances, and that communication involves selecting and combining available semiotic resources [17]. In this framework, music is not a decorative accompaniment to verbal text. It is a mode with perceptible features that can be re-represented in speech or writing.

The key process is transduction, or the movement of meaning from one mode to another. When a learner converts a sparse texture into an image of isolation, a rising intensity into a prediction of conflict, or a repeated motif into a claim about persistence, the learner must select lexical items, organize clauses, qualify certainty, and explain an inferential connection. The language is generated through interpretative work rather than copied from the stimulus. This process can make abstract language functions—describing degree, signaling contrast, expressing stance, explaining cause, and evaluating evidence—visible within a purposeful task.

Multimodality does not imply that every interpretation is equally defensible. Instrumental music permits plural meanings, but responses can still be evaluated according to the precision of description, relevance of audible evidence, coherence of explanation, and awareness of alternatives. This evidentiary requirement is essential because it separates interpretative freedom from unsupported assertion.

2.3 Shared music–language processing and its limits

Music and language unfold over time and require listeners to detect patterns, generate expectations, and relate smaller units

to larger structures. Patel identified a possible convergence in syntactic processing [4], while Asano et al. proposed hierarchical control as a shared neurocognitive mechanism [5]. These accounts support the possibility that attending to structured sound may recruit processes relevant to language. They do not, however, establish direct instructional transfer.

The OPERA hypothesis provides a useful boundary for the claim [6]. It proposes that music training can support speech processing when musical activity places greater precision demands on overlapping neural networks and when emotion, repetition, and attention promote plasticity. Its explanatory target is sustained training. Thus, OPERA can justify examining how musical engagement and language processing may interact, but it cannot justify predicting immediate gains from passive exposure to one composition. Reviews of music-training evidence and musical ability reinforce this distinction: positive associations or training effects concern enduring competencies and repeated practice, not necessarily brief foreground listening [7, 8].

For the present framework, shared-processing research establishes cognitive plausibility but is not the proposed classroom mechanism. Temporal and hierarchical features make music a structured object to which learners can attend and about which they can reason. The primary instructional mechanism is guided transduction: the task requires learners to notice an audible feature, represent its possible meaning in English, explain the connection, and evaluate alternatives. Teacher scaffolding is termed instructional mediation. Attention, affective engagement, and willingness to communicate are possible causal mediators that future studies may measure statistically. Keeping these meanings separate prevents a pedagogical process from being confused with a causal pathway.

2.4 Critical thinking as warranted judgment

Facione defined critical thinking through interpretation, analysis, evaluation, inference, explanation, and self-regulation [18]. This account is appropriate because the proposed activities do not seek one correct interpretation of a composition. They require judgments that are proportionate to the available evidence. A critical response identifies a claim, refers to an audible feature, explains the connection, considers another plausible interpretation, and revises or qualifies the conclusion where necessary.

Music education scholarship similarly suggests that open-ended musical problems can support critical inquiry when learners are required to articulate reasons and reflect on alternatives [19]. In ELT, this process has a dual purpose. It develops reasoning while also creating a genuine need for language resources such as causal connectives, stance markers, concession, comparison, counterargument, and evaluative vocabulary. The framework therefore treats critical thinking not as a general impression of depth but as an observable quality of language performance.

2.5 Integrated theoretical position

The four perspectives perform different functions in the conceptual model. Authenticity identifies the conditions under which communication becomes purposeful. Multimodality explains how sound can be transformed into verbal meaning. Shared-processing accounts explain why structured attention to music is cognitively plausible while defining the limits of transfer claims. Critical-thinking theory specifies the standards by which interpretations can be evaluated. Together, they support a mediated sequence:

audible feature → guided attention → transduction into language → evidence-based interpretation → critical evaluation. Figure 1 shows the illustrative of this framework.

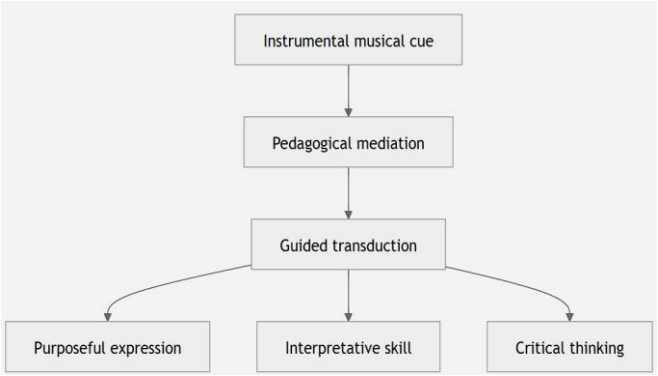

Figure. Graphical Presentation of the Conceptual Framework

This sequence is the theoretical core of the paper. Music supplies the object of attention, but the prompt and assessment criteria determine the intellectual and linguistic work.

3. REVIEW AND SYNTHESIS OF RELATED LITERATURE

3.1 Instrumental music as an authentic and culturally situated text

Instrumental music is capable of communicating affect and direction without specifying a verbal proposition. Tempo, intensity, articulation, pitch range, and timbre provide cues that listeners use when interpreting emotional expression [2]. These cues make a composition describable, but they do not make its meaning universal. Margulis et al. found that listeners imagined narratives while hearing instrumental excerpts and that narrative similarity was stronger within than across cultural groups [3]. Instrumental music therefore combines shared perceptual cues with culturally situated interpretation.

This dual character is pedagogically important. If meanings were entirely fixed, a music task would become another comprehension exercise with a predetermined answer. If meanings were entirely private, there would be no basis for evaluating an interpretation. Instrumental music occupies a productive middle position: learners may construct different meanings, yet they can be required to show how particular features support those meanings. Cultural variation also creates opportunities to compare assumptions, question stereotypes, and examine how prior experience influences interpretation.

Morricone's title theme illustrates these affordances [16]. Its recognizable motif, contrasting instrumental colors, changes in density, and associations with the Western film genre can prompt sensory description and narrative prediction. At the same time, its cultural location requires caution. Learners familiar with the film may reproduce established imagery, while unfamiliar learners may form different accounts.

Familiarity is therefore a moderator to be measured, not an uncontrolled advantage to be interpreted as learning.

3.2 Distinguishing forms of musical engagement

The literature becomes clearer when musical engagement is separated into distinct categories. Musical ability refers to relatively stable perceptual or productive capacities and has been positively associated with second-language outcomes in meta-analytic evidence [8]. Long-term musical training involves repeated practice and possible neuroplastic change; systematic evidence suggests benefits for some auditory and linguistic processes [7]. Active music-mediated language learning requires learners to sing, coordinate rhythm, discuss music, or complete language tasks organized around music and has produced encouraging cognitive, academic, or socio-emotional outcomes [10, 11]. Brief foreground listening presents music as an object of attention for a bounded activity. Continuous background music, by contrast, accompanies another focal task without becoming the object learners must analyze [12–15].

These exposures differ in duration, agency, attentional focus, and instructional purpose. A relationship between musical aptitude and pronunciation does not show that background music improves reading. A long-term training effect does not predict the result of one interpretative lesson. Likewise, a successful music-mediated unit may owe its benefit to collaborative discussion, repeated practice, teacher scaffolding, or emotional engagement rather than to sound alone. Treating these exposures as equivalent obscures the mechanism and weakens empirical design.

3.3 Affective engagement and language anxiety

Music may create emotional salience, enjoyment, or a less threatening classroom atmosphere. Kim et al. found favorable attitudes toward music use among university foreign-language learners and a modest inverse association between listening frequency and language anxiety [9]. Because the study was correlational, lower anxiety may have encouraged music use, music use may have reduced anxiety, or both may have reflected other learner characteristics. The finding supports anxiety as a relevant mediator, but not as an established causal pathway.

Music-mediated interventions provide additional support for affective engagement. Chen et al. reported gains in academic achievement, creative thinking, and self-esteem following an EFL intervention [10], while Cores-Bilbao et al. documented learners' awareness of socio-emotional skills during a music-mediated language experience [11]. These outcomes suggest that music can be embedded in activities that learners find meaningful and socially engaging. However, enjoyment, reduced anxiety, creativity, and linguistic performance remain different constructs. An enjoyable activity may increase participation without improving interpretative quality; reduced anxiety may help hesitant learners speak but may not strengthen the evidence used in their arguments. Future studies must therefore measure affective and performance outcomes separately.

3.4 Attention, interference, and the problem of background music

Evidence on background music establishes an important boundary condition. Cheah et al.'s systematic review found inconsistent effects across cognitive tasks and showed that outcomes depend on the task, music, and population [12]. Souza and Barbosa found that lyrics generally interfered with cognitive performance, whereas the effect of instrumental music was smaller and did not provide credible evidence of a general benefit [13]. Sun et al. likewise reported that lyrical background music could impair reading comprehension, particularly when the lyrics competed with the language of the text [14]. Xiao et al. demonstrated a dual effect in creativity tasks: preference may support positive experience, but cognitive interference can offset potential benefits [15].

These findings do not invalidate the educational use of instrumental music. They invalidate the assumption that music improves performance merely by being present. Background music divides attention because learners must ignore it while completing another task. In the proposed framework, music is foregrounded for a short period and is the evidence learners must analyze. The cognitive demand is therefore part of the task rather than an irrelevant competing stream. Even under this arrangement, volume, duration, familiarity, complexity, preference, sensory sensitivity, and language proficiency may alter the balance between engagement and overload.

3.5 Multimodal composing and second-language development

Research on multimodal L2 composing provides a broader pedagogical context. Kress's social-semiotic account positions meaning-making as the orchestration of multiple modes [17]. A review of 60 empirical studies in post-secondary L2 settings found growing interest in digital multimodal composing but also substantial variation in tasks, theoretical framing, and assessment [20]. Yu et al. similarly concluded that multimodal composing can support audience awareness, creativity, emotional expression, and identity work, while identifying missing links between research findings and routine classroom practice [21].

The relevance to instrumental music is not that learners must produce elaborate digital products. Rather, multimodal research demonstrates that language development can occur when learners interpret relationships among modes and make purposeful representational choices. It also exposes two risks. First, students may invest effort in aesthetic production without improving the target language. Second, teachers may value creativity without establishing reliable criteria for language and reasoning. The proposed framework responds by keeping the output linguistically explicit and assessable: a description, narrative, interpretative paragraph, oral defense, counterargument, or evaluative judgment must be linked to audible evidence.

3.6 Interpretation, creativity, and critical thinking as distinct outcomes

Interpretation, creativity, and critical thinking overlap but should not be collapsed. Interpretative skill concerns the construction of a coherent meaning from available evidence. Creativity concerns originality, flexibility, or elaboration in producing possibilities. Critical thinking concerns the quality of judgment, including the relevance of evidence, consideration of alternatives, and proportionality of a conclusion [18]. A learner may invent an original story that is

weakly connected to the music. Another may offer a conventional interpretation but defend it with precise evidence. A third may identify the cultural assumptions within the composition without producing an elaborate narrative.

This distinction has consequences for assessment. Originality alone cannot represent critical thinking, and grammatical accuracy alone cannot represent interpretation. A suitable rubric should therefore contain separate indicators for language control, specificity of audible evidence, coherence of reasoning, and consideration of alternatives. The framework proposes related domains, not a single composite trait.

3.7 Pedagogical beliefs, workload, and learner context

Teacher facilitation remains an implementation condition rather than an assumed constant. Teachers must be willing to invite open interpretation, permit defensible disagreement, and apply shared criteria instead of imposing a single intended meaning.

Feasibility is equally important. A model that demands specialized musical analysis or a new rubric for every excerpt is unlikely to be sustained. Reusable prompts, short excerpts, a common scoring guide, and shared departmental resources should therefore be examined as components of implementation fidelity.

3.8 Synthesis of the research gap

The literature supports four conclusions. First, instrumental music contains perceptible and culturally meaningful features that can prompt interpretation [2, 3]. Second, music and language share some processing demands, but evidence from musical ability or long-term training cannot be transferred directly to brief classroom listening [4–8]. Third, active music-mediated instruction may support affective and educational outcomes, whereas background music produces inconsistent effects and can interfere with verbal tasks [9–15]. Fourth, multimodal L2 learning is promising but requires clearer alignment among task, language output, and assessment [17, 20, 21].

What remains missing is a bounded account of how a lyric-free musical excerpt becomes a language-learning task and which observable performances should count as evidence of purposeful expression, interpretation, or critical thinking. The present framework addresses this procedural and explanatory gap. It does not claim that music itself produces the outcomes. It proposes that pedagogical mediation can transform audible features into evidence for purposeful language and reasoned judgment.

4. PROPOSED CONCEPTUAL FRAMEWORK

4.1 Construct boundaries

The model defines instrumental music as a short, lyric-free foreground stimulus selected for perceptible contrasts and used as the object of attention. Pedagogical mediation refers to teacher actions that direct attention, pose a language problem, provide scaffolds, require evidence, and specify assessment criteria. Purposeful expression is self-authored English used for an identifiable audience and purpose. Interpretative skill is the construction of a defensible meaning claim grounded in audible evidence. Critical thinking is an evaluative judgment that applies criteria, considers alternatives or counterevidence, and reaches a qualified conclusion.

These definitions exclude several adjacent constructs. Enjoyment and anxiety are affective conditions, not evidence of language achievement. Creativity may contribute to the range of interpretations but does not by itself demonstrate critical evaluation. Musical ability and prior training are learner characteristics, not forms of the proposed intervention. Background music is excluded because the music is not the focal evidence for the task.

Purposeful expression asks whether the learner communicates a personally constructed meaning for an audience and purpose. Evidence includes original wording, stance, elaboration, and audience awareness.

Interpretative skill asks whether a meaning claim is grounded in audible evidence. Evidence includes a specific cue, a defensible cue–claim connection, and a coherent explanation.

Critical thinking asks whether the learner evaluates competing explanations using stated criteria. Evidence includes counterargument, qualification, and justified revision.

4.2 The mediation chain

Consider a university English class in which students hear a 45-second excerpt twice. After the first listening, one student writes only that the music is ‘tense.’ Before the second listening, the teacher asks students to identify one audible change and explain what it suggests. The student then notes that the rhythm accelerates and the texture becomes denser, predicts an approaching confrontation, and later qualifies this interpretation after a classmate proposes a chase scene. The second response is not stronger because the music supplied vocabulary. It is stronger because the prompt converted an audible feature into a claim, evidence, explanation, and reconsideration.

The framework contains four linked elements:

Musical cue: a noticeable feature such as contrast in timbre, a change in dynamics, rhythmic acceleration, repetition, silence, or increased tension.

Mediating prompt: a teacher question or scaffold that directs attention and establishes a communicative problem.

Language performance: an observable oral or written response requiring description, interpretation, comparison, argument, or evaluation.

Assessment indicator: an explicit criterion showing whether the response uses precise language, relevant audible evidence, coherent reasoning, and appropriate qualification.

For example, a shift from sparse to dense texture may prompt learners to describe what changed and predict an event. Learners then compare predictions, cite the musical change that informed them, and revise their account after hearing an alternative. The music does not supply the words or the correct interpretation. It supplies evidence around which purposeful language is organized.

4.3 Progressive domains of performance

The framework progresses through three domains. In purposeful expression, sensory noticing, mood mapping, and sound–image connection elicit precise description, personal stance, and audience-directed comparison. In interpretative skills, narrative construction, pacing analysis, and metaphor

building require learners to connect audible features to a coherent meaning. In critical thinking, cultural deconstruction, perspective debate, and intentional-design evaluation require criteria, counterargument, and qualified judgment.

Progression does not mean that the domains are strictly linear. Description continues to support interpretation, and interpretation supplies the claims that critical evaluation examines. The sequence instead increases the epistemic demand placed on the learner: from identifying a feature, to explaining its possible meaning, to judging among alternatives.

4.4 Testable propositions

Proposition 1: Effect of instructional mediation. When time, teacher contact, and output requirements are held constant, learners who receive cue-focused prompts and explicit evidence criteria will produce more specific and better-supported English responses than learners who hear the same excerpt without interpretative scaffolding.

Proposition 2: Independent language formulation. When familiarity, task difficulty, and output requirements are controlled, instrumental prompts will result in less direct lexical reproduction and a higher proportion of independently formulated language than lyric-based prompts. Lexical independence, interpretative diversity, and response quality must be scored separately because diversity alone does not demonstrate learning.

Proposition 3: Music-specific and multimodal effects. When identical claim–evidence–reasoning requirements are used, mediated musical and mediated visual prompts will both outperform unmediated exposure if guided transduction is the operative mechanism. A musical advantage will be supported only if the mediated musical condition exceeds a comparable mediated nonmusical condition on prespecified language or reasoning outcomes.

Proposition 4: Conditional effects. Effects will vary according to language proficiency, musical familiarity and preference, cultural knowledge, task complexity, listening conditions, sensory sensitivity, and teacher facilitation. These variables should be measured prospectively rather than discussed only after results are known.

Proposition 5: Affective pathways. Enjoyment or reduced anxiety may improve participation through willingness to communicate, but any indirect effect on interpretative or critical-thinking quality should be tested rather than assumed. Stronger reasoning still requires a task that elicits evidence, alternatives, and qualification.

4.5 Rival explanations and boundary conditions

Any future positive finding must be evaluated against plausible alternatives. Novelty may temporarily increase effort. Additional discussion time may improve performance independently of music. Teacher enthusiasm may affect participation. Familiarity with the film may supply ready-made narratives. Musical training may make features easier to identify. Group interaction may improve responses through peer scaffolding. A comparable visual prompt may produce the same benefits, indicating a general multimodal rather than specifically musical mechanism.

The framework is bounded to short, foregrounded, lyric-free excerpts used in university ELT. It does not predict benefits from continuous background music, clinical music therapy, long-term musical training, or lyric memorization. It also does not assume that one composition is suitable for all learners. An accessible visual, tactile, or silent alternative should allow learners with hearing differences, sensory sensitivities, or negative associations to demonstrate the same targeted language and reasoning outcomes. Such alternatives should not be assumed to reproduce the musical stimulus; their comparability requires empirical examination.

5. DISCUSSION

The central contribution of the framework is the replacement of an exposure model with a mediation model. An exposure model assumes that music affects concentration, mood, or learning simply because it is present. This assumption cannot reconcile positive findings from structured music-mediated interventions [10, 11] with inconsistent or negative findings from background-listening studies [12–15]. The mediation model resolves the apparent contradiction by treating music as a nonverbal text. The teacher directs attention to a feature, establishes a communicative or interpretative problem, requires an observable English response, and evaluates the relationship between the claim and the evidence.

This distinction also clarifies the relevance of music–language processing theories. Shared temporal, syntactic, or hierarchical resources make interaction between the domains plausible [4–6], while findings concerning musical training and ability demonstrate that music-related competencies can be associated with linguistic processing [7, 8]. However, these bodies of evidence do not establish that brief exposure produces immediate language gains. In the proposed model, the most defensible mechanism is not automatic transfer but guided transduction: learners transform perceived sound into language and then explain the basis of that transformation.

The framework advances authentic-materials pedagogy by extending authenticity beyond print, speech, and video. Instrumental music is authentic at the level of the artifact, but the communicative quality of the task depends on learner agency, audience, purpose, and consequence [1]. Asking what emotion a composition creates may initiate engagement, but it remains limited unless learners specify the cue, compare interpretations, and communicate a defensible account. Purposeful expression is therefore defined as situated and self-authored performance, not unrestricted personal response.

The distinction among purposeful expression, interpretation, and critical thinking is another contribution. These outcomes are often combined under broad claims about creativity or higher-order learning. The framework shows why they require different evidence. A personal narrative demonstrates expression; a claim grounded in musical features demonstrates interpretation; an evaluation that applies criteria and addresses an alternative demonstrates critical thinking. Separating the constructs makes instrument development and hypothesis testing possible.

The model also addresses a persistent issue in multimodal L2 pedagogy. Reviews show that multimodal composing can support engagement, audience awareness, creativity, and

meaning-making, but classroom implementation and assessment remain uneven [20, 21]. The proposed cue–prompt–performance–indicator alignment prevents the nonverbal mode from displacing the language objective. Music is pedagogically useful only when the learner must produce and refine English in response to it. Accordingly, an attractive narrative, drawing, or video cannot substitute for evidence of the targeted language and reasoning performance. The choice of Morricone's composition makes the model concrete but also exposes the importance of cultural positioning. Listeners' narratives are partly culture-bounded [3]. Familiarity with Western film conventions may help some learners recognize tension or character types, whereas it may constrain them to established interpretations. Cultural deconstruction should therefore be part of the activity rather than an optional extension. Learners can examine whose perspective the soundtrack appears to privilege, which cultural images it evokes, and whether alternative instrumentation might represent the scene differently. In this way, cultural knowledge becomes an object of inquiry instead of an unmeasured source of advantage.

Implementation must remain feasible and inclusive. The framework therefore favors short excerpts, reusable prompts, and one common analytic rubric. Music should not be treated as an isolated remedy for unequal learning conditions. Accessibility requires alternative prompts, adjustable volume, written instructions, and flexible response modes that preserve the intended language and reasoning demands.

The conceptual paper is a necessary precursor to empirical research because it identifies what must be manipulated and measured. A simple music-versus-silence comparison would not test the model. The crucial contrast is between mediated and unmediated use of the same stimulus, with comparable time, teacher contact, and output requirements. The model will be supported only if cue-focused prompts improve the relevant language and reasoning performances after alternative explanations and moderators are considered. It will be weakened if equivalent visual prompts produce the same effects, if benefits are limited to enjoyment, or if no difference emerges between structured and unstructured listening. Its propositions are therefore falsifiable rather than rhetorical claims for the universal value of music.

6. LIMITATIONS OF THE CONCEPTUAL PAPER

This paper presents a selective theoretical synthesis rather than a systematic review or meta-analysis. It cannot estimate an overall effect, assess publication bias, or claim that all relevant studies were captured. Some evidence concerns musical ability or long-term training and is used only to establish theoretical boundaries; it should not be interpreted as direct support for the proposed classroom intervention. Studies of music-mediated instruction also vary in intervention content and outcomes, limiting direct comparison.

The framework uses one culturally recognizable film score as an illustration. Its affordances may be composition-specific, and learners' familiarity with the film or genre may influence their responses. The constructs of purposeful expression, interpretative skill, creativity, and critical thinking remain related and will require discriminant validation. The proposed mediators, moderators, and causal ordering have not been tested. The framework also does not establish an optimal excerpt length, volume, genre, repetition schedule, grouping arrangement, or lesson frequency.

Finally, the model is designed for university ELT. It should not be generalized to younger learners, clinical populations, other languages, or sustained musical training without evidence. Learners with hearing differences, sensory sensitivities, limited access to technology, or negative cultural associations may experience the activity differently. These limitations define the empirical agenda rather than invalidate the conceptual contribution.

7. CONCLUSION

Instrumental music has pedagogical potential in university ELT, but that potential does not arise from music merely being present. The literature distinguishes positive findings from sustained musical training, musical ability, and active music-mediated instruction from the inconsistent effects of continuous background listening. These forms of engagement cannot support one undifferentiated claim that music improves language learning.

The paper contributes to the body of knowledge by proposing a mediated multimodal-prompt model. The model explains how an audible feature can be transformed through guided attention into a self-authored language response and then evaluated through explicit criteria. It defines purposeful expression as purposeful language for an audience, interpretative skill as a meaning claim grounded in audible evidence, and critical thinking as qualified judgment that considers criteria and alternatives. It also identifies affective engagement, attention, and transduction as possible mechanisms and locates language proficiency, musical familiarity, preference, culture, task complexity, listening conditions, and teacher facilitation as moderators.

This conceptual precision provides what the existing literature lacks: an alignment among stimulus, instructional prompt, learner performance, and assessment indicator. It also supplies propositions that can be supported, qualified, or rejected through empirical research. The framework does not establish that instrumental music enhances learning. Its contribution is to specify when such a claim would be theoretically meaningful, how it should be tested, and what evidence would be required to sustain it.

8. RECOMMENDATIONS FOR FUTURE STUDIES

Future research should first validate the constructs and measures. Researchers should develop analytic rubrics with separate indicators for language control, specificity of audible evidence, coherence of interpretation, and consideration of alternatives. Content validity should be established through experts in ELT, assessment, critical thinking, and music education. Pilot studies should examine rubric structure, internal consistency where appropriate, and interrater reliability using independently scored learner responses.

Experimental studies should compare at least three conditions: a music-mediated prompt, the same musical excerpt without interpretative scaffolding, and a comparable nonmusical multimodal prompt. A silence-only condition may be added when the outcome involves concurrent task

performance. Conditions should be equivalent in duration, teacher contact, collaboration, and required output. Pretest measures or baseline tasks should be used so that changes in performance are not attributed to pre-existing language or musical ability.

Studies should measure language and reasoning outcomes separately from enjoyment, anxiety, engagement, and willingness to communicate. Mediation analysis may then examine whether affective engagement or reduced anxiety explains participation, while moderated models test whether proficiency, musical training, familiarity, preference, culture, and task complexity change the effect. Researchers should report effect sizes and confidence intervals, not only statistical significance.

Mixed-method and crossover designs are particularly suitable. Quantitative scores can determine whether the framework affects observable performance, while stimulated-recall interviews, learner journals, or focus groups can explain which cues learners noticed and why the same excerpt supported some learners but distracted others. Crossover designs can reduce stable individual differences by exposing the same learners to mediated musical and nonmusical conditions in counterbalanced order.

Cross-cultural and multi-site replication is necessary because imagined narratives and musical associations are culturally situated. Future studies should use several instrumental works varying in genre, familiarity, tempo, timbre, and emotional character rather than treating one composition as representative. Listening conditions—including excerpt length, sound level, repetition, equipment, and whether music is foregrounded or backgrounded—should be reported precisely.

Implementation studies should examine teacher preparation time, fidelity, acceptability, accessibility, and sustainability. Researchers should determine whether reusable prompts and a common rubric reduce workload and whether teachers with different pedagogical beliefs facilitate interpretative disagreement consistently. All trials should provide silent, visual, or tactile alternatives where appropriate and should examine whether accommodations yield equivalent opportunities to demonstrate the intended language and reasoning outcomes.

REFERENCES

[1] Gilmore, A., "Authentic materials and authenticity in foreign language learning," *Language Teaching*, **40**(2): 97–118(2007). https://doi.org/10.1017/S0261444807004144

[2] Juslin, P. N. and Laukka, P., "Communication of emotions in vocal expression and music performance: Different channels, same code?" *Psychological Bulletin*, **129**(5): 770–814(2003). https://doi.org/10.1037/0033-2909.129.5.770

[3] Margulis, E. H., Wong, P. C. M., Turnbull, C., Kubit, B. M. and McAuley, J. D., "Narratives imagined in response to instrumental music reveal culture-bounded intersubjectivity," *Proceedings of the National Academy of Sciences*, **119**(4): e2110406119(2022). https://doi.org/10.1073/pnas.2110406119

[4] Patel, A. D., "Language, music, syntax and the brain," *Nature Neuroscience*, **6**(7): 674–681(2003). https://doi.org/10.1038/nn1082

[5] Asano, R., Boeckx, C. and Seifert, U., "Hierarchical control as a shared neurocognitive mechanism for language and music," *Cognition*, **216**: 104847(2021). https://doi.org/10.1016/j.cognition.2021.104847

[6] Patel, A. D., "Why would musical training benefit the neural encoding of speech? The OPERA hypothesis," *Frontiers in Psychology*, **2**: 142(2011). https://doi.org/10.3389/fpsyg.2011.00142

[7] Neves, L., Correia, A. I., Castro, S. L., Martins, D. and Lima, C. F., "Does music training enhance auditory and linguistic processing? A systematic review and meta-analysis of behavioral and brain evidence," *Neuroscience & Biobehavioral Reviews*, **140**: 104777(2022). https://doi.org/10.1016/j.neubiorev.2022.104777

[8] Thompson, R. M., Salig, L. K. and Slevc, L. R., "Is musical ability related to second-language acquisition? A meta-analysis," *Royal Society Open Science*, **12**(1): 241193(2025). https://doi.org/10.1098/rsos.241193

[9] Kim, H. J., Chong, H. J. and Lee, M., "Music listening in foreign language learning: Perceptions, attitudes, and its impact on language anxiety," *Frontiers in Education*, **9**: 1372786(2024). https://doi.org/10.3389/feduc.2024.1372786

[10] Chen, M., Mohammadi, M. and Izadpanah, S., "Language learning through music on the academic achievement, creative thinking, and self-esteem of English as a foreign language learners," *Acta Psychologica*, **247**: 104318(2024). https://doi.org/10.1016/j.actpsy.2024.104318

[11] Cores-Bilbao, E., Fernández-Corbacho, A., Machancoses, F. H. and Fonseca-Mora, M. C., "A music-mediated language learning experience: Students' awareness of their socio-emotional skills," *Frontiers in Psychology*, **10**: 2238(2019). https://doi.org/10.3389/fpsyg.2019.02238

[12] Cheah, Y., Wong, H. K., Spitzer, M. and Coutinho, E., "Background music and cognitive task performance: A systematic review of task, music, and population impact," *Music & Science*, **5**: 20592043221134392(2022). https://doi.org/10.1177/20592043221134392

[13] Souza, A. S. and Barbosa, L. C. L., "Should we turn off the music? Music with lyrics interferes with cognitive tasks," *Journal of Cognition*, **6**(1): 24(2023). https://doi.org/10.5334/joc.273

[14] Sun, Y., Zhang, Y., Yang, T. and Luo, J., "Impact of background music on reading comprehension: Influence of lyrics language and listeners' habits," *Frontiers in Psychology*, **15**: 1363562(2024). https://doi.org/10.3389/fpsyg.2024.1363562

[15] Xiao, X., Tan, J., Liu, X. and Zheng, M., "The dual effect of background music on creativity: Perspectives of music preference and cognitive interference," *Frontiers in Psychology*, **14**: 1247133(2023). https://doi.org/10.3389/fpsyg.2023.1247133

[16] Morricone, E., *The Good, the Bad and the Ugly: Original Motion Picture Soundtrack* [soundtrack album], United Artists Records(1966).

[17] Kress, G., *Multimodality: A Social Semiotic Approach to Contemporary Communication*, Routledge, London(2010).

[18] Facione, P. A., *Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction*, American Philosophical Association, Newark, DE(1990).

[19] Kokkidou, M., "Critical thinking and school music education: Literature review, research findings, and perspectives," *Journal for Learning through the Arts*, **9**(1): 1–16(2013). https://doi.org/10.21977/D99112590

[20] Zhang, M., Akoto, M. and Li, M., "Digital multimodal composing in post-secondary L2 settings: A review of the empirical landscape," *Computer Assisted Language Learning*, **36**(4): 694–721(2023). https://doi.org/10.1080/09588221.2021.1942068

[21] Yu, S., Zhang, E. D. and Liu, C., "Research into practice: Digital multimodal composition in second language writing," *Language Teaching*, **59**(2): 197–213(2026). https://doi.org/10.1017/S0261444824000375