Journal of English Language and Education
ISSN 2597- 6850 (Online), 2502-4132 (Print)
Journal Homepage: https://jele.or.id/index.php/jele/index

OPEN ACCESS Article

AI-Aware Assessment of English Projects in Vocational Higher Education: A Conceptual Framework

https://doi.org/10.31004/jele.v11i6.3352

*Bontisesari, Irma Rasita Gloria Barus[ab]
[1]Universitas Pembangunan Nasional "Veteran" Jawa Timur, Indonesia
[2]Universitas IPB, Indonesia
Corresponding Author: bontisesari.bispro@upnjatim.ac.id

ABSTRACT

Generative artificial intelligence (GenAI) complicates judgments about individual English capability in collaborative multimodal projects, where group contributions, production quality, and AI assistance can obscure performance. This conceptual paper proposes an AI-aware assessment framework for a general English course serving 17 study programs in vocational higher education. A conceptual synthesis of literature on project-based language learning, multimodal composing, authentic assessment, collaborative learning, and GenAI assessment informs six principles: visible process, local specificity, multimodal and oral performance, individual accountability, declared and calibrated AI use, and reflective metacognition. The principles are illustrated through two existing course projects, including an alternative format for Project 2. Proposed adaptations appear in a principle-to-practice matrix and an illustrative assessment weighting scheme. The framework offers guidance for collecting complementary evidence of individual language capability while retaining vocationally relevant projects and shared language expectations. Its effectiveness and feasibility require empirical evaluation.

Keywords: *AI-Aware Assessment, Conceptual Framework, Generative AI, Project-Based Language Learning, Vocational Higher Education*

Article History:
Received 05th October 2026
Accepted 05th November 2026
Published 07th November 2026

INTRODUCTION

Generative artificial intelligence (GenAI) tools such as ChatGPT increase uncertainty about what take-home assignments reveal about students' abilities. Research has examined their learning opportunities, academic integrity implications, and effects on assessment in higher education (Bannert et al., 2023; Cotton et al., 2024; Xia et al., 2024). For language assessment, the concern is how much of a submitted text's quality can be attributed to the student's own composing ability. These developments are particularly relevant to second language (L2) writing because GenAI can produce grammatical accuracy, varied vocabulary, cohesion, and conventional text organization—features commonly assessed in language courses. Barrot (2023) discusses how ChatGPT can support writing through models and feedback while also replacing parts of the composing process students are expected to practice. In Indonesia, Puteri et al. (2026) found positive student perceptions of ChatGPT, although essay scores did not differ significantly between the experimental and conventional-instruction groups. For assessment purposes, the quality of an AI-assisted text and the student's independent language capability need to be considered separately.

AI-detection tools offer limited assistance in resolving this distinction. Weber-Wulff et al. (2023) identified substantial reliability limitations in the tools they tested, while Liang et al. (2023) reported frequent misclassification of texts written by non-native English writers. These findings caution against treating detector outputs as decisive evidence of misconduct. They also direct attention to a separate assessment question: what evidence does a task provide, and

OPEN ACCESS
© 2021 The Author.This article is licensed CC BY SA 4.0.
visit Creative Commons Attribution-ShareAlike 4.0 International License.

does that evidence support the intended judgment about a student's English? Following Dawson et al. (2024), this paper places the validity of assessment judgments at the center of its response to GenAI.

Existing frameworks already address aspects of this problem. The Artificial Intelligence Assessment Scale (AIAS) helps teachers communicate permitted levels of AI use (Perkins, Furze, et al., 2024), while institutional frameworks address governance and pedagogy (Chan, 2023). However, permitted-use statements alone cannot establish what students have learned. Corbin et al. (2025) argue for structural changes to assessment that reconsider the evidence tasks produce. This concern is particularly relevant to collaborative multimodal projects, where unequal group contributions, production quality, and AI assistance can jointly obscure individual language performance. Guidance is therefore needed on how these projects can retain their communicative value while providing evidence relevant to individual assessment.

The present paper addresses this problem in a compulsory general English course in vocational higher education. The course uses vocationally oriented tasks while maintaining common language outcomes across study programs. Its field-related content shares features with wide-angle English for Specific Purposes (ESP) course design, although it is not organized around the competencies of a single occupation (Basturkmen, 2010; Dudley-Evans & St. John, 1998). Students may explain how a financial report is prepared, an application is developed, or food quality is checked. These topics require different terminology and contextual knowledge, but can share language objectives such as sequencing information clearly and using grammatical structures appropriately. Assessment must accommodate this diversity while maintaining comparable language expectations.

Indonesian guidance provides a policy basis for responsible GenAI use in higher education (Kemendikbudristek, 2024). Translating that guidance into assessment requires decisions about task requirements, permitted assistance, individual performance, and scoring. This paper therefore asks: How can project-based English assessment in vocational higher education be designed to provide interpretable evidence of individual language capability when students have access to GenAI? It proposes six design principles and illustrates their application through two existing course projects, including an alternative format for Project 2, within a course serving 17 study programs. Its contribution is a context-specific synthesis linking assessment purposes with complementary evidence, a principle-to-practice matrix, and an illustrative assessment weighting scheme. These are conceptual proposals whose effectiveness requires empirical evaluation.

Project-Based Language Learning in Vocational Higher Education

Project-based language learning (PBLL) involves extended work towards a tangible product and integrates language, content, and skills development (Beckett & Slater, 2005; Stoller, 2006). Meaningful content, student decision-making, integrated language use, and communication with an audience make it relevant to vocationally oriented English tasks. Beckett and Beckett & Slater (2005) framework also emphasizes making the language-learning purposes of project work explicit. Genre instruction can further help students organize texts for particular purposes and audiences, such as explaining a process to readers unfamiliar with their field (Hyland, 2007). These features make projects suitable for connecting shared English outcomes with diverse vocational content, provided the language performance expected from students remains explicit.

The assessment challenge is that successful collaboration does not necessarily provide sufficient evidence of each member's English. One student may complete most of the writing while others contribute mainly to technical production. Research on collaborative learning assessment highlights the need to consider group performance and individual achievement together when interpreting assessment outcomes (Meijer et al., 2022). For English projects, this means identifying meaningful language contributions from every member. Selected drafts,

OPEN ACCESS
© 2021 The Author.This article is licensed CC BY SA 4.0.
visit Creative Commons Attribution-ShareAlike 4.0 International License.

Copyright (c) 2026 Bontisesari & Irma Rasita Gloria Barus.

attributable written sections, and individual responses can provide complementary evidence when their assessment purposes are clearly specified.

Multimodal Projects and Language Assessment

Multimodal projects combine resources such as written language, images, speech, sound, and video. Kress (2010) explains how different modes contribute to meaning-making, while research in language education has examined how digital video projects support learner autonomy and digital literacy (C A Hafner, 2014; Christoph A Hafner & Miller, 2011). Such projects are relevant to vocational communication because students must consider how language and other resources work together for a particular audience.

Multimodal production does not automatically resolve uncertainty about students' capabilities. GenAI can assist with scripts, storyboards, images, and other components, while technical skill or visual quality may make a product appear stronger than its language performance warrants. The assessment value of a multimodal project depends on the opportunities it creates for students to communicate, make decisions, and explain their work. Teachers therefore need to distinguish the quality of the overall product from the English capability demonstrated by individual students.

Multimodal assessment requires explicit decisions about which dimensions of communicative competence are being assessed (C A Hafner & Ho, 2020). In the proposed framework, graphic design and video-editing quality are distinguished from the English language performance targeted by the course. Visual and production features can remain relevant to communication, but their assessment should not obscure the basis of individual language judgments. Combining modes broadens the available evidence only when each component is connected to an appropriate learning outcome.

Authentic Assessment and Validity

Authentic assessment involves tasks that reflect meaningful uses of knowledge and skills beyond the classroom. Gulikers et al. (2004) identify five dimensions of authenticity: the task, physical context, social context, assessment result or form, and assessment criteria. Villarroel et al. (2018) emphasize realism, cognitive challenge, and evaluative judgement. These perspectives help explain why vocational English projects benefit from relevant field content, recognizable communicative purposes, and audiences beyond the teacher.

Authenticity and validity nevertheless require separate consideration. A project may resemble professional practice while providing limited evidence of an individual student's ability. For example, a group promotional video may communicate effectively even if only one member wrote the script. Assessment validity concerns the interpretations and uses of the evidence collected (Kane, 2013): whether the evidence supports the conclusions teachers draw about students' achievement. An authentic task therefore requires clear decisions about which capabilities are assessed, whose performance is represented, and how that performance is judged.

Locally situated tasks can support this process by requiring students to investigate and communicate information from their own study programs, facilities, or community settings. However, local content does not prevent AI assistance or guarantee valid assessment. Its value lies in creating opportunities to examine contextual accuracy, relevant language use, and students' explanations of their decisions. Following Dawson et al. (2024), assessment design needs to connect the evidence collected with the intended judgments about capability. For these projects, shared language criteria should therefore be applied alongside context-sensitive judgments about vocational content.

OPEN ACCESS

© 2021 The Author.This article is licensed CC BY SA 4.0.
visit Creative Commons Attribution-ShareAlike 4.0 International License.

Copyright (c) 2026 Bontisesari & Irma Rasita Gloria Barus.

Generative AI and Assessment Design

Existing AI policy and assessment frameworks address permitted assistance, institutional responsibilities, and assurance of learning (Chan, 2023; Lodge et al., 2023; Perkins, Roe, et al., 2024). Together, they inform decisions about how students may use AI and what evidence is needed to assess the intended outcomes. In this paper, *AI-aware assessment* means assessment that explicitly considers students' access to GenAI when setting task requirements, defining permitted assistance, and interpreting evidence of learning. This includes explaining which uses are allowed, asking students to document relevant assistance, and identifying where individual performance needs to be observed. Decisions about AI use should follow the learning outcome. A task assessing independent composing, for example, requires different conditions from one assessing a student's ability to evaluate and revise an AI-assisted text.

The present paper builds on these frameworks by examining their application to collaborative English projects in vocational higher education. Its focus is the relationship between permitted AI use and the evidence needed to assess individual language capability across different project formats. The six principles synthesize established practices for this context. Process documentation, individual accountability, and reflection already have established roles in teaching and assessment; access to GenAI makes their purposes and limitations particularly important to consider.

These perspectives address complementary assessment purposes. PBLL emphasizes project development, authentic assessment supports contextual engagement and evaluative judgement, and multimodal assessment distinguishes forms of communicative performance. Combined group and individual assessment and peer assessment inform judgments about individual contributions(Ion et al., 2024; Meijer et al., 2022). Their integration connects project development and contextual relevance with evidence of individual performance and the conditions under which work is produced.

These perspectives highlight several challenges in assessing individual English capability through collaborative vocational projects. Authenticity favors local, program-specific tasks, but lecturers must maintain comparable language standards across those different contexts. Within each task, multimodal production and collaboration can enrich communication while making individual language performance harder to distinguish from visual quality and group contributions. Permitted AI assistance adds another layer of uncertainty about what the final product demonstrates. Together, these challenges highlight the need to connect assessment judgments with clearly defined evidence of individual performance. The six principles proposed in this paper address this need and guide the adaptation of existing vocationally oriented English projects.

METHOD

This paper adopts a conceptual, design-oriented approach to develop an AI-aware assessment framework for project-based English in vocational higher education. Following Jaakkola (2020) emphasis on an explicit research design for conceptual articles, the paper explains how relevant literature and existing course documents inform framework development. The framework was developed in four steps: selecting relevant literature, identifying assessment problems and design responses, consolidating related ideas into six principles, and illustrating their application to existing course projects. The paper does not evaluate an intervention or measure changes in student performance.

Step 1: Identifying and selecting literature. Relevant literature was identified through searches of Scopus, ERIC, and Google Scholar, supplemented by backward citation tracking from key papers such as Dawson et al. (2024) and (Perkins, et al., 2024). Search terms combined GenAI and assessment terminology with terms related to project-based language learning, multimodal composing, ESP, and individual accountability. For GenAI-related literature, the

OPEN ACCESS
© 2021 The Author.This article is licensed CC BY SA 4.0.
visit Creative Commons Attribution-ShareAlike 4.0 International License.

Copyright (c) 2026 Bontisesari & Irma Rasita Gloria Barus.

search focused on publications from 2023 to 2026. No date limit was applied to the other perspectives to allow the inclusion of foundational theoretical works.

Sources were included if they (a) were peer-reviewed articles, scholarly books or chapters, or official guidance from higher education regulators or ministries; (b) were written in English or Indonesian; and (c) addressed at least one of three questions: what evidence a task produces about learning, how individual performance can be identified in group or project work, or how GenAI use should be managed in assessment. Opinion pieces without a clear argument about assessment, GenAI studies unrelated to assessment or writing, and literature focused on primary or secondary schooling were excluded. Selected sources informed the conceptual synthesis, while additional references supported the contextual and methodological discussion. Selection was guided by the authors' judgement of relevance to framework development; the synthesis was not intended to provide an exhaustive account of the literature.

Step 2: Extracting problems and responses**.** Sources selected for the synthesis were examined for two kinds of content: the assessment problem identified and the design response proposed, or a response inferred by the authors from the source's argument. Assessment problems included difficulty attributing individual contributions in group work, limitations of AI-detection tools, and visual quality obscuring weaknesses in language. Design responses included staged submissions, individual speaking roles, and declarations of AI use. These ideas were recorded in a working table, with one row per idea and its source identified.

Step 3: Consolidating the principles. Related responses were grouped to identify recurring assessment-design concerns. Drafts, outlines, source logs, and staged submissions, for instance, were consolidated under a principle concerned with how work develops, termed visible process. Some responses that initially appeared similar were deliberately kept separate. Process documentation and individual accountability both involve drafts and records, but they serve different evidentiary purposes: the first shows how a piece of work developed, whereas the second concerns which student contributed what. A group draft may therefore provide evidence of process without establishing each member's individual contribution. Each candidate principle was examined against two criteria: whether it addressed an assessment challenge identified in the literature and whether it could guide assessment across different project formats while allowing vocational content to vary. This process yielded six principles, organized according to the assessment purposes they address and the types of evidence they are intended to support.

Step 4: Applying the principles to course projects. The principles were applied conceptually to projects from a compulsory general English course serving 17 vocational study programs at a public university in Indonesia, where one author coordinates and teaches the course. Shared learning outcomes concern clear written and spoken communication, grammatical accuracy, appropriate field-related vocabulary, and communication for a specified audience.

The course has two main projects, including an alternative format for Project 2. Together, these represent the complete set of project assessments in the course and provide the basis for illustrating the framework's application. The project briefs and existing assessment rubrics document current teaching practice. The projects encompass written, visual, oral, video, and digital-game outputs. Project 1 involves a process explanation, a visual process representation, and an oral presentation. Project 2 involves a promotional video, while an educational digital game serves as an alternative to the video for students in two computing programs.

To examine how the principles could inform assessment design, the course documents—project briefs, content outlines, and assessment rubrics developed by the course team—were examined in three passes. First, each project was described in terms of its required outputs, language focus, intended audience, and assessment criteria. Second, each project was examined for gaps in the evidence available to assess individual English capability, including reliance on group scores and components that could be produced with substantial outside

OPEN ACCESS
© 2021 The Author.This article is licensed CC BY SA 4.0.
visit Creative Commons Attribution-ShareAlike 4.0 International License.

Copyright (c) 2026 Bontisesari & Irma Rasita Gloria Barus.

assistance. Third, the six principles were used to propose adaptations to the project requirements and assessment arrangements, including an illustrative assessment weighting scheme for Project 1. Existing requirements and proposed changes are distinguished in the subsequent discussion.

The analysis draws on selected literature and existing course documents to develop the framework and illustrate its application to project assessment. The projects provide contextual examples of how the principles can inform task requirements and assessment weighting. As a conceptual study, the paper does not involve student-level data; the proposed adaptations remain subject to implementation and empirical evaluation.

FINDINGS AND DISCUSSION

Overview of the Course Projects

The course documents describe two main group projects, with an alternative format for Project 2 for students in two computing programs. These projects are completed over several weeks and combine English communication with vocational or institutional content. Table 1 summarizes their requirements, language focus, intended audiences, and original assessment criteria.

Table 1. Overview of the Course Projects

Feature	Project 1: *My Field in Action*	Project 2: *My Study Program*	Project 2 Variant: Educational Digital Game
Study programs	All 17 programs	15 programs (12 in Semester 1, 3 in Semester 2)	2 computing programs (1 in Semester 1, 1 in Semester 2)
Weight in final grade	20%	30%	30%*
Main output	Process text (500–700 words), visual process representation, source list, and oral presentation (7–10 minutes)	English promotional video (3–4 minutes)	Digital learning game for a community tutoring program
Language focus	Process organization, active and passive voice, field-specific vocabulary	Spoken description, persuasion, pronunciation, audience engagement	Instructional language, game text, user guidance
Target audience	Classmates and teacher	Prospective students and the public (published on Instagram or YouTube)	Children in a community tutoring program
Original rubric criteria	Content and accuracy (25%), organization (25%), language use (20%, individual), vocational vocabulary (15%, individual), visual (15%)	Content (25%), organization and engagement (20%), language use (20%), speaking and pronunciation (15%), visual communication (10%), creativity and production (10%)	Adapted from Project 2

**For the two computing programs, the educational digital game is an alternative format for Project 2 and carries the same 30% course weight. It is not an additional project.*

Project 1 contributes 20% of the final grade, Project 2 contributes 30%, and peer assessment of individual contributions contributes a further 5%. Together, these components account for 55% of the final grade. This emphasis on collaborative project work is consistent with the orientation of Indonesia's higher education performance indicator for collaborative and participatory learning, IKU 7 (Kemendikbudristek, 2021). Because project-related assessment represents more than half of the course grade, the evidence it provides about individual students' English capability requires careful consideration.

Project 1, *My Field in Action*, asks each group to explain a process from its vocational field in English. Possible topics include developing a mobile application, checking food quality, or

OPEN ACCESS
© 2021 The Author.This article is licensed CC BY SA 4.0.
visit Creative Commons Attribution-ShareAlike 4.0 International License.

Copyright (c) 2026 Bontisesari & Irma Rasita Gloria Barus.

preparing a financial report. The project framework specifies common objectives, text length, language requirements, and assessment criteria while allowing the process topic to vary by program. Students organize their explanation around five questions: what the process is, why it matters, who and what is involved, how it takes place, and what it produces. They then create a visual process representation and deliver an oral presentation. This structure connects varied vocational content with shared expectations for explaining a process in English.

Project 2, *My Study Program*, requires a short English promotional video introducing the students' program to prospective students. Its content covers the program overview, learning experience, career prospects, and a closing invitation. Publication on Instagram or YouTube gives the task a potential audience beyond the classroom. Students in software engineering technology and computer engineering technology complete an educational digital game as an alternative. The game is designed for children attending a free community tutoring program near the campus, connecting students' technical skills with English instructions, feedback, and other user-facing text. Both formats require students to communicate for a specified audience, although they emphasize different uses of English.

The promotional video emphasizes spoken communication and persuasion, while the game emphasizes written guidance for young learners. Their equal weighting of 30% does not establish equivalent assessment demands. Course coordination therefore needs to identify the shared English outcomes assessed through both formats and distinguish these from format-specific requirements. Technical sophistication, visual quality, and production effort should be judged according to their relevance to the stated outcomes and assessment criteria.

GenAI creates potential uncertainty within each format. Students may use it to generate a process explanation, prepare a promotional script, or write game instructions. These possibilities do not establish that students used GenAI or failed to learn from their projects. They indicate that the final output alone may provide an incomplete account of individual language capability. The design question is how to retain the projects' vocational relevance while collecting evidence that supports judgments about each student's performance.

Six Design Principles for AI-Aware Project Assessment

The conceptual synthesis informed six design principles. Each addresses an assessment challenge that arises when judgments about individual English capability rely mainly on a final group product. Table 2 summarizes the assessment problem addressed by each principle, its main theoretical basis, and examples of relevant evidence. The principles fall into three pairs. Visible process and local specificity concern the *work*: how it developed and how it connects to a real setting. Multimodal and oral performance and individual accountability concern the *student*: what each person can communicate and what each person contributed. Declared and calibrated AI use and reflective metacognition concern the *conditions and judgement* behind the work: what assistance was used and how students evaluate language choices and tool outputs. These pairs organize the framework's emphasis; their evidentiary functions may overlap. Their value lies in bringing together complementary sources of evidence relevant to individual performance, while recognizing the limits of each source.

OPEN ACCESS

© 2021 The Author.This article is licensed CC BY SA 4.0.
visit Creative Commons Attribution-ShareAlike 4.0 International License.

Copyright (c) 2026 Bontisesari & Irma Rasita Gloria Barus.

Table 2. Conceptual Basis of The Six Principles

Principle	Assessment problem addressed	Main theoretical basis	Examples of assessment evidence
1. Visible process	The final product may obscure how the work developed.	Project-based language learning (Beckett & Slater, 2005; Stoller, 2006)	Outlines, drafts, revision notes, checkpoint discussions
2. Local specificity	Generic content can be produced without engaging with the vocational setting	Authentic assessment (Gulikers et al., 2004)	Accurate, setting-specific information communicated in English
3. Multimodal and oral performance	Prepared products may provide limited evidence of spontaneous communication; production quality may obscure language performance.	Multimodal composing (Hafner, 2014; Kress, 2010); multimodal assessment (Hafner & Ho, 2020)	Written and multimodal outputs, prepared spoken performance, and responses to unprepared questions.
4. Individual accountability	Group scores may obscure differences in individual contributions and language performance.	Combined group and individual assessment (Meijer et al., 2022); self- and peer-assessment (Ion et al., 2024)	Attributable written sections, individual speaking performances, and peer ratings of contributions
5. Declared and calibrated AI use	The final product alone does not reliably reveal what AI assistance was used or whether it complied with the task rules	AI assessment frameworks (Corbin et al., 2025; Perkins et al., 2024)	AI-use declarations, in-class writing samples
6. Reflective metacognition	The product may not reveal how students evaluate language, including AI-generated language.	Evaluative judgement (Tai et al., 2018; Villarroel et al., 2018)	Reflections that justify language choices and revisions

Principle 1. Make The Learning Process Visible

Assessment should include selected evidence of project development alongside the final output. This may comprise a topic proposal, source log, outline, draft, or revision note. Such evidence is consistent with PBLL's emphasis on extended work and the integration of language, content, and skills (Beckett & Slater, 2005; Stoller, 2006). It can help teachers examine how students organize information, respond to feedback, and revise their language. However, a sequence of submitted documents does not establish independent authorship because GenAI can also produce drafts and revision explanations. Process evidence is more informative when connected to scheduled checkpoints at which students discuss specific choices or complete a short task under teacher observation. Teachers should select a manageable number of checkpoints, with each providing evidence relevant to the intended assessment judgments.

Visible process should also be distinguished from individual accountability (Principle 4). Process evidence shows how a piece of work developed over time and may belong to the whole group; accountability evidence concerns which student contributed what and the language performance each student demonstrates. A well-documented group draft may therefore provide evidence of development without establishing each member's individual contribution.

Principle 2. Build Local Specificity into Projects

Projects should require students to investigate and communicate information from settings relevant to their studies. Depending on the task, this may involve a laboratory procedure, a campus facility, a field site, or a community partner. For a process explanation, students could describe an observed procedure and identify details that distinguish it from a generic account. This connects the communicative task with the vocational context and intended audience, consistent with authentic assessment (Gulikers et al., 2004). Local information is valuable because students must select, interpret, and communicate it accurately.

OPEN ACCESS

© 2021 The Author.This article is licensed CC BY SA 4.0.
visit Creative Commons Attribution-ShareAlike 4.0 International License.

Copyright (c) 2026 Bontisesari & Irma Rasita Gloria Barus.

It does not prevent AI assistance, and contextual accuracy alone does not demonstrate English proficiency. Assessment should therefore consider both the reliability of the information and how effectively students communicate it in English. Requirements should also reflect students' access to facilities and information so that contextual investigation does not become an avoidable barrier to completing the language task.

Local specificity raises a concern in a course serving 17 programs: if groups describe different processes in different settings, can their work be judged against comparable standards? The framework addresses this concern by proposing shared language criteria while allowing vocational content to vary. For Project 1, these criteria include whether process stages are sequenced clearly, whether field terms are used accurately and explained for the audience, and whether active and passive voice are used appropriately. Content accuracy can be considered against documented sources and observations, with clarification from students where needed. Comparability can be further supported by shared annotated exemplars from different programs and moderation meetings in which lecturers score the same samples and discuss their differences (Bloxham et al., 2016). Such moderation could help lecturers interpret shared criteria across different vocational contexts.

Principle 3. Include Multimodal and Oral Performance

Written and visual outputs can be accompanied by individual presentations, demonstrations, or responses to questions. These components provide opportunities to observe different aspects of students' communication, consistent with attention to multimodal performance in language education (C A Hafner, 2014). A written process text provides evidence of the organization, grammatical accuracy, and field vocabulary present in the submitted work; attributing these features to an individual student requires consideration of the assistance and assessment conditions involved. A rehearsed presentation or video narration provides evidence of pronunciation, fluency, and audience-appropriate delivery for prepared language. Responses to unprepared questions provide evidence of listening comprehension, spontaneous speaking, and content understanding. An infographic shows how information is condensed and labelled through language and visual resources. These formats should therefore be interpreted according to the competencies they demonstrate, without treating oral performance as proof of independent written authorship.

Assessing multimodal projects requires explicit decisions about which dimensions of communicative competence are being evaluated (C A Hafner & Ho, 2020). In the proposed framework, language performance is assessed separately from visual and production quality to make the basis of each score clear. The weighting of visual and production criteria should reflect their relevance to the course outcomes, while language descriptors should identify the features students are expected to demonstrate. Individual oral responses should also be assessed using questions of comparable difficulty and clearly defined criteria. These arrangements connect each component with an identifiable assessment purpose.

Principle 4. Assess Individual Contributions

Group assessment should include evidence attributable to each student. Research on collaborative learning assessment highlights the importance of considering both group performance and individual achievement when making assessment judgments (Meijer et al., 2022). Possible mechanisms include identifiable written contributions, individual speaking roles, brief explanations, and peer assessment. Each student's responsibility should involve meaningful English use. Assigning a student only to formatting, for example, may document participation without providing sufficient evidence of language capability. Self- and peer-assessment can help differentiate individual contributions and inform students' perceptions of fairness in group assessment (Ion et al., 2024). In the proposed framework, these accounts are considered alongside submitted work and observed performance. Peer ratings inform judgments about contributions, while relevant language evidence supports judgments about

OPEN ACCESS
© 2021 The Author.This article is licensed CC BY SA 4.0.
visit Creative Commons Attribution-ShareAlike 4.0 International License.

English capability. This arrangement combines assessment of the shared product with individual assessment of the intended language outcomes.

Principle 5. Specify and Document Permitted AI Use

Project instructions should explain which forms of GenAI assistance are permitted and at which stages. These decisions should follow the learning outcomes, in line with the AIAS approach (Perkins et al., 2024). Brainstorming or language feedback may support one task, while generating a complete text would undermine a task intended to assess independent composing. Rules alone, however, do not secure valid evidence (Corbin et al., 2025). This principle therefore combines clear expectations and declarations with assessment conditions that do not depend entirely on the accuracy of students' self-reports.

In practice, a declaration can be kept short. Students state (a) which tool they used, (b) at which stage and for which section, (c) what they asked it to do, and (d) how they used the output, including an example where relevant. Lecturers should interpret this information in relation to the assessed outcome and the features affected by assistance. AI-supported planning may leave evidence of sentence construction available for assessment, while AI correction of grammar limits what the revised text demonstrates about independent grammatical accuracy. Substantial AI rephrasing or reorganization similarly requires caution when attributing language and organizational choices to the student. The relevant question is which assessed features the assistance affected and what additional evidence is needed to interpret individual performance.

Where independent writing is assessed, a brief observed writing task can provide complementary written evidence. Oral explanations can support judgments about spoken communication and content understanding, but do not substitute for evidence of writing. Suspected use of prohibited assistance should be handled through the institution's academic-integrity procedures. Assessment scores should be based on evidence relevant to the stated outcomes, and declarations should serve as a basis for interpretation and discussion.

Principle 6. Require Reflective Metacognition

Students should provide brief individual reflections on their project decisions, difficulties, revisions, and learning. Where GenAI was used, reflection should go beyond describing that use. Students can be asked to identify a suggestion they rejected and explain why, point out an error or inappropriate term the tool introduced, or compare an AI-suggested sentence with their final version and justify the change. A student who notes that the tool replaced *sterilize* with *clean*, losing a technical meaning needed for a laboratory procedure, provides a specific explanation of language awareness that may not be visible in the final text alone. Reflection prompts can also ask students to explain their own communicative choices, such as why a passive structure suits a procedure or how they simplified vocabulary for children in the game variant. These are forms of evaluative judgement, the capacity to make decisions about the quality of one's own and others' work (Tai et al., 2018). Reflection provides additional evidence of students' reasoning about language, but cannot establish whether a product was generated by AI.

Integrating the Six Principles

Taken together, the six principles connect complementary assessment evidence with the intended language outcomes. An observed writing sample can inform judgments about writing, an individual oral response can demonstrate spoken performance, and drafts and reflections can provide context for interpreting decisions and revisions. Their integration is intended to distinguish product quality, participation, and individual language capability. Table 3 translates this approach into proposed adaptations for the three project formats, building on existing requirements and identifying opportunities to collect additional evidence.

OPEN ACCESS
© 2021 The Author.This article is licensed CC BY SA 4.0.
visit Creative Commons Attribution-ShareAlike 4.0 International License.

Copyright (c) 2026 Bontisesari & Irma Rasita Gloria Barus.

Table 3. Principle-To-Practice Matrix for The Three Design Cases

Principle	Project 1: Process text and visual representation	Project 2: Promotional video	Project 2 variant: Educational game
1. Visible process	Submit topic, source log, outline, and draft before the final text, followed by a short revision note.	Submit storyboard and script draft, with teacher feedback before filming.	Submit game concept, learner profile, and draft in-game texts before development.
2. Local specificity	Ask students to explain a process in an accessible teaching facility or field setting, supported by documented sources and at least one relevant local example.	Build on the suggested campus footage by requiring accurate, program-specific information and original footage where access and consent permit.	Require students to explain how the game's content and language suit the intended tutoring context.
3. Multimodal and oral performance	Retain the visual process representation and presentation; add unprepared questions directed to each member.	Require each member to contribute spoken English, followed by individual responses to questions after screening.	Include a live demonstration in which each member explains a design decision and responds to questions in English.
4. Individual accountability	Identify each member's written contribution and presentation role; assess relevant language evidence individually.	Identify each member's script contribution and speaking role; assess individual spoken performance.	Identify each member's contribution to instructions, feedback, or dialogue; assess the relevant written language individually.
5. Declared and calibrated AI use	Require an initial student-written draft before permitted AI-assisted editing. Collect an AI-use declaration and a brief observed writing sample when independent writing is assessed.	Permit declared brainstorming assistance and specify conditions for script development. Use recorded and live performance for speaking judgments, with observed written evidence if independent script writing is assessed.	Permit declared code assistance while specifying separate conditions for English text production. Use brief observed writing where independent written English is assessed.
6. Reflective metacognition	Individual reflection (150–200 words) on language choices, revisions, and evaluation of any AI suggestions.	Individual reflection on speaking performance, project decisions, and AI use.	Individual reflection on writing for young learners, design decisions, and AI use.

For Project 1, the proposed adaptations make the development of the process text more available for assessment. The existing brief requires a written explanation, sources, a visual representation, and a presentation, but these outputs alone provide limited information about how each student developed the language. Selected outlines, drafts, and revision notes would provide context for interpreting the final text, while a local example would connect its content to a setting students have investigated. Where independent writing is an assessed outcome, a brief observed writing sample would provide complementary evidence. These additions connect the existing product with evidence of development and individual performance.

For Project 2, the existing video task already includes spoken communication and encourages footage of the students' learning environment. The proposed changes would make individual contributions more explicit by identifying script responsibilities, requiring a speaking contribution from each member, and adding questions after screening. A storyboard and script checkpoint would also provide opportunities to discuss content and language decisions before production. Recorded delivery and live responses would then support distinct judgments about prepared and spontaneous speaking, without being treated as proof of independent script writing.

OPEN ACCESS

© 2021 The Author.This article is licensed CC BY SA 4.0.
visit Creative Commons Attribution-ShareAlike 4.0 International License.

Copyright (c) 2026 Bontisesari & Irma Rasita Gloria Barus.

The game variant illustrates how permitted AI assistance can differ across components of the same project. Programming support may be allowed because technical production is not the primary English outcome being assessed. Conditions for producing instructions, feedback messages, and dialogue would be specified separately according to the intended language outcomes. The intended audience of children in a community tutoring program also gives students a concrete communicative purpose, consistent with authentic assessment (Gulikers et al., 2004). Assessment would therefore consider how clearly students communicate with young users and explain their language choices, alongside any individually assessed writing evidence.

An Illustrative Weighting Scheme for Project 1

Project 1 is used to illustrate how the principles can inform assessment weighting because its existing rubric explicitly distinguishes group and individual criteria. The example focuses on Project 1; weighting for the promotional video and game variant would require separate consideration of their respective language outcomes. Table 4 presents one possible redistribution of assessment weight for Project 1. Under the proposed scheme, the individually assessed proportion increases from 35% to 60%. This includes direct assessment of language performance alongside process evidence, reflection, and AI-use declaration. The 60% therefore represents the individually assessed component, not the weighting for language proficiency alone.

Table 4. Original and Proposed Assessment Weightings for Project 1

Criterion	Original weight	Proposed weight	Assessment level
Process content and accuracy	25%	20%	Group
Process organization	25%	15%	Group
English language use	20%	15%	Individual
Vocational vocabulary	15%	10%	Individual
Infographic or visual representation	15%	5%	Group
Process evidence, reflection, and AI-use declaration	–	20%	Individual
Oral explanation and response to questions	–	15%	Individual
Total	100%	100%	
Individually assessed	35%	60%	

The proposed weighting scheme redistributes assessment weight toward evidence of students' individual language performance, engagement with the project process, and reflective judgement. A student's use of permitted GenAI assistance should not, by itself, lower the assessment score. Judgments about independent writing should draw on relevant written evidence, while oral explanations and responses should be scored against the relevant speaking criteria. Process documentation, reflection, and AI-use declarations should be evaluated according to their own stated criteria. Difficulty explaining a particular decision may inform the score for that specific assessed outcome, but should not automatically be treated as evidence of AI misuse or used to reduce scores for unrelated language outcomes. The proposed 60% individual weighting illustrates one possible allocation; appropriate weightings will depend on the learning outcomes and assessment context.

The Rationale for AI-Aware Assessment Design

The framework prioritizes assessment design as a response to GenAI in general English courses within Indonesian vocational higher education. Three arguments support this approach. The first concerns fairness. Weber-Wulff et al. (2023) identified reliability limitations in the detection tools they tested, while Liang et al. (2023) reported misclassification affecting non-native English writers. These findings raise concerns about using detector outputs as decisive evidence in assessments of Indonesian students' English. Academic-integrity

OPEN ACCESS
© 2021 The Author.This article is licensed CC BY SA 4.0.
visit Creative Commons Attribution-ShareAlike 4.0 International License.

procedures remain necessary, but routine scoring should rest on relevant performance evidence and clearly communicated assessment conditions.

The second argument concerns validity. Dawson et al. (2024) emphasize the importance of evidence that supports judgments about student capability. Collecting drafts, contextual information, individual contributions, and observed performance can broaden the basis for those judgments, provided each source is interpreted within its limits. Local knowledge does not independently establish English proficiency, and oral performance does not establish writing ability. The framework therefore connects each assessment judgment with evidence relevant to the particular language outcome.

The third argument concerns preparation for professional practice. Guidance on AI in higher education recognizes the importance of responsible tool use and clear expectations for teaching and assessment (Chan, 2023; Kemendikbudristek, 2024). Declared and calibrated assistance can create opportunities for students to evaluate AI outputs while preserving conditions for assessing independent language performance where required. The AIAS approach supports aligning permitted assistance with assessment purposes (Perkins, Furze, et al., 2024). Within these projects, students would therefore encounter both opportunities to work with AI and responsibilities for explaining the communication they produce.

The framework organizes established practices according to their assessment purposes, providing a basis for connecting the evidence collected with the judgments that evidence can support. The project illustrations suggest a practical advantage of coordinating assessment across a multi-program course. Shared project requirements could allow course teams to introduce the six principles through coordinated revisions to project briefs and assessment criteria, with adjustments for different project formats. Each program could retain its own vocational content while following common expectations for evidence of individual English performance. This approach may ease coordination across classes, although its feasibility remains to be examined.

Implications for Course Coordination

At the same time, the principles involve additional workload. Collecting staged drafts, conducting question-and-answer sessions, and reviewing individual reflections require teacher time, particularly in large classes. Teachers may therefore need to prioritize the principles according to their available resources and assessment purposes. Local specificity (Principle 2) can connect existing tasks with relevant vocational settings, while individual oral responses (Principles 3 and 4) can provide evidence of spoken communication and content understanding. Where independent writing is assessed, a brief observed writing task under specified AI-use conditions (Principle 5) can provide relevant written evidence. Additional process checkpoints (Principle 1) and structured reflections (Principle 6) can be introduced progressively as course coordination and workload permit. This staged approach keeps implementation focused on the intended language outcomes while allowing course teams to examine the workload involved.

Limitations and Directions for Future Research

The framework draws on literature and project documents from a single multi-program English course. Its application illustrates possible assessment arrangements, while implementation is needed to examine their practical value. Transfer to other settings may require adjustments to language outcomes, student proficiency, project formats, and available resources. Programs emphasizing spoken service interaction, for example, may need different evidence from those emphasizing written process explanations. The framework consequently offers a design basis that requires contextual adaptation and empirical evaluation.

Several next steps follow. First, the proposed arrangements could be piloted in selected programs representing different project formats. Second, lecturers could use shared performance descriptors to score the same samples, examine inter-rater agreement, and

OPEN ACCESS
© 2021 The Author.This article is licensed CC BY SA 4.0.
visit Creative Commons Attribution-ShareAlike 4.0 International License.

discuss differences through moderation. Third, studies could compare evidence from drafts, observed writing, prepared speaking, live responses, and reflections, considering the particular outcomes each component assesses. Agreement may be relevant when components target the same capability; differences may be expected when they assess distinct skills.

Further research could examine students' perceptions of fairness, clarity of AI-use expectations, and experiences of disclosure. Where ethically and practically feasible, these accounts could be considered alongside records of tool use to explore the completeness of declarations. Teacher time and coordination demands should also be documented. Together, these investigations would help determine which elements provide useful evidence and can be sustained across different vocational teaching contexts.

CONCLUSION

This paper examines how project-based English assessment in vocational higher education can provide reliable evidence of individual language capability when students use generative artificial intelligence (GenAI). It proposes assessment design as the primary response, emphasizing evidence aligned with intended language outcomes rather than attempting to identify AI involvement in final products. Six principles guide the collection and interpretation of complementary evidence related to project development, vocational context, individual performance, permitted assistance, and reflective judgement. The framework can be integrated into existing English projects through selected checkpoints, individual speaking roles, AI-use declarations, and student reflections. Shared language and communication criteria may also promote consistent assessment across study programs while accommodating different vocational contexts. Project illustrations and a proposed weighting scheme demonstrate practical applications. However, the framework's effectiveness and feasibility require empirical evaluation. Future research should pilot the adaptations, examine scoring consistency, investigate students' experiences of AI-use disclosure, and assess teacher workload to refine the framework and determine its suitability for wider implementation.

ACKNOWLEDGEMENTS

The authors thank the English teaching team for contributing to the course projects and providing insights into their design. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sector.

DECLARATION

Author Contributions: Bontisesari was responsible for developing the central research concept, conducting the literature synthesis, formulating the proposed assessment framework, and drafting the manuscript. Irma Rasita Gloria Barus contributed to the refinement of the conceptual framework, critical evaluation of relevant literature, academic supervision, and manuscript revision. Both authors participated in reviewing the arguments and approving the final manuscript.

Conflict of Interest: The authors confirm that there are no financial, professional, or personal relationships that could be perceived as having influenced the development or presentation of this study.

Funding: The authors conducted this research independently and did not receive financial assistance, research grants, or institutional funding from external organizations.

Ethical Approval: This article is a conceptual study based on the synthesis and critical examination of previously published literature. It did not involve experiments, human participants, or the collection of personal data. Therefore, formal ethical clearance was not applicable.

OPEN ACCESS
© 2021 The Author.This article is licensed CC BY SA 4.0.
visit Creative Commons Attribution-ShareAlike 4.0 International License.

Informed Consent: Informed consent was not applicable because this study did not recruit participants, conduct interviews, administer questionnaires, or collect data directly from individuals.

Use of Generative AI: Generative AI tools were used as supporting instruments for language editing and improving the clarity and readability of the manuscript. The development of the conceptual framework, interpretation of the literature, formulation of the six assessment principles, and academic arguments remained the responsibility of the authors. All AI-assisted suggestions were critically reviewed, and the authors take full responsibility for the content and scholarly integrity of the article.

REFERENCES

Bannert, M., Kasneci, E., Sessler, K., Stefan, K., Fischer, F., Gasser, U., Groh, G., Stephan, G., Poquet, O., Sailer, M., Schmidt, A., & Seidel, T. (2023). C*hatGPT for good? On opportunities and challenges of large language models for education.* 1–13.

Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. Assess*ing Writing, 57,* Art*i*cle 100745. https://doi.org/10.1016/j.asw.2023.100745

Basturkmen, H. (2010). *Develop*ing courses in English for specific purposes*. Palgrave Macmillan https://doi.org/10.1057/9780230290518

Beckett, G. H., & Slater, T. (2005). The project framework: A tool for language, content, and skills integration. *EL*T Journal,* 59*(2), 108–116. https://doi.org/10.1093/eltj/cci024

Bloxham, S., Hughes, C., & Adie, L. (2016). What's the point of moderation? A discussion of the purposes achieved through contemporary moderation practices. *Assessment & Evaluation in Higher Education,* 41*(4), 638–653. https://doi.org/10.1080/02602938.2015.1039932

Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. *Inte*rnational Journal of Educational Technology in Higher Education, 20*, Art*i*cle 38. https://doi.org/10.1186/s41239-023-00408-3

Corbin, T., Dawson, P., & Liu, D. (2025). Talk is cheap: Why structural assessment changes are needed for a time of GenAI. *Assessment & Evaluation in Higher Education, 50*(7*), 1087–1097. https://doi.org/10.1080/02602938.2025.2503964

Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *In*novations in Education and Teaching International,* 61*(2), 228–239. https://doi.org/10.1080/14703297.2023.2190148

Dawson, P., Bearman, M., Dollinger, M., & Boud, D. (2024). Validity matters more than cheating. *Assessment & Evaluation in Higher Education, 49*(7*), 1005–1016. https://doi.org/10.1080/02602938.2024.2386662

Dudley-Evans, T., & St. John, M. J. (1998). *De*velopments in English for specific purposes: A multi-disciplinary approach*. Cambridge University Press.

Gulikers, J. T. M., Bastiaens, T. J., & Kirschner, P. A. (2004). A five-dimensional framework for authentic assessment. *Educational Technology Research and Development*, *52*(3), 67–86. https://doi.org/10.1007/BF02504676

Hafner, C A. (2014). Embedding digital literacies in English language teaching: Students' digital video projects as multimodal ensembles. *TESOL Quarterly*, *48*(4), 655–685. https://doi.org/10.1002/tesq.138

Hafner, C A, & Ho, W. Y. J. (2020). Assessing digital multimodal composing in second language writing: Towards a process-based model. *Journal of Second Language Writing*, *47*, 100710. https://doi.org/10.1016/j.jslw.2020.100710

Hafner, Christoph A, & Miller, L. (2011). Fostering le*arner autonomy in English for science : a collaborative digital video project in a technological learning environment 1. 15(3), 68–86 https://doi.org/10.64152/10125/44263*

Hyland, K. (2007). Genre pedagogy: Language, literacy and L2 writing instruction. *Journal of*

OPEN ACCESS
© 2021 The Author.This article is licensed CC BY SA 4.0.
visit Creative Commons Attribution-ShareAlike 4.0 International License.

Second Language Writing, *16*(3), 148–164. https://doi.org/10.1016/j.jslw.2007.07.005

Ion, G., Díaz-Vicario, A., & Mercader, C. (2024). Making steps towards improved fairness in group work assessment: The role of students' self- and peer-assessment. *Active Learning in Higher Education*, *25*(3), 425–437. https://doi.org/10.1177/14697874231154826

Jaakkola, E. (2020). Designing conceptual articles: Four approaches. *AMS Review*, *10*(1–2), 18–26. https://doi.org/10.1007/s13162-020-00161-0

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, *50*(1), 1–73. https://doi.org/10.1111/jedm.12000

Kemendikbudristek. (2021). *Keputusan Menteri Pendidikan dan Kebudayaan Nomor 3/M/2021 tentang indikator kinerja utama perguruan tinggi negeri dan lembaga layanan pendidikan tinggi*. Kementerian Pendidikan dan Kebudayaan Republik Indonesia.

Kemendikbudristek. (2024). *Panduan penggunaan generative artificial intelligence (GenAI) pada pembelajaran di perguruan tinggi*. Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia.

Kress, G. (2010). *Multimodality: A social semiotic approach to contemporary communication*. Routledge.

Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against non-native English writers. *Patterns*, *4*(7), 100779. https://doi.org/10.1016/j.patter.2023.100779

Lodge, J. M., Howard, S., Bearman, M., & Dawson, P. (2023). *Assessment reform for the age of artificial intelligence*. Tertiary Education Quality and Standards Agency.

Meijer, H., Brouwer, J., Hoekstra, R., & Strijbos, J.-W. (2022). Exploring construct and consequential validity of collaborative learning assessment in higher education. *Small Group Research*, *53*(6), 891–925. https://doi.org/10.1177/10464964221095545

Perkins, M., Furze, L., Roe, J., & MacVaugh, J. (2024). The Artificial Intelligence Assessment Scale (AIAS): A framework for ethical integration of generative AI in educational assessment. *Journal of University Teaching and Learning Practice*, *21*(6). https://doi.org/10.53761/q3azde36

Perkins, M., Roe, J., Postma, D., McGaughran, J., & Hickerson, D. (2024). Detection of GPT-4 generated text in higher education: Combining academic judgement and software to identify generative AI tool misuse. *Journal of Academic Ethics*, *22*(1), 89–113. https://doi.org/10.1007/s10805-023-09492-6

Puteri, C. G., Cahyono, B. Y., Widiati, U., & Suryati, N. (2026). Using ChatGPT as a writing assistant: A study on essay quality development among Indonesian EFL university students. *Indonesian Journal of Applied Linguistics*, *15*(2), 332–346. https://doi.org/10.17509/3ebmy158

Stoller, F. L. (2006). Establishing a theoretical foundation for project-based learning in second and foreign language contexts. In G. H. Beckett & P. C. Miller (Eds.), *Project-based second and foreign language education: Past, present, and future* (pp. 19–40). Information Age Publishing.

Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: Enabling students to make decisions about the quality of work. *Higher Education*, *76*(3), 467–481. https://doi.org/10.1007/s10734-017-0220-3

Villarroel, V., Bloxham, S., Bruna, D., Bruna, C., & Herrera-Seda, C. (2018). Authentic assessment: Creating a blueprint for course design. *Assessment & Evaluation in Higher Education*, *43*(5), 840–854. https://doi.org/10.1080/02602938.2017.1412396

Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O., Šigut, P., & Waddington, L. (2023). Testing of detection tools for AI-generated text. *International Journal for Educational Integrity*, *19*, 26. https://doi.org/10.1007/s40979-023-00146-z

Xia, Q., Weng, X., Ouyang, F., Lin, T. J., & Chiu, T. K. F. (2024). A scoping review on how generative artificial intelligence transforms assessment in higher education. *International*

OPEN ACCESS
© 2021 The Author.This article is licensed CC BY SA 4.0.
visit Creative Commons Attribution-ShareAlike 4.0 International License.

Copyright (c) 2026 Bontisesari & Irma Rasita Gloria Barus.

Journal of Educational Technology in Higher Education, 21, 40. https://doi.org/10.1186/s41239-024-00468-z

OPEN ACCESS

© 2021 The Author.This article is licensed CC BY SA 4.0.
visit Creative Commons Attribution-ShareAlike 4.0 International License.