

Preparing Pre-Service English Language Teachers to Assess Speaking and Writing Through Digital Analytic Rubrics: A Competency-Based Methodological Framework

Ahmadjonova Shahnoza Mahamadjon qizi

Tashkent State Technical University, Department "Foreign Languages" Uzbekistan

Received: 29 August 2026; **Accepted:** 25 September 2026; **Published:** 22 October 2026

Abstract: Digital assessment environments create new opportunities for evaluating second-language speaking and writing, yet technology alone cannot secure defensible scores or useful feedback. Pre-service English language teachers need integrated preparation in defining assessment constructs, developing analytic descriptors, applying digital rubrics, calibrating judgments, and using results to support revision. This conceptual and methodological article synthesizes foundational work on rubric quality, language assessment literacy, teacher learning, and the Common European Framework of Reference for Languages (CEFR). It proposes a six-stage, competency-based training framework and illustrates its application through task-specific speaking and writing rubrics, moderation activities, and a performance-based assessment plan for teacher candidates. The framework treats digital rubrics as instruments for making evidence visible and supporting professional judgment, not as automated substitutes for that judgment. Its potential contributions are explicit alignment between communicative constructs and rubric criteria, calibration as a teachable practice, and feedback uptake as an integral component of assessment competence. No participant data were collected and no effectiveness claims are made. The paper concludes with implementation safeguards and a feasible research agenda for examining validity, rater agreement, feedback quality, and transfer to teaching practice.

Keywords: Digital analytic rubric; pre-service English teachers; speaking assessment; writing assessment; language assessment literacy; CEFR; rater calibration; formative feedback.

Introduction: Evaluating productive language skills is a demanding professional task. A learner may speak fluently while offering weak arguments, or write an accurate paragraph that does not address its communicative purpose. Teachers must therefore decide what performance represents, which features matter for a particular task, and how observed evidence supports a judgment. Digital learning management systems can store criteria, ratings and comments, but they do not resolve these underlying assessment questions.

Teacher assessment literacy is better understood as situated professional practice than as possession of isolated testing knowledge. Xu and Brown (2016)

describe an evolving relationship among knowledge, conceptions, contexts and teacher learning. For future English teachers, this suggests that rubric design must be taught together with the interpretation of performances, discussion of rating disagreements and pedagogical follow-up. Research on scoring rubrics likewise indicates that clarity and assessor preparation matter for dependable use (Jonsson & Svingby, 2007; Knoch, 2009).

The present paper asks: (1) What competencies are needed to use digital analytic rubrics for speaking and writing? (2) How can teacher educators sequence activities to develop these competencies? (3) What evidence could establish whether such preparation works? The article develops a transparent conceptual

proposal rather than reporting an intervention.

CONCEPTUAL FOUNDATIONS

1. Assessment literacy and professional judgment

Assessment literacy encompasses the selection of appropriate methods, understanding of evidence, interpretation of results and ethical use of judgments. The Teacher Assessment Literacy in Practice perspective locates these decisions within institutional and sociocultural conditions (Xu & Brown, 2016). Accordingly, teacher candidates should be asked not only to create a rubric, but to defend why each criterion belongs in it, whether descriptors are distinguishable, and what consequences follow from its use. An intervention study in pre-service language teacher education also highlights the importance of linking explicit assessment instruction with practical evaluation activities (Damacena-Dutra & Borges-Almeida, 2024).

2. What analytic rubrics can and cannot do

An analytic rubric separates a complex performance into dimensions and describes performance levels for each. This structure can support diagnostic comments and make judgments more transparent. Yet descriptors that repeat vague adjectives such as “good” and “poor” can disguise disagreement rather than reduce it. Knoch (2009), comparing writing rating scales, found that more detailed descriptors supported more reliable distinctions in that study. This does not mean that more detail is always better: excessively elaborate rubrics can overload novice raters and fragment the communicative whole. Rubric development therefore involves balancing construct coverage, observability and usability.

Rubrics may support formative processes when learners understand criteria and have opportunities to act on feedback (Panadero & Jönsson, 2013). Conversely, a digital gradebook that simply displays numbers remains primarily a reporting tool. Teachers should therefore plan the next learning action at the same time as the assessment.

3. CEFR-informed, task-specific assessment

The CEFR Companion Volume (Council of Europe, 2020) offers illustrative descriptors for communicative language activities, including spoken production,

interaction and written production. These descriptors are a reference for curriculum alignment, not a substitute for local task specifications or a validated scoring instrument. A paired discussion, an individual presentation and an argumentative essay call for partly different criteria. An assessment that uses the same dimensions for every task risks construct underrepresentation or the inclusion of irrelevant features.

4. Digital affordances and risks

Rubric functions within learning management systems can organize criterion-level judgments, attach comments to work, archive successive performances and facilitate moderation. Audio recordings and annotated texts also allow evidence to be revisited. The same features raise concerns about data protection, consent, access inequality and the temptation to equate machine-generated scores with warranted judgments. The teacher educator should select technology only after clarifying the assessment purpose and local privacy requirements.

METHOD

This article uses a theory-informed conceptual synthesis and pedagogical modelling approach. It is not a systematic review or an empirical evaluation. The sources were selected for their relevance to four complementary areas: teacher assessment literacy, analytic rubric quality, language performance assessment, and formative use of feedback. Ideas from these areas were translated into a training sequence and observable competence indicators. The proposed rubrics are illustrative author-designed teaching instruments; they have not undergone psychometric validation.

The framework is guided by three design principles. Construct alignment requires criteria to represent the communicative ability targeted by a task. Evidential accountability requires raters to justify judgments with observable performance features. Feedback usability requires students to receive information that can guide a subsequent attempt. These principles are revisited throughout the training sequence.

PROPOSED SIX-STAGE TRAINING FRAMEWORK

Stage	Teacher-candidate activity	Observable output
-------	----------------------------	-------------------

1. Construct orientation	Compare tasks, CEFR-informed objectives and evidence requirements	Task-specific assessment blueprint
2. Descriptor design	Draft and critique analytic criteria and performance bands	Rubric with distinguishable descriptors
3. Digital implementation	Build rubric and feedback workflow in an approved platform	Functional digital assessment activity
4. Rater calibration	Independently score common performances; discuss discrepancies	Evidence-based moderation notes
5. Feedback and uptake	Give criterion-linked guidance and design a revision opportunity	Feedback plus learner action plan
6. Reflective refinement	Audit rubric performance and revise problematic descriptors	Revised rubric and rationale

The sequence is iterative rather than strictly linear. A calibration discussion may reveal an ambiguous descriptor, sending candidates back to design; learner revisions may reveal that comments were too general, prompting feedback redesign. The essential unit of professional learning is therefore an assessment cycle, not a single platform operation.

1. Stage 1: Construct orientation

Teacher candidates begin with two contrasting tasks: a two-minute persuasive monologue and a paired problem-solving conversation. They discuss whether interactional responsiveness belongs in both assessments. In the monologue, organization and intelligibility may be central; in the paired task, turn management and response to a partner become more salient. Candidates create a one-page blueprint specifying task, target learners, intended construct, evidence and anticipated feedback.

2. Stage 2: Descriptor design

Candidates transform the blueprint into a manageable analytic rubric. They replace unobservable judgments (“excellent English”) with evidence-linked descriptions (“develops a position with relevant supporting examples”). Adjacent levels should be distinguishable without pretending that proficiency develops in perfectly uniform steps. Candidates also check whether accent, speed, topic familiarity or handwriting could unfairly influence a criterion unrelated to the target construct.

3. Stage 3: Digital implementation

Candidates transfer their rubric into an institutional learning management system, verify that each criterion and performance level is displayed correctly, and test the workflow using anonymized sample performances. They practice attaching comments and exporting feedback. The platform is treated as a recordkeeping and communication environment; the assessor retains responsibility for judgments.

4. Stage 4: Calibration and moderation

Two or more candidates rate the same recordings or texts independently. They compare criterion-level judgments, cite evidence and document why their scores differ. A facilitator supplies anchor examples and asks participants to revise ambiguous descriptors. In a subsequent empirical study, agreement can be assessed with statistics appropriate to the rating scale and design; simple percentage agreement should not be presented as a complete validity argument.

5. Stage 5: Feedback and learner action

Each candidate writes one strength statement and one actionable next step tied to a rubric criterion. For example: “Your position is clear; add a concrete example in paragraph two to strengthen the supporting argument.” Learners then revise a text or repeat an oral task. The revision provides evidence of feedback uptake, which is distinct from merely delivering a comment.

6. Stage 6: Reflective refinement

Candidates audit whether criteria were consistently understood, whether feedback matched evidence and

whether learners could use it. A short reflective memo documents changes to descriptors and identifies risks such as halo effects, excessive attention to grammar or insufficient consideration of task achievement.

ILLUSTRATIVE ASSESSMENT INSTRUMENTS

The following four-level scales are examples for

Criterion	4 — Accomplished	3 — Competent	2 — Emerging	1 — Limited
Argument	Clear position with relevant, developed support	Position clear; support generally relevant	Position present; support thin	Position unclear or unsupported
Organization	Ideas sequenced with effective signposting	Mostly logical progression	Uneven progression	Sequence difficult to follow
Fluency	Sustains message with manageable pauses	Some hesitation, message maintained	Frequent hesitation disrupts flow	Pauses impede communication
Language resources	Range and control support precise meaning	Adequate range; errors rarely obscure meaning	Limited range; some errors affect meaning	Restricted resources often obscure meaning
Intelligibility	Consistently easy to understand	Generally intelligible	Occasional difficulty understanding	Frequent difficulty understanding

For paired interaction, the educator should add or substitute a criterion on responsiveness and turn management. Pronunciation should be evaluated in

methodology workshops, not official CEFR scales. Before classroom or high-stakes use, they should be adapted to task purpose, language level, context and local validation evidence.

1. Speaking: short persuasive presentation

relation to intelligibility and communicative purpose, not conformity to a single prestige accent.

2. Writing: short argumentative essay

Criterion	4 — Accomplished	3 — Competent	2 — Emerging	1 — Limited
Task response	Fully addresses prompt with developed reasoning	Addresses prompt with adequate support	Partially addresses prompt	Substantially misses task
Organization	Purposeful paragraphing and progression	Generally coherent progression	Uneven organization	Ideas difficult to follow
Evidence	Specific, well-explained support	Relevant support, sometimes general	Limited or weakly connected support	Little relevant support
Lexical choice	Precise and appropriate choices	Mostly appropriate vocabulary	Repetition or imprecision affects expression	Restricted or frequently inappropriate vocabulary
Language control	Varied structures with good control	Errors present but meaning clear	Recurring errors sometimes obscure meaning	Errors frequently obscure meaning

These criteria deliberately distinguish organization from

supporting evidence. The rubric should be revised when

the writing genre changes: a formal email, reflective journal and argumentative essay do not share identical communicative priorities.

ASSESSMENT OF TEACHER-CANDIDATE COMPETENCE

The framework proposes a performance-based

portfolio rather than a single knowledge test. Each candidate submits an assessment blueprint, a digital rubric, two annotated sample ratings, moderation notes, two feedback examples and a reflective revision memo. These artifacts make professional reasoning visible.

Competence domain	Evidence of proficiency	Potential weakness
Construct alignment	Criteria directly represent task purpose	Irrelevant or missing criteria
Descriptor quality	Observable, differentiated bands	Vague or overlapping bands
Digital workflow	Rubric usable; feedback retrievable	Technical setup obstructs use
Rating justification	Claims supported by performance evidence	Scores based on impression
Feedback usefulness	Specific guidance plus opportunity to act	Generic praise or score only
Reflection and ethics	Revision rationale; secure data handling	No audit of fairness or privacy

For a future evaluation, independent expert review could examine content relevance and clarity; calibration tasks could measure agreement; feedback could be coded for specificity and actionability; and follow-up teaching observations could assess transfer. No numeric cutoffs are proposed here because thresholds require defensible empirical and contextual justification.

DISCUSSION: CONTRIBUTION AND BOUNDARY CONDITIONS

The distinctive contribution of the proposal is not the invention of analytic rubrics, which are well established, but the integration of rubric design, digital implementation, moderated scoring and feedback uptake into one assessable preparation cycle. This integration responds to the gap between knowing assessment terminology and exercising informed judgment in classroom conditions. It is consistent with the view of assessment literacy as practice (Xu & Brown, 2016) and with evidence that detailed rating criteria and rater preparation can matter for performance assessment (Knoch, 2009).

The framework should not be interpreted as guaranteeing objective ratings. Rater training can improve shared understanding while contextual factors and legitimate interpretive differences remain. Nor should digital platforms be assumed to increase

learning by themselves. Where access is unequal, teacher educators should provide alternative submission routes and avoid making digital fluency a hidden measure of language proficiency.

A further consideration is emerging AI assistance in rubric construction. Generative systems may produce draft descriptors rapidly, but those drafts still require scrutiny for construct relevance, fairness, distinguishability and task alignment. Teacher candidates can use such outputs as material for critical comparison while retaining full responsibility for the final rubric and for decisions about learner performance.

LIMITATIONS AND FUTURE RESEARCH

The article is conceptual. It does not report a sampled literature review, classroom implementation, participant interviews, reliability estimates or learning gains. The sample rubrics are instructional prototypes and should not be used for consequential certification decisions without appropriate validation. The proposed stages may also require adjustment to institutional infrastructure and local assessment traditions.

A feasible next study would compare teacher candidates receiving conventional rubric instruction with candidates participating in the full six-stage cycle. Both groups could complete equivalent pre- and post-training tasks using unseen speaking recordings and

writing samples. Outcomes could include rubric quality, justified scoring decisions, agreement between raters and actionable feedback. Interviews and reflective journals would illuminate why particular changes occurred. Ethical approval, consent and secure handling of learner work would be essential.

CONCLUSION

Digital analytic rubrics can make assessment criteria visible and support structured feedback, but their value depends on informed professional use. Preparing pre-service English language teachers to assess speaking and writing requires explicit work on construct definition, descriptor quality, digital implementation, calibration, feedback and reflection. The six-stage framework offered here provides a coherent sequence of learning activities and portfolio evidence for teacher education programs. Its practical promise is plausible but remains a hypothesis to be tested through careful empirical research.

Declarations

Funding: [Insert verified funding information or “No external funding was received.”]

Conflicts of interest: [Author to confirm.]

Ethics statement: No human participants or personal data were collected for this conceptual article.

Data availability: No empirical dataset was generated or analyzed.

AI-use disclosure: [Complete according to actual use and target journal policy.]

REFERENCES

1. Andrade, H. G. (2005). Teaching with rubrics: The good, the bad, and the ugly. *College Teaching*, 53(1), 27–31. <https://doi.org/10.3200/CTCH.53.1.27-31>
2. Council of Europe. (2020). *Common European Framework of Reference for Languages: Learning, teaching, assessment—Companion volume*. Council of Europe Publishing. <https://www.coe.int/en/web/common-european-framework-reference-languages/cefr-companion-volume-and-its-language-versions>
3. Damacena-Dutra, M., & Borges-Almeida, V. (2024). Exploring evidence of assessment literacy development in preservice language teacher

education: Reliability and conceptions of assessment. *Revista X*, 19(2). <https://doi.org/10.5380/rvx.v19i2.94734>

4. Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2(2), 130–144. <https://doi.org/10.1016/j.edurev.2007.05.002>
5. Knoch, U. (2009). Diagnostic assessment of writing: A comparison of two rating scales. *Language Testing*, 26(2), 275–304. <https://doi.org/10.1177/0265532208101008>
6. Panadero, E., & Jönsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9, 129–144. <https://doi.org/10.1016/j.edurev.2013.01.002>
7. Panadero, E., & Jönsson, A. (2020). A critical review of the arguments against the use of rubrics. *Educational Research Review*, 30, Article 100329. <https://doi.org/10.1016/j.edurev.2020.100329>
8. Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58, 149–162. <https://doi.org/10.1016/j.tate.2016.05.010>