Journal of Inclusive Methodology and Technology in Learning and Teaching
ISSN 2785-5104
Anno 6 n. 3 (2026)

EDIZIONI UNIVERSITARIE ROMANE
editoria scientifica dal 1952

Engaging Teachers in Assessment Design: RubricaBot as a Reflective Mediator for Assessment Literacy and AI Literacy

Federica Emanuel [1] *, Marco Rondonotti [1], Angelo Basta [1] and Paolo Raviolo [1]

[1] eCampus University, Department of Humanities and Social Sciences (DiSUS); federica.emanuel@uniecampus.it; marco.rondonotti@uniecampus.it; angelo.basta@uniecampus.it; paolo.raviolo@uniecampus.it
* Correspondence: federica.emanuel@uniecampus.it

Abstract: Generative Artificial Intelligence is increasingly entering assessment practices, raising questions about how teachers engage with AI-generated responses while retaining professional judgement. This study investigates teachers' perceptions of RubricaBot, a chatbot designed to support rubric construction. A total of 956 participants enrolled in teacher qualification programmes first received training on rubric-based assessment, then constructed a rubric related to their own teaching context and subsequently used RubricaBot to review it. A post-activity questionnaire examined assessment practices, perceptions of autonomous rubric design, and the usefulness, ease of use, limitations, reflective potential, and formative value of RubricaBot. Autonomous rubric construction was perceived as moderately demanding. Participants with previous assessment-specific training reported significantly greater confidence, ease in defining criteria and levels, rapidity, and perceived structure of the process. RubricaBot was generally perceived as useful, easy to use, and supportive of reflection, while concerns about reliability and contextual adequacy remained. The findings suggest that AI-supported rubric design may provide a useful setting for integrating assessment literacy and AI literacy while preserving teachers' responsibility for assessment decisions.

Keywords: assessment literacy; AI literacy; generative artificial intelligence; rubric-based assessment; teacher education

Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction[1]

1.1 Assessment literacy, GenAI, and professional judgement

Generative Artificial Intelligence (GenAI) is reshaping educational practices and the professional competences required of teachers. Assessment is a particularly sensitive area: generative tools can propose criteria, performance descriptors, feedback, and assessment instruments, but assessment requires interpretation of evidence, alignment with learning goals, contextual understanding, and responsibility for evaluative decisions.

[1] The paper represents the result of a collaborative effort by the authors; specifically, FE authored the sections 1.1, 3 and 4; MR the sections 2 and 5; AB the sections 1.2; PR the section 6.

EDIZIONI UNIVERSITARIE ROMANE
editoria scientifica dal 1952

This distinction matters in formative assessment, where evidence about learning informs subsequent teaching and learning rather than merely certifying performance (Black & Wiliam, 2009). Explicit criteria, standards, and feedback make expectations visible and support revision. Teachers therefore need to articulate and justify the reasoning underlying assessment decisions.

Teacher assessment literacy extends beyond familiarity with assessment techniques to designing, selecting, interpreting, and using assessment within specific educational contexts. Pastore and Andrade (2019) conceptualise it as a multidimensional competence encompassing conceptual knowledge, practical capabilities, and sociocultural dimensions.

GenAI expands these requirements. Cui et al. (2025) highlights the interconnection between cognitive, professional, contextual, and ethical competences in GenAI-enhanced assessment. Teachers must evaluate the pedagogical appropriateness of assessment proposals alongside the reliability and contextual relevance of generated outputs.

Technology-acceptance models provide useful constructs for investigating teachers' perceptions of new technologies, particularly perceived usefulness and ease of use (Davis, 1989; Venkatesh et al., 2003). However, usefulness and ease of use alone are insufficient to establish the pedagogical value of AI-supported assessment.

Professional judgement is therefore central to AI-supported assessment. Bearman et al. (2024) argue that GenAI increases the importance of evaluative judgement. Although their analysis primarily concerns learners, the principle also applies to teachers: AI-generated outputs remain objects of evaluation, not authoritative solutions. Xia et al. (2024) likewise identify the joint development of assessment, AI, and digital literacy as necessary for responding to GenAI-related transformations in assessment.

1.2 RubricaBot and professional rubric design

Rubrics provide a particularly relevant context in which assessment reasoning becomes explicit. Through criteria, performance levels, and descriptors, they can clarify expectations, increase transparency, and support feedback and formative assessment processes (Brookhart, 2018; Panadero & Jonsson, 2013).

Effective rubric design requires criteria aligned with learning outcomes, distinguishable performance levels, observable descriptors, and coherence between the task and expected performance.

Constructing rather than simply using a rubric requires teachers to translate assessment intentions into explicit criteria and performance standards, making their professional reasoning available for examination.

GenAI can provide alternative formulations against which teachers can examine their assessment choices.

RubricaBot is a customised chatbot built on ChatGPT to support rubric construction and revision in teacher education (di Padova et al., 2026). Its configuration uses structured instructions to define the assessment task and guide the formulation of dimensions, indicators, and performance levels. The design draws on information about the school level, subject, learning objectives, activity, and product or task to be assessed. These inputs provide the context for proposing observable descriptors and distinguishing levels of performance.

The interaction is designed to be incremental: information supplied in earlier exchanges informs subsequent proposals, which teachers can refine through further dialogue. The configuration aims to support internal coherence and clear language; it does not establish the validity of the resulting rubric. Teachers remain responsible for checking whether descriptors capture the intended learning, whether differences between levels are meaningful, and whether proposed criteria are appropriate for their learners. Generic formulations require revision rather than automatic adoption.

In the present study, participants constructed a rubric autonomously before using RubricaBot to review it. This sequence preserved an initial phase of independent assessment reasoning and positioned generated suggestions as material for comparison and revision.

RubricaBot is considered here as a potential reflective mediator. Its pedagogical role lies in supporting teachers as they examine alternatives, justify revisions, or reject unsuitable suggestions. The study investigates perceptions of this interaction rather than assuming that chatbot use improves assessment competence.

The study therefore addressed the following research questions:

RQ1. What assessment practices and perceptions of autonomous rubric design characterize participating teachers?

RQ2. How do teachers perceive the usefulness, ease of use, limitations, reflective potential, and formative value of RubricaBot in rubric design?

RQ3. Do perceptions of autonomous rubric design differ according to previous assessment-specific professional development?

2. Materials and Methods

2.1 Research design and participants

Data were collected from 956 participants enrolled in Italian university teacher qualification programmes who took part in the assessment-training activity.

Women represented 64.3% of respondents and men 34.2%; 0.3% identified as non-binary and 0.6% preferred not to answer. The participants' mean age was 41.99 years (SD = 9.36), ranging from 24 to 64 years. Most participants worked in upper-secondary education (51.2%) or lower-secondary education (33.1%), while 8.3% were not yet teaching. The average length of teaching experience was 5.21 years (SD = 4.48, range = 0-42).

Participants came from all areas of Italy: 49.0% were from Northern Italy, 21.9% from Central Italy, and 29.1% from Southern Italy and the Islands. Only 19.8% (n = 189) reported previous participation in professional-development activities specifically focused on assessment, while 79.6% (n = 761) reported no previous specific training.

Participation in the questionnaire was voluntary. Before accessing it, participants received an information notice concerning the study and the processing of collected data and were required to provide explicit consent to participate.

2.2 Training activity and procedure

The training activity was implemented as part of the assessment component of the teacher qualification programmes.

EDIZIONI UNIVERSITARIE ROMANE
editoria scientifica dal 1952

Before the practical activity, participants received instruction on rubric-based assessment, including the function and structure of rubrics and the role of learning objectives, criteria, descriptors, and performance levels.

During the training session, participants were then asked to construct an assessment rubric autonomously. Each participant developed a rubric related to their own teaching context and disciplinary area.

In the subsequent phase, participants used RubricaBot to review the rubric they had independently constructed. At the end of the activity, participants completed the questionnaire through the QuestionPro platform.

The questionnaire collected demographic and professional information, previous experience of assessment training, information about current assessment practices, and participants' perceptions of the autonomous and AI-supported phases of the activity. The questionnaire investigated the following areas: current assessment practices; perceived confidence and ease in autonomous rubric construction; critical engagement with RubricaBot-generated suggestions; perceived ease of use and clarity of interaction; perceived usefulness and efficiency of the chatbot; adaptability of AI-generated outputs to the teaching context; intention to use and recommend RubricaBot; perceived reliability and contextual limitations; and the reflective, metacognitive, and formative value of the activity.

2.3 Data analysis

Descriptive statistics were used to summarise participant characteristics, assessment practices, and responses concerning autonomous and RubricaBot-supported rubric design. Differences between participants with and without previous assessment-specific professional development were examined using Welch's independent-samples t-tests. Effect sizes were estimated using Cohen's d.

Analyses were conducted using IBM SPSS Statistics 30.

3. Results

3.1 Assessment practices

Participants reported the frequent use of both conventional and teacher-led assessment practices and some feedback-oriented approaches. Teacher-developed assessments were the most frequently reported (M = 4.96), followed by classroom discussion (M = 4.48) and oral feedback during the learning process (M = 4.38).

Traditional written tests (M = 4.21) and multiple-choice tests (M = 4.03) were also frequently used. Rubrics occupied an intermediate position (M = 3.60), whereas digital assessment tools (M = 3.05), metacognitive diaries (M = 2.68), and portfolios/e-portfolios (M = 2.30) were used less frequently.

These results describe a context in which participants were familiar with several assessment practices, while rubric use was neither uncommon nor among the most frequently adopted approaches.

3.2 Autonomous rubric construction

The autonomous phase was perceived as moderately demanding. Participants reported relatively limited confidence in independently constructing a rubric (M = 2.90) and similarly modest ease in defining criteria and performance levels (M =

2.92). The construction process was not perceived as particularly rapid (M = 2.78). Its organisation received a somewhat more positive evaluation, with participants reporting that the process was relatively well structured (M = 3.30).

Previous assessment-specific professional development was associated with more positive perceptions of autonomous rubric construction across all four aspects considered. Participants with previous training reported greater confidence in constructing a rubric, greater ease in defining criteria and performance levels, higher perceived rapidity of the construction process, and a more positive evaluation of its structure (Table 1).

Table 1. Autonomous rubric design according to previous assessment training

	Previous training M (n = 189)	**No previous training** M (n = 761)	**Welch's t**	**Cohen's d**
Rubric confidence	3.26	2.80	5.32, p < .001	0.43
Criteria and levels	3.21	2.85	4.07, p < .001	0.34
Perceived rapidity	3.06	2.71	3.59, p < .001	0.31
Process structure	3.49	3.26	2.65, p = .008	0.23

The largest difference concerned confidence in rubric construction, followed by ease in defining criteria and performance levels. Smaller differences emerged for perceived rapidity and process structure.

3.3 Perceptions of interaction with RubricaBot

Overall, participants evaluated the RubricaBot-supported activity positively (Table 2). RubricaBot was perceived as easy to use and useful for rubric construction, while participants also reported critical engagement with the assessment process. In particular, participants reported that comparison with the chatbot stimulated reflection on the quality of their rubric, whereas acceptance of chatbot responses without modification was more moderate.

Table 2. Selected perceptions of the RubricaBot-supported activity

Area	**Selected item**	**M**
Ease of use	RubricaBot was easy to use	4.07
Usefulness	RubricaBot was useful for rubric construction	4.09
Critical engagement	Comparison stimulated critical reflection on rubric quality	3.82
	Responses were accepted without modification	3.00
Limitations	Doubts about content reliability	3.10
	Difficulty capturing the specific educational context	3.13
Reflective value	Active comparison with the chatbot-supported rubric	3.89
Formative value	Overall experience was formative	3.98

Note. Items were rated on a five-point scale.

Positive ratings were also observed for generating new ideas or strategies (M = 3.88), deeper reflection (M = 3.86), making assessment criteria more explicit (M = 3.85), and identifying previously overlooked aspects (M = 3.82).

Overall, positive ratings of ease of use and usefulness coexisted with moderate concerns about reliability of generated content and its ability to capture the specific educational context. At the same time, participants positively evaluated the reflective and formative contribution of the activity, particularly the active comparison between their original rubric and the chatbot-supported version.

4. Discussion

The study examined teachers' perceptions of an activity in which autonomous rubric construction preceded interaction with RubricaBot. The findings concern difficulties in autonomous design, prior assessment training, and positive chatbot perceptions accompanied by critical attention to its outputs.

Rubric use occupied an intermediate position among participants' assessment practices, while confidence in construction and ease in defining criteria and performance levels remained modest. Familiarity with assessment practice may therefore differ from confidence in making criteria and standards explicit. Translating intended outcomes into descriptors and meaningful performance levels can support assessment literacy development (Brookhart, 2018; Pastore & Andrade, 2019).

Previous assessment-specific professional development was associated with more positive perceptions of autonomous rubric construction. Differences were small to moderate and largest for confidence. The findings suggest a possible contribution of specific training to the perceived manageability of rubric design, without establishing a causal effect. The limited proportion reporting such training reinforces the importance of professional development focused on assessment design, not only tools (Rondonotti & Emanuel, 2025).

Participants perceived RubricaBot as easy to use and useful for rubric construction, consistent with technology-acceptance constructs (Davis, 1989; Venkatesh et al., 2003). Positive perceptions coexisted with reported critical reflection and evaluation of chatbot responses, while acceptance without modification remained moderate. Favourable perceptions therefore need not imply unconditional reliance. Professional judgement remains essential for evaluating generated criteria, descriptors, and performance levels (Bearman et al., 2024).

Participants reported that the interaction helped make criteria explicit, identify overlooked aspects, stimulate reflection, and generate ideas or strategies. These are perceptions, not evidence of an objective improvement in assessment competence. Their interpretation must also consider the structure of the training activity.

Training on rubric-based assessment and autonomous construction preceded chatbot use. This sequencing may preserve independent assessment reasoning and make GenAI a resource for comparison and revision. Teachers retain responsibility for the assessment process rather than using generation solely as a shortcut to produce materials.

The activity may connect assessment literacy with AI literacy: teachers need assessment knowledge to judge proposed descriptors and levels, and AI literacy to recognise the limitations of generated content. Its educational value may reside in the judgement exercised while examining suggestions.

5. Limitations and future research

The study relies on self-reported perceptions collected immediately after the activity, so it provides no objective evidence of improved assessment literacy. The sequential design included no experimental or control conditions, and rubric quality was not independently evaluated before and after chatbot use. Conclusions about improved alignment, criteria, descriptors, or performance levels are therefore not warranted.

RubricaBot also requires scrutiny as a generative tool. Its intended contextualisation depends on the information supplied by the teacher, and plausible wording may conceal generic criteria or insufficiently differentiated levels (di Padova et al., 2026). Self-report data do not establish how teachers handled these issues in individual exchanges. Accordingly, perceived ease of use cannot establish the validity, fairness, or accessibility of generated rubrics.

Future studies should compare independently evaluated rubrics before and after AI-supported revision, and examine transfer to classroom assessment. Recording the model version, configuration instructions, prompts, and exchanges would allow analysis of suggestions accepted, modified, or rejected. Such analysis could relate revision choices to prior assessment expertise and test output consistency across repeated requests. Evaluating clarity, curricular alignment, and accessibility across learner needs would help determine whether adaptations preserve the intended construct rather than standardising assessment.

6. Conclusions

The study's findings suggest that AI-supported rubric design provides a significant context in which to observe the intersection between assessment literacy and AI literacy.

The autonomous construction of rubrics remains a relatively challenging task for teachers, particularly with regard to defining criteria and performance levels. Participants with previous assessment-specific training expressed more positive perceptions of the process, albeit with small to moderate differences. This finding highlights the importance of continuing to invest in training programmes specifically dedicated to assessment.

RubricaBot was generally perceived as useful, easy to use, and supportive of assessment design. At the same time, participants did not express unconditional trust in its outputs: some reservations remain regarding its reliability and contextual adequacy. This finding appears consistent with the literature, which identifies professional judgement as an essential component of teachers' AI literacy and emphasises the need to maintain a balance between AI support and human responsibility in assessment processes (Cui et al., 2025; Prilop et al., 2025; Sperling et al., 2024).

Participants also reported that the activity encouraged them to reflect on their own rubric, to clarify assessment criteria, to identify previously overlooked aspects, and to acquire new ideas and strategies. These findings support the interpretation of RubricaBot as a potential reflective mediator, provided its use is integrated into a training programme that preserves the teacher's role in examining, validating, and contextualising the proposals generated by AI.

The educational significance of GenAI in assessment may therefore lie in its capacity to create opportunities for making professional reasoning more visible,

subjecting it to critical discussion, and reconsidering the assumptions underlying assessment decisions. From this perspective, initial teacher education and continuing professional development have a strategic role to play: educational institutions should integrate AI not as an isolated technical competence, but within learning experiences that foster the joint development of assessment literacy, AI literacy, and professional responsibility.

References

Bearman, M., Tai, J., Dawson, P., Boud, D., & Ajjawi, R. (2024). Developing evaluative judgement for a time of generative artificial intelligence. *Assessment & Evaluation in Higher Education, 49*(6), 893–905. https://doi.org/10.1080/02602938.2024.2335321

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*(1), 5–31. https://doi.org/10.1007/s11092-008-9068-5

Brookhart, S. M. (2018). Appropriate criteria: Key to effective rubrics. *Frontiers in Education,* 3, Article 22. https://doi.org/10.3389/feduc.2018.00022

Cui, Y., Meng, Y., & Tang, L. (2025). Reconsidering teacher assessment literacy in GenAI-enhanced environments: A scoping review. *Teaching and Teacher Education, 165*, 105163. https://doi.org/10.1016/j.tate.2025.105163

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340. https://doi.org/10.2307/249008

di Padova, M., Basta, A., Tinterri, A., & Dipace, A. (2026). RubricaBot: Un supporto intelligente alla progettazione valutativa nella formazione dei docenti. In L. Giannandrea, F. Gratani, & L. M. Capolla (Eds.), *Valutazione e Intelligenza Artificiale. Prospettive didattiche, etiche e metodologiche* (pp. 311–324). Pensa Evolution.

Panadero, E., & Jonsson, 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

Pastore, S., & Andrade, H. L. (2019). Teacher assessment literacy: A three-dimensional model. *Teaching and Teacher Education, 84*, 128–138. https://doi.org/10.1016/j.tate.2019.05.003

Prilop, C. N., Mah, D.-K., Jacobsen, L. J., Hansen, R. R., Weber, K. E., & Hoya, F. (2025). Generative AI in teacher education: Educators' perceptions of transformative potentials and the triadic nature of AI literacy explored through AI-enhanced methods. *Computers and Education: Artificial Intelligence, 9,* 100471. https://doi.org/10.1016/j.caeai.2025.100471

Rondonotti, M., & Emanuel, F. (2025). Assessment literacy e formazione continua: Le posture valutative del personale docente. *Pedagogia Oggi, 23*(2), 226–233. https://doi.org/10.7346/PO-022025-27

Sperling, K., Stenberg, C.-J., McGrath, C., Åkerfeldt, A., Heintz, F., & Stenliden, L. (2024). In search of artificial intelligence (AI) literacy in teacher education: A scoping review. *Computers and Education Open, 6*, 100169. https://doi.org/10.1016/j.caeo.2024.100169

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly, 27*(3), 425–478. https://doi.org/10.2307/30036540

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