STIAEM

Slow Tech Innovation

AI, Ethics and

New Teaching Methods

Designing Authentic Assessment Tasks Rooted in Real-World Contexts

Co-funded by the European Union

2025-1-IT02-KA220-SCH-000365213

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them... Project reference code: 2025-1-IT02-KA220-SCH-000365213

Università di Foggia

Studi Umanistici, Lettere, Beni Culturali, Scienze della Formazione

İSTANBUL - ESENLER

İLÇE MİLLİ EĞİTİM MÜDÜRLÜĞÜ

1474

STIAEM

Slow Tech Innovation
AI, Ethics and
New Teaching Methods

Co-funded by
the European Union

2025-1-IT02-KA220-SCH-000365213

eduOpen

CC BY-NC-SA 4.0 International

This license requires that reusers give credit to the creator. It allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, for noncommercial purposes only. If others modify or adapt the material, they must license the modified material under identical terms.

Università di Foggia
Studi Umanistici, Lettere, Beni Culturali,
Scienze della Formazione

OER 6.3.1: Designing Authentic Assessment Tasks Rooted in Real-World Contexts
Reference Module: Module 6 - Classroom Practice and Authentic Assessment
Related Video-Pill: Video 6.3 - Designing Authentic Assessment Tests
Authoring partner: University of Foggia - Italy
Topic / Thematic Focus: Authentic assessment tasks and ecological validity in real-world contexts

Introduction and Learning Objectives

This resource accompanies, together with OER 6.3.2, video-pill 6.3 'Designing Authentic Assessment Tests' and constitutes its theoretical-methodological in-depth study. The text defines authentic assessment starting from the formulation of Grant Wiggins (1990) and organises its design criteria in the five-dimension framework of Gulikers, Bastiaens and Kirschner (2004), with the aim of guiding the teacher in the construction of tasks anchored in real contexts.
The framework connects to the European Framework for the Digital Competence of Educators (DigCompEdu; Redecker, 2017), with specific reference to Area 4 - Assessment (4.1 Assessment strategies, 4.2 Analysing evidence, 4.3 *Feedback* and planning) and in connection with Area 3 - Teaching and Learning and with Area 5 - Empowering Learners.
By the end of the reading, the teacher will be able to:

- distinguish 'surrogate measures' from authentic performances and recognise the construct of ecological validity;
- apply the five-dimension framework of Gulikers, Bastiaens and Kirschner to design or redesign a test;
- design tasks rooted in the real context of the class, structurally resistant to automated output alone.

1. The docimological discrepancy: surrogate tests and ecological validity

A recurring experience in secondary-school classrooms highlights a discrepancy that directly questions our assessment tools. Students who achieve high scores in multiple-choice tests with apparent confidence then freeze in front of an open task: they do not know where to begin, they cannot formulate a hypothesis or build an argument. The opposite phenomenon is just as frequent: pupils who are not brilliant in closed tests reason effectively about real problems, connect ideas and propose original solutions. Two profiles, mirror-image outcomes: what are traditional tests really measuring?
The answer of docimological research is that such tools operate as surrogate measures: efficient and simplified items, convenient substitutes from which we believe we can validly infer something about the student's competences (Wiggins, 1990). These are often reliable tests - two markers award the same score - but characterised by low ecological validity: they tell us little about how that student will tackle the authentic contexts of learning, work and citizenship. Real life, in fact, does not place before us five options of which only one is correct; it places before us open problems that require hypotheses, argumentation and decision.
Recognising this fracture does not mean disowning the value of standardised tests, but placing them within their perimeter: they are adequate for monitoring a performance, not for orienting learning that aspires to mastery. It is from this awareness that the proposal of authentic assessment arises.

2. Wiggins and authentic assessment: intellectually worthy tasks

In 1990 Grant Wiggins formulates the definition that has become the reference of the field: assessment is authentic when it directly examines the student's performance on intellectually worthy tasks. To true performance, Wiggins counterposes the logic of

conventional tests, which rely on indirect measures: efficient and simplified items, substitutes from which we think we can deduce the pupil's abilities.

> *«Assessment is authentic when we directly examine student performance on worthy intellectual tasks. Traditional assessment, by contrast, relies on indirect or proxy 'items' - efficient, simplistic substitutes from which we think valid inferences can be made» (Wiggins, 1990).*

The distinction has a precise operational consequence. Traditional tests reveal whether the student recognises, remembers or mechanically applies something learned outside its context of use; authentic assessment, by contrast, tests the ability to orchestrate one's cognitive resources in front of an open situation that is meaningful in itself. If we want to assess the ability to problematise and solve problems, to conduct a documentary investigation, to sustain a discussion or to revise a text until it is effective, then assessment must be built exactly on these exemplary intellectual challenges (Wiggins, 1990).

However, a recurring misunderstanding must be cleared up: authentic assessment does not abolish knowledge. Wiggins is explicit in specifying that disciplinary knowledge is not the goal of the assessment, but the tool with which the performance is carried out. Communicative competence is not measured by having pupils recite dictionary entries from memory, just as content serves to reason, not to be repeated. Hence the principle, taken up by Lauren Resnick, that sums up the stakes: what the school assesses is what it gets - if it does not assess it, it will not get it (Wiggins, 1990).

3. The five-dimension framework of Gulikers, Bastiaens and Kirschner

On the design level, Wiggins's contribution identifies some recurring characteristics of the authentic task: the challenge is poorly structured, lacking a ready-made solution and a single answer key; it simulates the adult or professional role; it requires a tangible product or an observable performance; and it adopts transparent assessment criteria, known to the student in advance. On this basis, Gulikers, Bastiaens and Kirschner (2004) have built a five-dimension framework that works as a genuine checklist for those designing a test.

The first dimension is the task: how close it is to what the professionals of the field concretely do and how many complex competences it mobilises. The second is the physical context: does the environment of the test resemble the real one? Are the relevant tools available? The third is the social context: does the test reproduce the relational dynamics of the field - cooperation among peers, presentation to an audience, negotiation of meanings? The fourth is the expected outcome: is a complex product required, assessable on multiple dimensions, or a bare grade? The fifth, finally, consists of the criteria: are the rubrics anchored to the standards of expert practice and shared with the students?

The following table summarises the five dimensions and the related operational indicators, offering itself as a grid of analysis for classroom design.

Dimension	Guiding question	Operational indicator
1. Task	Does the task resemble what the experts of the field do?	Degree of complexity and affinity with expert practice
2. Physical context	Are the environment and the tools of the test similar to the real ones?	Fidelity of the environment; availability of tools
3. Social context	Does the test reproduce the relational dynamics of the field?	Cooperation among peers; presentation to an audience
4. Expected outcome	Is a complex and multidimensional product required?	Tangible product vs bare grade
5. Criteria	Are the rubrics anchored to the standards and shared?	Analytic rubrics known in advance

The framework is not a purely theoretical construction. Gulikers, Bastiaens and Kirschner (2004) subjected it to empirical verification in a qualitative study conducted in a vocational institute for nursing care, questioning students and teachers on the relative weight of the five dimensions. The investigation confirms the completeness of the framework and shows that the perception of authenticity does not depend on a single dimension, but on the coherence with which all of them combine to simulate expert practice: neglecting the social context or the assessment criteria compromises the authenticity of the whole task, even when the disciplinary content is sound.
The consequence is significant on the pedagogical level: with this matrix, assessment ceases to be an episodic and almost punitive event and becomes a serious simulation of competent action. The student does not feel interrogated: they feel at work. And the transparency of the criteria, rather than weakening the rigour of the measurement, increases its equity, orienting the learner's self-regulation.

4. Anchoring the task in the real context

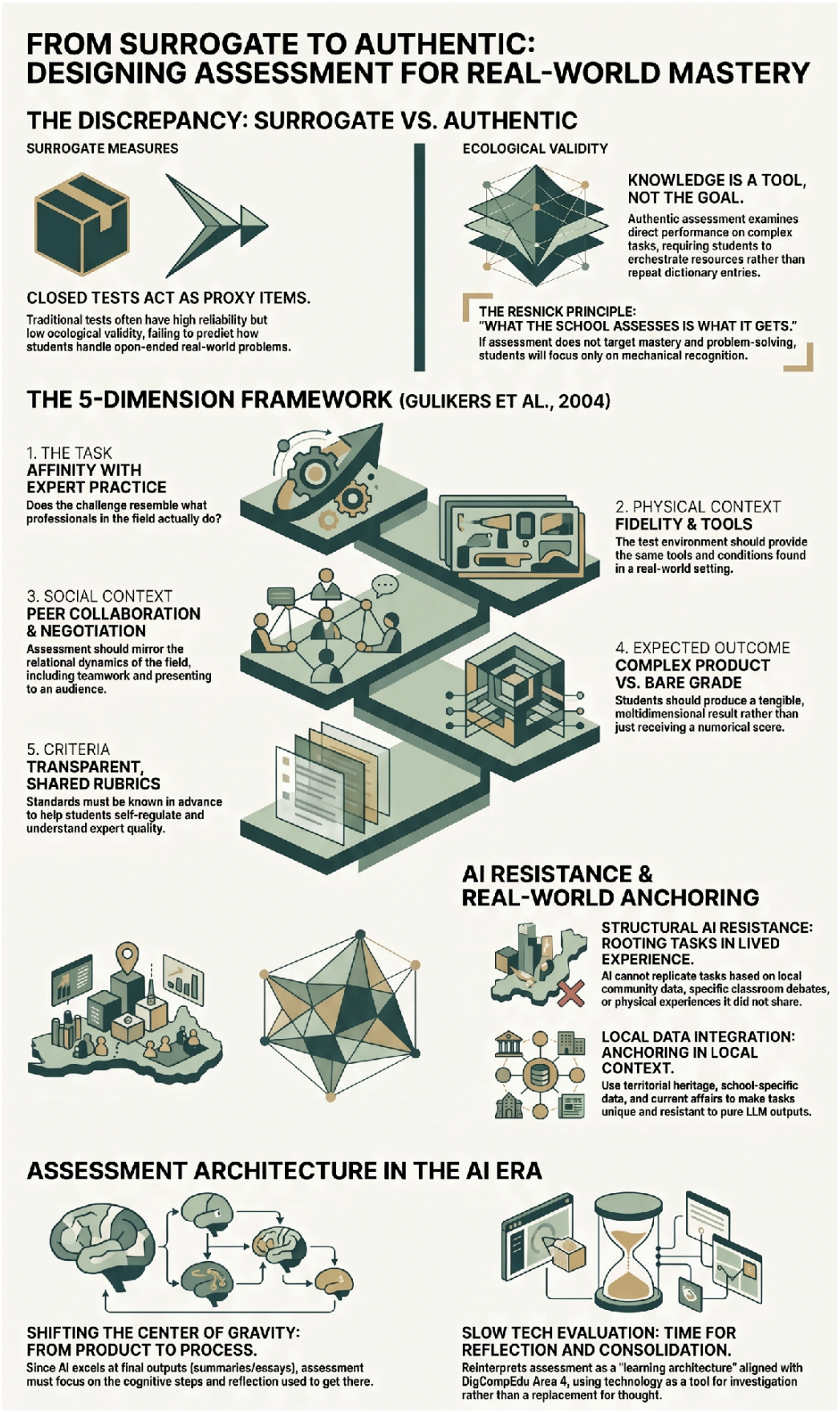
Anchoring tasks in the real context is, for authentic assessment, a constitutive choice and not an embellishment. It means rooting the test in local history, in current affairs, in the experiences and data that the class produces: an investigation into a territorial heritage, a documentary dossier on a controversial topic, the analysis of data collected by the school community, the design of a solution for a concrete problem of the context. The disciplinary content remains central, but ceases to be an object of output and becomes the instrument of a situated performance.
This choice produces a decisive effect in the age of generative models. An artificial intelligence system can synthesise notions, compare concepts and produce fluent texts on general topics; it possesses, however, no physical or social anchoring in our class: it did not take part in the discussion, it did not collect that data, it did not live that experience. Rooting the test in the real context makes it, by construction, resistant to automated output alone: the model's output, however fluent, cannot satisfy a performance that requires what it does not have.
Contextualisation, finally, intertwines with the transparency of the criteria: if the student knows in advance what the test requires - the dimensions of the product, the quality standards, the level descriptors - they can consciously orient their work. Assessment thus becomes a learning device, not merely a moment of certification.

5. Designing in the age of AI: assessment as a learning architecture

Designing authentic tests in the age of generative artificial intelligence means acknowledging a shift in the centre of gravity of assessment. Language models excel in the performances that schools have long measured: summarising, synthesising, expounding notions, answering standardised questions. It follows that any test founded on the pure output of information has lost discriminating capacity: it no longer distinguishes the competent student from one who has not understood; it measures, if anything, the intermediation of a software.
Authentic assessment represents the antidote to this drift, because it shifts attention from the final product to the cognitive processes that generate it. In terms of *Slow Tech*, it reinterprets the time of assessment as time for reflection and consolidation, and not as an acceleration of performance; it brings technology back to being a tool of investigation, not of substituting thought. Assessment, in this perspective, is not a methodological luxury: it is a necessity for a school that intends to assess what really counts.
Villarroel et al. (2018) translate these principles into a course-design scheme centred on authentic assessment. In a later work, the same authors apply these principles to the redesign of written examinations and traditional tests (Villarroel et al., 2019).

A design responsibility therefore follows for the school as a whole. If authentic assessment requires complex tasks, rich contexts and shared criteria, its adoption cannot be an isolated initiative of the individual teacher: it demands times for collegial design, coherence across subjects and a rethinking of the assessment calendar. In this perspective, assessment ceases to be an administrative duty and becomes an occasion for professional development for the whole educational community.
On the operational level, the OER 6.3.2 resource provides the methodological translation of these principles: how a traditional test is concretely redesigned, how the learning process is made observable and how the real constraints of time, subjectivity of marking and size of classes are governed.

Visual Scaffolding



Conclusions and Metacognitive Reflection Prompts

The articulation of this resource converges on a clear core: authentic assessment does not measure a surrogate performance, but the student's ability to orchestrate their cognitive resources in front of intellectually worthy tasks, rooted in real contexts and accompanied by

transparent criteria. The five-dimension framework of Gulikers, Bastiaens and Kirschner offers the teacher an operational grid for designing and verifying their tests, while Wiggins's formulation provides its deontological foundation: to assess what counts means to orient learning towards mastery.

To consolidate these core ideas and transfer them into one's own practice, the following reflection prompts are proposed:

- In which tests of your repertoire is the anchoring to a real context absent, and how could you reformulate the stimulus so that it requires an authentic performance?
- With reference to the five dimensions, which one turns out, in your subject, to be the most neglected (the social context? the criteria? the expected outcome?) and what consequences does it have on the ecological validity of your tests?
- How would the relationship between assessment and the use of generative models change if you shifted attention from the final product to the cognitive processes that generate it?

Technical Metadata Sheet

Metadata Field	Specification / Parameter Value
Identifier	OER 6.3.1
Title	Designing Authentic Assessment Tasks Rooted in Real-World Contexts
Creator	University of Foggia - Italy
Description	The resource defines authentic assessment from Wiggins' formulation and organises its design criteria in the five-dimension framework of Gulikers, Bastiaens and Kirschner. It contrasts surrogate measures with authentic performance and introduces the construct of ecological validity. The practical focus is on designing tasks rooted in the real context of the classroom that are structurally resistant to automated reproduction alone, treating assessment as a learning architecture.
Keywords	authentic assessment; ecological validity; surrogate measures; Gulikers framework; real-world tasks; task design; assessment architecture; AI-resistant tasks
Language	EN (English)
Affiliation	University of Foggia - Italy
Contributor	STIAEM Transnational Scientific Board (peer review, Rhodes validation protocol)
Media Format	Text; Infographics
Genre	Training Manual; Educational Guide
URL	
Citation	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://link.springer.com/article/10.1007/BF02504676 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(8), 1287-1301. https://doi.org/10.1080/02602938.2017.1412396 Villarroel, V., Boud, D., Bloxham, S., Bruna, D., & Bruna, C. (2019). Using principles of authentic assessment to redesign written examinations and tests. Innovations in Education and Teaching International, 58(1), 38-49. https://doi.org/10.1080/14703297.2018.1564882 Wiggins, G. (1990). The case for authentic assessment (ERIC Digest ED328611). ERIC Clearinghouse on Tests, Measurement, and Evaluation. https://eric.ed.gov/?id=ED328611 Redecker, C. (2017). European framework for the digital competence of educators: DigCompEdu (Y. Punie, Ed.). Publications Office of the European Union. https://doi.org/10.2760/159770 CAST. (2024). Universal Design for Learning guidelines version 3.0. CAST. https://udlguidelines.cast.org/ European Commission. (2022). Ethical guidelines on the use of artificial intelligence (AI) and data in teaching and

	learning for educators. Publications Office of the European Union. https://doi.org/10.2766/153756
Educational Level	Secondary School
Associated Module / MOOC Video Lesson	Module 6 - Classroom Practice and Authentic Assessment / Video 6.3 - Designing Authentic Assessment Tests
Learning Objectives	distinguish 'surrogate measures' from authentic performances and recognise the construct of ecological validity; apply the five-dimension framework of Gulikers, Bastiaens and Kirschner to design or redesign a test; design tasks rooted in the real context of the classroom, structurally resistant to automated reproduction alone.
Rights	Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)
Date Available	13/09/2026
Date Modified	

References and Webography

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://link.springer.com/article/10.1007/BF02504676

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(8), 1287-1301. https://doi.org/10.1080/02602938.2017.1412396

Villarroel, V., Boud, D., Bloxham, S., Bruna, D., & Bruna, C. (2019). Using principles of authentic assessment to redesign written examinations and tests. Innovations in Education and Teaching International, 58(1), 38-49. https://doi.org/10.1080/14703297.2018.1564882

Wiggins, G. (1990). The case for authentic assessment (ERIC Digest ED328611). ERIC Clearinghouse on Tests, Measurement, and Evaluation. https://eric.ed.gov/?id=ED328611

Redecker, C. (2017). European framework for the digital competence of educators: DigCompEdu (Y. Punie, Ed.). Publications Office of the European Union. https://doi.org/10.2760/159770

CAST. (2024). Universal Design for Learning guidelines version 3.0. CAST. https://udlguidelines.cast.org/

European Commission. (2022). Ethical guidelines on the use of artificial intelligence (AI) and data in teaching and learning for educators. Publications Office of the European Union. https://doi.org/10.2766/153756