TeSLA e-Assessment Model & Framework

TeSLA

Trust-based Authentication & authorship e-assessment analysis

E-assessment model & framework

March, 2019

Author: José Janssen (OUNL), Anna Guerrero Roldan (UOC), Henry Hermans (OUNL), Ingrid Noguera (UOC)

This project has been co-funded by the HORIZON 2020 Programme of the European Union. This publication reflects the views only of the author, and the Commission cannot be held responsible for any use, which may be made of the information contained therein.

Contents

1. Introduction

This report describes the educational assessment model and further TeSLA assessment framework as iteratively developed in the course of the TeSLA project with the aim to establish a shared language. A shared vocabulary and understanding are paramount in collaboratively developing the system, as well as in communicating the system to prospective users and facilitating its deployment.

Section 2 focuses on the educational assessment model that was developed with the aim to establish a shared language concerning e-assessment (see also the glossary in Appendix 1). The model also provides the basis for clarification of the further e-assessment framework: general positioning of authentication and authorship verification in e-assessment (section 3), deployment of specific authentication and authorship verification instruments in connection with specific assessment activities (section 4), as well as the scope of these instruments in relation to various categories of academic dishonesty (section 5).

2. Educational assessment model

The model described in this section was developed iteratively: it informed and in turn was informed by the implementation and evaluation of TeSLA instruments in various pilots. For a more detailed description of the theoretical and methodological background of the model see: Janssen, Guerrero Roldan, Hermans & Noguera (2019)[1]. The Educational Assessment Model (Figure 1) describes an ***Assessment*** as an assembly of one or more ***Activities*** designed to measure ***LearningOutcomes*** with the aim to establish (collect evidence of) a person's competences at a particular moment in time. An activity is to be conceived as any stimulus provided to a learner, in order to evoke a response (***ActivityResponse***).

The general term ***Activity***, rather than 'assessment task', was deliberately chosen to reflect the fact that learning and assessment are not always clearly distinct activities as is the case for instance in embedded assessments or stealth assessments where they blend into one so that assessment does not disrupt learning. A ***LearningOutcome*** is a specification of what a learner should know or be able to do at the end of an activity or series of activities (e.g. a course). The result a learner achieves by carrying out activities demonstrates to what extent learning outcomes have been attained. Learning outcomes are measurable operationalizations (indicators) of one or more competences: sets of knowledge, skills, and attitudes. These may be acquired at various levels of proficiency and may be domain (job/sector) specific or transversal, i.e. competences that are relevant across a broad range of jobs and sectors, such as communication and collaboration, also referred to as 21st-century skills. This alludes to the fact that an assessment is designed within (implies) a particular ***Context***. Other context attributes to be considered in assessment design, besides employment-related requirements, are for instance, institutional assessment principles and policies; departmental, disciplinary and personal norms, the overall program and the role of the unit/module, class size, learning environment, and delivery mode (online/face-to-face/blended).

The attribute *responseType* indicates what the learner is expected to do: the learner may simply need to select a correct answer in a multiple choice exam, to answer an open

[1] Janssen, J., Guerrero Roldan, A., Hermans, H. & Noguera J. (2019) Authentication and authorship verification in e-assessment: addressing the need for a generic conceptual model of educational assessment. Manuscript submitted for publication

Funded by the European Commission

question, to create a product or to otherwise perform an activity to demonstrate attainment of learning outcomes.

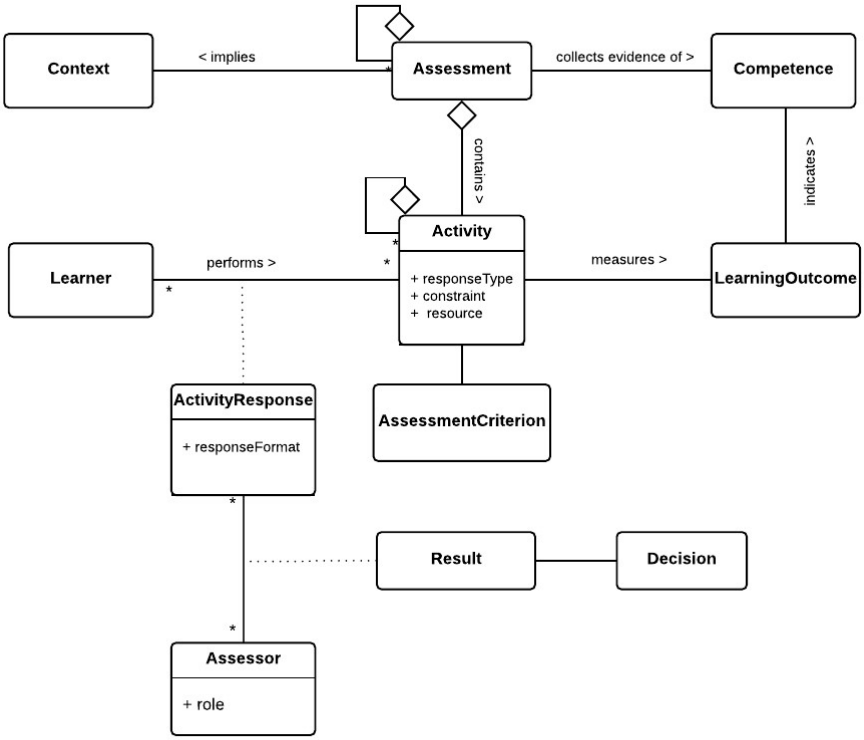

Figure 1 High level, generic model of educational assessment

Legend:

Core concept
https://www.lucidchart.com/documents/edit/e80fbba9-5a38-4c96-a905-59af1d346f17/0?callback=close&name=docs&callback_type=back&v=2358&s=464

_____ Relation between concepts

text> Characterisation of the relationship + direction (>) in which to read the relationship, e.g. assessment collects evidence of competence

--------- Association: indicates a concept which evolves from a relationship, e.g. when the learner performs an activity this leads to an ActivityResponse

* Multiplicity: asterix indicates that more than one instances of the class object can be involved in the relationship, e.g. one or more activities can be performed by one or more learners.

Nested concept: an assessment can consist of multiple assessments, e.g. a final exam consisting of a multiple choice exam and an assignment.

The actual response provided by the learner (**ActivityResponse**) may be delivered in various formats (*responseFormat*): a text may be typed or generated using speech to text software, performance may be delivered 'real-time' or submitted as a video and/or audio

recording. An Activity may be designed to be carried out (performed) using particular resources (e.g. a book in an open book exam or specific references for an essay) or tools (e.g. communication tools such as a blog or discussion forum or more concrete tools like a microscope, calculator, or software programme) as intrinsic parts of the assessment design.

An **Assessor** applies an activity's **AssessmentCriterion(s)** to the ActivityResponse of a Learner or a group of learners. The Assessor is a crucial actor in bringing the assessment process to completion, not only in terms of evaluating the response but also of providing feedback to the learner. The Assessor may be a teacher, a peer, or the learner (self), or may be automated. The Assessor determines the **Result**: a judgement that basically takes the form of an appraisal of the demonstrated competence (mark, score, grade) and/or a reflection on the demonstrated competence (description), i.e. feedback. The result is meant to inform a decision, which can take many forms: additional reading or assignments, enrolment to a course, approving a candidate for a job.
The appropriateness of various authentication and authorship verification technologies depends partly on what the assessment is used for, i.e. is it meant to inform a low-stake or high stake decision. Low stake assessments such as formative tests require no authentication/authorship checks.

The model indicates that an assessment, as well as an activity, can be nested, i.e. may contain a number of separate assessments/activities, as is expressed by the aggregation (recursive loops) attached to these concepts. Thus an assessment can - in its simplest form - consist of one single activity. On the other hand, it may consist of a series of activities, within a course or across courses as in curriculum-based assessment, and even beyond a curriculum as is the case with assessment of prior learning (APL). Moreover, an assessment can consist of a series of assessments, as in continuous assessment. This nested structure is of particular relevance for authentication and authorship verification as will also be further elaborated in the next section.

3. Authentication and authorship verification in e-assessment

Authentication and authorship verification are important to guarantee the quality of assessment and ultimately certification, irrespective of whether the assessment takes place in a face-to-face or online setting. Authentication takes place prior to and during execution of assessment activities, while authorship verification takes place in hindsight, through analysis of submitted products (essay, report, thesis). Note that, in terms of the model presented in Figure 1, authentication is connected to the Learner, whereas authorship verification probes the ActivityResponse (Figure 2). Authentication merely establishes whether the right person is taking the assessment, which in itself, provides no guarantee the activity is performed without any 'unauthorized support'. The latter inevitably requires some form of (human or technological) supervision on top of authentication.
Verification of a person's identity may rely on various means. In educational assessment, a student pass represents one possible means, i.e. what you have (a pass, ID card, cell phone, etc.). Other means involve what you know (user id and password), who you are (fingerprint, face, iris, voice), and what you do (typing rhythm or writing style) (cf. McBride, 2005). The authentication instruments integrated into TeSLA typically involve the latter two means of authentication: face recognition (FR) and voice recognition (VC) rely on who you are, whereas keystroke dynamics (KD) relies on how you type. Forensic analysis (FA), another of the TeSLA instruments, is a way to verify authorship which in fact includes authentication based on a person's writing style: it compares the writing style of a submitted text with a previous registration of this person's writing style. All authentication methods

Funded by the European Commission

based on who you are or what you do require baseline measurements of (sensitive) personal data that provide the basis for comparison/authentication.

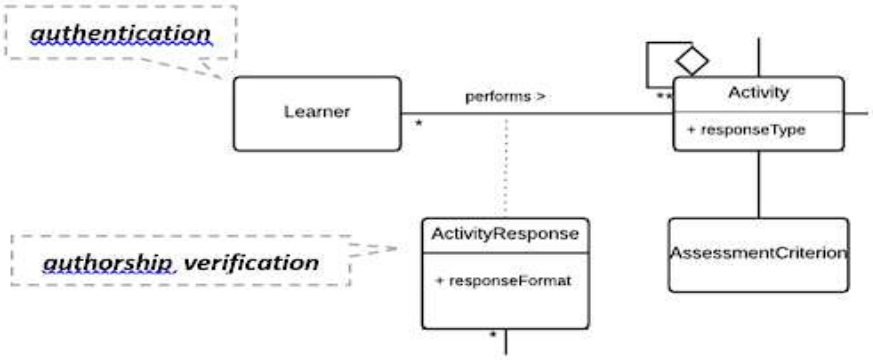

Figure 2. Authentication and authorship verification in assessment

A further TeSLA instrument aiming at authorship verification is plagiarism detection software (PD). This instrument is agnostic of any user. It simply compares a document with existing documents and thus relies on what others wrote (did)' as a source to verify authorship. The use of TeSLA instruments is most likely preceded by authentication through a user id and password (what you know) required to log in to the learning environment. Though this might also be done based on biometric data (who you are/what you do), currently this is far from common practice. Specific requirements apply to all of the TeSLA instruments regarding the baseline measurement and or assessment activity as is indicated in Table 1.

Table 1: TeSLA instrument requirements

Instrument	Baseline measurement	Assessment activity
Face recognition (FR)	Approximately 7 sec. video capture following detailed instructions	
Voice recognition (VR)	3 x 5 speech samples (10-15 sec.), preferably on three different days	Speech sample(s) of 10 sec.
Keystroke dynamics (KD)	15 samples of 125 keys	125 keys
Forensic analysis (FA)	Natural language - 3 samples adding up to 1000 words	Natural language min. 1000 words
Plagiarism detection (PD)	N/A	Natural language or code

Authorship verification methods deployed in TeSLA apply only to texts, but some further dependencies exist. Whereas forensic analysis (writing style analysis) can be used only for natural language texts, plagiarism detection can be carried out both on natural language texts and code. As mentioned above, forensic analysis requires a baseline writing style registration to which the writing style of the activity response can be compared.

Funded by the European Commission

4. Authentication and authorship verification related to assessment activities

The deployment of authentication and authorship verification instruments relates to the *type* of response that is required and the *format* of the response, as is indicated in Table 2. Authentication instruments are typically designed to be deployed while the learner is taking the assessment, so they depend on real-time input. However, instruments relying on who you are (FR and VR) might also be applied retrospectively on recordings a learner has made during assessment execution and submitted afterward. This is reflected in the table by two separate columns indicating two stages in the assessment process (execution and submission). Note that in Table 2 the format in which the response is given is not indicated in a separate column, but implicit in the columns Execution and Submission. For instance, authentication of a learner providing a short answer during activity execution based on keystroke dynamics requires a particular format. However, if authentication is to be carried out retrospectively, this can only be done based on a recording of the face (or possibly, though less likely, a voice recording).

Table 2: TeSLA instruments by ResponseType at different stages of the assessment process

ResponseType		**Execution** (authentication based on streaming data)	**Submission** (authentication and/or authorship verification based on uploaded data)
Select		FR	FR
Create	*Short answer (min. 125 keys)*	FR, VR, KD	FR, VR
	Long answer natural language or code (no minimum requirements)	FR, VR, KD	FR, VR, PD
	Written assignment natural language (>1000 words)	_	FA, PD
	Artifact, not text based (presentation, model, etc.)	_	FR, VR
Perform	*Online oral exam*	FR, VR	_
	Online game, role play, simulation	FR, VR, KD	FR, VR
	Online presentation	FR, VR	FR, VR

Unlike authentication, authorship verification takes place after the assessment activity has been completed and subsequently submitted. Actually, this is precisely why authorship verification is desirable: more elaborate writing assignments require a longer period of time and are therefore typically carried out in an unsupervised context. An interesting, serendipitous result from the TeSLA project is that in those instances where learners are required to create an artifact which does not take too much time (e.g. create a summary, a conceptual model, a few lines of code), combining a think-aloud instruction with retrospective use of voice recognition can be considered a powerful e-assessment design, which not only establishes authentication and authorship verification in one go, but also provides much more detailed insight in a learner's competence and thought processes than a mere artefact.

5. TeSLA instruments in relation to various categories of academic dishonesty

Having addressed the relation between responseType and deployment of the various authentication and authorship verification instruments, a final mapping of the instruments on the various categories of academic dishonesty in connection with various ResponseTypes is presented in Table 3 to clarify the scope and limits of this set of instruments.
We distinguish the following categories of dishonest practices in relation to e-assessment: *impersonation* (having someone else take an exam for you), *cheating* (the intentional (attempted) use of unauthorised materials or information in an examination), *collusion* (unauthorised cooperation), *fabrication* (making up/inventing data and/or information), *plagiarism* (passing someone else's work as one's own) and *contract cheating* (submitting a product that was created by someone else). Though the distinction appears subtle in some cases, the various categories pose different challenges for authentication and authorship verification depending on whether or not the assessment activity is performed in a face-to-face or an online context. Although face-to-face settings can be more directly supervised, even in these settings there are no watertight solutions to guarantee the right person is providing his/her own response. Here too, the supervisor will have to check student identity, for instance through a student pass that shows a picture.

Table 3. Authentication/authorship verification in connection with academic dishonesty categories

responseType		**Academic dishonesty category**				
		Impersonation	*Cheating / Collusion*	*Plagiarism*	*Contract cheating*	*Fabrication*
Select		FR, VR	S	X	X	X
Create	online	FR, VR, KD	S	X	X	X
	upload	X	FA, PD	PD	FA	-
Perform	online	FR, VR, KD	S	X	X	-
	upload	FR, VR	-	X	X	-

X : academic dishonesty category not relevant in connection with this responseType
S : supervision required
- : not covered by any existing instrument

Connecting academic dishonesty categories to various responseTypes is not a straightforward task, due to, for instance, the fact that the responseType perform covers a broad variety of possible activities. Besides, the categories of cheating and collusion appear somewhat problematic: they both involve seeking 'outside' help, the main distinction being that collusion may occur more or less intentionally (for instance due to lack of transparency on what is allowed), whereas cheating always involves a conscious, deliberate act against the rules. However, as this distinction is less relevant for our purposes, they have been combined in Table 3.

Another complication arises from the fact that the responseType perform comprises a wide variety of activities. Though table 3 suggests plagiarism and contract cheating are unlikely to occur in connection with this responseType we cannot rule this out entirely. However, the

Funded by the European Commission

main purpose of table 3 is to illustrate that authentication and authorship verification instruments target particular categories of academic dishonesty: prevention/detection of cheating in online assessments requires a supervised context as provided, for instance, by e-proctoring solutions. Finally, the 'ultimate case' of fabrication, which may occur in connection with more elaborate assessment activities, is not covered so far.

6. Worked example

We will now present a worked example to demonstrate application of the above model and framework. For this purpose we have chosen a case that represents some, but not too much, complexity: continuous assessment in a master course "Research Methods in Education".

The course is part of a curriculum which is designed according to didactic principles of active learning including activation through authentic tasks and collaborative learning [CONTEXT]. On average the course attracts 75 students [CONTEXT].
To complete the course students [LEARNER] have to demonstrate they are able to write a research proposal [COMPETENCE]. To this end they need to develop their knowledge about research questions and how they relate to particular research methods, how to write a Method section, how to evaluate research designs etc. [LEARNING OUTCOMES]. Students work towards these ends performing a series of tasks [ACTIVITIES], which are designed to develop and/or measure learning outcomes.

At the start of the course students are required to provide a brief description of a research problem [ACTIVITY / RESPONSETYPE = CREATE]. This activity was designed for multiple purposes: to activate students' prior learning, to measure entry levels (diagnostic), and to inform later group formation. The teacher [ASSESSOR] reads the assignments [ACTIVITYRESPONSE] and provides feedback [RESULT] based on a set of criteria [ASSESSMENTCRITERION], which may lead to further steps (additional reading, request for re-submission) [DECISION].

Though this can be considered an example of a low stake assessment, the teacher wants to make sure the activity response can be attributed to the individual student, as the activity was intended to inform her of an individual student's entry level. Only too often, the submitted assignments look very similar or even identical. She therefore wants to start deploying authentication and/or authorship verification in relation to this activity. The writing assignment is not that extensive, that forensic analysis could be applied. The most straightforward options here would be to deploy keystroke dynamics if students edit the text online, or plagiarism detection in case the assignment is completed by uploading a text file [ACTIVITYRESPONSE / RESPONSEFORMAT = text].

The first weeks of the course students develop their knowledge about various research methods. To this end they read learning materials, take part in virtual classes and discuss the relevance of each method for the research problem they formulated. Besides, formative assessment activities are in place for this knowledge component of the course that take the form of quizzes: students are presented with cases (problems and research questions) for which they have to select the most adequate research method [ACTIVITY / RESPONSETYPE SELECT]. For each case they also have to provide a brief explanation of their answer [RESPONSETYPE CREATE]. The results [RESULT] of these formative activities are used in two ways [DECISION] by the teacher: firstly, they help to identify specific misunderstandings as input for additional discussion and instruction on a group level in the virtual classroom and secondly, they help to identify students who appear to be struggling most with the course contents, who can then be given more adequate support. For these reasons the quizzes are an obligatory part of the course, and it is essential that students perform these quizzes relying on their own effort. To assure this, the quizzes are

Funded by the European Commission

accessible only for a short period of time and students type their explanations in text fields [ACTIVITYRESPONSE / RESPONSEFORMAT = keystroke] so that keystroke dynamics can be applied to verify the students identity.

Finally, students collaborate in small groups [LEARNER>1] to write a research proposal [ACTIVITY / RESPONSETYPE = CREATE]. This activity will be graded [RESULT] to inform whether the student has passed/failed the course requirements [DECISION]. The research proposal handed in by a group [ACTIVITYRESPONSE / RESPONSEFORMAT = text] is a jointly created product, rather than an assembly of texts that can be attributed to individual students, forensic analysis cannot be applied. This leaves plagiarism detection as an option for authorship verification.

Appendix 1 - Glossary

This glossary provides definitions - in alphabetical order - of model classes (in bold) as well as attributes (+) related to each class.

Activity	A designed stimulus that is provided to a learner or a group of learners to evoke a response (aimed at attainment and/or assessment of learning outcomes).
+ responseType	Specification of the *expected* response to the stimulus (Activity). Possible ResponseTypes: 1. select answer 2. create answer or product 3. perform/enact/demonstrate.
ActivityResponse	The *registered* response to an Activity by a (group of) learner(s).
+ responseFormat	The format in which a learner has provided the response, e.g. selection of an answer option may be done through speech rather than a mouse click by learners with special needs.
Assessment	One or more Activities designed to measure LearningOutcomes, i.e. to collect evidence of a person's competences.
AssessmentCriterion	Specification of 'rules' to be applied in evaluating an ActivityResponse.
Assessor	Actor (human or machine) responsible for assessing the ActivityResponse according to related criteria.
+ role	Specification of the human Assessor, e.g. peer, teacher, tutor, self.
Competence	Combination of knowledge, skills, and/or attitudes.
Context	External conditions influencing assessment design, e.g.: program and the role of the unit/module, student numbers, employment-related requirements, institutional assessment policies, delivery mode.
Decision	Action following the Result of an Assessment.
Learner	Person or group of persons (e.g. in collaborative learning) engaged in activities directed at attainment and/or measurement of learning outcomes.
LearningOutcome	What a learner is expected to know or to be able to do upon successful completion of an Activity: a measurable operationalisation of a Competence level.
Result	Outcome of the application of the ActivityCriteria to an ActivityResponse by an Assessor.