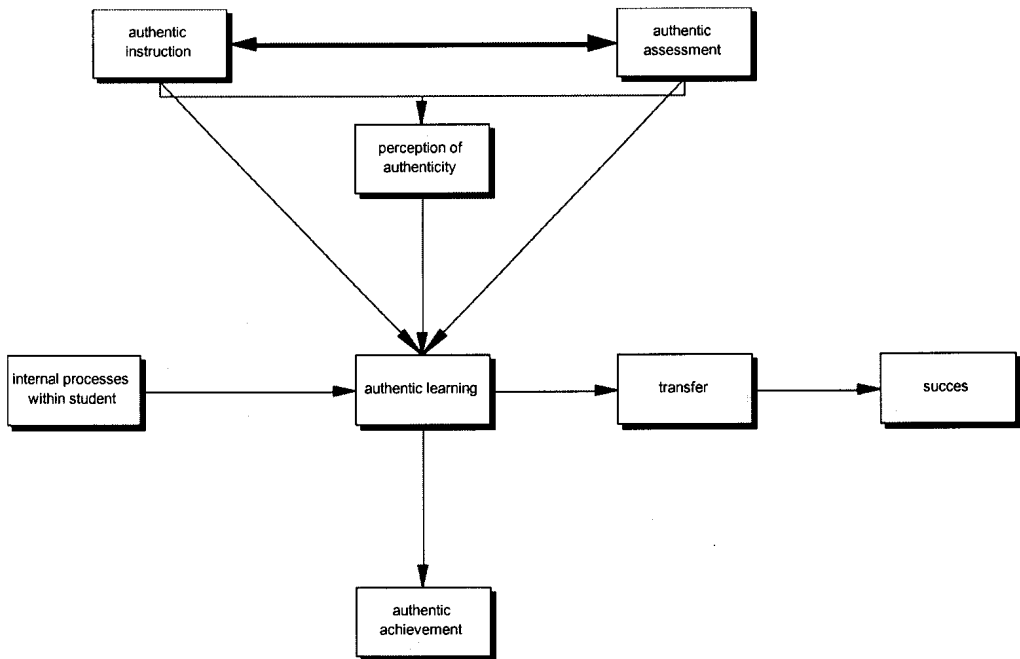


A five-dimensional framework for authentic assessment

Judith T. Gulikers, Theo J. Bastiaens & Paul A. Kirschner

Educational Technology Research and Development, 52 (2004)

Figure 1 □ General framework.



1997), but rather that we see it as a specific subset within a specific field of endeavor, namely becoming an academic. In this we concur with Brown, Collins and Duguid (1989) who, too, saw authentic achievement to be more than authentic academic achievement.

The following section discusses five dimensions (a theoretical framework) that can vary in their degree of authenticity in determining the authenticity of an assessment. The purpose of this framework is to shed light on in the concept of assessment authenticity and to provide guidelines for implementing authenticity elements into competency-based assessment.

TOWARD A FIVE-DIMENSIONAL FRAMEWORK FOR AUTHENTIC ASSESSMENT

To define authentic assessment, we carried out a review of literature on authentic assessment, on authenticity and assessment in general, and on student perceptions of (authentic) assessment elements. Five dimensions of authentic assess-

ment were distinguished: (a) the assessment task, (b) the physical context, (c) the social context, (d) the assessment result or form, and (e) the assessment criteria. These dimensions can vary in their level of authenticity (i.e., they are continuums). It is a misconception to think that something is either authentic or not authentic (Cronin, 1993; Newmann & Wehlage, 1993), because the degree of authenticity is not solely a characteristic of the assessment chosen; it needs to be defined in relation to the criterion situation derived from professional practice. For example: carrying out an assessment in a team is authentic *only* if the chosen assessment task is also carried out in a team in real life. The main point of the framework is that each of the five dimensions can resemble the criterion situation to a varying degree, thereby increasing or decreasing the authenticity of the assessment.

Because authentic assessment should be aligned to authentic instruction (Biggs, 1996; Van Merriënboer, 1997), the five dimensions of a framework for authentic assessment are also applicable to authentic instruction. Even though the focus of this article is on authentic assess-

ment, an interpretation of the five dimensions for authentic instruction is included in this article to show how the same dimensions can be used to create an alignment between authentic instruction and authentic assessment. The dimensions and the underlying elements of authentic instruction as presented in Figure 2 and Figure 3 do the same for authentic assessment.

As the figures show, learning and assessment tasks are a lot alike. This is logical, because the learning task stimulates students to develop the competencies that professionals have and the assessment task asks students to demonstrate these same competencies without additional support (Van Merriënboer, 1997). Schnitzer (1993) stressed that for authentic assessment to be effective, students need the opportunity to practice with the form of assessment before it is used as an assessment. This implies that the learning task must resemble the assessment task, only with different underlying goals. Learning tasks are for learning, and assessment tasks are for evaluating student levels of learning in order to improve (formative), or in order to make decisions (summative). These models show how a five-dimensional framework can deal with a (conceptual) alignment between authentic instruction and assessment. The interpretation and validation of the five dimensions for authentic assessment will be further explained and examined in the rest of this article.

An Argumentation for the Five Dimensions of Authentic Assessment

As stated, there is confusion and there exist many differences of opinions about what authenticity of assessment really is, and which assessment elements are important for authenticity. To try to bring some clarity to this situation, the literature was reviewed to explicate the different ideas about authenticity. Many sub-concepts and synonyms came to light, which were conceptually analyzed and divided into categories, resulting in five main aspects of authenticity. The notion of authenticity as a continuum (Newmann & Wehlage, 1993) resulted

in a conceptualization of these five aspects as dimensions that can vary in their degree of authenticity.

Task. An authentic task is a problem task that confronts students with activities that are also carried out in professional practice. The fact that an authentic task is crucial for an authentic assessment is undisputed (Herrington & Herrington, 1998; Newmann, 1997; Wiggins, 1993), but different researchers stress different elements of an authentic task. Our framework defines an *authentic task* as a task that resembles the criterion task with respect to the integration of knowledge, skills, and attitudes, its complexity, and its ownership (see Kirschner, Martens, & Srijbos, 2004). Furthermore, the users of the assessment task should perceive the task, including above elements, as representative, relevant, and meaningful.

An authentic assessment requires students to integrate knowledge, skills, and attitudes as professionals do (Van Merriënboer, 1997). Furthermore, the assessment task should resemble the *complexity* of the criterion task (Petraglia, 1998; Uhlenbeck, 2002). This does not mean that every assessment task should be very complex. Even though most authentic problems are complex, involving multidisciplinary, ill-structuredness, and having multiple possible solutions (Herrington & Herrington, 1998; Kirschner, 2002; Wiggins, 1993), real-life problems can also be simple, well structured with one correct answer, and requiring only one discipline (Cronin, 1993). The same need for resemblance holds for *ownership* of the task and of the process of developing a solution. Ownership for students in the assessment task should resemble the ownership for professionals in the criterion task. Savery and Duffy (1995) argued that giving students ownership of the task and the process to develop a solution is crucial for engaging students in authentic learning and problem solving. On the other hand, in real life, assignments are often imposed by employers, and professionals often use standard tools and procedures to solve a problem, both decreasing the amount of ownership for the employer. Therefore, the theoretical framework argues that in order to make students competent in dealing with professional

problems, the assessment task should resemble the complexity and ownership levels of the real-life criterion situation.

Up to this point, task authenticity appears to be a fairly objective dimension. This objectivity is confounded by Sambell, McDowell, and Brown (1997), who showed that it is crucial that students perceive a task as relevant, that (a) they see the link to a situation in the real world or working situation; or (b) they regard it as a valuable transferable skill. McDowell (1995) also stressed that students should see a link between the assessment task and their personal interests before they perceive the task as meaningful. Clearly, perceived relevance or meaningfulness will differ from student to student and will possibly even change as students become more experienced.

Physical context. Where we are, often if not always, determines how we do something, and often the real place is dirtier (literally and figuratively) than safe learning environments. Think, for example, of an assessment for auto mechanics for the military. The capacity of a soldier to find the problem in a nonfunctioning jeep can be assessed in a clean garage, with all the conceivably needed equipment available, but a future physical environment may possibly involve a war zone, inclement weather conditions, less space, and less equipment. Even though the task itself is authentic, it can be questioned whether assessing students in a clean and safe environment really assesses their ability to wisely use their competencies in real-life situations.

The physical context of an authentic assessment should reflect the way knowledge, skills, and attitudes will be used in professional practice (Brown et al., 1989; Herrington & Oliver, 2000). *Fidelity* is often used in the context of computer simulations, which describe how closely a simulation imitates reality (Alessi, 1988). Authentic assessment often deals with high-fidelity contexts. The presentation of material and the amount of detail presented in the context are important aspects of the degree of fidelity. Likewise, an important element of the authenticity of the physical context is that the number and kinds of resources available (Segers, Dochy, & De Corte, 1999), which mostly

contain relevant as well as irrelevant information (Herrington & Oliver), should resemble the resources available in the criterion situation. For example, Resnick (1987) argued that most school tests involve memory work, while out-of-school activities are often intimately engaged with tools and resources (calculators, tables, standards), making such school tests less authentic. Segers et al. (1999) argued that it would be inauthentic to deprive students of resources, because professionals do rely on resources. Another important characteristic crucial for providing an authentic physical context is the *time* students are given to perform the assessment task (Wiggins, 1989). Tests are normally administered in a restricted period of time, for example two hours, completely devoted to the test. In real life, professional activities often involve more time scattered over days or, on the contrary, require fast and immediate reaction in a split second. Wiggins (1989) said that an authentic assessment should not rely on unrealistic and arbitrary time constraints. In sum, the level of authenticity of the physical context is defined by the resemblance of these elements to the criterion situation.

Social context. Not only the physical context, but also the social context, influences the authenticity of the assessment. In real life, working together is often the rule rather than the exception, and Resnick (1987) emphasized that learning and performing out of school mostly takes place in a social system. Therefore, a model for authentic assessment should consider social processes that are present in real-life contexts. What is really important in an authentic assessment is that the social processes of the assessment resemble the social processes in an equivalent situation in reality. At this point, this framework disagrees with literature on authentic assessment that defines collaboration as a characteristic of authenticity (e.g., Herrington & Herrington, 1998). Our framework argues that if the real situation demands collaboration, the assessment should also involve collaboration, but if the situation is normally handled individually, the assessment should be individual. When the assessment requires collaboration, processes such as social interaction, positive

interdependency and individual accountability need to be taken into account (Slavin, 1989). When, however, the assessment is individual, the social context should stimulate some kind of competition between learners.

Assessment result or form. An assessment involves an assessment assignment (in a certain physical and social context) that leads to an assessment result, which is then evaluated against certain assessment criteria (Moerkerke, Doorten, & de Roode, 1999). The assessment result is related to the kind and amount of output of the assessment task, independent of the content of the assessment. In the framework, an authentic result or form is characterized by four elements. It should be an (a) quality product or performance that students can be asked to produce in real life (Wiggins, 1989). This product or performance should be a (b) demonstration that permits making valid inferences about the underlying competencies (Darling-Hammond & Snyder, 2000). Since the demonstration of relevant competencies is often not possible in one single test, an authentic assessment should involve a (c) full array of tasks and multiple indicators of learning in order to come to fair conclusions (Darling-Hammond & Snyder, 2000). Uhlenbeck (2002) showed that a combination of different assessment methods adequately covered the whole range of professional teaching behavior. Finally, students should (d) present their work to other people, either orally or in written form, because it is important that they defend their work to ensure that their apparent mastery is genuine (Wiggins, 1989).

Criteria and standards. *Criteria* are those characteristics of the assessment result that are valued; *standards* are the level of performance expected from various grades and ages of students (Arter & Spandel, 1992). Setting criteria and making them explicit and transparent to learners beforehand is important in authentic assessment, because this guides learning (Sluijsmans, 2002) and, after all, in real life, employees usually know on what criteria their performances will be judged. This implies that authentic assessment requires criterion-referenced judgment. Moreover, some criteria should be related to a

realistic outcome, explicating characteristics or requirements of the product, performance, or solutions that students need to create. Furthermore, criteria and standards should concern the development of relevant professional competencies and should be based on criteria used in the real-life situation (Darling-Hammond & Snyder, 2000).

Besides basing the criteria on the criterion situation in real life, criteria of an authentic assessment can also be based on the interpretation of the other four dimensions of the framework. For example, if the physical context determines that an authentic assessment of a competency requires five hours, a criterion should be that students need to produce the assessment result within five hours. On the other hand, criteria based on professional practice can also guide the interpretation of the other four dimensions of authentic assessment. In other words, the framework argues for a reciprocal relationship between the criterion dimension and the other four dimensions.

Some Considerations

~~What does all of this mean when teachers or instructional designers try to develop authentic assessments? What do they need to consider?~~

~~The first consideration deals with predictive validity. If the educational goal of developing competent employees is pursued, then increasing the authenticity of an assessment will be valuable. More authenticity is likely to increase the predictive validity of the assessment because of the resemblance between the assessment and real professional practice. However, one should not throw the baby out with the bath water. Objective tests are still very useful for certain purposes as high-stakes summative assessments on individual achievement, where predicting student ability to function competently in future professional practice is not the purpose.~~

~~Another consideration in designing authentic assessment is that we should not lose sight of the educational level of the learners. Lower level learners may not be able to deal with the authenticity of a real, complex, professional situation. If they are forced to do this, it may result in cogni-~~