
A Five-Dimensional Framework for Authentic Assessment

Author(s): Judith T. M. Gulikers, Theo J. Bastiaens, Paul A. Kirschner

Source: *Educational Technology Research and Development*, Vol. 52, No. 3 (2004), pp. 67-86

Published by: [Springer](#)

Stable URL: <http://www.jstor.org/stable/30220391>

Accessed: 03/09/2010 12:54

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/action/showPublisher?publisherCode=springer>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Springer is collaborating with JSTOR to digitize, preserve and extend access to *Educational Technology Research and Development*.

A Five-Dimensional Framework for Authentic Assessment

□ Judith T. M. Gulikers
Theo J. Bastiaens
Paul A. Kirschner

Authenticity is an important element of new modes of assessment. The problem is that what authentic assessment really is, is unspecified. In this article, we first review the literature on authenticity of assessments, along with a five-dimensional framework for designing authentic assessments with professional practice as the starting point. Then, we present the results of a qualitative study to determine if the framework is complete, and what the relative importance of the five dimensions is in the perceptions of students and teachers of a vocational college for nursing. We discuss implications for the framework, along with important issues that need to be considered when designing authentic assessments.

□ It is widely acknowledged that in order to meet the goals of education, a constructive alignment between instruction, learning and assessment (ILA) is necessary (Biggs, 1996). Traditional frontal classroom instruction for learning facts, assessed through short-answer or multiple-choice tests, is an example of such an alignment. The ILA-practices in this kind of education can be characterized as instructional approach—knowledge transmission; learning approach—rote memorization; and assessment procedure—standardized testing (Birenbaum, 2003). This approach to assessment is also known as the testing culture (Birenbaum & Dochy, 1996) and consists primarily of decontextualized, psychometrically designed items in a choice-response format to test for knowledge and low-level cognitive skill acquisition. The tests are primarily used in a summative way to differentiate between students and rank them according to their achievement. However, the alignment compatible with present-day educational goals has changed over the years. Current educational goals focus more on the development of competent students and future employees than on simple knowledge acquisition. The ILA-practices that characterize these goals are instructional-approach—focused on learning and competence development; learning-approach—reflective-active knowledge construction; and assessment-procedure—contextualized, interpretative, and performance assessment (Birenbaum, 2003). Here, the goal of assessment is the acquisition of higher-order thinking processes and competencies instead of factual knowledge and basic skills. The function of the assessment changes from being summa-

tive to also serving a formative goal of promoting and enhancing student learning. This view requires alternative assessments because standardized, multiple-choice tests are not suitable for this (Birenbaum & Dochy, 1996; Segers, Dochy, & Cascallar, 2003). Birenbaum and Dochy (1996) characterized alternative assessments as follows: Students have a responsibility for their own learning; they reflect, collaborate, and conduct a continuous dialogue with the teacher. Assessment involves interesting real-life or authentic tasks and contexts as well as multiple assessment moments and methods to reach a profile score for determining student learning or development. Increasing the authenticity of an assessment is expected to have a positive influence on student learning and motivation (eg., Herrington & Herrington, 1998). Authenticity, however, is only a vaguely described dimension of assessment, because it is thought to be a familiar and generally known concept that needs no explicit defining (Petraglia, 1998). This article focuses on defining authenticity in competency-based assessment, without ignoring the importance of other characteristics of alternative assessment.

Based on an extensive literature study, a theoretical framework consisting of five dimensions of assessment that can vary in their degree of authenticity is presented. After the description of this framework, the results of a qualitative study are discussed. This study explored whether the framework is a complete description of authenticity or is missing important elements, and what the relative importance of the dimensions is in the perceptions of students and teachers at a nursing college.

The Importance of Authentic Competency-Based Assessment

The two most important reasons for using authentic competency-based assessments are (a) their construct validity and (b) their impact on student learning, also called consequential validity (Gielen, Dochy, & Dierick, 2003). *Construct validity* of an assessment is related to whether an assessment measures what it is supposed to measure. With respect to competency

assessment this means that (a) tasks must appropriately reflect the competency that needs to be assessed, (b) the content of an assessment involves authentic tasks that represent real-life problems of the knowledge domain assessed, and (c) the thinking processes that experts use to solve the problem in real life are also required by the assessment task (Gielen et al., 2003). Based on these criteria, authentic competency-based assessments have a higher construct validity for measuring competencies than so-called objective or traditional tests have.

Consequential validity describes the intended and unintended effects of assessment on instruction or teaching (Biggs, 1996) and student learning (Dochy & McDowell, 1998). As stated, Biggs's (1996) theory of constructive alignment stresses that effective education requires instruction, learning, and assessment to be compatible. If students perceive a mismatch between the messages of the instruction and the assessment, a positive impact on student learning is unlikely (Segers, Dierick, & Dochy, 2001). This impact of assessment on instruction and on student learning is corroborated by researchers as Frederiksen (1984, "The Real Test Bias"), Prodromou (1995, "Backwash Effect"), Gibbs (1992, "Tail Wags the Dog"), and Sambell and McDowell (1998, "Hidden Curriculum"). Frederiksen and Prodromou implied that tests have a strong influence on what is taught, because teachers teach to the test, even though the test might focus on things the teacher does not find most important. Gibbs emphasized that student learning is largely dependent on the assessment and on student perceptions of the assessment requirements. Sambell and McDowell held that the effects of instruction and assessment on learning are largely based on teacher and student perceptions of the curriculum, which can deviate from the actual intentions of the curriculum. All four ideas support the proposition that learning and assessment are two sides of the same coin, and that they strongly influence each other. To change student learning in the direction of competency development, authentic competency-based instruction aligned to authentic competency-based assessment is needed.

Defining Authentic Assessment

The question is thus, What is authenticity? Different researchers have different opinions about authenticity. Some see authentic assessment as a synonym for performance assessment (Hart, 1994; Torrance, 1995), while others argue that authentic assessment puts a special emphasis on the realistic value of the task and the context (Herrington & Herrington, 1998). Reeves and Okey (1996) pointed out that the crucial difference between performance assessment and authentic assessment is the degree of *fidelity* of the task and the conditions under which the performance would normally occur. Authentic assessment focuses on high fidelity, whereas this is not as important an issue in performance assessment. These distinctions between performance and authentic assessment indicate that every authentic assessment is performance assessment, but not vice versa (Meyer, 1992).

Savery and Duffy (1995) defined authenticity of an assessment as the *similarity* between the cognitive demands—the thinking required—of the assessment and the cognitive demands in the criterion situation on which the assessment is based. A criterion situation reflects or simulates a real-life situation that could confront students in their internship or future professional life. Darling-Hammond and Snyder (2000) argued that dealing only with the thinking required is too narrow. In their view, students need to develop competencies because real life demands the ability to integrate and coordinate knowledge, skills, and attitudes, and the capacity to apply them in new situations (Van Merriënboer, 1997). Birenbaum (1996) further specified the competency concept by emphasizing that students need to develop not only cognitive competencies such as problem solving and critical thinking, but also meta-cognitive competencies such as reflection, and social competencies such as communication and collaboration.

The definition of *authentic assessment* used in this study is: an assessment requiring students to use the same competencies, or combinations of knowledge, skills, and attitudes, that they need to apply in the criterion situation in professional life. The level of authenticity of an assessment is thus defined by its degree of

resemblance to the criterion situation. This idea is extended and specified by the theoretical framework that describes that an assessment can resemble a criterion situation along a number of dimensions.

Complicating matters is the fact that authenticity is subjective (Honebein, Duffy & Fishman, 1993; Huang, 2002; Petraglia, 1998) and is dependent on perceptions. This implies that what students perceive as authentic is not necessarily the same as what teachers and assessment developers see as authentic. If these perceptions do indeed differ, then the fact that teachers usually develop authentic assessments according to their own view causes a problem: Although we may do our best to develop authentic assessments, this may all be for nothing if the learner does not perceive them as such. This process, known as *preauthentication* (Huang, 2002; Petraglia, 1998), can be interpreted either as that it is impossible to design an authentic assessment, or that it is very important to carefully examine the experiences of the users of the authentic assessments, before designing authentic assessments (Nicaise, Gibney & Crane, 2000). We chose the latter interpretation.

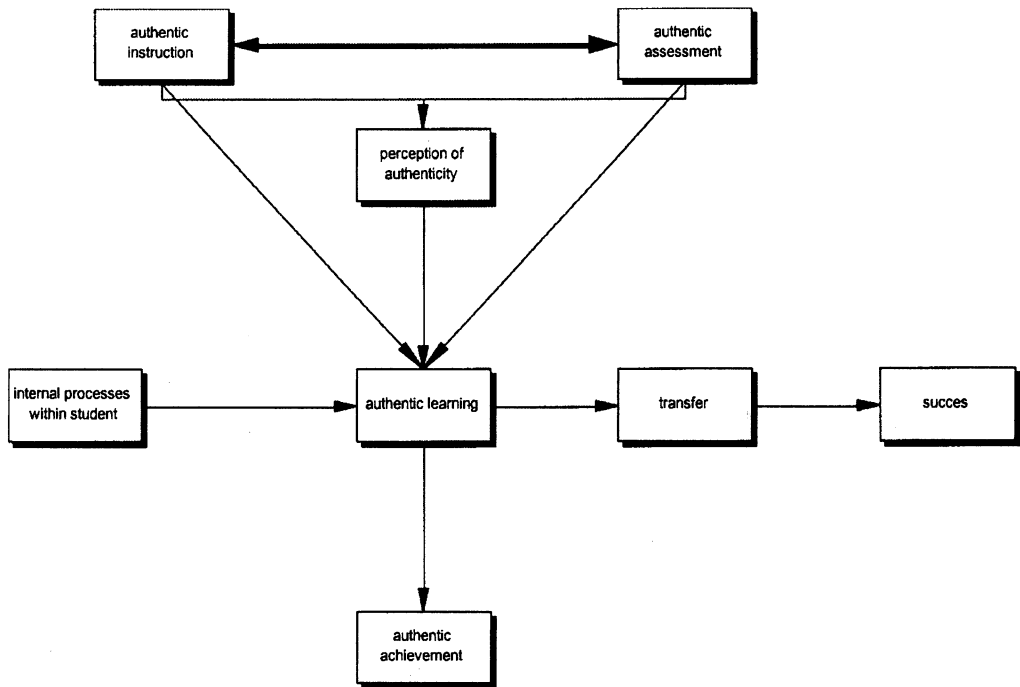
This discussion about authentic assessment and validity shows that:

1. In light of the constructive alignment theory (Biggs, 1996) authentic assessment should be aligned to authentic instruction in order to positively influence student learning.
2. Authentic assessment requires students to demonstrate relevant competencies through a significant, meaningful, and worthwhile accomplishment (Resnick, 1987; Wiggins, 1993).
3. Authenticity is subjective, which makes student perceptions important for authentic assessment to influence learning.

These three elements led to the following general framework (Figure 1) for the place of authentic assessment in educational practices.

The concept of *authentic achievement*, as we use it here, requires a note of explanation. This article deals with authentic assessment in general, regardless of the level or field of endeavor. This does not mean that we dismiss the concept of authentic *academic* achievement (Newmann,

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, & Strijbos, 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

Figure 2 □ Five-dimensional model for authentic instruction.

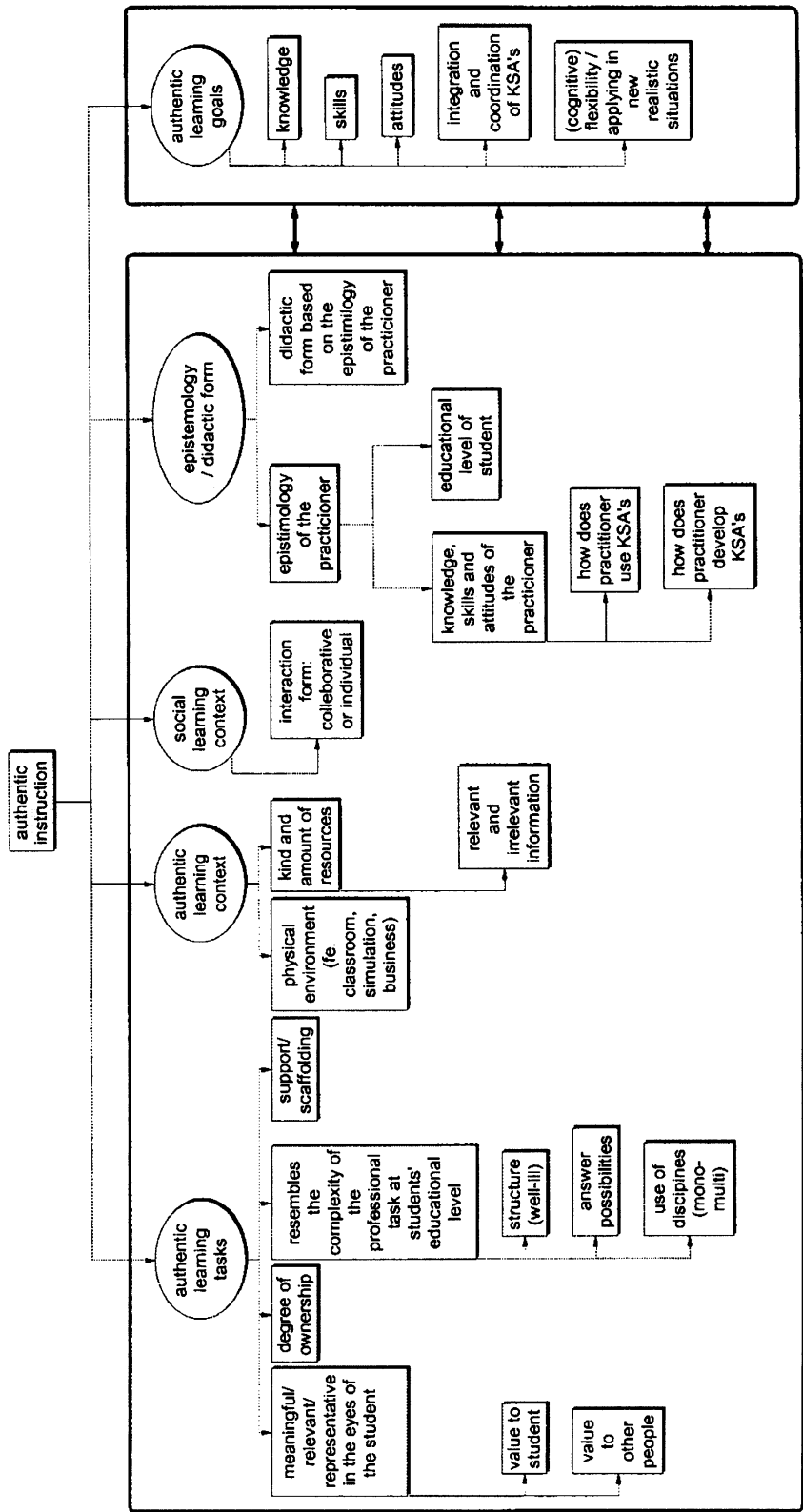
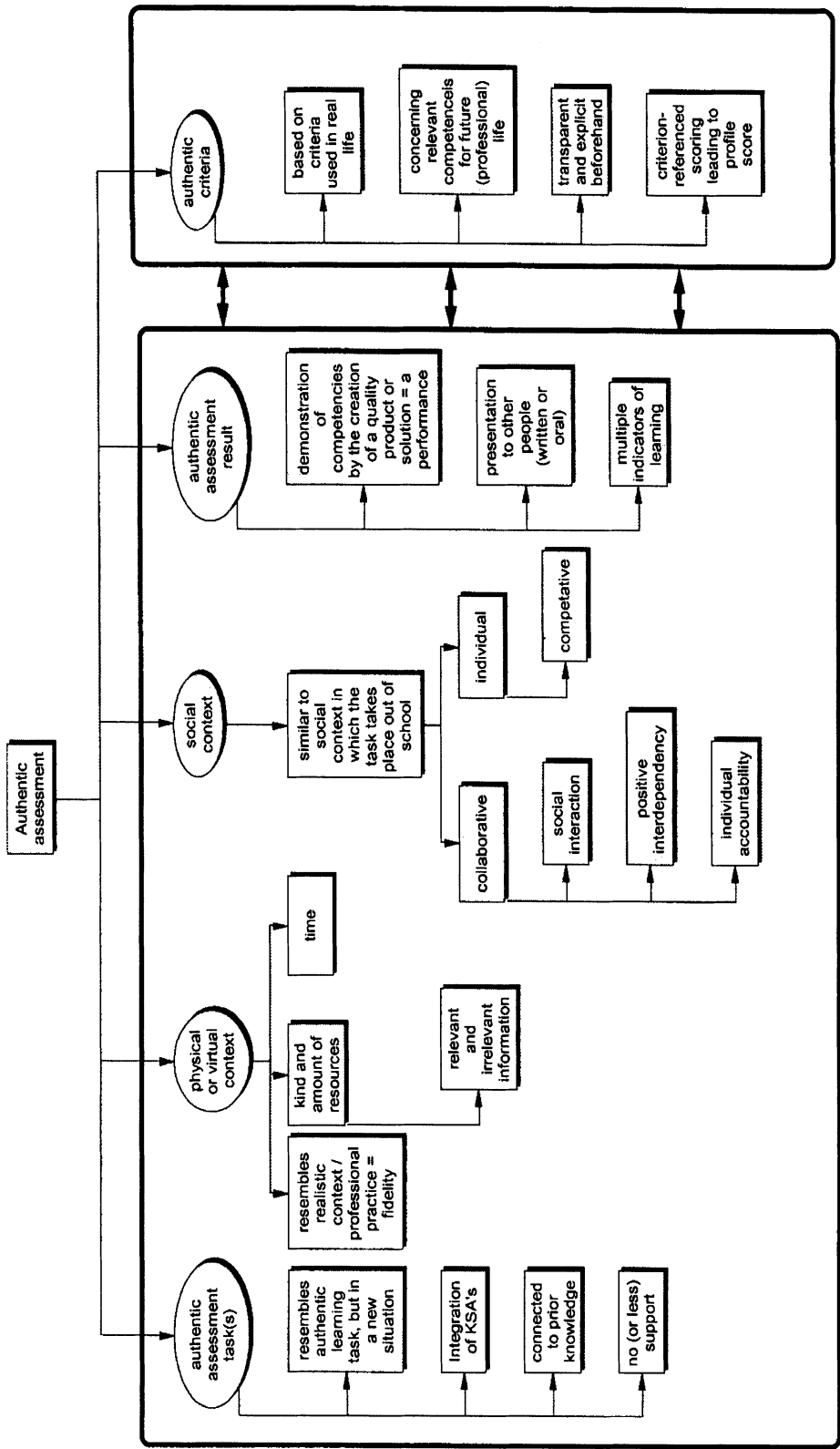


Figure 3 □ Five-dimensional model for authentic assessment.



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 a (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-

tive overload and, in turn, have a negative impact on learning (Sweller, Van Merriënboer, & Paas, 1998). As a result, a criterion situation will often need to be an abstraction of real professional practice in order to be attainable for students at a certain educational level. The question that immediately comes to mind in this context is How do you create an authentic assessment for students who are not prepared to function as beginning professionals? The answer is that the authenticity of an assessment should be defined by its degree of resemblance to the criterion situation (i.e., an abstraction from professional practice) and not necessarily to real professional practice. Van Merriënboer (1997) argued that an abstraction of real professional practice (i.e., the criterion situation) can still be authentic as long as the abstracted situation requires students to perform the whole competency as an integrated whole of constituent competencies. The abstraction results from simplifying contextual factors that complicate the performance of the whole competency.

A third consideration also sheds a light on the question stated in the previous sections, namely the subjectivity of authenticity. The perception of what authenticity is may change as a result of educational level, personal interest, age, or amount of practical experience with professional practice (Honebein et al., 1993). This implies that the five dimensions that are argued in the framework for authentic assessment are not absolute but, rather, variable. It is possible that assessing professional competence of students in their final year of study, when they have often served internships and have a better idea of professional practice, requires more authenticity of the physical context than when assessing first year students, who usually or often have little practical experience. Designers must take changing student perspectives into account when designing authentic assessment.

The qualitative study described in the rest of this article has two main goals. First, it explores whether our five-dimensional framework completely describes authenticity or whether important elements may be missing. Second, it explores the relative importance of the five dimensions. A subgoal of this study was to explore if the perception of (the importance of)

the authenticity dimensions differed between students and teachers and between students with different amounts of practical and educational experience. The differences and similarities along a limited number of dimensions can give insight into what is crucial for defining and designing authentic assessments.

METHOD

Participants

Students and teachers from a nursing college took part in this study. One session of the study involved only teachers, one session involved sophomore students (second year), and one session involved senior students (fourth year). The student groups could be further divided into a group of students studying nursing in a vocational training program (VTP) where they are primarily in school and make use of short internships, and a group that studied nursing in a block release program (BRP) where learning and working are integrated on an almost daily basis. This resulted in five groups of participants: (a) 8 sophomore VTP students (M age = 18.5 years), (b) 8 sophomore BRP students (M age = 20.9 years), (c) 8 senior VTP students (M age = 19.7 years), (d) 4 senior BRP students (M age = 31.4 years), and (e) 11 teachers (M age = 42.8 years). The number of participants per session was limited because of the practical possibilities of the group support system used in this study.

Materials

An electronic group support system (GSS) at the Open University of the Netherlands was used as research tool. A GSS is a computer-based information processing system designed to facilitate group decision making. It is centered on group productivity through idea generation, preference, and opinion exchange of people involved in a common task in a shared environment. The GSS allows collaborative and individual activities such as brainstorming, idea generation, sorting, rating, and clustering via computer

communication. To prevent participants (especially students) from feeling inhibited in expressing their ideas and opinions, the GSS was a good option because it is completely anonymous. Furthermore, it was a practical and valuable method because it made it possible to collect a lot of information in a structured way in a short period of time.

To examine the relative importance of the five dimensions, four case descriptions of assessments that varied in their amount of authenticity based on the five dimensions of the model were designed. They described competencies from the nursing competency profile, which were validated by two employees of the nursing college. To check the influence of the GSS session itself on the perceptions of the authenticity of the cases, the descriptions were used in a pre- and a posttest. To do this, a second set of different but comparable case descriptions was designed, which resulted in two sets of four cases. Cases A and E were completely authentic except for the task; Cases B and F were completely authentic except for the physical context; Cases C and G were completely authentic except for the result or form; and Cases D and H were completely authentic (see Appendix for a full description of a completely authentic case description).

Procedure

All participants had access to a GSS computer. During a two-hour session, participants carried out both individual and collaborative activities.

At the beginning and end of the GSS session, participants were presented four case descriptions (ABCD or EFGH). In six paired comparisons ($4 \times 3/2$), they chose the case that they considered to be a more authentic assessment. This activity was meant to determine the relative importance of the different dimensions of authentic assessment in the eyes of the different groups of participants. A second underlying purpose of this activity was to bring participants in a specific reference frame for the rest of the session, and to focus their thinking toward authenticity of assessment instead of assessment in general.

A distinction was made between VTP students and BRP students; it was possible that

because of the differences in their studies, they would have different perceptions of what determines authenticity. VTP students, BRP students, and teachers were randomly divided in two halves, one that received Cases ABCD in the pretest and EFGH in the posttest, and one that received the cases in the reverse order.

After the initial rating of the case descriptions, the participants were appraised of the purpose of the study. In order to create a specific frame of mind, a very general description was given of the term *authenticity* (i.e., true to life). The GSS part of the study consisted of four activities. The first activity required the participants to enter into the system their own statements that described authenticity of an assessment. This was a free brainstorm, and participants were encouraged to generate as many statements as possible. Statements were anonymously entered into the GSS, where it was also possible to respond to statements made by others. After this electronic brainstorm, the contributions were discussed in order to clarify them. This was recorded for later use and analysis.

The second activity required respondents to specify (voting is a feature of a GSS) the 10 most important statements for designing authentic assessments that were generated during the brainstorm. The purpose of this activity was to determine which elements the participants perceived as being especially important for authentic assessment. After completing these two activities, a prototype five-dimensional framework for authentic assessment was presented as a framework for assessing professional behavior. The five dimensions were explained to the participants in an attempt to create mutual understanding about the meaning of the dimensions. The five dimensions were characterized as follows:

1. Task: What do you have to do?
2. Physical context: Where do you have to do it?
3. Social context: With whom do you have to do it?
4. Result or form: What has to come out of it? What is the result of your efforts?
5. Criteria: How does what you have done have to be evaluated or judged?

The third and fourth activities consisted of

paired comparisons to determine the relative importance of the dimensions. Activity three consisted of 10 paired comparisons of the five dimensions ($5 \times 4/2$). Participants had to choose the dimensions of the framework that they perceived to be more important for authentic assessment. The fourth activity was the same as the activity at the beginning of the experiment: The participants were again required to carry out paired comparisons of case descriptions that varied in their amount of authenticity according to the five-dimensional framework. Each group received the counterbalanced set of case descriptions to those compared at the beginning of the experiment.

Analysis

A characteristic of the GSS is that the answers, statements, choices, and so forth, of each individual participant are anonymous. This means that scores per participant were not available, which precluded the possibility of carrying out statistical tests. On the other hand, the anonymity inhibited socially accepted answering behavior, and has been shown to stimulate response in idea generation and increase the reliability of answers. The data, thus, were qualitatively analyzed. The tapes of the discussions were transcribed. Both discussion statements and the statements keyed in during the brainstorm were analyzed to discern which of the five dimensions of the framework they fit. Statements that did not fit were classified as *other*.

The paired comparison data of the five dimensions, that is, the number of times that a dimension in the paired comparisons was rated as more important than another dimension, were tallied per participant group. The absolute scores were then translated into rankings. The paired comparisons of the case descriptions were analyzed in the same way.

RESULTS

In general, the task, the result or form, and the criteria were rated as most important for the authenticity of the assessment. The social context was clearly considered to be least important for authenticity, and the importance of the physical context was strongly discussed.

The Relative Importance of the Five Dimensions: Paired Comparisons

The paired comparisons of the dimensions and of the case descriptions gave insight into the relative importance of the five dimensions for designing authentic assessments. The comparisons of the dimensions resulted in five rankings (sophomore VTP students, sophomore BRP students, teachers, senior VTP students, and senior BRP students) from 1 to 5. The paired comparisons of the case descriptions were analyzed for the same groups, but were measured in pre- and posttests, which resulted in ten rankings from 1 to 4.

Table 1 □ Ranking of dimensions by group.

	Task	Physical context	Social context	Result or form	Criteria
Sophomore VTP students	2.0	4.5	4.5	1.0	3.0
Sophomore BRP students	1.0	3.5	5.0	3.5	2.0
Teachers	1.0	4.0	5.0	2.0	3.0
Senior VTP students	2.0	5.0	3.5	3.5	1.0
Senior BRP students	2.0	4.0	5.0	1.0	3.0

Total	8.0	21.0	23.0	11.0	12.0
-------	-----	------	------	------	------

Note. 1 = most important, 5 = least important

Table 1 shows rankings per group of the five dimensions, based on their perceived importance in providing authenticity to an assessment (1 = *most important*, 5 = *least important*). Table 1 shows that all groups perceived the task as important (score 1 or 2), and all groups except the senior VTP students (score 3.5), perceived the social context as the least important. Furthermore, the result or form and criterion dimensions received more than average importance, whereas all groups perceived the physical context as relatively unimportant (score about 4). In short, independent of the group (see totals in Table 1), the task was perceived as most important, followed by the result or form and criterion dimensions; the physical context and especially the social context lagged far behind.

The results of the paired comparisons of the case descriptions, in pre- and posttests, also gave insight into the relative importance of the dimensions. Table 2 shows rankings per group of the four case descriptions.

A 1 meant that this case was perceived as the most authentic case description and a 4 referred to the least authentic case description. An important finding, for the framework, was that the case that described a completely authentic assessment based on the presence of all five

dimensions was perceived as most authentic (score 1) by all, except the senior BRP students on the posttest (score 2.5). The other three kinds of cases showed an interesting pattern. The case that was authentic except for the task received mostly a score of 2, which meant that this case was perceived as relatively authentic, which in turn meant that the task (which was not authentic in this case) was *not* perceived as very important in designing an authentic assessment. This is contrary to the findings of the paired comparisons of the dimensions in which the task was perceived as very important in providing authenticity to an assessment. Finally, the participant groups disagreed about the authenticity of the remaining two kinds of cases. All sophomore students ranked the case that was authentic expect for the result as 4, meaning that they perceived this case to be the least authentic. In other words, they perceived the result or form dimension as most important for designing an authentic assessment. Teachers, on the other hand, ranked the case that was authentic except for physical context as the least authentic case (score 4), which meant that teachers perceived physical context to be most important in designing an authentic assessment. Senior students did not appear to differentiate, meaning

Table 2 □ Ranking of case descriptions by group.

	<i>All authentic except for the task</i>	<i>All authentic except for the physical context</i>	<i>All authentic except for the result or form</i>	<i>All authentic</i>
Sophomore VTP, pretest	2.0	3.0	4.0	1.0
Sophomore BRP, pretest	2.0	3.0	4.0	1.0
Sophomore VTP, posttest	3.0	2.0	4.0	1.0
Sophomore BRP, posttest	2.0	3.0	4.0	1.0
Teachers pretest	3.0	4.0	2.0	1.0
Teachers posttest	2.0	4.0	3.0	1.0
Senior VTP pretest	2.0	3.5	3.5	1.0
Senior BRP pretest	2.0	3.5	3.5	1.0
Senior VTP posttest	2.0	3.5	3.5	1.0
Senior BRP posttest	1.0	4.0	2.5	2.5

Note. 1 = *most authentic*, 4 = *least authentic*

that they perceived the cases with no authentic physical context or with no authentic result or form as equally inauthentic (score 3.5). To sum, the findings of the paired comparisons of the case descriptions indicated that when all of the dimensions in the framework are present in a case, the case was unequivocally seen as most authentic. In addition, there appear to be contradictory results with respect to task authenticity compared to the results of the paired comparisons of the dimensions. Finally, when evaluating assessment cases, teachers and students appear to differ with respect to the importance of the authenticity of physical context versus result authenticity.

Completeness and Relative Importance: What Do Participants Say?

Table 3 shows that all dimensions received attention in the brainstorm and discussion sessions. Furthermore, these results corroborated the earlier findings, in that social context received the least attention in all groups. Besides the five dimensions, almost all subelements of the dimensions, described in the framework, were reviewed.

Based on the number of statements and the ratios of the statements compared to each other, as shown in Table 3, sophomores place primary interest on task, followed by physical context. Seniors and teachers place equal emphasis on task and result. Teachers differ from all students, regardless of level, with respect to the emphasis on physical context. Teachers devoted a lot of time to discussing the required fidelity level of the physical context in an effective authentic assessment. Especially emphasized was the question of whether the physical context

should be real professional practice or a simulation in school.

A closer look at the content of the brainstorm statements gave the impression that teachers and seniors agreed more with each other and with the idea of the framework, than with the sophomore students, especially when it came to task and result or form dimensions. Teachers and seniors agreed with the framework that an authentic task required an integration of professional knowledge, skills, and attitudes, and they acknowledged that the task should resemble real-life complexity. On the other hand, sophomore students were preoccupied with knowledge testing, they had problems picturing the idea of integrated testing, and were primarily concerned with making assessment easier and clearer (e.g., “assignments should be less vague, not more than one answer should be possible”) instead of simulating real-world complexity. In the result or form dimension, teachers and seniors agreed that more assessment moments and methods should be combined for a fairer and more authentic picture of students’ professional competence. Sophomores did not discuss the result or form dimension much; they only mentioned that reshaping current tests in the form of cases would make them more realistic. In other words, every kind of assessment could be made more authentic by adding realistic information.

A specification of the *other* statements (see Table 4) showed, first, that all groups made statements emphasizing the alignment between instruction and assessment, and between school and real-life practice. This is in agreement with the theoretical ideas behind the framework for authentic assessment. Second, Table 4 shows that issues concerning the assessor of an authentic assessment, and organizational or pre-

Table 3 □ Number of statements per dimension of each group.

	<i>Task</i>	<i>Physical context</i>	<i>Social context</i>	<i>Result or Form</i>	<i>Criteria</i>	<i>Other</i>
Sophomore students	24	19	6	7	13	45
Senior students	34	21	9	36	12	26
Teachers	16	39	5	19	21	56

Table 4 □ Variables in the *other* category, per group.

	<i>Sophomore students</i>	<i>Senior students</i>	<i>Teachers</i>
General statements applicable to all five dimensions	6	1	2
Instruction	28	7	5
Alignment instruction—assessment	2	3	3
Alignment school—practice	6	3	3
Assessor	3	3	6
Organization or preconditions	–	–	7
Influence on the learning process	–	–	4
Not defined or nonsense	–	9	26

conditional issues, should be taken into account in a framework for authentic assessment. Issues related to the assessor dealt with the realization that people from professional practice should be involved in defining and using criteria and standards. Organizational issues involved statements about conditions that should be met *before* authentic assessment can be implemented in school. For example, teachers talked about placing students in professional practice sooner and more often for the purpose of assessing them in this professional context. Finally, Table 4 shows that sophomores took the opportunity to talk and complain about the instruction. Although instruction was not being evaluated (i.e., it was about assessment), 28 statements dealt with what was taught and not with what was assessed. Seniors were more focused, and teacher statements were spread over different *other* variables and the 26 statement of the *not defined* variable included mostly jokes or questions they asked each other.

CONCLUSION

Overall, the five-dimensional framework gave a good description of what dimensions and elements should be taken into account in an authentic assessment; the participants discussed all dimensions and almost all elements described in the framework. However, elements concerning the assessor and organization issues should be added to complete the framework, as these elements turned out to be important to all participant groups.

A combination of the results of the GSS activities led to the conclusion that task, result or form, and criteria were perceived as very important for authentic assessment. Physical context was most important in the eyes of teachers. Social context was perceived as the least important dimension.

Furthermore, not all groups perceived the dimensions and elements in the same way. Teachers and seniors mostly agreed with each other and with the theoretical framework; however, sophomores often deviated from the other groups. There were no differences between VTP and BRP students.

DISCUSSION

To reiterate: The two questions with which we began were (a) Is the framework complete? (b) Do students differ from teachers with respect to what they perceive as important for authenticity? Both of these questions shed light on possible guidelines for designing authentic assessments.

The answer to Question 1 appears to be yes. The five dimensions appear to adequately define authenticity, as demonstrated by both the brainstorming and the high ranking of those cases that were authentic on all five dimensions. The adequacy of the framework is corroborated by the finding that during the brainstorming, most subelements of the dimensions as described by the framework were seen as important when designing authentic assessment. The paired comparisons showed some subtle differ-

ences in the importance of the five dimensions for providing authenticity. While the task, the result or form, and the criterion dimensions turned out to be very important for authenticity, the physical context and especially the social context dimensions were perceived as less important. Social context is unequivocally perceived as the least important dimension of authenticity. All groups stressed the need for individual testing, although both students and teachers stressed that most nursing activities in real life are collaborative. Teachers explained that “assessing in groups is a soft spot, we just don’t know how to assess students together, because at the end we want to be sure that every individual student is competent.” It should not be concluded, based on these findings, that social context is not important for authentic assessment, but if choices have to be made in designing an authentic assessment, social context is probably the first dimension to leave out.

The findings on importance of task are sometimes contradictory. Although the brainstorming and the paired comparisons of the dimensions show that task was perceived as very important by all, the paired comparisons of the cases made task seem less important. It is possible, thus, that although the respondents consider task (as an abstracted concept) to be most important, they are not able to identify (i.e., they do not perceive) an authentic task. A possible explanation for this is that the all-authentic-except-for-the-task case resembles current assessment practices. Because previous experiences are found to strongly influence perceptions (Birenbaum, 2003), the familiarity of these cases may have influenced the paired comparisons of the cases. If this is the case, the paired comparisons of the five dimensions were probably a more objective measure of the importance of the five dimensions.

Finally, it might be the case that assessor-related issues would complete the framework. This could be done by adding a sixth dimensions called “the assessor,” or by adding the issues concerning who should use and develop authentic criteria and standards as subelements to the criterion dimension.

With respect to Question 2, concerning the differences between students and teachers in

their perception of authenticity, some interesting findings came to light. The most differences were found between sophomores and teachers, while seniors agreed with teachers more often. Moreover, the perceptions of teachers and seniors agreed more with the ideas of the theoretical framework. Possibly, the perceptions of older students have changed during their college career as a result of having had experience with professional practice; the perceptions of sophomores—who have less practical experience—seemed to be based primarily on their previous experiences with assessment, which explained the focus on knowledge and in-school testing. In other words, it appears that sophomore students have different conceptions and possibly misconceptions of real professional practice and, thus, authenticity of assessment.

Furthermore, the brainstorming and the paired comparisons of the case descriptions showed differences between teachers and students in the perception of physical context. Teachers focused on the importance of increasing the authenticity of physical context by placing the assessment in professional practice, whereas students, especially sophomores, mostly focused on in-school testing with, for example, simulated patients and realistic equipment.

Finally, all groups agreed on the relative unimportance of the social context and on the importance of using criteria that resemble the criteria used in real professional practice. Teachers and students agree that, at this point, the criteria used in school differ too much from criteria used in professional institutes, and that school criteria are often unknown or misinterpreted by assessors at the professional institutes.

Future Implications

The findings of the study allow for some critical questions and guidelines concerning the design of authentic assessment. First, student perceptions should be considered in designing effective authentic assessments. The qualitative results of this study showed that students, especially at the beginning of their study and with little practical experience, have different conceptions (possibly misconceptions) of what authenticity

means than do older, more experienced students and teachers. For authentic assessment to work, two options need to be considered: either (a) the assessment should meet the expectations of the sophomores, for example, by sticking to explicit knowledge testing in the name of authentic assessment, which is likely to confirm unwanted learning behavior; or (b) explicit attention should be given to changing student perceptions and, thereby, opening the possibilities to change their learning behavior toward professional competency development, when implementing authentic assessment.

Second, we might be able to save precious time and money in the design, development and implementation of authentic assessment with respect to the physical context and the creation of social contexts. Research should examine if assessing students in a real professional context has additional value for students, or if assessing in an (electronic) simulation in school is authentic enough as long as students are confronted with an authentic task, result or form, and criteria. Simulation in school, virtual or not, is probably easier and less expensive to implement, and, therefore, warrants careful consideration.

The exploratory nature of this study, without the possibility of quantitative statistical analyses owing to the nature of the GSS, makes firm conclusions impossible. However, the electronic GSS efficiently delivered a lot of qualitative data in a short period of time. What the data of this study *do* show is that authenticity is definitely a multifaceted concept, and that a number of the facets (dimensions) appear to be of more importance than others. This can have far-reaching implications for educational design.

The actual effectiveness of this framework for designing authentic assessments, however, should be examined by evaluating the influences of different kinds and levels of authenticity of assessment on student learning and motivation. Because implementing authenticity elements in assessment requires a lot of time, money, and energy (Martens, Bastiaens, & Gulikers, 2002), research should examine which elements of the framework are crucial for affecting student learning in the direction of the development of professional competencies.

Finally, as stated at the beginning of this article, authenticity is only one of the elements and quality criteria of competency-based (alternative) assessment (Birenbaum & Dochy, 1996; Dierick, Dochy, & Van de Watering, 2001). Making decisions about implementing authentic elements in an assessment should be considered in the broader context of quality criteria for assessment (i.e., reliability or generalizability), and in the context of other assessment goals (i.e., timeliness, affordability, and accountability). However, a thorough discussion of these other assessment goals and criteria is beyond the scope of this article.

The argumentation of the theoretical framework and the qualitative study gave some interesting impulses to further theoretical and practical research concerning authentic assessments and student perceptions, and especially the focus on vocational college is interesting, because most assessment research is done in higher education. All participants in this study agreed that instruction and assessment in school should be aligned with each other and that developing education that focuses on the development of competencies and takes professional practice as a starting point, requires assessments that are also competency based and based on professional practice. In other words, it requires authentic assessment. □

Judith T. M. Gulikers [judith.gulikers@ou.nl], Theo J. Bastiaens, and Paul A. Kirschner are with the Educational Technology Expertise Center at the Open University of the Netherlands, P.O. Box 2960, 6401 DL Heerlen, The Netherlands.

The authors would like to thank Marijke Bijns for her help in organizing the participation of teachers and students in the GSS. They would also like to thank Dr. Robert Schuur for his assistance in setting up and carrying out the GSS sessions.

REFERENCES

- Alessi, S. M. (1988). Fidelity in the design of instructional simulations. *Journal of Computer-Based Instruction*, 15(2), 40–47.
- Arter, J. A., & Spandel, V. (1992). An NCME instructional module on: Using portfolio of student work in instruction and assessment. *Educational Measurement: Issues and Practice*, 11(1), 36–45.
- Biggs, J. (1996). Enhancing teaching through construc-

- tive alignment. *Higher Education*, 32, 347–364.
- Birenbaum, M. (1996). Assessment 2000: Towards a pluralistic approach to assessment. In M. Birenbaum & F. J. R. C. Dochy (Eds.), *Alternatives in assessment of achievements, learning processes and prior knowledge* (pp. 3–29). Boston, MA: Kluwer Academic Publishers.
- Birenbaum, M. (2003). New insights into learning and teaching and their implications for assessment. In M. Segers, F. Dochy, & E. Cascallar (Eds.), *Optimising new modes of assessment: In search of quality and standards* (pp. 13–36). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Birenbaum, M., & Dochy, F. J. R. C. (1996). *Alternatives in assessment of achievements, learning processes and prior knowledge*. Boston, MA: Kluwer Academic Publishers.
- Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, 18(1), 32–42.
- Cronin, J. F. (1993). Four misconceptions about authentic learning. *Educational Leadership*, 50(7), 78–80.
- Darling-Hammond, L., & Snyder, J. (2000). Authentic assessment in teaching in context. *Teaching and Teacher Education*, 16, 523–545.
- Dierick, S., Dochy, F., & Van de Watering, G. (2001). Assessment in het hoger onderwijs: Over de implicaties van nieuwe toetsvormen voor de edumetrie. [Assessment in higher education: About the implications of new assessment forms for edumetrics] *Tijdschrift voor Hoger Onderwijs*, 19(1), 2–18.
- Dochy, F. J. R. C., & McDowell, L. (1998). Assessment as a tool for learning. *Studies in Educational Evaluation*, 23(4), 279–298.
- Frederiksen, N. (1984). The real test bias, influences of testing and teaching on learning. *American Psychologist*, 39(3), 193–202.
- Gibbs, G. (1992). *Improving the quality of student learning*. Bristol, UK: Technical and Educational Services.
- Gielen, S., Dochy, F., & Dierick, S. (2003). The influence of assessment on learning. In M. Segers, F. Dochy, & E. Cascallar (Eds.), *Optimising new modes of assessment: In search of quality and standards* (pp. 37–54). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Hart, D. (1994). *Authentic assessment: A handbook for education*. Menlo Park, CA: Addison-Wesley Publishing Company.
- Herrington, J., & Herrington, A. (1998). Authentic assessment and multimedia: How university students respond to a model of authentic assessment. *Higher Educational Research & Development*, 17(3), 305–322.
- Herrington, J., & Oliver, R. (2000). An instructional design framework for authentic learning environments. *Educational Technology Research and Development*, 48(3), 23–48.
- Honebein, P. C., Duffy, T. M., & Fishman, B. J. (1993). Constructivism and the design of learning environments: Context and authentic activities for learning. In T. M. Duffy, J. Lowyck, & D. H. Jonassen (Eds.), *Designing environments for constructive learning* (pp. 88–108). Berlin: Springer-Verslag.
- Huang, H. M. (2002). Towards constructivism for adult learners in online learning environments. *British Journal of Educational Technology*, 33, 27–37.
- Kirschner, P. A. (2002). Three worlds of CSCL: Can we support CSCL? Heerlen: Open University of the Netherlands.
- Kirschner, P. A., Martens, R. L., & Strijbos, J. W. (2004). CSCL in higher education? A framework for designing multiple collaborative environments. In P. Dillenbourg (Series Ed.) & J. W. Strijbos, P. A. Kirschner & R. L. Martens (Vol. Eds.), *Computer-supported collaborative learning: Vol. 3. What we know about CSCL ... And implementing it in higher education* (pp. 3–30). Boston, MA: Kluwer Academic Publishers.
- Martens, R., Bastians, Th., & Gulikers, J. (2002). Leren met computergebaseerde authentieke taken: motivatie, gedrag en resultaten van studenten [Learning with computer-based authentic tasks: student motivation, behavior and results]. *Pedagogische Studiën*, 79(6), 469–482.
- McDowell, L. (1995). The impact of innovative assessment on student learning. *Innovations in Education and Training International*, 32(4), 302–313.
- Meyer, C. (1992). What's the difference between authentic and performance assessment? *Educational Leadership*, 49(8), 39–40.
- Moerkerke, G., Doorten, M., & de Roode, F. A. (1999). *Constructie van toetsen voor competentiegericht curricula [Construction of assessments for competency-based curricula] (OTEC report 1999/W02)*. Heerlen, The Netherlands: Open Universiteit Nederland, Educational Technology Expertise Center.
- Newmann, F. M. (1997). Authentic assessment in social studies: Standards and examples. In G. D. Phye (Ed.), *Handbook of classroom assessment: Learning, achievement, and adjustment* (pp. 359–380). San Diego, CA: Academic Press.
- Newmann, F. M., & Wehlage, G. G. (1993). Five standards for authentic instruction. *Educational Leadership*, 50(7), 8–12.
- Nicaise, M., Gibney, T., & Crane, M. (2000). Toward an understanding of authentic learning: Student perceptions of an authentic classroom. *Journal of Science Education and Technology*, 9, 79–94.
- Petraglia, J. (1998). *Reality by design: The rhetoric and technology of authenticity in education*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Prodromou, L. (1995). The backwash effect: From testing to teaching. *ELT Journal*, 49(1), 13–25.
- Reeves, T. C., & Okey, J. R. (1996). Alternative assessment for constructivist learning environments. In B. G. Wilson (Ed.), *Constructivist learning environments: Case studies in instructional design* (pp. 191–202). Englewood Cliffs, NJ: Educational Technology

- Publications.
- Resnick, L. B. (1987). Learning in school and out. *Educational Leadership*, 16(9), 13–20.
- Sambell, K., & McDowell, L. (1998). The construction of the hidden curriculum: Messages and meanings in the assessment of student learning. *Assessment and Evaluation in Higher Education*, 23(4), 391–402.
- Sambell, K., McDowell, L., & Brown, S. (1997). But is it fair?: An exploratory study of student perceptions of the consequential validity of assessments. *Studies in Educational Evaluation*, 23(4), 349–371.
- Savery, J., & Duffy, T. (1995). Problem based learning: An instructional model and its constructivist framework. *Educational Technology*, 35, 31–38.
- Schnitzer, S. (1993). Designing and authentic assessment. *Educational Leadership*, 50(7), 32–35.
- Segers, M., Dierick, S., & Dochy, F. (2001). Quality standards for new modes of assessment. An exploratory study of the consequential validity of the OverAll test. *European Journal of Psychology of Education*, 16(4), 569–586.
- Segers, M., Dochy, F., & Cascallar, E. (2003). *Optimising new modes of assessment: In search of qualities and standards*. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Segers, M., Dochy, F., & De Corte, E. (1999). Assessment practices and students' knowledge profiles in a problem-based curriculum. *Learning Environments Research*, 2, 191–213.
- Slavin, R. E. (1989). Research on cooperative learning: An international perspective. *Journal of Educational Research*, 33, 231–243.
- Sluijsmans, D. (2002). *Student involvement in assessment: the training of peer assessment skills*. Unpublished doctoral dissertation, Open University of the Netherlands, Heerlen, The Netherlands.
- Sweller, J., Van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10(3), 251–296.
- Torrance, H. (1995). *Evaluating authentic assessment*. Buckingham, UK: Open University Press.
- Uhlenbeck, A. (2002). *The development of an assessment procedure for beginning teachers of English as a foreign language*. Unpublished doctoral dissertation, University of Leiden, Leiden, The Netherlands.
- Van Merriënboer, J. J. G. (1997). *Training complex cognitive skills: A four-component instructional design model for technical training*. Englewood Cliffs, NJ: Educational Technology Publications.
- Wiggins, G. (1989). Teaching to the (authentic) test. *Educational Leadership*, 46(7), 41–47.
- Wiggins, G. P. (1993). *Assessing student performance: Exploring the purpose and limits of testing*. San Francisco, CA: Jossey-Bass/Pfeiffer.

See Appendix, overleaf

Appendix ☐ Authentic case description: Authentic on five dimensions.

You're working as an intern in a nursing home. During your internship you have to show that you are capable of providing basic care to geriatric patients with different kinds of problems. You have to be able to help them with their own personal care, such as washing themselves, getting dressed, and combing their hair. In school you learned what basic care is, the care that you have to be able to provide to different kinds of geriatric patients in a nursing home, how you should do this, and what you should take into account while doing this (e.g., the different kinds of disabilities or dysfunctions, privacy issues, and rules). In other words, you know which criteria will be used to judge your competence in providing this basic care to the geriatric patients. To judge whether you are competent, your mentor at the nursing home observes you while you are taking care of three patients, each with different problems. The first patient is suffering from dementia, but is physically in good health, the second is paraplegic, and the third is obese. You are allowed to take care of these patients on your own, but you can also ask colleagues for assistance if you think the care is too trying or difficult to carry out alone. You are required to show that you are sensitive to the different disabilities of the patients, and that you take this into account in your actions. Furthermore, you have to be able to communicate well with the patients, explain to them what is going to happen, and answer their questions. Your mentor will write a report on your performance in each of the three cases, and finally draw a conclusion about your competence in providing personal care to different kinds of geriatric patients.

Authentic task:

Provide personal care to geriatric patients (washing, getting dressed, combing) while taking into account their disabilities, dysfunctions, and privacy rules and issues.

Authentic physical context:

Performing the tasks during an internship at a nursing home for geriatric patients.

Authentic social context:

Allowing the student to take care of the patients alone, but also allowing the student to ask for assistance if the task is too trying or difficult to carry out alone.

Authentic result:

Judging the competence of the student by observing the student demonstrating competence by performing the whole task. Competence is judged based on multiple (three) performances in different contexts (different patients).

Authentic criteria:

Judging the relevant professional competency in relation to different aspects of the complete competency (washing, dressing, combing, but also communicating, taking disabilities and dysfunctions into account, taking privacy rules into account). Criteria are known to the students—they learned them in school before beginning the internship—and are based on the basic care that nurses have to provide in real life.