Paper presented at the second biannual joint Northumbria/EARLI SIG assessment conference, Bergen, 2004

Perceptions of authentic assessment

Five dimensions of authenticity

Judith T.M. Gulikers, Theo J. Bastiaens and Paul A. Kirschner

Open University of the Netherlands

Correspondence concerning this paper should be addressed to Judith T.M. Gulikers, Open University of the Netherlands, Educational Technology Expertise Center, PO Box 2960, 6401 DL Heerlen, The Netherlands. E-mail: judith.gulikers@ou.nl

Most of this paper will be published in:
Gulikers, J., Bastiaens, Th., & Kirschner, P. (in press). A five-dimensional framework for authentic assessment. *Educational Technology Research & Development*

Abstract

Based on intuition, everybody agrees that authentic assessment is 'the way to go', because authenticity is expected to positively influence learning and motivation. This enthusiasm, however, has led to different interpretations of authenticity, and intuitive authentic assessment practices. To lift authentic assessment beyond intuition, this paper first presents a review of the literature on authenticity of assessments along with a five-dimensional framework for designing authentic assessments with professional practice as the starting point. These dimensions are: task, physical context, social context, result/form, and criteria. After a thorough description of these dimensions and their sub-characteristics, a qualitative study explored the completeness of this framework and the relative importance of the five dimensions in the eyes of students (with different amounts of educational and practical experience) and teachers. The framework and explorative study showed that authenticity is a multi-faceted concept and that a number of facets (dimensions) appear to be of more importance than others. Furthermore, differences in perceptions between different student groups and teachers came to light. Implications for future research are discussed along with important issues that need to be considered in the design of authentic assessments.

Defining Authentic Assessment: Five Dimensions of Authenticity and Perceptions Thereof

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). Current educational goals, especially in Europe, focus on the development of competent students and future employees (Dochy, 2001). The ILA-practices that characterize these goals are: instruction that focuses on learning and competence development; learning based on reflective-active knowledge construction; and assessment that is contextualized, interpretative and performance-based (Birenbaum, 2003). The need to contextualize assessment in interesting, real-life and authentic tasks is described as *one of the crucial elements* of alternative, or competency-based assessment that suits current educational goals (Birenbaum & Dochy, 1996). Dochy (2001) describes the assessment of the application of knowledge to actual, real-life (authentic) cases as the core goal of alternative assessments. Gielen, Dochy and Dierick (2003) even argue that authenticity of the assessment tasks is an imperative condition to achieve the expert level of problem solving. Moreover, increasing the authenticity of an assessment is expected, and experienced by students, to have a positive influence on student learning and motivation (eg. Herrington & Herrington, 1998; Sambell, McDowell, & Brown, 1997). Authenticity, however, remains a vaguely described characteristic of assessment, because it is thought to be a familiar and generally known concept that needs no explicit defining (Petraglia, 1998). Because authenticity is thought to be such an important element of competency-based assessment on the one hand, and the fact that what authenticity actually is, is only vaguely described on the other hand, this study focuses on defining authenticity in competency-based assessment.

Based upon 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, which explored whether the framework is a complete description of authenticity or is missing important elements and what the relative importance of the dimensions are in the perceptions of students and teachers at a vocational college for nursing.

The Importance of Authentic Competency-Based Assessment

The two most important reasons for using authentic competency-based assessments are their construct validity and 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 tasks must appropriately reflect the competency that needs to be assessed, that the content of an assessment involves authentic tasks that represent real-life problems of the knowledge domain assessed, and that 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 upon these criteria, authentic competency-based assessments are expected to have a higher construct validity for measuring competencies than so-called objective or traditional tests. Messick (1994) argues that construct underrepresentation is one of the major threats to construct validity, which is countered by increasing the authenticity of the assessment. Authenticity, he argues, deals with not leaving anything out of the assessment of a certain construct, leading to minimal construct underrepresentation.

Consequential validity describes the intended and unintended effects of assessment on instruction or teaching (Biggs, 1996) and student learning (Dochy & McDowell, 1998). The expected positive influence of authentic assessment on student learning is twofold (Gielen, Dochy, & Dierick, 2003). First, it is expected to stimulate the development of professional competencies, and second it is likely to increases students' motivation to learn through the fact that students experience authentic assessments as more interesting and meaningful, because they realize the relevancy and usefulness of it for their future lives.

Defining Authentic Assessment

The question that remains is, what is authenticity? Different researchers have different opinions about authenticity. Some see authentic assessment as a synonym to 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) point 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)

Messick (1994) focuses our attention to the fundamental ambiguity that pervades all authentic assessment practices, namely, *authentic to what*? Honebein, Duffy and Fishman (1993) strengthen the importance of this question by saying that authenticity is a relative concept. In other words, the authenticity of something can only be defined in relation to something else. A test can either be authentic to the school or to the real world. For example, an assessment task can be authentic with respect to school problems, but inauthentic with respect to everyday life experience, because school problems do not relate to everyday life. The point taken in this study is that the authenticity of an assessment is defined by its resemblance to the real world, specifically, to the professional real world. Because current educational goals stress the importance of developing competent professionals, we argue that it is important to design assessments that resemble situations that starting professionals can be confronted with in the working life. The situation, according to which the authenticity of an assessment in this paper is defined, is called criterion situation. A *criterion situation* reflects a real-life situation that students can be confronted with in their internship or future professional life, which serves as a basis for designing an authentic assessment.

Another issue in defining authentic assessments that logically follows from the previous section deals with *what* students need to learn or develop from working with authentic assessments that resemble professional, real-life situations. Savery and Duffy (1995) define 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. In other words, students need to develop professional thinking skills. Darling-Hammond and Snyder (2000) argue that dealing only with the thinking required is too narrow, because real life demands the ability to integrate and coordinate knowledge, skills, and attitudes, and the capacity to apply them in new situations. In their view, authentic assessment includes opportunities for the development and examining of students' thinking *and* actions. This implies that authentic assessment requires students to *demonstrate* their learning. Birenbaum (1996) deepens this idea of assessing thinking *and* action by emphasizing that students not only need to develop 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. In other words, real life (reflected in the criterion situation) involves different kinds of competencies that all should be taken into account in designing authentic assessments for developing competent future employees.

The definition of authentic assessment used in this study is *an assessment requiring students to demonstrate the same (kind of) 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 here is the fact that authenticity is subjective (Honebein, Duffy & Fishman, 1993; Huang, 2002; Petraglia, 1998) and is dependent on perceptions. Entwistle and Entwistle (1991) already showed that students' perceptions of the learning environment influence how they learn, not necessarily the context itself. A recent literature review of Struyven, Dochy, and Janssens (2003) showed that this is also true for students' perceptions of alternative assessments. The fact that authenticity is subjective 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, namely: although we may do our best to develop authentic assessments, this may all be for nothing if the learner does not perceive it as such. This process, known as *pre-authentication* (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). Obviously, we chose for the latter interpretation.

Figure 1 summarizes the important elements of the above discussion: to positively influence student learning authentic assessment should be *aligned to* authentic instruction; authentic assessment requires students to *demonstrate* their competencies in a situation that resembles professional practice; and authenticity is *subjective*, which makes it important to take students' perceptions into account when designing an authentic assessment.

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.

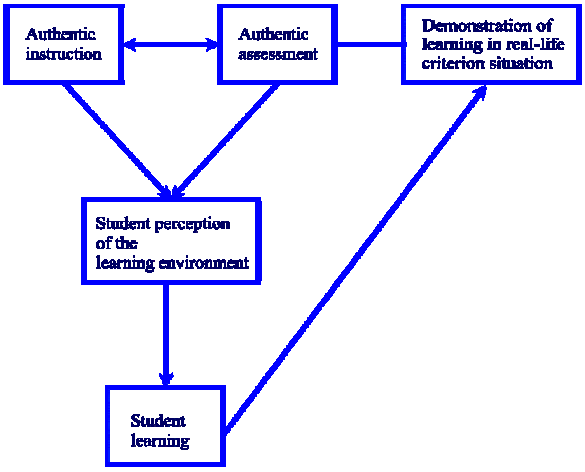

Figure 1. The general outline

Towards a Five-Dimensional Framework for Authentic Assessment

For defining 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. Many sub-concepts and synonyms came to light, which were conceptually analysed and divided into categories, resulting in five main aspects of authenticity, namely: the task, the physical context, the social context, the assessment result or form and the criteria. We argue that these aspects are dimensions that 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). The degree of authenticity is not solely a characteristic of the assessment chosen; it needs to be defined in relation to the (professional) criterion situation in real life. 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 in real life to a varying degree, thereby increasing or decreasing the authenticity of the assessment. Figure 2 shows the five dimensions of authentic assessment and their sub-elements. The interpretation of the five dimensions will be further explained and examined in the rest of this paper.

An Argumentation for the Five Dimensions of Authentic Assessment

As our definition for authentic assessment showed, we argue that the authenticity of all five dimensions is defined by its resemblance to the criterion situation and, to recapitulate, a *criterion situation* reflects a real-life situation that students can be confronted with in their internship or future professional life. This sets the frame for the argumentation of the five dimensions 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 (e.g., Herrington & Herrington, 1998; Wiggins, 1993), but different researchers stress different elements of an authentic task. Brown, Collins and Duguid (1989) stress that the authenticity of the task can be expressed on a continuum; the degree of authenticity of an educational task can be gauged by the degree to which the activities undertaken by students are like those activities undertaken by practicing communities in the real world beyond the learning institution. Our framework defines the degree of authenticity of the task by its degree of resemblance to the criterion task with respect to the integration of knowledge, skills and attitudes, its complexity and its ownership. Furthermore, the users of

the assessment task should perceive the task, including above elements, as representative, relevant and meaningful.

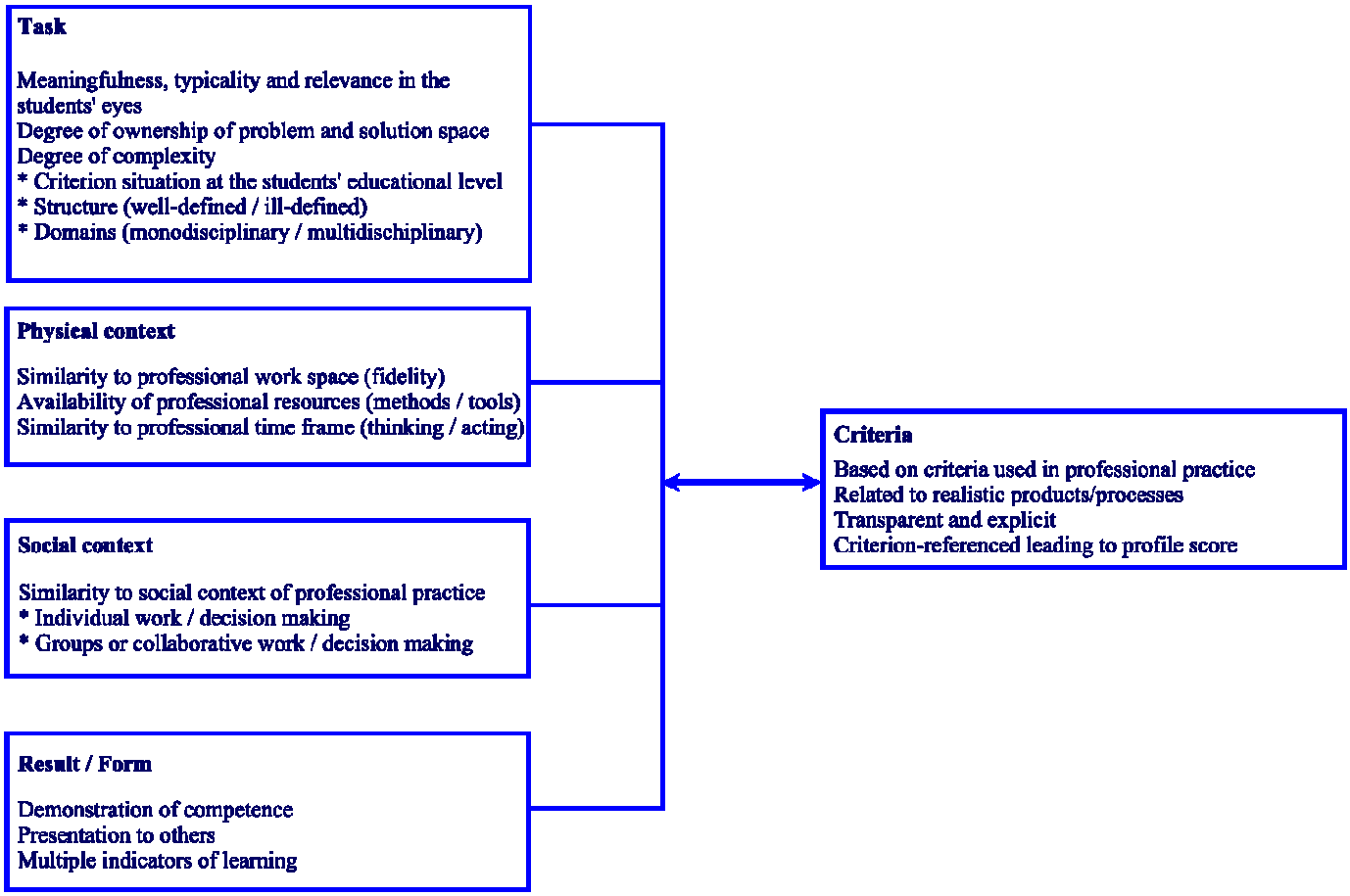

Figure 2. Five dimensions of authentic assessment

First of all, an authentic assessment requires students to *integrate knowledge, skills and attitudes* as professionals do (Gielen, Dochy, and Dierick, 2003; Darling-Hammond & Snyder, 2000). Moreover, we argue that 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, as is often argued by most advocates of authentic assessment (f.e. Herrington & Herrington, 1998; Wiggins, 1993). Even though most authentic problems are complex, involving multidisciplinarity, ill-structuredness, and having multiple possible solutions, 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 real-life task. Savery and Duffy (1995) argue 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 (Resnick, 1987), both decreasing the amount of ownership for the employee. 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. Messick (1994) and (Stein, Isaacs, & Andrews 2004) argue that meaningfulness to students is at the heart of authenticity. They stress that merely providing assessment tasks representative of the professional discipline is not enough for creating an authentic experience. This needs to be combined with the perception of meaningfulness from the learners' perspective. 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 capability of a soldier to find the problem in a non-functioning jeep can be assessed in a clean garage, with the availability of all the possibly needed equipment, but the future physical environments will 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 capacity 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 (Herrington & Oliver, 2000) or the wayb. *Fidelity* is often used in the context of computer simulations, which describes 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*, which mostly contain *relevant as well as irrelevant information* (Herrington & Oliver, 2000), should resemble the resources available in the criterion situation (Arter & Spandel, 1992; Segers, Dochy, & De Corte, 1999). For example, Resnick (1987) argues that most school tests involve memory work, while out-of-school activities are often intimately engaged with tools and resources (calculators, tables, standards), making these school tests less authentic. Segers, Dochy and De Corte (1999) argue that it would be inauthentic to deprive students from resources as professionals also 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, which is 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) says that an authentic assessment should not rely on unrealistic and arbitrary time constraints. In sum, the level of the 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. Brown, Collins and Duguid (1989) argue that an authentic activity should reflect practices of the culture. In real life, working together is often the rule rather than the exception and Resnick (1987) emphasizes 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 like social interaction, positive interdependency and individual accountability need to be taken into account (Slavin, 1989). When, on the other hand, the assessment is individual, the social context should stimulate some kind of competition between learners.

Assessment Result / 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 or form 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/form is characterized by three elements. It should require students to *demonstrate* their learning by creating a *quality product or performance* that they can be asked to produce in real life (Wiggins, 1993). The rationale behind requiring student to demonstrate their learning in a real life situation is that this makes it more valid to make inferences about underlying competencies and to predict future functioning in professional career (Klarus, 2003). Besides, a demonstration gives more insight into the problem solving processes that students use to handle a situation and taking the product as well as the process into account is an important characteristic of alternative assessment in general (Dochy, 2001). Since the demonstration of relevant competencies is often not possible in one single test, an authentic assessment should involve a *full array of tasks and multiple indicators of learning* in order to come to fair conclusions about (professional) competence (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 *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). This characteristic serves another goal as well. It signals to students that their work is important to other people, which increases the perception of relevance and meaningfulness (Darling-Hammond, 1994).

Criteria and Standards. Criteria are those characteristics of the assessment result that are valued; standards are the levels of performance expected from various grades and ages of students (Arter & Spandel, 1992). Criteria and standards should *concern the development of relevant professional competencies* and should be *based upon criteria used in the real-life (i.e., criterion) situation* (Darling-Hammond & Snyder, 2000). 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. Setting criteria and making them *explicit and transparent to learners* beforehand is important in authentic assessment (Darling-Hammond, 1994), because this guides learning (Sluijsmans, 2002) and after all, in real life, employees usually know on what criteria their performances will be judged. Moreover, this implies that authentic assessment requires *criterion-referenced judgment*.

Figure 2 shows that the criterion dimension has a special status in the five-dimensional framework. This dimension has a reciprocal relationship with the other dimensions. On the one hand, criteria based on professional practice, which is often the starting point for developing authentic assessments, should guide the *interpretations* of the other four dimensions. On the other hand, criteria can also be based on the interpretation of another dimension of the framework. For example, if the physical context requires the use of certain resources and tools, the criteria should specify how these should be used in the demonstration of competence, as these criteria guide students' learning.

Some Considerations

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

The first consideration deals with *predictive validity* and the goal of the assessment. Predictive validity is the degree to which future performance can be predicted by current assessment performance. 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 very useful for certain purposes as high-stakes summative assessment on an individual achievement, where predicting students' ability to function competently in future professional practice is not the purpose. Hence, the goal of the assessment partly determines the importance of increasing the authenticity of an assessment.

Furthermore, as mentioned in the introduction, authenticity is only one of the (crucial) elements of alternative assessments. Moreover the framework, as it is described in the theoretical argumentation, shows an ideal picture of authentic assessment practices. In real educational practice, one has to deal with other quality criteria (eg., reliability) and practical possibilities as well. For example, a certain criterion situation describes that an authentic assessment should allow students to work on the assessment task for several hours spread over one week, while practical possibilities or reliability or accountability considerations make it impossible to completely comply with this timeframe. Every educational assessment requires a compromise between different quality criteria, goals and practical possibilities. However, we argue that increasing the authenticity of the assessment has to be carefully considered in this debate, especially when it comes to competency-based assessments.

Another consideration in designing authentic assessment is that we should not lose sight of the educational level of the learners. Students who are at the beginning of their studies possibly cannot deal with the authenticity of a real complex professional situation. If they are forced to do this, it will result in cognitive overload and in turn will 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) argues 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.

A third consideration sheds a light on 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, Duffy & Fishman, 1993). This implies that the optimal levels of authenticity of each of the five dimensions 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 done internships and have a better idea of professional practice, requires more authenticity of the physical context than when assessing first year students, who often have little practical experience. Designers must take student perspectives and the changes therein into account when designing authentic assessment.

The exploratory, qualitative study described following sections of this paper has two main goals. First, it explores whether our five-dimensional framework completely describes authenticity or whether important elements are missing. Second, it explores the relative importance of the five dimensions. A sub goal of this study is 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 insights in 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: (1) eight sophomore VTP students (Mean age 18.5 years), (2) eight sophomore BRP students (Mean age 20.9 years), (3) eight senior VTP students (Mean age 19.7 years), (4) four senior BRP students (Mean age 31.4 years), and (5) eleven teachers (Mean 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. 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 since 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 post-test. For this reason, a second set of different but comparable case descriptions were 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/form; and cases D and H were completely authentic according to the five-dimensional framework.

Procedure

At the beginning and end of the GSS session, participants were presented four case descriptions (ABCD or EFGH). In six paired comparisons (4x3/2), they chose the case that they considered to be a more authentic assessment. A distinction was made between VTP students and BRP students because it was possible that due to the differences in their studies, they would have different perceptions of what determines authenticity.

After the initial rating of the case descriptions, the participants were informed of the purpose of this study. To create a common frame of reference, a general description of the

terms authenticity (i.e., true to life) and authentic assessment, as a means of evaluating professional behavior, was given. Then, participants were required to brainstorm about what authenticity of an assessment meant to them and they had to enter their own statements into the GSS. After this electronic brainstorm, the contributions were discussed in order to clarify them. This was recorded for later use and analysis.

After the brainstorm and discussion, 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/form: What has to come out of it/ What is the result of your efforts?
5. Criteria: How does what you've done have to be evaluated/judged?

The following two activities consisted of paired comparisons to determine the relative importance of the dimensions. One activity consisted of 10 paired comparisons of the five dimensions (5x4/2). Participants had to choose the dimensions of the framework that they perceived as more important for authentic assessment. The last 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, and choices of each individual participant are anonymous. This means that scores per participant were not available. This precludes the possibility of carrying out statistical tests. On the other hand, this anonymity is likely to stimulate response in idea generation and to increase the reliability of answers since socially acceptable answering behavior is inhibited. The data, thus, were qualitatively analyzed. The tapes of the discussions were transcribed. The resulting discussion statements and the statements keyed in during the brainstorms were analyzed as to 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, was 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/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 in dispute.

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 students VTP and BRP, teachers, senior students VTP and BRP) from 1 to 5. The paired comparisons of the case

descriptions were analyzed for the same groups, but were measured in pre- and post-tests, which resulted in ten rankings from 1 to 4.

Table 1. Rankings of the five dimensions by the different groups (1 = most important, 5 = least important)

	Task	Physical context	Social context	Result/ form	Criteria
Sophomore VTP students	2	4.5	4.5	1	3
Sophomore BRP students	1	3.5	5	3.5	2
Teachers	1	4	5	2	3
Senior VTP students	2	5	3.5	3.5	1
Senior BRP students	2	4	5	1	3
Total	8	21	23	11	12

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

Table 2. Rankings of the case descriptions by the different groups (1 = most authentic, 4 = least authentic)

	All authentic except for the task	All authentic except for the physical context	All authentic except for the result/form	All authentic
Sophomore VTP, pre test	2	3	4	1
Sophomore BRP, pre test	2	3	4	1
Sophomore VTP, post test	3	2	4	1
Sophomore BRP, post test	2	3	4	1
Teachers pre test	3	4	2	1
Teachers post test	2	4	3	1
Senior VTP pre test	2	3.5	3.5	1
Senior BRP pre test	2	3.5	3.5	1
Senior VTP post test	2	3.5	3.5	1
Senior BRP post test	1	4	2.5	2.5

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 for the senior BRP students on the post-test (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 with a "4" meaning that they perceived this case to be the least authentic. In other words, they perceived the result/form dimension as most important for designing an authentic assessment. Teachers, on the other hand, ranked the case that was authentic except for the physical context as the least authentic case (score 4), which meant that teachers perceived the physical context to be most important in designing an authentic assessment. Senior students did not appear to differentiate, meaning that they perceived the cases with no authentic physical context or with no authentic result/form as equally inauthentic (score 3.5). To sum up, 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, that the case was unequivocally seen as the most authentic. Second, there appear to be contradictory results with respect to task authenticity compared to the results of the paired comparisons of the dimensions. Finally, teachers and students appear to differ with respect to the importance of the authenticity of the physical context versus result authenticity when evaluating assessment cases.

Completeness and Relative Importance: What Do Participants Say?

Table 3 shows that all dimensions received attention in the brainstorms and discussions. Furthermore, these results corroborated the earlier findings in that the social context received the least attention in all groups. Besides the five dimensions, almost all sub-elements of the dimensions, described in the framework, were reviewed.

Table 3. Number of statements per dimension of each group

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

Based upon the number of statements and the ratios of the statements compared to each other as shown in Table 3, sophomores placed primary interest on the task followed by the physical context. Seniors placed most emphasis on the task and the result/form. Teachers perceived the task, the result/form and the criteria dimensions as equally important for authentic assessment. Teachers differed from all students, regardless of the year, with respect to the emphasis on the physical context. Teachers devoted a lot of time discussing the required fidelity level of the physical context in an effective authentic assessment. Especially the question whether the physical context should be real professional practice or a simulation in school was discussed.

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 the sophomore students, especially when it comes to the task and the result/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 more clear and easy (e.g., "assignments should be less vague, not more than one answer should be possible") instead of simulating real-world

complexity. In the result/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/form dimension much; they only mentioned that reshaping current tests in the form of cases would make it more realistic. In other words, according to their perceptions, every kind of assessment could be made more authentic by adding realistic information.

Table 4. Variables in the 'other' category per group

	Sophomore students	Senior students	Teachers
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 / preconditions	-	-	7
Influence on the learning process	-	-	4
Not defined / nonsense	-	9	26

A specification of the *other* statements (see Table 4) showed, first, that all groups used 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 showed that issues concerning the assessor of an authentic assessment and organizational or preconditional 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 students in this professional context. These are important issues, but are beyond the purpose of the five-dimensional framework for designing authentic assessments. Finally, Table 4 showed that sophomores took the opportunity to talk and complain about the "instruction". Although this was not asked of them (i.e., it was about assessment) 28 statements dealt with what was taught and not with what was assessed. Seniors were more focused and teachers 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 sub-elements described in the framework. However, issues concerning the assessor of authentic assessments turned out to be important to all participant groups and should be considered in designing effective authentic assessments.

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

Furthermore, not all groups perceived the dimensions and their sub-elements in the same way. The teachers and seniors mostly agreed with each other and with the theoretical

framework, while the sophomores often deviated from the other groups. There were no differences between VTP and BRP students.

Discussion

At this point it is necessary to restate the perspective of this study. The three questions with which we began were: (1) Is the framework complete? (2) What is the importance of the five dimensions? (3) Do students differ from teachers with respect to what they perceive as important for authenticity? These questions shed a light on possible guidelines for designing authentic assessments.

With respect to the first question, the answer appears to be yes. The five dimensions appear to adequately define authenticity as seen in both the brainstorms 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 brainstorms most sub-elements of the dimensions as described by the framework were seen as important when designing authentic assessment. However, on the basis of the brainstorms and discussions it can be argued that including 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 a sub-elements to the criteria dimension. We chose for the second option.

The paired comparisons of the five dimensions showed some subtle differences in the importance of the five dimensions for providing authenticity. While the task, the result/form and the criteria dimensions turned out to be very important for authenticity, the physical context and especially the social context were perceived as less important. All groups stressed the need for individual testing, while on the other hand students as well as 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 the social context is not important for authentic assessment, but if choices have to be made in designing an authentic assessment, the social context is probably the first dimension to leave out.

The findings on importance of the task are sometimes contradictory. While the brainstorms and the paired comparisons of the dimensions showed that the task was perceived as very important by all, the paired comparisons of the cases made the task seem less important. It is possible, thus, that while the respondents consider the task (as 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 was the case, the paired comparisons of the five dimensions were probably a more objective measure of the importance of the five dimensions.

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

sophomore students have different conceptions and possibly misconceptions of real professional practice and thus of authenticity of assessment.

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

Future Implications

The findings of this explorative study allow for some critical considerations 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 older, more experienced students and teachers. For authentic assessment to work, two options need to be considered in this matter. Either the assessment meets the perceptions 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 changing student perceptions and thereby opening the possibilities to change their learning behavior towards professional competency development should be given explicit attention when implementing authentic assessment. In addition, the (mis)conceptions of sophomore students can also indicate that the importance of authentic assessment, in general, increases as students proceed through their studies.

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. Previous research (De Bock, Verschaffel, Janssens, Van Dooren, & Claes, 2003; Gulikers, Bastiaens, & Martens, in press) shows that a more authentic physical context does not automatically lead to improvements in students' performance. Research should examine if assessing students in a real professional context has additional value for students, or if assessing with a simulation in school is authentic enough as long as students are confronted with an authentic task, result/form and criteria. Simulation in school is probably easier and less expensive to implement and therefore warrants careful consideration.

The explorative nature of this study without the possibility of quantitative statistical analyses due to the nature of the GSS makes firm conclusions impossible. However, what the data of this study DO show is that authenticity is definitely a multi-faceted 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, 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 said in the beginning of this paper, authenticity is only one of the elements and competency-based (alternative) assessment (Birenbaum & Dochy, 1996; Dierick & Dochy 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 generalisability), 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 paper.

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.

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 constructive alignment. *Higher Education, 32*, 347-364.

Birenbaum, M., & Dochy, F. J. R. C. (1996). *Alternatives in assessment of achievements, learning processes and prior knowledge*. Boston, MA: Kluwer Academic Publishers.

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 the implications for asessment. In M. Segers, F. J. R. C. Dochy, & E. Cascallar (Eds.), *Optimising new modes of assessment: In search of qualities and standards*. Dordrecht: 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. (1994). Setting standards for students: The case for authentic assessment. *The Educational Forum, 59*, 14-21.

Darling-Hammond, L., & Snyder, J. (2000). Authentic assessment in teaching in context. *Teaching and Teacher Education, 16*, 523-545.

De Bock, D., Verschaffel, L., Janssens, D., Van Dooren, W., & Cleas, K. (2003). Do realistic contexts and graphical representations always have a beneficial impact on students' performance? Negative evidence from a study on modelling non-linear geometry problems. *Learning and Instruction, 13*, 441-463.

Dierick, S., & Dochy, F. (2001). New lines in edumetrics: New forms of assessment lead to new assessment criteria. *Studies in Educational Evaluation, 27(4),* 307-329.

Dochy, F. (2001). A new assessment era: different needs, new challenges. *Research Dialogue in Learning and Instruction, 10*(1), 11-20.

Dochy, F. J. R. C., & McDowell, L. (1998). Assessment as a tool for learning. *Studies in Educational Evaluation, 23*(4), 279-298.

Entwistle, N. J., & Entwistle, A. (1991). Contrasting forms of understanding for degree examinations: the student experience and its implications. *Higher Education, 22*, 205-227.

Gielen, S., Dochy, F., & Dierick, S. (2003). Evaluating the consequential validity of new modes of assessment: The influence of assessment on learning, including pre-, post- and true assessment effects. In M. Segers, F. Dochy, & E. Cascallar (Eds.), *Optimising new modes of assessment: In search of quality and standards*. Dordrecht: Kluwer Academic Publishers.

Gulikers, J., Bastiaens, Th., & Martens, R. (in press). The surplus value of an authentic learning environment. *Computers in Human Behavior*

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.), *Desinging environments for constructive learning* (pp. 88-108). Berlin: Springer-Verslag.

Huang, H. M. (2002). Towards constructivism for adult learners in online learning environments. *Britisch Journal of Educational Technology, 33*, 27-37.

Klarus, R. (2003). *Instrument voor competentiegericht toetsen en beoordelen.[Instrument for competency-based assessment and evaluation]*. Retrieved December, 2, 2003, from: **http:/www.kenniscentrumevc.nl/site/documenten/KC3.pdf**

McDowell, L. (1995). The impact of innovative assessment on student learning. *Innovations in Education and Training International, 32*(4), 302-313.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. Educational Researcher, 23(2), 13-23

Meyer, C. (1992). What's the difference between authentic and performance assesment? *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, Onderwijstechnologisch Expertisecentrum.

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:Tthe rhetoric and technology of authenticity in education*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

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., & Brown, S. (1997). But is it fair?: An exploratory study of student perceptions of the consequential validity of assessment. *Studies in Educational Evaluation, 23*(4), 349-371.

Savery, J. R., & Duffy, T. M. (1995). Problem based learning: An instructional model and its constructivist framework. In B. G. Wilson (Ed.), *Constructivist learning environments*. Englewood Cliffs, NJ: Educational Technology Publications.

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.

Stein, S.J., Isaacs, G., & Andrews, T. (2004). Incorporating authentic learning experiences within a university course. *Studies in Higher Education, 29(2),* 239-258.

Struyven, K., Dochy, F., & Janssens, S. (2003). Students' perceptions about new modes of assessment in higher education: A review. In M. Segers, F. Dochy, & Cascallar E. (Eds.), *Optimising new modes of assessment: In search of qualities and standards* (pp. 171-224). Dordrecht: Kluwer Academic Publishers.

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: Open University Press.

Uhlenbeck, A. (2002). *The development of an assessment procedure for beginning teachers of englich as a foreign language.* Unpublished doctoral dissertation. University of Leiden, Leiden.

Van Merriënboer, J. J. G. (1997). *Training complex cognitive skills: a four-component instructional design model for technical training*. Englewood Cliffs, NJ: Educational Technology Publishers.

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, US: Jossey-Bass/Pfeiffer.