

Authentic assessment, student and teacher perceptions:

Citation for published version (APA):

Gulikers, J., Bastiaens, T., & Kirschner, P. A. (2006). Authentic assessment, student and teacher perceptions: the practical value of the five dimensional-framework. *Journal of Vocational Education & Training*, 58(3), 337-357. <https://doi.org/10.1080/13636820600955443>

DOI:

[10.1080/13636820600955443](https://doi.org/10.1080/13636820600955443)

Document status and date:

Published: 01/01/2006

Document Version:

Peer reviewed version

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

<https://www.ou.nl/taverne-agreement>

Take down policy

If you believe that this document breaches copyright please contact us at:

pure-support@ou.nl

providing details and we will investigate your claim.

Downloaded from <https://research.ou.nl/> on date: 10 May. 2021

Open Universiteit
www.ou.nl



Catchline *Journal of Vocational Education and Training*
Vol. X, No. X, Month 200X, pp. 000–000

Running head Authentic assessment, student and teacher

This is a pre-print of the article that was published as

Gulikers, J. T. M., Bastiaens, Th. J., & Kirschner, P. A. (2006). Authentic assessment, student and teacher perceptions: the practical value of the five dimensional-framework. *Journal of Vocational Education and Training*, 58, 337-357

Copyright Taylor & Francis, available online at

<http://www.tandf.co.uk/journals/rjve>

Authentic assessment, student and teacher perceptions: the practical value of the five-
dimensional framework

J.T.M. Gulikers*, Th. J. Bastiaens, and P.A. Kirschner

Open University of the Netherlands

* Corresponding Author. Open University of the Netherlands, Educational Technology Expertise Center, PO Box 2960, 6401 DL Heerlen, The Netherlands.
Email: Judith.Gulikers@ou.nl

Abstract

This study focused on determining the facets of assessment authenticity by exploring the perceptions of both Vocational Education and Training students and teachers. It elaborates on a theoretical five-dimensional framework (5DF) that differentiates between five dimensions and several sub-elements of authenticity. This framework led to the development of a questionnaire for examining if the facets of the 5DF are recognized by students and teachers in practice. Reliability and factor analysis as well as readability scores were used. Teachers recognised both the dimensions and the sub-elements as facets that determine assessment authenticity. In the eyes of the students, four of the five dimensions (Task, Physical Context, Form and Result/Criteria) determine authenticity, while students do not perceive the Social Context as a characteristic of assessment authenticity, neither do they differentiate the several sub-elements. Implications for the using the 5DF to develop or evaluate authentic assessments are discussed.

Keywords: Authentic assessment; student and teacher perceptions; perception questionnaire; vocational education

Authentic assessment, student and teacher perceptions: the practical value of the five-dimensional framework

Assessment, in the past often referred to as “testing”, has been an important aspect of educational practice for a long time (eg., Bloom, 1956) and the idea that assessment is a salient variable in determining what and how students learn has become an often examined subject of research in the last two decades (eg., Scouller, 1997; Scouller & Prosser, 1994; Thomas & Bain, 1984). Segers, Dochy, and Cascallar (2003) argue that several aspects characterise tests or assessments, one of which being the authenticity continuum. This continuum shows that an assessment can span the gap between artificial and decontextualized on the one hand or authentic and situated on the other. New modes of assessment, that focus on competencies needed for future jobs, tend to lean towards the authentic side of the continuum, since authenticity is expected to be crucial for preparing students for the dynamic world of work that characterises current society (Boud, 1995; Newmann & Associates, 1996; Segers et al.). This article focuses on authenticity operationalised as a correspondence between the assessment situation and the working or professional practice situation (Gulikers et al, 2004; Messick, 1994). By creating this correspondence, authentic assessment is expected to stimulate students to develop skills or competencies relevant for their future world of work. But authenticity is not an ‘objective’ quality as such; it is subjective and dependent on who is judging the authenticity. In this study it is argued that assessment authenticity depends on the resemblance between the assessment and the working situation it aims to reflect, but more importantly on a person’s (e.g., teacher or student) perception of this resemblance. This implies that what one person perceives as authentic is not necessarily authentic in the eyes of someone else.

This study focuses on determining the facets of authenticity by exploring perceptions of authentic assessment in a Vocational Education and Training context. We argue that it is important to explore the concept of authenticity from two different angles, namely a theoretical and a practical angle. In a literature review, described in Gulikers, Bastiaens, and Kirschner (2004), the theoretical approach resulted in a five-dimensional framework (5DF) for describing assessment authenticity from an objective viewpoint. The *practical angle* focuses on examining what determines authenticity in the perception of different users (e.g., developers or assessees). Boud (1995) argued that assessment research should focus on carefully examining assessments as students see it. Little is known about how students experience various assessment characteristics (Learning & Skills Research Centre (LSRC), 2004), while this provide us with the most relevant information for developing assessments that are helpful for student learning (Boud, 1995). Therefore, this study focuses on the practical side by examining how students and teachers perceive the authenticity of an assessment? More specifically, are the theoretical facets of authenticity, as described in the 5DF, recognised and valued in practice?

Theoretical Background

A major problem of education is the fact that there are gaps between teaching in school and the real world and between assessment tasks and what occurs in the world of work (Achieve, 2006; Boud, 1990). In the last decade of the previous century, the educational culture changed from knowledge-based towards competency-based education and the educational goal became to develop competent students and future employees (Segers et al, 2003). However, recent review reports in several countries showed that students are not prepared for the real world (i.e., work or college) after leaving school (Achieve; American College Testing, 2006; LSRC, 2004; Ministry of Education, Culture and Science of the Netherlands, 2005). The problem is that school standards are not aligned to the expectations

of the world of work (Achieve; Messick, 1994). Bridging the gap between learning and working is a salient issue in the 21st century.

In Vocational Education and Training (VET) in the Netherlands, the context of this study, the link between learning and working is even more crucial, because of several characteristics of this type of education (Biemans et al, 2004): (a) VET is primarily a final (for a job) type of education, (b) there is a strong focus on becoming a practitioner and less on pure theoretical development, and (c) students at these schools are often more practically oriented and directed towards working instead of studying. As a result, ‘authentic learning’ has become an important issue in these schools (Kerka, 1995). Ideas from cognitive apprenticeship (Collins et al, 1989) and situated learning (Brown et al, 1989), in which authentic learning plays a central role, have resulted in educational practices in which learning activities are contextualized in realistic situations. Literature on authentic practices (e.g., Newmann & Associates, 1996; Wiggins, 1993), but also professional development and assessment literature (e.g., Boshuizen et al, 2004; Segers et al. 2003) argues that schools need to provide students with authentic real-life learning experiences, with their complexity and limitations, early in their educational trajectory to stimulate students to more higher-order thinking processes and active learning. To foster authentic learning and to improve student achievement it is imperative that authentic assessment is aligned to authentic instruction as well as to real world expectations (Achieve, 2006; Biggs, 1996; Linn et al., 2002).

The need to contextualize assessment in interesting, real-life and authentic tasks is considered *one of the crucial elements* of new modes of assessment (Birenbaum & Dochy, 1996). Dochy (2001) described the assessment of the application of knowledge to actual, real-life (authentic) cases as the core goal of alternative assessments. Gielen et al (2003) even argued that authenticity of the assessment tasks is imperative to achieve the expert level of problem solving, because this addresses the competencies of the discipline (Gielen), as well

as the thinking processes that characterise expert performance (Newmann & Associates, 1996) Messick (1994, p.21) stresses the importance of assessment authenticity, because authenticity deals with “not leaving anything relevant out”. By focusing on resembling real-world performances, authentic assessment are likely to cover the relevant aspects (e.g., knowledge, skills, processes) of this real-world performance. In short, assessment authenticity is argued to be important for preparing students for the unexpected world of work.

The problem, however, is that what is meant by authenticity is often not clearly defined. A literature study (Gulikers et al, 2004) showed that authenticity can be, and often is described in very different ways. In addition, no clear and very diverse guidelines exist for developing authentic assessments. This results in authentic assessment practices that ‘dress up’ existing assessments with some ‘real world’ elements (Cummings & Maxwell, 1999), without a clear understanding of what these real-world elements are and how these are appropriately implemented. To get a better insight into what authenticity really is and what facets are important to consider in developing authentic assessment (i.e., what these real world elements exactly are), we first present a theoretical reference frame and definition of authenticity and then describe a study examining whether the theoretical hypothesised facets of assessment authenticity are recognised and supported in practice.

The Theoretical Angle: Objective Authenticity

A crucial question in defining authenticity is: “authentic to what?” (Messick, 1994, p. 18). Honebein, Duffy and Fishman (1993) argued that the authenticity is a relative concept, meaning that the authenticity of something can only be defined by its resemblance to *something else* and it is the specification of this something else that is crucial for a further discussion about and examination of the concept of authenticity. For example, Messick argued that an assessment can be very authentic with respect to school objective but completely not authentic with respect to the real world, because school objectives do not

reflect real-world requirements (Achieve, 2006). Since the goal of education is to prepare students for the world of work, at least in vocational education (Wonacott, 2000), the point taken in this study is that the authenticity of an assessment should be defined by its *resemblance to students' current or future professional practice*. Resembling professional practice, however, means more than merely implementing some superficial 'realistic' elements, which only leads to, as Cumming and Maxwell (1999) call it, 'camouflage' (p. 188).

Several researchers described standards for authenticity that should be applied to all authentic practices to stimulate the kind of learning that students need for their future lives (Herrington & Herrington, 1998; Newmann & Associates, 1996; Wiggins, 1993). Newmann and Associates identified three key criteria characterising successful authentic and intellectual performance in most fields of work: construction of knowledge, disciplined inquiry, and value beyond school. Wiggins (1993) suggests several indicators for authentic assessment including observable performances and dealing with ill-structured and complex issues. This literature argues that these kinds of criteria allow for judging the quality of assessment authenticity and hold for *all* authentic assessments, independent of a discipline or occupation. In fact, the American College Testing Group (2006) and Wilson, Lizzio and Ramsden (1997) argued that, independent of a student's choice for work of college after finishing high school, preparing students for today's economy, requires educating *all* students according to a common academic standard. Mueller (2006) calls these discipline-independent standards 'process standards', which refer to generic skills relevant for all kinds of jobs. By taking these discipline-independent standards into account, an authentic assessment can help students succeed in their future lives (i.e., college or work; different fields of work). However, in occupational types of education where students are educated for a specific profession or where students are more practically instead of theoretically/thinking-oriented, an authentic

assessment should *also* be aligned to more occupation-specific content and value performance standards that define knowledge, skills and attitudes characteristic of a specific field of work (Achieve, 2006; Darling-Hammond & Snyder, 2002; Gulikers et al; Mueller). These kinds of discipline-dependent standards are likely to be important for showing students to link between the assessment and their future work, which is important for student motivation and learning, especially for more practically-oriented students (Kasworm & Marienau, 1999).

Furthermore we previously argues that authenticity is a relative concept and that the authenticity of an assessment needs to be judged by its resemblance to the working situation it aims to reflect. This implies that requiring ‘dealing with ill-structured and complex issues’ is *only* authentic when the working situation also requires ‘dealing with ill-structured and complex issues’ (Gulikers et al, 2004; Messick, 1994). Developing an authentic assessment should start with an analysis of the professional practice situation to find out what kind of knowledge, skills and attitudes (or competencies) experts use when handling this situation and how they use them. This analysis provides up-to-date performance standards for developing an authentic assessment or for evaluating the authenticity of the assessment. The definition used in this study is that an authentic assessment should *require students to use and demonstrate the same (kind of) competencies, or combinations of knowledge, skills and attitudes, that are applied in this situation in professional life* (Gulikers et al, 2004).

Additionally, Cronin (1993) and Newmann and Wehlage (1993) argued that authenticity is a dimensional construct. This means that an assessment can be more or less authentic depending on its *degree of resemblance* to the professional practice situation. Gulikers and colleagues (2004) expanded this idea by describing a literature study on assessment authenticity that resulted in a five-dimensional framework (5DF). This framework argues that the degree of authenticity of an assessment depends on the degree of resemblance between five assessment characteristics and the professional practice situation. These five

dimensions are (a) the assessment task(s), (b) the physical context in which the assessment takes place, (c) the social context of the assessment, (d) the result or form that defines the output of the assessment, and (e) the assessment criteria. From now on, these will be referred to as the five dimensions of authenticity. These five dimensions can be further described by several sub-elements. Figure 1 shows the five dimensions and their sub-elements (see Gulikers et al, 2004 for a full description of this framework). The main idea behind this framework is that different kinds of authentic assessments can be developed by varying the degree of resemblance between the five dimensions with their sub-elements and the professional situation that the assessment aims to reflect.

*** INSERT FIGURE 1 ***

Up to this point, one could argue that authenticity seems to be an objective concept. We argue, however, that authenticity is not purely objective, but that an important part of authenticity is in the eye of the beholder, as will be described in the next section.

The Practical Angle: Subjective Authenticity

Just as the concept ‘expensive’ has an ‘objective’ side, as in a Rolls Royce is expensive compared to an Austin Mini-Cooper (the original), it also has a ‘subjective’ side in that a Rolls Royce is expensive for an average person, but not for a billionaire. The same is true for assessment authenticity. It has an objective side, which was just handled, but also a subjective one, namely how the assessee/assessor perceives the authenticity of an assessment.

It is important to investigate this subjective side of authenticity for two reasons. First, student perceptions of assessment characteristics are found to determine what and how students learn (Entwistle, 1991; McDowell, 1995; Scouller, 1997; Struyven et al, 2003; Van Rossum & Schenk, 1984). This implies that it is not the objective authenticity, but rather student perceptions thereof that influence their learning. Before an authentic assessment will stimulate students to develop professionally relevant skills or competencies, students have to

perceive a resemblance between the assessment and their future professional life. Second, student and teacher perceptions of assessment characteristics are found to differ (MacLellan, 2001; Ngar-Fun, 2005). With respect to authenticity, Honebein and colleagues (1993) argued that its perception can change as a result of age, kind of and amount of schooling or practical experience. As a result, students and teachers are likely to differ in how they perceive authenticity. When this is indeed so, problems for educational practices might arise, since teachers are mostly the ones to develop the authentic assessment and they do so according to what they think is authentic. This process is called 'pre-authenticitation' and shows the relevance of exploring teacher as well as student perceptions of authenticity, as it is important to develop assessments that are perceived as being authentic not only by teachers, but by students as well.

This Study

The goal of this study is to determine the facets of authentic assessment by examining if the elements proposed in the 5DF are recognised in practice. For this purpose, we needed to develop a way to measure if and how students and teachers value and differentiate various facets of authenticity in an assessment. This led to the development of a perception questionnaire based on the 5DF. If different facets are perceived, the idea that an assessment can be made more or less authentic in various ways by manipulating those facets in an assessment, is supported. On the other hand, if students and/or teachers do not recognise various facets, the use of the 5DF in theory and in practice, has to be reconsidered. Hence, this study might provide practical guidelines for developing and evaluating authentic assessments.

Method

Participants

One hundred and fifteen students (mean age = 18.1, SD = 3.27) of a Vocational Education and Training (VET) for Social Work and 18 of their teachers enrolled in this study. The students were in their second year of study and studied Social Work in a vocational training programme in which learning and working were alternated on a regular basis. The 18 teachers were all assessors in the authentic assessment that was the object of this study.

Materials

The Authentic Assessment. This study was designed around an authentic assessment on the topic of “determining care needs”. Students received a case description of a handicapped client, living in a social home, who has been in a car accident and had physical and behavioural problems afterwards. This caused problems for the client himself as well as for the people around him. The goal was to draw up an activity and guidance plan to help the client and to improve his functioning in the community. The assessment consisted of two parts. First, students received the case description of the client and had to formulate, in writing, their ideas and the actions they were planning to analyse and observe the client. Second, students were confronted with the client in a ten minute role-play in which they had to observe the client and discuss their activity and guidance plans with this client. Both activities took place in school. The second part was a simulated role-play in which a teacher played the role of the client. Every student performed the assessment individually. Moreover, student performance was observed and graded by two assessors (teachers) on a list of ten performance criteria known to the students, and the goal of the assessment was summative.

The Instructional Phase. A competency-based instructional period of 9 weeks preceded the authentic assessment. This period focused on planning activities within the social work institution. During 8 weeks, students worked in groups on critical professional problem situations, for example “analysing client needs”, “planning individual activities based on

client needs”, or “planning groups activities”. They had to set learning goals focussing on knowledge as well as skills/attitudes. After a period of self-study and skill-training, all group members had to perform a formative assessment. This was a role-play assignment based on a new, but related problem case. The summative assessment (in this case “determining care needs”) was based on a selection of course objectives and performance criteria. The performance criteria for the summative assessment were made known to students one week prior to the assessment in which students were freed from obligatory educational activities and in which they could choose for themselves how to prepare for the assessment

The Perception Questionnaire. To study student and teacher perceptions of authenticity, a 5-point Likert scale perception questionnaire was developed based on the authenticity facets of the 5DF. The idea behind the questionnaire was to first measure perceptions at a dimensional level (the resemblance of the five dimensions to professional practice) and then, to measure the perception of four sub-elements (the resemblance of sub-elements to professional practice). In addition to measuring perceptions of the authenticity of different facets of the assessment, the Vocational Relevance scale of the Course Perception Questionnaire (Entwistle & Ramsden, 1983) was used to measure perceptions of the overall authenticity of the assessment. This scale implicitly assumes that the authenticity of an assessment one-dimensional (this assessment [as a whole] is based on professional practice) instead of a multidimensional as is presumed by the 5DF. An additional scale was developed to measure the kind of learning that was stimulated by this assessment. This scale was developed to examine if it was indeed perceived that the assessment assessed the capability to apply knowledge and skills to real-life situations. Table 1 shows the scales of the student questionnaire, accompanied by an example of an item and the number of items per scale. The teacher questionnaire was almost identical, except that the word ‘I’ was replaced with the word ‘the student’.

*** INSERT TABLE 1 ***

All items were contextualized in a specific assessment (here: “determining care needs”) and referred to students’ future professional practice as a social worker (an example of an item: “This assessment task resembled the tasks of a real social worker”). Thomas and Bain (1984) showed that examining study approaches for or perceptions of a specific assessment requires a contextualized questionnaire otherwise respondents report their preferred learning approach or their perceptions of assessments in general.

After the items were constructed, four teachers of different VET schools were asked to review the items. They were asked if the items were readable, understandable and clear for students at this educational level. Where necessary, the questionnaire was adapted in line with their suggestions.

Procedure

During one week, all students took part in the assessment. Every student completed the perception questionnaire after finishing the assessment. Teachers filled in the questionnaire at the end of the assessment week.

Analysis

Because the questionnaire was developed based on theoretical insights, confirmatory reliability analyses were done to find out if the scales of the perception questionnaire reliably measure the authenticity facets that were intended. Cronbach Alpha was calculated for all scales, for students and for teachers separately, with $\alpha = .6$ as lower limit. The decision to examine reliabilities for teachers and students separately was guided by the possibility that the reliabilities differ between these groups since the groups represent two populations that differ on various aspects (e.g., age and amount of practical experience and schooling [see Honebein et al, 1993]) that can influence their way of answering questions about authentic assessment.

In addition, an exploratory factor analysis (Principal Component Analysis with Varimax rotation) was conducted on the 115 student questionnaires. The factor analysis was used to:

1. Explore the assumption of multi-dimensionality of the authenticity construct.
2. Explore how students structured the construct of assessment authenticity, and
3. Explore if this factor structure corroborated the authenticity facets of the perception questionnaire and the 5DF.

For further validation, three reviewers were asked to interpret the factors. These reviewers were researchers who were not involved in this study and were both unfamiliar with the scales of the perception questionnaire and the 5DF. The factors and their interpretations were compared with the theoretical ideas of the 5DF.

Results

Table 2 shows the results of the reliability analysis on the scales of the perception questionnaire for students and for teachers.

*** INSERT TABLE 2 ***

This table presents some interesting findings. First, all reliabilities were higher in the teacher group than in the student group. Second, all scales (except for the task ownership subscale) were reliable in the teachers group, while 5 out of the 11 scales did not exceed .6 for the student group. Third, at the dimensional level, the task, the physical context and the result/form dimensions were reliable in both groups, while the criterion dimension showed marginal reliability and the social context showed very low internal consistency in the student group. Fourth, at the sub-element level, only the criterion transparency subscale was reliable in the student group.

Two explanations could be given for the unreliabilities in the student group: (a) There was no straightforward fit between the scales and the underlying constructs in this questionnaire, resulting in a different clustering of the items than in the pre-defined scales; or (b) the items were too difficult for VET students to understand. Both these options were examined.

First, an explorative factor analysis was done to examine if the underlying factor structure in the questionnaire could explain the unreliabilities. An initial factor analysis resulted in 15 factors possessing eigenvalues of 1.0 or more. However, these factors were impossible to interpret and the scree-plot suggested a six-factor solution as a more appropriate structuring of the student perceptions (Cattell, 1966; Lizzio & Wilson, 2004). When a reliability analysis was conducted per factor, the first six factors turned out to have a Cronbach's alpha of more than .6. The comprehensibility argument (Dunteman, 1989) corroborated selecting these six factors, because they were readily interpretable in the eyes of the three reviewers. Then, a new factor analysis was conducted on the remaining items that primarily loaded on these six factors. Table 3 shows the results of the final factor analysis. These six factors accounted for 63 % of the variance.

*** INSERT TABLE 3 ***

A closer look, from a more qualitative point of view, at the distribution of the items over the remaining factors and the items that fell out of the final factor analysis showed an interesting pattern. First, almost all items of the task, physical context, result/form and criteria dimension fell in the final factors. Second, none of the social context items loaded on the final factors. Third, the task items clustered in Factor 3 and the physical context items clustered in Factor 2. On the other hand, the result/form items and the criterion items did not cluster in the expected way. The original result/form scale contained three result items and three form items. The factor analysis showed that these items did not belong together, since the form

items clustered in Factor 1 while the result items clustered with the criterion items in the Factors 5 and 6. Fourth, the criterion transparency scale was the only sub-element that was represented in the final factors (Factor 4). Fifth, three items of the overall authenticity scale loaded on the final factors, two of which clustered with the task items on Factor 3 and the other one clustered with the form items on Factor 1.

Three reviewers (1,2,3) named the final six factors as follows:

1. Connection of assessment form with the profession (1, 2), assessment method (3).
2. Professional context (1, 3), perception of fidelity (2).
3. Content authenticity (1, 2, 3).
4. Clear expectations (1, 2, 3).
5. Job-relevant criteria (1, 3), job-related judgement (2).
6. Relevance of the output for the profession (2), job-related judgement (1, 3).

The internal consistency of the factors was calculated, for both the student and teacher group, to examine if the factors represented reliable constructs. If the factors are internally consistent, interpreting these factors is more valid.

*** INSERT TABLE 4 ***

The results in Table 4 showed that the reliability of the factors was much better than the original scales in the student group, while the reliabilities remained high in the teacher group. These findings might mean that these final six factors more adequately described the facets that determine assessment authenticity in the eyes of the students.

The second possible explanation for the unreliability of a number of questionnaire scales in the student group could be that the items in those scales were accurate for assessing the intended variables, but too difficult for students to fully understand. This could be a reasonable explanation, since the scales were reliable in the teacher group. To assess the reading difficulty of the questionnaire, Flesch-Kincaid Grade Level scores were calculated per

scale of the questionnaire. These scores are based on technical aspects of the reading material (sentence length and word length), without looking at the content or meaning of the words. Still et al (2005) showed that these scores can be used to test surveys at the item level, thereby giving valuable input for questionnaire development and improvement. Based on the Flesch Kincaid scores, the *minimal suggested reading age* (MSRA) per scale could be calculated by adding a value of 5 to the Flesch-Kincaid Grade level score resulting in the formula:

$$\text{MSRA} = 0.39L + 11.8N - 15.59 + 5$$

where L stands for average number of words per sentence and N for average number of syllables per word. It turned out that the MSRA for the unreliable scales was 17.58, while the MSRA for the reliable scales was 14.85. The mean age of the student participants was 18.1.

Two extenuating circumstances need to be addressed at this point. First, Klare (1963) showed that people prefer to read below their MSRA and for a pupil to properly comprehend what (s)he is reading, the reading level should at least be two years below the MSRA. Our findings are in agreement with this. Second, age alone does not give enough information, since not all students of the same age have the same intellectual capacities and as a result different reading levels can be expected from them. For example, even though VET students and pre-university students are of the same age, it is likely that their reading levels differ. VET is a form of vocational education does not allow entry to further higher academic education (i.e., university) and is primarily populated by students who are work-oriented and/or not capable of successfully following an academic, pre-university, curriculum. Pre-university education is theoretical type of education and prepares pupils for university (Eurydice, 2004). To make a statement about whether or not the questionnaire scales were too difficult for VET students to understand, insight in the reading level that is normally expected from VET students, compared to pre-university students, is needed. A thorough search of literature and institutions in the Netherlands (e.g., Educational Council, Inspectorate of

Education, Ministry of Education, Culture and Science, etc.), however, did not turn up data on differences between reading levels of VET students and pre-university students. To compensate for this, we gathered a representative sample of VET study material, pre-university study material, and university level course materials. For each category of material, four samples were taken and then the MSRA was calculated (see Table 5). This resulted in an absolute and relative norm for the suggested reading age for VET students, which was compared to the mean suggested reading ages of the questionnaire scales.

*** INSERT TABLE 5 ***

Table 5 shows that the mean suggested reading age for the sample of VET study material is 15.41 and for pre-university material 18.1; university level course materials had an MSRA of 21.49. In other words, the reliable scales required a reading age that is normal for VET students, while the unreliable scales had an MSRA at a pre-university level.

Conclusion and discussion

The main goal of this study was to determine the facets of assessment authenticity from a practical angle by exploring how students and teachers perceive the authenticity of an assessment. More specifically, we examined if the theoretical facets of authenticity, as described in the 5DF were recognised and corroborated in practice. The main conclusion could be that authentic assessment is indeed perceived as a multidimensional construct, but some reconceptualisations are in order. Moreover, teachers and students do seem to differ in what they perceive as determining facets of authentic assessment. More specifically, teachers distinguish all the dimensions as well as the sub-elements as described in the theoretical 5DF, while students only differentiate four of the five dimensions and do not differentiate at the

sub-element level. These findings have implications for the 5DF, for future use of the questionnaire, for practice and for future research.

Students versus Teachers

Teachers recognised both the dimensions and the sub-elements as facets that determine assessment authenticity. In the eyes of the students, four of the five dimension (Task, Physical Context, Form and Result/Criteria) determine authenticity, while students do not perceive the Social Context as a characteristic of assessment authenticity, neither do they distinguish the several sub-elements. Two possibilities were examined to explain these findings. First, students perceive authenticity differently than the 5DF (and thus, the questionnaire) proposes which is reflected in new factor structure. Second, the questionnaire scales were too difficult for students to understand.

The factor structure found in the student group suggests that students have a much less elaborate perception of assessment authenticity than the 5DF proposes, while teachers seem to support the more elaborate conceptualisation of authenticity. This might be explained by the fact that students are “consumers” of the assessment, while teachers are the developers and in addition have much more experience with assessment practices and development. As a result, teachers are likely to have given assessments and the ideas behind assessments much more thought than students have. Teachers also have much more educational and practical experience, which might have changed their perception of assessment authenticity compared to students, who might not even be aware of the existence of some characteristics (Honebein et al, 1993). Because of this increased degree of experience, teachers are likely to have more developed schemata for thinking about assessments (Sternberg, 1999). These results support the idea of Honebein et al that having more practical and/or educational experience changes how one thinks of assessment authenticity.

The results of the readability analyses showed that the scales that were above the normally expected reading level of students, were identical to the unreliable scales. This might have caused the differences between the recognition of elements of authenticity of students on the one hand, and teachers and the 5DF on the other hand. The readability data also showed that teachers appear to have difficulties placing themselves in the position of students, as can be seen by the fact that a number of VET teachers rated the questionnaire scales as clear and understandable for VET-students prior to the administration of the questionnaire.

The differences between students and teachers suggests that we, as educators and instructional designers, should not automatically assume that students see an assessment as teachers see it. This stresses the relevance of investigating student perceptions of assessment characteristics as these are argued to be the motor behind student learning (e.g., Boud, 1995; Scouller, 1997).

Implications for the 5DF

The results of the factor analysis suggest that students structure authenticity partially different from what the 5DF proposes. Even though drawing firm conclusions would be inappropriate, because the number of participants (115) was relatively small for a factor analysis and it was only conducted within one student group, we think that this study does point out several interesting indications concerning the dimensions of authentic assessment most of which are corroborated by other research.

First of all, it supports the theoretical idea behind the 5DF that authenticity is multidimensional (Gulikers et al, 2004), meaning that the authenticity of an assessment depends on several assessment characteristics instead of on an overall resemblance between the assessment and professional practice.

Second, the Task (Factor 3) and the Physical Context (Factor 2) were perceived as two dimensions of authenticity. This distinction has also been made in previous theoretical research (e.g., Cumming & Maxwell, 1999) and empirical research already pointed at the individual impact of both these elements on student learning and motivation (Gulikers et al, 2005).

Third, the result/form and the criterion dimensions showed a deviating pattern. Result and form were not perceived of as one dimension. The clustering of items and the interpretation of the reviewers argued for a separate Form dimension (Factor 1). The Result items clustered with the criterion items in Factors 5 and 6. When looking at these factors semantically and at the interpretations of the reviewers, the following possible interpretation could be given: Both factors referred to job-relevant judgement, but Factor 5 refers to judgement in professional practice in more general terms, while Factor 6 refers to judgement in the social work profession in specific. A reasonable interpretation could be that these two factors can be combined in a Result/Criterion dimension that focuses on judging students on job-relevant aspects. As a check, the internal consistency was calculated over the six items of Factor 5 and 6 combined, which was $\alpha = .68$ (students) and $\alpha = .77$ (teachers). This makes the interpretation of combining factor 5 and 6 into one Result/Criterion dimension more plausible.

Fourth, students did not recognize the social context as a dimension of authenticity. They apparently do not (yet) realize that much work involves social activities like collaborating and working in teams (Boshuizen et al, 2004; Boud, 1998; Kochan et al, 1999). The traditional school situation tends to isolate academic (individual, on-task) skills from social (group, off-task) skills, which feeds the belief that schoolwork is individual. This is corroborated by a previous qualitative study that examined teacher and student perceptions of the five dimensions of the 5DF (Gulikers et al, 2004). This study showed that students as well

as teachers perceived the social context as the least important dimension of authentic assessment. The reason for this was *not* that they felt that the social context was unimportant, but that they had the strong belief, based on many assessment experiences (Boud, 1995; Samuelowicz & Bain, 2002), that assessment is an individual affair. This leaves us with three possibilities: (a) the social context is not perceived by students as one of the dimensions of authenticity and should therefore be deleted from the 5DF; (b) the social context might still be a facet of authenticity, but the items of the questionnaire were too difficult for students to result in an internally consistent scale. This explanation is supported by the high MSRA for the unreliable scales; or (c) the strong belief that assessment by definition is individual should be changed first before students will ever be able to perceive the social context as a facet of authentic assessment. Because teachers in this study recognised the social context, the literature study that led to the 5DF and other authenticity literature (Herrington & Herrington, 1998; Newmann and Associates, 1996) supports the social context as a dimension of authenticity, and analyses of successful work performances argues for the importance of social processes in professional practice (Kochan et al, 1999), we argue that the social context should be kept into the framework for assessment authenticity. However, the Social Context, or the questionnaire used to examine perceptions of the social context, should be reconsidered in future research

Fifth, as said before, students perceived authenticity at a dimensional level, while teachers recognised several sub-elements within these dimensions. This suggests that the 5DF appropriately describes several assessment characteristics that are important for assessment authenticity. However, we should be aware that to influence student perceptions of authenticity, the five-dimensional framework should be used at the dimensional level.

In short, examining the theoretically hypothesised dimensions of authenticity from an practical angle suggest a new conceptualisation of the dimensions of assessment authenticity

that involves a Form (Factor 1), a Physical context (Factor 2), a Task (Factor 3) and a Result/ Criterion dimension (Factor 5 and 6). These factors are corroborated by both students and teachers as seen in the reliability scales in Table 4 and are therefore viewed as crucial elements of assessment authenticity. Contrary to the 5DF, student perceptions seem not to support the social context as a dimension of authentic assessment. In addition, students seem to think about authenticity on a dimensional level, while teachers also recognise the sub-elements as described in the 5DF.

Practical Implications

Two problems need to be addressed with respect to the development and use of authentic assessments in educational practice (Cummings & Maxwell, 1999; Cooper, 1994; Roelofs & Terwell, 1999). First, teachers develop authentic assessments without thinking through and explicating what this authenticity means and how it is operationalised in the assessment, and second, assessments that teachers developed to be authentic are not automatically perceived as being authentic by students.

The results of this study have two practical implications that help dealing with these problems. First, the 5DF is a helpful tool for teachers or educational developers to talk and think about authentic assessments, to make implicit beliefs about authentic assessment explicit, and to develop various kinds of authentic assessment. Second, if we want to influence student learning with authentic assessments, we have to change one or more of the following assessment characteristics: the task, physical context, the assessment form, or the result and criteria of the assessment. These are the assessment characteristics that influence student perception of assessment authenticity, as these four characteristics can resemble professional practice to varying degrees. Thus, instead of ‘dressing up’ traditional assessments with some ‘real world’ elements at a superficial level (Cummings & Maxwell,

1999) making assessments authentic in the eye of the student requires changing more fundamental aspects (i.e., task, physical context, form, or result/criteria) of an assessment.

Next to students and teachers, professional practice is an important party in authentic assessment practices, especially in vocational education where students are relatively close to working. Wiggins (1993) argues that the validity of authentic assessments depends in part upon whether the test simulates "tests" of ability used in professional practice. To increase the predictive validity of authentic assessments for future work performance (i.e., the degree to which assessment performance predicts future work performance), assessment practices should involve professional practice in the development of authentic assessments and find out what practitioners perceive as authentic assessments. In the Netherlands, practitioners are involved in the development of performance standards for authentic assessments of work performance (Tillema et al, 2000) and a review of the LSRC (2004) in the UK showed that practitioners have clear ideas of what kind of assessment features are important for assessing work performance or competence. Also the high-stakes accountability systems in the US could benefit from involving practice in setting and assuring high-quality standards. This would be valuable for both improving student achievement as well as for making more valid inferences about success in school, college and workplace (Linn, 2000; Linn et al, 2002). The 5DF allows for examining practitioners perceptions of authenticity and comparing them to student and teacher perceptions, which gives concrete guidelines for adapting assessment practices for both school as well as workplace assessments or assessment on-the-job (Gulikers, 2005).

Future Research

To increase the generalisability of the findings of this study and strengthen the practical implications, future research should evaluate if the results found in this study also hold in other student groups in vocational education. It should be examined if the factor

structure found in this study is corroborated in other groups. A new perception questionnaire for evaluating assessment authenticity should be developed and tested. This questionnaire should focus on the dimensions of authenticity reflected in the found scales in this study, as these seem to determine student perceptions of authenticity, and a new social context scale that meets the reading level of the students should be developed.

Future research should examine how student perception of the authenticity of the assessment dimensions can be increased. What makes an assessment look more like professional practice in the eye of the student? This probably depends on what students think professional practice looks like (Lizzio & Wilson, 2004). Furthermore, it should be examined what the influence is of increased student perception of authenticity on their learning. Are students stimulated to deeper learning or development of professional skills or competencies when they perceive an assessment as being more authentic? Additionally, it should be examined if different kinds of authentic assessment in vocational education have different effects on student learning and development of professional skills? Previous studies (Gulikers et al, 2004; Gulikers et al, 2005) suggested that the various dimensions of authenticity might be of differing importance. For example, the assessment task seemed to be more important than the physical context. This raises questions such as: is increasing the authenticity of the task enough to stimulate students to deeper learning or should more than one dimension resemble professional practice?

In addition, this study showed that teachers and student differ in how they perceive authenticity. A previous explorative study (Gulikers et al, 2004) corroborated that students and teachers differ in how they perceive authenticity, but additionally, this study indicated that student groups that differ in the kind and amount of working and assessment experience can differ in their perception of authenticity as well. Thus, previous experiences with professional practice as well as with using or developing authentic assessment, might

influence what someone sees as determining elements of authenticity. Future research should differentiate between VET student groups (e.g. students with little versus a lot of experience in professional practice), examine if these different groups perceive authentic assessments differently, and study what kind of operationalisation of the dimensions are most effective for the learning of these different student populations.

Besides these content related directions for future research, this study might also give suggestions for future research methodology. Future research that uses questionnaires, at least with students at a educational level that is comparable to VET, should consider taking student reading levels into account in the development of new questionnaires or translation of existing questionnaires. Furthermore, complementing quantitative data with qualitative data might give more insight into student, and also teacher, perceptions of assessment authenticity.

In conclusion, this study showed that exploring assessment authenticity from a practical viewpoint, by examining the perceptions of the users, has additional value over only a theoretical examination of assessment authenticity. This corroborates the idea that authenticity is, at least partly, subjective.

Authors' Biographies

References

- Achieve, Inc. (2006) Closing the expectations gap. An annual progress report on the alignment of high school policies with the demands of college and work. Retrieved July 12, 2006, from the Achieve, Inc. Web site: <http://www.achieve.org>
- American College Testing (2006, May) *Ready for college and ready for work: same or different?* Available online at: <http://www.act.org/path/policy/pdf/ReadinessBrief.pdf> (accessed 24 July 2006).
- Biggs, J. (1996) Enhancing teaching through constructive alignment, *Higher Education*, 32, 347-364.
- Biemans, H., Nieuwenhuis, L., Poell, R., Mulder, M. & Wesselink, R. (2004) Competence-based VET in the Netherlands: background and pitfalls, *Journal of Vocational Education and Training*, 56(4), 523-538.
- Birenbaum, M. & Dochy, F. J. R. C. (1996) *Alternatives in assessment of achievements, learning processes and prior knowledge* (Boston, MA, Kluwer Academic Publishers).
- Bloom, B. S. (1956) *Taxonomy of educational objectives. The classification of educational goals* (London, Longman Group Limited).
- Boshuizen, H. P. A., Bromme, R., & Gruber, H. (2004) *Professional learning: Gaps and transitions on the way from novice to expert*. (Dordrecht, Kluwer Academic Press).
- Boud, D. (1990) Assessment and the promotion of academic values, *Studies in Higher Education*, 15(1), 101-111.
- Boud, D. (1995) Assessment and learning: contradictory or complementary?, in: P. Knight (Ed.) *Assessment for learning in Higher Education* (Londen, Kogan Page).
- Brown, J.S., Collins, A. & Duguid, P. (1989) Situated cognition and the culture of learning, *Educational Researcher*, 18(1), 32-42.

- Calderón, J. L., Hays, R. D., Lui, H. & Morales, L. S. (2005) Variation in readability within surveys. Available online at:
<http://www.chime.ucla.edu/measurement/presentations/Variation%20in%20Readability%20Within%20Surveys.pdf> (accessed 10 February 2005)
- Cattell, R.B. (1966) The scree test for number of factors, *Multivariate Behavioral Research*, 1, 245-276.
- Collins, A., Brown, J.S. & Newman, S.E. (1989) Cognitive apprenticeship: Teaching the craft of reading, writing and mathematics, in: L.B. Resnick (Ed.) *Knowing, learning and instruction: Essays in honor of Robert Glaser* (Hillsdale, NJ, Erlbaum).
- Cooper, B. (1994) Authentic testing in mathematics? The boundary between everyday and mathematical knowledge in national curriculum testing in English school,. *Assessment in Education: Principles, Policy & Practice*, 1(2), 143-166.
- Cronin, J. F. (1993) Four misconceptions about authentic learning, *Educational Leadership*, 50(7), 78-80.
- Cummings, J. J. & Maxwell, G. S. (1999) Contextualising authentic assessment, *Assessment in Education: Principles, Policy & Practice*, 6(2), 177-194.
- Dochy, F. (2001) A new assessment era: different needs, new challenges, *Research Dialogue in Learning and Instruction*, 10(1), 11-20.
- Dunteman, G. H. (1989) *Principal Component Analysis* (Thousand Paks, CA, Sage Publications).
- Entwistle, N.J. (1991). Approaches to learning and perceptions of the learning environment. Introduction to the special issue, *Higher Education*, 22, 201-204.
- Entwistle, N. J. & Ramsden, P. (1983) *Understanding Student Learning* (London, Croom Helm).

- Eurydice (2004) *The educational system in the Netherlands 2004*. (Dutch Eurydice Unit, Netherlands Ministry of Education, Culture and Science, The Hague).
- Gielen, S., Dochy, F. & Dierick, S. (2003) Evaluating the consequential validity of new modes of assessment: The influence of assessment on learning, including the 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.(2005) Ontwerpen van authentieke toetsen met de beroepspraktijk als uitgangspunt [Designing authentic assessment with professional practice as starting point]. *Handboek Effectief Opleiden*, 39, 33-53.
- Gulikers, J., Bastiaens, Th. & Martens, R. (2005) The surplus value of an authentic learning environment, *Computers in Human Behavior*, 21(3), 509-521.
- Gulikers, J., Bastiaens, Th. & Kirschner, P. (2004) A five-dimensional framework for authentic assessment, *Educational Technology Research and Development*, 52(3), 67-85.
- Herrington, J. & Herrington, A. (1998) Authentic assessment and multimedia: How university students respond to a model of authentic assessment, *Higher Educational Research and Development*, 17(3), 305-322.
- 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* (Berlin, Springer-Verslag).
- Huang, H. M. (2002) Towards constructivism for adult learners in online learning environments, *British Journal of Educational Technology*, 33, 27-37.
- Kasworm, C. E., & Marienau, C. A. (1997) Principles for assessment of adult learning. *New Directions of Adult and Continuing Education*, 75, 5-16.

- Kerka, S. (1995) *Authentic assessment in vocational education*. Available online at:
<http://www.calpro-online.org/ERIC> (accessed 12 October 2005).
- Klare, G. R. (1963) *The Measurement of Readability* (Iowa, Iowa State University Press).
- Learning and Skills Research Centre (2004) Do summative assessments and testing have a positive or negative effect on post-16 learners' motivation for learning in the learning and skills sector? A review of the research literature on assessment in post-compulsory education in the UK (London, Learning and Skills Research Centre).
- Linn, R. L. (2000) Assessment and accountability. *Educational Researchers*, 29(2), 4-16.
- Linn, R. L., Baker, E. L., & Betebenner, D. W. (2002) Accountability systems: Implications of requirements of the No Child Left Behind Act of 2001. *Educational Researcher*, 31, 3-16.
- Lizzio, A. & Wilson, K. (2004) First-year students' perceptions of capability. *Studies in Higher Education*, 29(1), 109-128.
- Maclellan, E. (2001) Assessment for learning: the differing perceptions of tutors and students. *Assessment and Evaluation in Higher Education*, 26(4), 307-318.
- 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.
- Ministry of Education, Culture and Science of the Netherlands (2005, May.) *Leren en werken: Plan van aanpak 2005-2007 [Learning and working: Action plan for 2005-2007]*. Available online at: www.minocw.nl (accessed 14 March 2006).

- Mueller, J. (2006) The authentic assessment toolbox. Available online at: <http://jonathan.mueller.faculty.noctrl.edu/toolbox/standardtypes.htm> (accessed 15 December 2002).
- Newmann, F. M., & Associates (1996) *Authentic achievement: Restructuring schools for intellectual quality* (San Francisco, Ca: Jossey-Bass)
- Newmann, F. M., Marks, H. M., & Gamoran, A. (1996) Authentic pedagogy and student performance. *American Journal of Education*, 104, 280-312.
- Newmann, F. M. & Wehlage, G. G. (1993) Five standards for authentic instruction, *Educational Leadership*, 50(7), 8-12.
- Ngar-Fun, L. (2005). *Hong Kong Academics' and Students' Perceptions of Assessment Purposes and Practices*. Hong Kong: The Hong Kong Institute of Education. Available online at: <http://www.ied.edu.hk/loap/index.html> (accessed 16 June 2005).
- Scouller, K. M. (1997). Students' perceptions of three assessment methods: Assignment essay, multiple choice question examination, short answer examination, *Research and Development in Higher Education*, 20, 646-653.
- Scouller, K. M. & Prosser, M. (1994) Students' experiences in studying for multiple choice question examinations, *Higher Education*, 19, 267-279.
- Segers, M., Dochy, F. & Cascallar, E. (2003) The era of assessment engineering: Changing perspectives on teaching and learning and the role of new model of assessment, in: M. Segers, F. Dochy & Cascallar E. (Eds.) *Optimising new modes of assessment: In search of qualities and standards* (Dordrecht, Kluwer Academic Publishers).
- 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* (Dordrecht, Kluwer Academic Publishers).

- Thomas, P. R. & Bain, J. D. (1984) Contextual dependence of learning approaches: The effects of assessments, *Human Learning*, 3, 227-240.
- Tillema, H. H., Kessels, J. W. M., & Meijers, F. (2000) Competencies as building blocks for integrating assessment with instruction in vocational education: a case from the Netherlands. *Assessment and Evaluation in Higher Education*, 25, 265-278.
- Van Rossum, E. J. & Schenk, S. M. (1984) The relationship between learning conceptions, study strategies and learning outcome, *British Journal of Educational Psychology*, 54(1), 73-83.
- 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).
- Wilson, K. L., Lizzio, A., & Ramsden, P. (1997) The development, validation and application of the course experience questionnaire. *Studies in Higher Education*, 22, 33-53.
- Wonacott, M. E. (2000) Vocational Education Research Trends, *Trends and Issues Alert*, 15. Available online at: <http://www.calpro-online.org/ERIC> (accessed 3 April 2005).

Figure Caption

Figure 1. The five-dimensional framework for assessment authenticity.

Table Caption

Table 1. The scales of the perception questionnaire.

Table 2. The reliabilities of the scales of the perception questionnaire.

Table 3. Factor loading of the final items.

Table 4. Reliabilities of the final factors from the perception questionnaire.

Table 5. Mean minimal suggested reading age (MSRA) for three educational levels.

Table 1. The scales of the perception questionnaire

Main scale	Subscale	Number of items	Example
Task		5	The task of this assessment is an important aspect of a social workers' job
	Task complexity	4	The task of this assessment was more complicated than the tasks I have to perform in my work placement
	Task ownership	5	The responsibility that I got in solving this assessment task is different from the responsibility I get in my work placement
Physical context		5	The context in which I had to perform this assessment resembles the professional practice of a social worker
	Resources availability	5	In this assessment I could use of all the resources/ equipment that are used in professional practice
Social context		4	If I had to perform this task in my work placement, I would have cooperated more with my colleagues
Result/form		6 (3 / 3)	The result that I had to produce in this assessment was something that a social worker in professional practice has to produce also
Criteria		4	The criteria that were used in this assessment are different than the criteria that are used in professional practice
	Transparent criteria	4	It was hard to find out what was expected of me in this assessment
Overall		5	This assessment was oriented to my future profession of social worker
authenticity Kind of learning		5	In this assessment I had to apply the thing I'd learned in a professional practice situation

Table 2. The reliabilities of the scales of the perception questionnaire.

	Students (<i>n</i> = 115)	Teachers (<i>n</i> = 18)
Task	.62	.86
Task complexity	.49	.81
Task ownership	.41	.59
Physical context	.80	.89

Resource availability	.51	.83
Social context	.35	.84
Result/form	.72	.71
Criteria	.55	.74
Criteria transparency	.73	.72
Overall authenticity	.70	.96
Kind of learning	.61	.86

Table 3. Factor loading of the final items.

Item	1	2	Factor 3	4	5	6
<i>Scale: Overall authenticity</i>						
This assessment was oriented to my future profession of social work			.64			
This assessment was clearly oriented to professional requirements	.58					
This assessment prepared me for my future profession			.56			
<i>Scale: Task</i>						
The task of this assessment resembled the tasks of a real social worker			.77			
The task in this assessment was an important part of the social work profession			.60			
The task of this assessment differed from the tasks of a real social worker			.60			
<i>Scale: Physical context</i>						
The context in which I had to perform the assessment was fake		.79				
The context in which I had to perform the assessment looked like the professional practice of a social worker		.66				
The context in which I had to perform this assessment looked just like the real world (seemed real?)		.85				
The context in which I had to perform this assessment was realistic		.79				

Item	Factor					
	1	2	3	4	5	6
<i>Scale: Result/form</i>						
The result (output) that I had to produce in this assessment is part of the job of a social worker			.57			
That on which I was being evaluated in this assessment is different from what I am evaluated on in practice					.85	
The result that I had to produce in this assessment is something that a real social worker also has to produce in practice						.68
This way of assessing (authentic assessment) is an effective way of assessing professional skills	.62					
This way of assessing (authentic assessment) fits well with the social work profession	.76					
<i>Scale: Criteria</i>						
The criteria resembled the criteria that I have to meet in practice					.70	
The criteria that I had to meet in this assessment resembled the criteria that I have to meet in practice					.72	
In this assessment I was evaluated on criteria that are important for the profession of social worker						.72
In this assessment I was evaluated on things that I never have to use in real professional practice of a social worker						.55
<i>Scale: Criterion transparency</i>						
The criteria that I had to meet in this assessment were clear enough					.61	
Before I started with the assessment it was clear to me what was expected of me					.65	
It was hard to find out what was expected on me in this assessment					.82	
<i>Scale: Kind of learning</i>						
In this assessment, both on knowledge and professional skills were important	.75					

Table 4. Reliabilities of the final factors from the perception questionnaire.

Factors	Students	Teachers
	(<i>n</i> = 115)	(<i>n</i> = 18)
Factor 1 (Form)	.76	.81
Factor 2 (Physical context)	.83	.91
Factor 3 (Task)	.79	.90
Factor 4 (Criterion transparency)	.76	.92
Factor 5 (Result/Criteria)	.68	.76
Factor 6 (Result/Criteria)	.69	.82

Table 5. Mean minimal suggested reading age (MSRA) for three educational levels.

	Mean MSRA
VET study material	15.41
VWO study material	18.10
Adult education study material	21.49

Note. The means of the three kinds of study materials are based on four examples of study material from different disciplines each.

VET = Vocational Education and Training; VWO = Dutch abbreviation for pre-university education.

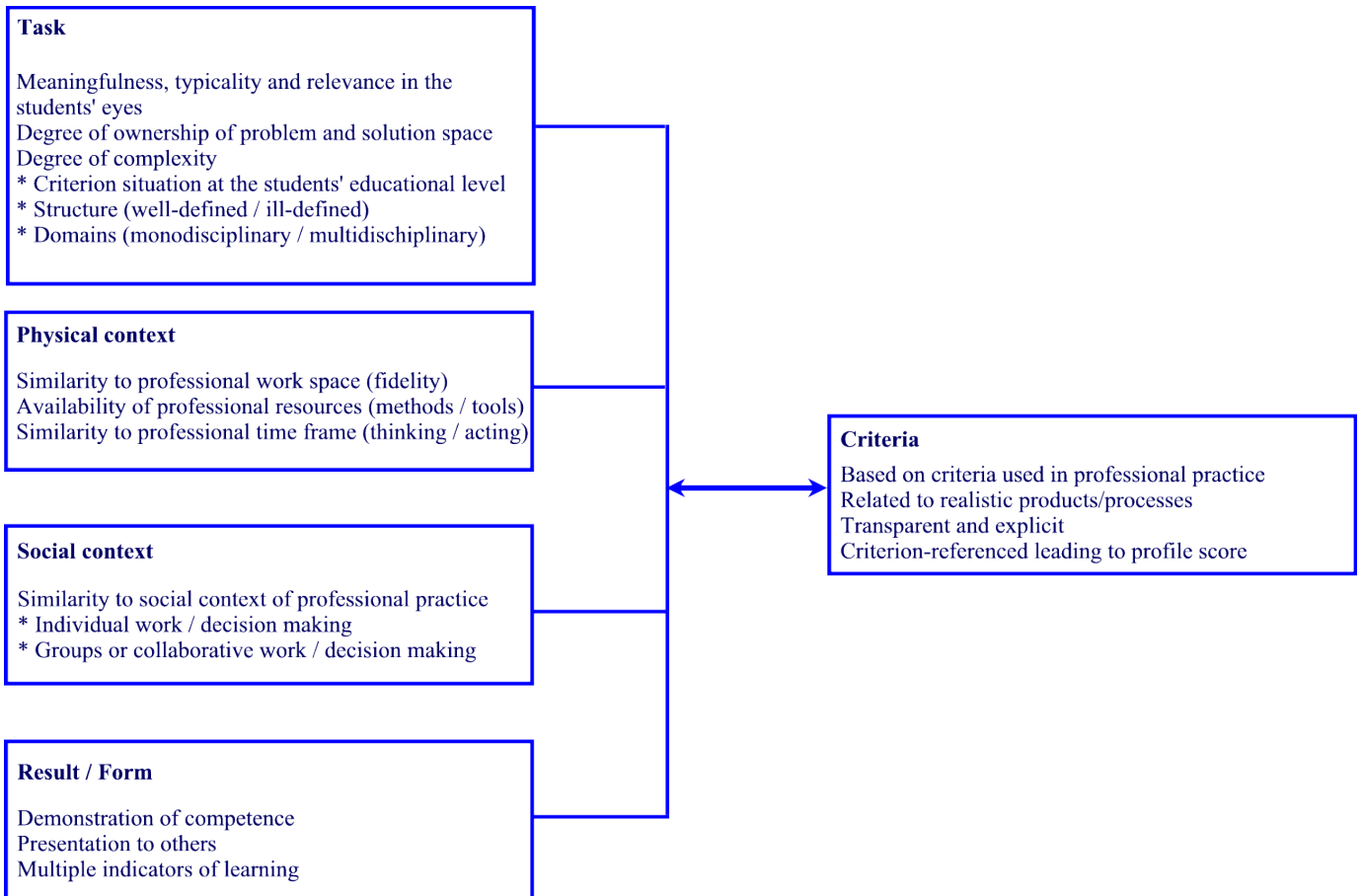


Figure 1.