Relations between student perceptions of assessment authenticity, study approaches and learning outcome

Open Universiteit
www.ou.nl

Running head: STUDENT PERCEPTIONS OF ASSESSMENT AUTHENTICITY

Relations between student perceptions of assessment authenticity, study approaches and

learning outcome

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

Open University of the Netherlands

Abstract

This paper examines the relationships between perceptions of authenticity and alignment on study approach and learning outcome. Senior students of a vocational training program performed an authentic assessment and filled in a questionnaire about the authenticity of various assessment characteristics and the alignment between the assessment and the instruction. Deep or surface study activities and the development of generic transferable skills were measured with a questionnaire as well. Correlational analysis and structural equation modeling were used to examine the hypothesis that more perception of authenticity and alignment resulted in more deep learning and development of generic skills. Results showed that when the <u>task</u>, <u>physical context</u> and assessment <u>form</u> are more authentic and when there is more alignment there is also evidence of more deep learning and/or an increase in generic skill development. Authenticity perceptions did not affect surface learning. Contrary to expectations, more authentic assessment <u>criteria</u> resulted in a decrease in deep learning and generic skill development. The explanation might be that authentic, but too concrete criteria, focusing on specific actions, might hamper motivation and learning at least for more experienced students.

Keywords: Authentic Assessment, Vocational Education, Student Perceptions, Student learning

The More the Better? Relations between Student Perceptions of Assessment Authenticity,

Study Approaches and Learning Outcome

Boud (1990, p. 101) stated that "there is often a gap between what we require of students in assessment tasks and what occurs in the world of work" and Gibbs (1992) argued that "the tail wags the dog" in that student learning is very much guided by the ways in which the learning is assessed. These two ideas show the background of this study that deals with making assessment look more like professional practice (i.e. authentic assessments) in order to stimulate students to learn and develop the knowledge, skills, and attitudes (i.e. competencies) they need for their future working lives.

An important goal of education, at least in vocational education, is to prepare students for a professional life. In the industrial era, working class people were educated for efficient functioning as skilled workers at the assembly line (Birenbaum, 2003). Schooling focused on acquiring factual knowledge and basic skills mainly through drill and practice. Current society, however, is dynamic and characterized by rapid developments in information and communication technologies and their effects on the size and sustainability of our knowledge base. Jobs have changed and different requirements are placed on graduates. Successful performance in this society demands not only a profound knowledge base and routine skills, but rather the ability to flexibly adapt knowledge and integrate it with skills and attitudes to solve new problems and handle unknown situations. To prepare students for the jobs that characterize modern society, students need to learn different "things" in a different way. As a reaction to this, the last 15 years have witnessed a lot of educational practices, at least in vocational education in the Netherlands (Tillema, Kessels, & Meijers, 2000). Schools changed their curricula and pedagogy towards more competency-based education. But changing teaching is not enough. According to the constructive alignment theory (Biggs, 1996), to change learning, both instruction *and* assessment practices need to change. Changing the assessments might be even more important as learning is so driven by assessment that the form and nature of assessment can swamp the effect of any other aspect of the curriculum (Boud, 1990).

Changing Assessments

As a reaction to societal developments that took place in the last two decades, the assessment paradigm has shifted from a testing culture to an assessment culture (Birenbaum, 1996). The testing culture is characterized by so-called objective, standardized test instruments that focus on measuring atomized bits of knowledge at the expense of more complex, higher-order knowledge and skills. The main function of testing is to rank and grade

students. These standardized tests have been broadly criticized (Birenbaum, 1996; Glaser & Silver, 1993; Hambleton & Murphy, 1992) as being disconnected from the real world and not suitable for assessing students' ability to be flexible in adapting and applying knowledge, skills and attitudes in context. In contrast, the assessment culture stresses the importance of competencies, understanding and application (Birenbaum, 1996, Perkins & Blythe, 1994). It is characterized by the integration of instruction and assessment and the function of assessment is not only to grade students but also to stimulate their learning and competency development. The assessment formats and instruments characteristic to the assessment culture are performance-based, integrated and contextualized methods (Birenbaum, 2003). The goal of these assessments is to increase the correspondence between what students need to do in school and what is expected from them after finishing their studies (Boud, 1995). In this light, authenticity became one of the crucial elements of new kinds of assessment that focus on professional competency development (Boud 1990; 1995; Dochy, 2001; Segers, Dochy, & Cascallar, 2003). By bringing assessment "in context" authentic assessments are thought to help bridge the gap between learning and working (Cummings & Maxwell, 2002) and are expected to stimulate students to develop competencies that are relevant for their future professional life. The study reported upon here examined what elements determine the authenticity of an assessment in the eyes of students and how this (perceived) authenticity influences student learning.

Assessment Authenticity

One important change in assessments, characteristic of the assessment culture, is that assessments are not decontextualised and atomistic, but more contextualized or authentic (Segers et al., 2003) focusing on the use of skills in context. Assessment practices shifted from mainly using standardized tests such as multiple-choice or short answer to the use for example of performance assessment or portfolio assessment. The purpose of these kinds of

assessment is usually to stimulate or evaluate students' capability to handle professional problem situations (Birenbaum, 1996). In these kinds of assessments, authenticity is important with respect to the assessment's (1) construct validity, and (2) consequential validity (Boud, 1990; Dierick & Dochy 2001; Gielen, Dochy, & Dierick, 2003; Messick, 1994). Construct validity means that the assessment measures what it is supposed to measure (Messick). In the case of competency-assessment this means that the assessment must be an appropriate reflection of the underlying construct (i.e. competency) that it wants to assess. To evaluate if students are capable of integrating knowledge, skills and attitudes into an effective performance in real life, more emphasis should be placed on letting students actually perform the task in a realistic situation, instead of asking them to write down what they think they would do in a hypothetical situation (Miller, 1990; Van Merriënboer, 1997). Therefore, Gulikers, Bastiaens, and Kirschner (2004) argue that when the purpose of the assessment is competency development or evaluation, the assessment should require students to demonstrate the same competencies as experts would use in the real-life situation. This is more likely to occur when there is a greater correspondence between the assessment situation and the professional practice situation on which the assessment is based. When the assessment task is more representative of real life situations and the assessment requires students to think and work as experts would, the assessment is more likely to actually measure the "things" (i.e., the competencies) that it is supposed to measure (Gielen, Dochy, & Dierick, 2003; Messick, 1994). In other words, authenticity is imperative for valid competency-assessment, since validity means that the assessment measures what it is supposed to measure (e.g., Messick).

Consequential validity means that an assessment has an impact on student learning (Boud, 1995; Dierick & Dochy, 2001; Messick, 1994). Realistic assessments that resemble what students will encounter in their future jobs are expected to stimulate and motivate

students to employ more deep learning and develop the competencies relevant for their future working lives (Boud, 1990; 1995; Gulikers et al., 2004). Several qualitative studies (Herrington & Herrington, 1998; McDowell, 1995; Sambell, McDowell, & Brown, 1997) asked students which characteristics of assessments they experience as positive for their learning. The authors concluded that students favored assessments that relate to authentic tasks, encourage them to apply knowledge in realistic contexts, show them relevance for their life outside school, and emphasize the use and development of skills that are needed in professional life. Thus, from a theoretical as well as from a student point of view, increasing the authenticity of an assessment is expected to have a positive effect on student learning and help students prepare for their working life.

Cumming and Maxwell (2002) showed that in many educational practices the importance of authenticity is recognized, but the operationalization of this authenticity is far from optimal. Making an assessment more authentic is mostly translated into making the assessment more 'realistic' (e.g., having a higher fidelity) without careful consideration of what elements make the assessment more realistic or authentic. This superficial approach has resulted in assessments that sometimes damaged students' learning (Cooper, 1994). Students could not appreciate the increased authenticity; instead they perceived the assessment to be more artificial, which only distracted them from an effective learning process. Despite good intentions of the developers, the assessments did not encourage students to adopt the kind of study approaches that were intended. This suggests that it is important to be careful when developing new modes of assessment, otherwise the results can be counter-productive for learning (Boud, 1990).

In order to carefully examine what makes an assessment authentic and how this influences student learning, this study builds on a literature review (Gulikers et al., 2004) that

unraveled the concept of authenticity. This resulted in a five-dimensional framework (5DF; Figure 1) that describes which assessment characteristics determine its authenticity.

***   INSERT FIGURE 1 ABOUT HERE   ***

In the 5DF, authenticity depends on the resemblance of five assessment facets (the task, physical context, social context, form, and result/criteria) and the professional practice situation on which the assessment is based. The five assessment facets can be described as follows:

1. Task. The assessment assignment that defines the content of the assessment

2. Physical context. The environment in which students have to perform the assessment task

3. Social context. The interaction (im)possibilities during the assessment

4. Form. The assessment method, independent of the content

5. Criteria. The characteristics of the performance (product/process) that are valued

The framework argues that authenticity is a multidimensional construct (i.e., the five facets) and that an assessment can be made more authentic in different ways by varying the resemblance of one or more of the facets of the 5DF and professional practice. This means that there is not an 'authentic - not-authentic' dichotomy, but rather an authenticity continuum. This study examines how the authenticity of the five assessment facets influence student learning.

Student Perceptions and the Impact on Learning

The previous section argues for the importance of making an assessment more authentic. However, making an assessment more authentic in the eyes of the developer is not enough, since the effect of assessment on student learning seems to be mediated by students' perceptions of the assessment requirements (Boud, 1995; Entwislte, 1991; Sambell et al., 1997; Scouller, 1997; Van Rossum & Schenk, 1984). These studies show that how students

perceive the assessment, rather than the actual assessment or teachers' intentions, affects to a large extent student learning. To be more specific, student perceptions of the assessment requirements influence their study approach (how they learn) and their learning outcomes (what they learn).

The 3P model (Biggs, 1989) addresses the relationships between perceptions of the learning environment, study approaches and learning outcomes. Biggs argues that the influence of students' perceptions can be very pervasive and that they can influence student learning in two ways. Perceptions of the learning environment can have a <u>direct</u> influence on learning outcomes, but the influence of perceptions of the learning outcome can also be <u>indirect</u> through study approach. Empirical results (Lizzio, Wilson, & Simons, 2002) supported both these relationships. They showed that positive perceptions of the learning environment had a direct positive effect on learning outcomes as well as an indirect effect on learning outcomes through stimulating a deep study approach. In addition, studies of Scouller (1997; 1998) and Sambell et al. (1997) showed that students adapted their study approach when they perceived assessments as having different requirements. With respect to perceptions of assessment authenticity, McDowell (1995) and Herrington and Herrington (1998) showed that students say that an assessment positively influences their learning when they perceive it as relevant or as having a connection to reality. These results show that students' perceptions are very important to consider when developing assessments. If increasing the authenticity of an assessment is thought to stimulate deep learning and help students develop professional competencies, then it is imperative that students perceive the assessment as authentic, which in turn should make students decide that a deep study approach would give the best learning outcomes.

Even though authentic assessment is expected to positively influence student learning, there has not been much (quantitative) research on the impact of <u>perceptions of authenticity</u>

on student learning. This study tries to get more insight into the actual influences of

perceptions of authenticity on study approach and learning outcomes. For this purpose, this

study builds on the 5DF described in the previous section. By splitting up the concept of

authenticity in the different facets described by the 5DF (Figure 1), it becomes possible to

gain a detailed picture of what influences students' perceptions of assessment authenticity and

how the perceptions of these different facets influence study approach and learning outcome.

An empirical study of Gulikers, Bastiaens, and Martens (2005) that manipulated two

dimensions (i.e., task and physical context) showed that the authenticity of a task and the

physical context have a differential impact on student learning. This supports the idea that it is

important to split up the concept of assessment authenticity into different facets and to

examine their individual impact on student learning.

Alignment between Instruction, Learning and Assessment

Biggs' constructive alignment theory (1996) suggests that assessments should be

considered as part of the learning environment. More specifically, to elicit a certain type of

learning, instruction and assessment should <u>both</u> be directed towards this kind of learning

(i.e., rote learning pedagogy should match rote learning assessment and competency learning

pedagogy should match competency learning assessment). Empirical research by Segers,

Dierick, and Dochy (2001) supported this. They showed that when students perceived a

mismatch between a new kind of assessment that focused on applying knowledge to realistic

problems and instruction that primarily valued memorization, a positive effect of the

assessment on students' study activities and learning outcomes failed to appear. Theoretical as

well as empirical evidence indicates that the effects of new modes of assessment should be

examined in the light of the entire learning environment (Struyven, 2005). To this end, this

study considers students' perception of alignment between the instruction and the assessment

next to examining the influence of students' perceptions of the authenticity of the five

assessment characteristics. It is expected that when students perceive a match between assessment and instruction, this will positively influence their study activities and learning outcomes, or at least will not be detrimental to them.

<u>Research Questions</u>

In this study, authentic assessments are defined as assessments that require students to demonstrate the same combinations of knowledge, skills and attitudes (i.e., competencies) that are applied in the professional practice situation on which the assessment is based (Gulikers et al., 2004). Authenticity is operationalized in five assessment characteristics, namely the resemblance of the task, physical context, social context, form, and criteria to the professional practice situation. The two research questions of this study are: (1) How do students' perceptions of the authenticity of an assessment influence study approach and learning outcome? More specifically, what are the direct and indirect influences of students' perceptions of the authenticity of five assessment characteristics on their study approach and their learning outcomes?, and (2) What is the impact of perception of alignment between assessment and instruction on study approach and learning outcome?

The quantitative, empirical study described here tries to determine whether the expected connections between perceptions of authenticity, study approach and learning outcomes do exist and if they do, how these connections work. Perception of authenticity is divided into perception of authenticity of the five assessment facets as defined by the 5DF (task, physical context, social context, form, and result/criteria) and it is hypothesized that these five perceptions affect study approach and learning outcomes individually. Furthermore, it is expected that perception of increased assessment authenticity stimulates students to use a deep approach to studying and develop generic, professional skills and that the employment of a surface study approach negatively influences the development these skills.

With respect to the second research question it is hypothesized that when students perceive more alignment between assessment and instruction, meaning that they experience that the instruction and the assessment are aimed at the same kind of learning, they will employ more deep learning and reach a better learning outcome.

Method

Participants

One hundred and eighteen senior students (mean age = 19.16, SD = 1.14) studying Social Work at a vocational education and training institute (VET) enrolled in this study. The students were final year students and had been studying Social Work in a competency-based learning environment combined with authentic assessments for three years. In other words, they were familiar with the kind of authentic assessment used in this study.

Materials

The assessment. This study made use of an existing assessment in a vocational education and training institute for social work, which was designed to be an authentic assessment. The topic of the assessment was "applying for a job". From a teachers' point of view, this was thought to be very authentic for senior students, since they would finish school within four months, leaving them at the mercy of real professional practice. The assessment consisted of two parts (a) writing a letter of application and a curriculum vitae for one of three social work related vacancies, and (b) taking part in a job interview based on the application letter. Both activities took place in school and the job interview was simulated in a role-play with a teacher playing the role of employer. One week before the assessment, students received a list of seven assessment criteria that focused on observable behavioral aspects. At the start of the assessment students received three descriptions of social work vacancies, one of which they could choose to be the task of their assessment. During the interview, students had to show that they could deal competently with the problem situation at hand. Students had

to perform the assessment individually and their performance was observed and scored by two independent assessors on the set of criteria.

The instructional phase. A competency-based instructional period of 9 weeks preceded the authentic assessment. This period focused on the students' role as a professional. During eight weeks, students worked in groups on critical professional problem situations, for example "rights and obligations of employees", "dealing with the selection committee", or "coaching of participants". They had to set learning goals focusing on knowledge as well as skills and attitudes. During this training phase of self-study and skills training, students had to perform several formative assessments. These were all role-play assignments based on a new, but related problem case. The summative assessment (in this case "applying for a job") was based on a selection of course objectives that was translated into the assessment criteria. Although the course objectives were available from the beginning of the course, the assessment criteria were revealed one week prior to the assessment in which students were freed from obligatory educational activities.

Perception questionnaire. A questionnaire based on the five-dimensional framework (Gulikers et al., 2004) was developed. Its scales examined whether and how the students perceived the authenticity of the task, the physical context, the social context, the form, and the criteria. The 24 items of the questionnaire all assessed the perception of the resemblance of these five assessment characteristics to (future) professional practice (e.g., "The task of this assessment prepared me for my future professional life"). The items were scored on a 5-point Likert scale ranging from 1 (totally disagree) to 5 (totally agree), resulting in a score for the perceived degree of resemblance between the assessment characteristics and professional practice. All scales, except for the social context scale, had a reasonable internal consistency, shown in Cronbach's alpha ranging from .69 to .83. Due to its low reliability ($\alpha = .35$) the social context scale was excluded from further analysis.

Perception of alignment. The perception of alignment was measured by a 5-item questionnaire, examining whether students perceived the instruction to convey the same message as the assessment with regard to what kind of learning is valued (e.g.,. "During the instructional phase I had to use my knowledge in the same way as during the assessment" or "Based on the instruction, I expected a different kind of assessment"). Cronbach's alpha for this scale was .73.

Study approach. Study approach was measured with the Revised-Study Process Questionnaire 2 Factors (R-SPQ-2F; Biggs, Kember, & Leung, 2002), a revision of the Study Process Questionnaire (Biggs, 1987). The R-SPQ-2F is a 20-item questionnaire that is more adapted to current society and modern ideas of education than the original. It was used to distinguish between two study approaches, namely a deep study approach (DSA) and a surface study approach (SSA). DSA is characterized by study activities that focus on understanding and constructing meaning of the content to be learned. SSA involves activities associated with memorization and reproduction of atomized bits of factual information (Biggs, 1987). Several studies indicated reliable coefficients for the two scales, the items were short, all positively stated and without difficult wording. These were important considerations, since the research population involved students at the VET level and not at the higher professional or academic education level, which was the research population involved in most previous research done to validate study approach questionnaires. Moreover, the questionnaire was successfully used in previous research (e.g., Scouller 1997) to examine relationships between study approaches and learning outcomes. The original questionnaire was translated into Dutch and contextualized to the authentic assessment that was the object of this study. This contextualization was needed to examine students' study approach for a particular assessment, instead of their default or preferred study approach (Entwistle, McCune, & Hounsell, 2002; Thomas & Bain, 1984). Results indicated that the two scales of

the translated version had a reasonable internal consistency in the VET context (Cronbach's alpha = .65 for SSA, and= .81 for DSA).

Qualitative learning outcome. The qualitative learning outcome was measured with a Dutch translation of the Generic Skill Development (GSD) scale of the Course Experience Questionnaire (CEQ) (Wilson, Lizzio, & Ramsden, 1997). This scale measured the extent to which students felt that a certain learning activity (in this case, studying for the authentic assessment) contributed to the development of six transferable generic skills (i.e., problem-solving, analytic skills, teamwork, confidence in tackling unfamiliar situation, ability to plan work, and written communication skills). This scale was added to the CEQ in 1997 as a reaction to the requirements of society in which students not only need to acquire content knowledge, but also need to possess skills relevant to employability and lifelong learning. Lizzio et al. (2002) showed that this scale could be used as a qualitative learning outcome measure. In addition, teachers at three VET institutes confirmed that these skills were very relevant and part of the learning objectives for their students. The translated version revealed a good internal consistency in the VET context (Cronbach's alpha = .72)

Quantitative learning outcome. The quantitative learning outcome was measured by two independent assessors who, during the assessment, scored students' performance during the assessment on several criteria that were placed in a scoring rubric. After the performance, both assessors discussed their scorings, which resulted in one final grade. Due to practical reasons, it was only possible to collect data on the final grade for 77 of the 118 students.

Analysis

To examine the relationships between the various variables, first correlational analyses were used. Correlations were calculated between all perception scales, deep and surface study approaches and both learning outcomes. To test the hypothesis about the influences of perceptions of authenticity and alignment on a deep study approach and the development of

generic skills, Structural Equation Modeling (SEM) with AMOS was used. Contrary to regression analysis, this method is appropriate for examining direct as well as indirect effects on a dependent variable and this method can detect small changes within one group (Joreskog, 1993). The study examined the direct and indirect relationships between five independent variables (i.e., the four reliable authenticity scales and the perception of alignment scale), the intermediate variable DSA and the dependent variable GSD. The perception variables were not expected to influence SSA, but a negative influence of SSA on GSD was added to the model. These variables and their on theory based hypothesized relationships together make up the hypothesized model.

SEM was used to assess the extent to which the hypothesized model adequately fitted or described the empirical data. In this technique, several indices were used as criteria to examine the fit of the model with the data (Byrne, 2001; Joreskog, 1993). This meant that the chi-square needed to be small relative to the degrees of freedom and non-significant, the comparative fit index (CFI), normed fit index (NFI) and the goodness-of-fit index (GFI) should be large ($> .95$), and the root mean square error of approximation (RMSEA) should be small ($< .05$). To explore possible misfits of the model, the modification indexes (MI) for the regression weights could be examined (Byrne). MIs give information about the relationships that were set to zero in the tested model. High MI scores can indicate that an important link is missing in the model. The theoretical model in this study did not incorporate relations between perceptions of authenticity and a surface study approach. Therefore, these missing links were of particular interest.

For SEM purposes only 77 values of the dependent variable grade were available, compared to 118 valid values of the other variables. Therefore, the grade was not used in the structural model.

Results

Table 1 displays the correlations between the perception scales, study approaches and both learning outcomes.

*** INSERT TABLE 1 ***

This table reveals several things: First, almost all significant correlations were in the expected direction. (The probability of a Type I error was maintained at .05 for all subsequent analyses.) They all stressed a positive relationship between perceptions, DSA and the learning outcomes. All perception scales correlated positively with both outcome measures, except for the physical context and grade. In addition, perception of authenticity of the <u>physical context</u> and the <u>task</u> showed a significant positive correlation with a DSA, $\underline{r}(118) = .20$, $\underline{p} < .05$ and $\underline{r}(118) = .23$, $\underline{p} < .01$ respectively. This meant that when students perceived the assessment task and/or the physical context as more authentic, they reported more use of a deep study approach. The only unexpected relationship was a positive correlation between SSA and GSD, meaning that more surface study activities improved the development of generic skills. Second, there was a significant correlation between the GSD (qualitative learning outcome), measured with a student self-report questionnaire and the more objective grade (quantitative learning outcome), $\underline{r}(77) = .25$, $\underline{p} < .05$. This would imply that a higher grade coincides with more generic skill development. Third, as expected, there were no significant correlations between SSA and the perception scales, which supports the idea that authenticity perceptions do not influence surface learning. Fourth, DSA as well as SSA correlated positively with GSD, $\underline{r}(118) = 52$, $\underline{p} < .01$ and $\underline{r}(118) = .33$, $\underline{p} < .01$ respectively, while DSA correlated negatively with grade, $\underline{r}(77) = .23$, $\underline{p} < .05$. In other words, the employment of more deep study activities but also more surface study activities positively influence the development of generic skills, while more deep learning results in a lower grade. Finally, the perception of alignment did not correlate with study approach but showed significant positive correlations

with both outcome measures, r(118)   = 27, p < .01 for GSD; r(77) = .30, p < .05 for grade.

This would mean that perception of alignment does not influence how students learn (study

approach), but that more perception of alignment between assessment and instruction does

lead to better learning outcomes.

Figure 2 shows the hypothesized model and the found values of the relationships

between the variables (path coefficients). This model had a good fit with the sample data,

$\chi 2(10, N = 118) = 7,71, p = .26$, CFI = .99, NFI = .97, GFI = .98 and RMSEA was .05, and

seven of the 12 path coefficients were significant (p < .05).The model explained 54% of the

qualitative learning outcome, and the perceptions explained 20% of the deep study approach.

In addition, the SEM output showed no MIs for the regression weights, indicating that there

were no important links missing in the model, supporting the nonexistence of relationships

between perceptions of authenticity and a surface study approach.

***   INSERT FIGURE 2 ABOUT HERE   ***

This model reveals several things about the influences of authenticity perceptions on

study approach and learning outcome. At least four of the five authentic assessment

dimensions show significant relations with a DSA (the fifth dimension, social context, was

not included due to the insufficient reliability of the scale). In line with the results of the

correlations, an increase in reported DSA was seen when the assessment task and physical

context were perceived as being more authentic. However, contrary to the correlations (Table

1), the structural model showed an unexpected significant relationship between perception of

criterion authenticity and a DSA ($\beta = -.44$); that is, the more authentic the assessment criteria

were perceived, the less deep the students reported having studied. Because the perception of

criterion authenticity negatively influences a DSA, it indirectly influences the GSD in a

negative way as well ($\beta = -.22$). The positive effect between perceived criterion authenticity

and GSD that was found in the correlational analysis disappeared ($r = .23, p < .05$) when the

indirect effect of perceived criterion authenticity on GSD through DSA was added to the equation.

Concerning the relationships between the perception scales and the study approaches on GSD it appears that almost all influences of authenticity perceptions were indirect through study approach. Only the authenticity of the assessment <u>form</u> influenced GSD in a positive and direct way ($\beta = .25$). Moreover, in agreement with the correlational analysis, both study approaches positively influenced GSD indicating that an increase in surface or deep learning both result in the development of more generic skills. However, the influence of a DSA was almost twice as big as the influence of a SSA ($\beta = .47$ and $\beta = .28$ respectively).

The second research question dealt with the influence of perception of alignment on study approach and learning outcome. The hypothesized model shows that perception of alignment does indeed add to explaining the variance of the learning outcome ($\beta = .15$, $\underline{p} < .05$), but in line with the correlations, it does not significantly influence the study approach ($\beta = -.01$, $\underline{p} = .89$)

Conclusion and Discussion

The main hypothesis was that an increased perception of authenticity would result in more deep learning and improved learning outcomes, especially in the development of professionally relevant competences. This was –for the most part- supported by the data. All significant correlations and the hypothesized structural model revealed positive relationships between perceptions of authenticity, a deep study approach and/or the learning outcome. More authenticity of the <u>task</u> and the <u>physical context</u> of the assessment increased the use of a deep study approach. Increased perception of the authenticity of the <u>task, physical context,</u> and the <u>form</u> of the assessment all appeared to positively influence generic skills development and grade. In addition, an important finding was that there are no significant correlations found between perceptions and a surface study approach. This supports the adequacy of the

theory-based hypothesized model used in this study that describes positive relationships

between authenticity perceptions and deep learning and no relationships between perceptions

and surface learning.

However, some unexpected relations were found as well. First, the structural model

showed that an increase in the perception of authenticity of the assessment criteria negatively

influenced a deep approach to studying. As a result, perceived criterion authenticity has a

negative, indirect effect on generic skills development as well. This is contrary to the

significant positive correlation between perceived criterion authenticity and generic skills

development (Table 1). This positive relationship disappeared when the indirect relationship

through a deep study approach was added. This finding shows the additional value of

structural equation modeling over correlations or regression that only examine direct

relationships. Second, not only deep studying, but also surface studying (to a lesser extent)

resulted in more generic skills development.

Two possible explanations for the negative relationship between authentic criteria and

deep studying might be that the criteria were too specific and were revealed only one week

before the assessment. Even though the criteria were perceived as authentic (i.e., important

criteria in professional practice), the criteria focused on very specific and concrete behavioral

actions in an unrealistic time frame of ten minutes. For example "the student makes eye

contact with the client" or "the student asks at least one open questions". Previous research

(e.g., Govearts, Van der Vleuten, Schuwirth, & Muijtjens, 2005) showed that senior students

prefer more holistic assessment criteria. Specific criteria that focus on small concrete actions

might be demotivating and as a result inhibit learning. Second, students did not receive the

assessment criteria at the beginning of the course, but only one week prior to the assessment.

In other words, students did not get the criteria when studying, but rather when they had one

week off to focus only on preparing for the assessment. This might stimulate students to focus

especially on these selected criteria as Boud (1990) argued that assessment encourages students to focus on the topics that are assessed at the expense of those that are not. The concreteness of the assessment criteria and the fact that students received them only one week before the assessment possibly stimulated learning the criteria by heart and practicing in demonstrating only these specific actions, instead of learning in a more holistic way focused on understanding (deep study approach). The assessment culture advocates transparent and concrete criteria (e.g., Dierick & Dochy, 2001) to let students know what is expected of them, but this study shows that this can have a negative effect on student learning if not implemented or perceived correctly. Moreover, one can question if performance criteria in real life are always that specific and concrete?

The finding that deep learning as well as surface learning positively influenced generic skills development shows that students employing a deep study approach were able to effectively deal with the assessment, but that students who mainly used surface activities could get by as well. Thus, succeeding in this assessment did not require deep learning. This result was found in previous studies as well (Biggs, 1987; Gijbels, 2005; Scouller & Prosser, 1994). These studies showed that although a deep study approach is expected to lead to higher achievement (both in terms of quality as well as quantity) and new kinds of assessments are expected to require a deep study approach, assessment does not always reward the deep approach. On the other hand, the positive influence of deep learning on the learning outcome is stronger than the influence of surface learning, which seems to indicate an advantage for deep learners. Another explanation for the positive effect of both surface and deep learning on generic skill development is given by the four-component instructional design model for developing complex skills (Van Merriënboer, 1997). This model states that acquiring complex skills requires deep understanding of the non-routine aspects of a complex skill as well as memorization and drill and practice of the routine aspects of a complex skill. In other words,

surface study activities as well as deep study activities are required in complex skill development.

The second research question considered the positive influence of perception of alignment on learning. The study showed that if students perceive that both instruction and assessment focus on the same kind of learning, this does not influence their study approach but it does positively influence their learning outcomes. This seems to hold for any kind of learning, since the alignment scale measured if students thought that the instruction and the assessment required the same kind of learning, without referring to a specific kind of learning. This corroborates the theory and empirical data on the need for constructive alignment (Biggs, 1996; Segers et al., 2001) and suggests that it is always valuable to examine the effect of assessment on students in the light of the whole learning environment of which the assessment is part.

The significant correlations and path coefficients seem small indicating a small to moderate effect size (correlations ranging from $r = .20$ to $r = .52$ and path coefficients ranging from $\beta = .15$ to $\beta = .47$). Cohen (1988) argues that we should compare the values to other, comparable studies in the field, since the found effects in behavioral sciences will always be much smaller compared to the effects found in, for example, the physical sciences. If we look at the data from this point of view, the found effects are not that small at all. Lizzio et al. (2002) examined direct and indirect relationships between perceptions of the learning environment (including perceptions of the assessment), deep and surface study approaches, and learning outcomes (grade and generic skills development). They reported path coefficients ranging from $\beta = .07$ to $\beta = .32$. Tang (1991) described relations between general study approach, assessment preparations strategies and learning outcomes and found coefficients ranging from $\beta = .10$ to $\beta = .43$.

An additional, but important finding in this study was the significant positive correlation between the qualitative and quantitative learning outcomes (general skill development and grade). Lizzio et al. (2002) argued that the general skill development scale was a valid indicator of learning outcome. However, this is a self-report questionnaire, which is not always considered to be a reliable indicator for actual behavior. The grade, on the other hand, was based on a rating of student performance by two independent assessors. The positive and significant correlation between the grade, based on assessor evaluation, and the self-reported development of generic skills corroborates the validity of the generic skill development scale as a measure of qualitative learning outcome. This might be a valuable finding, because evaluating students qualitatively and on their development of generic skills relevant for employment will become even more important in competency-based education.

Limits and Future Directions

Some notes of caution should be drawn here. The results of the structural equation modeling should be treated with caution. Structural equation modeling was used to get a deeper insight into the (values of the) direct and indirect relationships between perceptions, study approach and learning outcomes than is possible with correlational analysis or regression. However, the model was tested with a group of only 118 students who all worked with the same authentic assessment. The smaller the group of participants, the more the structural equation modeling results are dependable on the specific dataset (Byrne, 2001; Joreskog, 1993). This implies that the relationships and values found in the hypothesized model are indicative and only applicable to this student group. Future research should replicate this kind of study to examine the stability of the relationships found in this study in other cases (other students and other assessments).

Since the data set for grade (quantitative learning outcome) was 77, structural equation modeling was only used to test the influences of perceptions on study approach and generic

skills development (qualitative learning outcome). The correlation matrix (see Table 1) showed that the pattern of significant correlations between perceptions and both learning outcomes look alike, while the relationships between the study approached and both learning outcomes are different (i.e., significant positive correlations between the study approaches and generic skills development; a negative, significant correlation between a deep study approach and grade and a negative but not significant correlation between a surface study approach and grade). Previous research also showed that the influences of perceptions or study approaches on a qualitative or quantitative learning outcome differed (e.g., Gijbels, 2005; Lizzio et al., 2002; Scouller & Prosser, 1994). Future research should examine the relationships between authenticity perceptions, study approach and grades. Especially as long as grades stay one of the most often used measurement of learning.

To gain a deeper insight into the relationships between student perceptions of authentic assessment and the way students study for this assessment and what they learn from it, qualitative research should be used in addition to quantitative studies. Previous research showed that asking students how they studied for a certain (kind of) assessment revealed a lot of useful and valuable information (Sambell et al., 1997; Segers, et al., 2001), which was necessary to build a rich, and contextualized picture of the assessment under investigation. Semi-structured interviews with groups of students can reveal several explanations for the (un)expected relationships found in this study (Morgan, 1997).

Theoretical and Practical value

The hypothesized structural model corroborates the premise that different facets of authenticity influence study approach and learning outcome differently. This supports the 5DF (Gulikers et al., 2004) which argues that an assessment can be made more authentic in different ways and that there is not an 'authentic - not-authentic' dichotomy, but rather an authenticity continuum. In agreement with Gibbs (1999), the 5DF and this study show that

making small changes to the assessment (e.g., increasing the authenticity of the task) can positively influence student learning. In practical terms, this research shows that changing from traditional to competency-based education with authentic assessments need not to be a "one-shot deal". This would make the transition a lot easier for schools and their teachers. For example, much could be gained by first increasing the authenticity of the task, since increasing task-authenticity stimulated deep learning and resulted in better learning outcomes. In the end, however, the results of this study would argue that increasing the authenticity of the task, the physical context, ánd the assessment form results in the most benefits in terms of learning and outcomes.

The study does support the idea that authentic assessment is a multidimensional concept and that various aspects of authenticity influence what and how students learn in a competency-based curriculum in which assessment, instruction and learning are in alignment and aim at bridging the gap between learning and working.

Reference List

Biggs, J. (1996). Enhancing teaching through constructive alignment. Higher Education, 32, 347-364.

Birenbaum, M. (1996). Assessment 2000: Towards a pluralistic approach to assessment. In M. Birenbaum & F. J. R. C. Dochy (Eds.), Alternatives in assessment of achievements, learning processes and prior knowledge (pp. 3-29). Boston: Kluwer Academic Publishers.

Birenbaum, M. (2003). New insights into learning and teaching and the implications for assessment. 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.

Boud, D. (1990). Assessment and the promotion of academic values. Studies in Higher Education, 15, 101-111.

Boud, D. (1995). Assessment and learning: contradictory or complementary? In P. Knight (Ed.), Assessment for learning in Higher Education (pp. 35-48). Londen: Kogan Page.

Byrne, B. M. (2001). Structural Equation Modeling with AMOS. Basic concepts, applications and programming. Mahwah, NJ: Lawrence Erlbaum Associates.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cooper, B. (1994). Authentic testing in mathematics? The boundary between everyday and mathematical knowledge in national curriculum testing in English schools. Assessment in Education: Principles, Policy & Practice, 1, 143-166.

Cummings, J. J. & Maxwell, G. S. (1999). Contextualising authentic assessment. Assessment in Education: Principles, Policy & Practice, 6, 177-194.

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

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

Dochy, F. J. R. C. & McDowell, L. (1997). Assessment as a tool for learning. Studies in Educational Evaluation, 23, 279-298.

Entwistle, N., McCune, V., & Hounsell, J. (2002). Approaches to studying and perceptions of university teaching-learning environments: Concepts, measures and preliminary findings (Occasional Report 1). United Kingdom: University of Edinburgh, ETL-project, Higher and Community Education, the School of Education.

Entwistle, N. J. & Ramsden, P. (1983). Understanding Student Learning. London: Croom Helm.

Entwistle, N. J. (1991). Approaches to learning and perceptions of the learning environment. Introduction to the special issue. Higher Education, 22, 201-204.

Gibbs, G. (1992). Improving the quality of student learning. Bristol: Technical and Educational Services.

Gibbs, G. (1999). Using assessment strategically to change the way students learn. In S. Brown & A. Glasner (Eds.), Assessment matters in higher education (pp. 41-53). Buckingham: Open University Press.

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.

Gijbels, D. (2005). Effects of new learning environments. Taking students' perceptions, approaches to learning and assessment into account. Unpublished doctoral dissertation, University of Maastricht, The Netherlands.

Glaser, R., & Silver, E. (1993). Assessment, testing and instruction: Retrospect and prospect. Review of Research in Education, 20, 393-419.

Govaerts, M. J. B., Van der Vleuten, C. P. M., Schuwirth, L. W. T., & Muijtjens, A. M. M. (2005). The use of observational diaries in in-training evaluation: Student perceptions. Advances in Health Sciences Education, 10, 171-188.

Gulikers, J., Bastiaens, Th., & Martens, R. (2005). The surplus value of an authentic learning environment. Computers in Human Behavior, 21, 509-521.

Gulikers, J., Bastiaens, Th., & Kirschner, P. (2004). A five-dimensional framework for authentic assessment. Educational Technology Research and Development, 52, 67-85.

Hambleton, R. K., Murphy, E. (1992). A psychometric perspective on authentic measurement. Applied Measurement in Education, 5, 1-16.

Herrington, J. & Herrington, A. (1998). Authentic assessment and multimedia: how university students respond to a model of authentic assessment. Higher Educational Research & Development, 17, 305-322.

Joreskog K.G. (1993). Testing stuctural equation modeling. In K. A. Bollen & J. S. Long (Eds.), Testing structural equation models (pp. 294-316). Newbury Park, CA: Sage.

Lizzio, A., Wilson, K., & Simons, R. (2002). University students' perceptions of the learning environment and academic outcomes: implications for theory and practice. Studies in Higher Education, 27, 27-51.

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

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

Morgan, D. L. (1997). Focus groups as qualitative research. London: Sage Publications.

Perkins, D. & Blythe, T. (1994). Putting understanding up front. Educational Leadership, 60, 4-7.

Ramsden, P. (1992). Learning to teach in higher education. London: Routledge.

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, 349-371.

Scouller, K. M. & Prosser, M. (1994). Students' experiences in studying for multiple choice question examinations. Higher Education, 19, 267-279.

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.

Segers, M., Dierick, S., & Dochy, F. (2001). Quality standards for new modes of assessment. An exploratory study of the consequential validity of the OverAll test. European Journal of Psychology of Education, 16, 569-586.

Segers, M., Dochy, F., & Cascallar, E. (Eds.) (2003). Optimising new modes of assessment: In search of qualities and standards. Dordrecht: Kluwer Academic Press.

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

Struyven, K. (2005). The effects of student-activated teaching/learning environments of students' perceptions, student performance and pre-service teachers' teaching. Unpublished doctoral dissertation, University of Leuven, Belgium.

Tang, C. (1994). Effects of modes of assessment on students' preparation strategies. In G. Gibbs (Ed.), Improving student learning - Theory and practice (pp. 151-170). Oxford: Oxford City Centre for Staff Development.

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. Journal of Curriculum Studies, 25, 265-278.

Thomas, P. R. & Bain, J. D. (1984). Contextual dependence of learning approaches: The effects of assessments. Human Learning, 3, 227-240.

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.

Van Rossum, E. J. & Schenk, S. M. (1984). The relationship between learning conceptions, study strategies and learning outcome. British Journal of Educational Psychology, 54, 73-83.

Author Note

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

Educational Technology Expertise Center of the Open University of the Netherlands.

Correspondence concerning this article should be addressed to Judith T.M. Gulikers who is now at the Educational Technology Expertise Center of the Open University of the Netherlands, PO Box 2960, 6401 DL Heerlen, The Netherlands. Electronic mail may be sent via Internet to Judith.Gulikers@ou.nl.

Table 1.

Correlations between Student Perceptions, Study Approaches, and Learning Outcomes

	Surface Study Approach (SSA) ($n = 118$)	Deep Study Approach (DSA) ($n = 118$)	Generic Skill Development (GSD) ($n = 118$)	Grade ($n = 77$)
Task	-.02	.23**	.33**	.33**
Physical context	.15	.20*	.39**	.15
Form	.15	.12	.48**	.47**
Criteria	.05	-.14	.23*	.32**
Alignment	-.08	.05	.27**	.30*
Surface Study Approach	-	.02	.33**	-.10
Deep Study Approach	.02	-	.52**	-.23*
Generic Skill Development	.33**	.52**	-	.25*
Grade	-.10	-.23*	.25*	-

* $p < .05$, two-tailed. ** $p < .01$, two-tailed.

Figure Captions

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

Figure 2. The hypothesized structural model describing the standardized path coefficients of the direct and indirect relationships between perceptions, study approach and generic skills development.
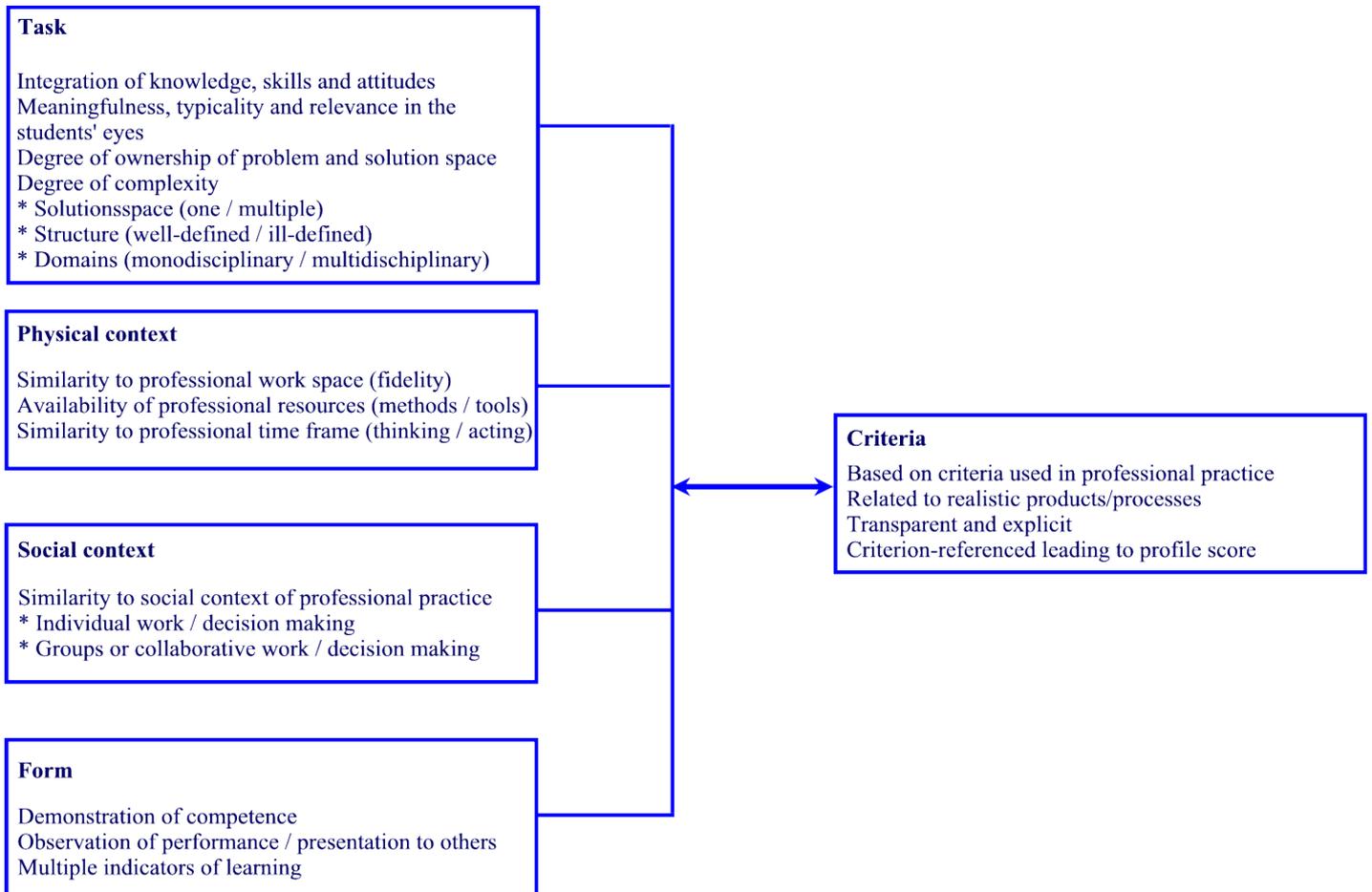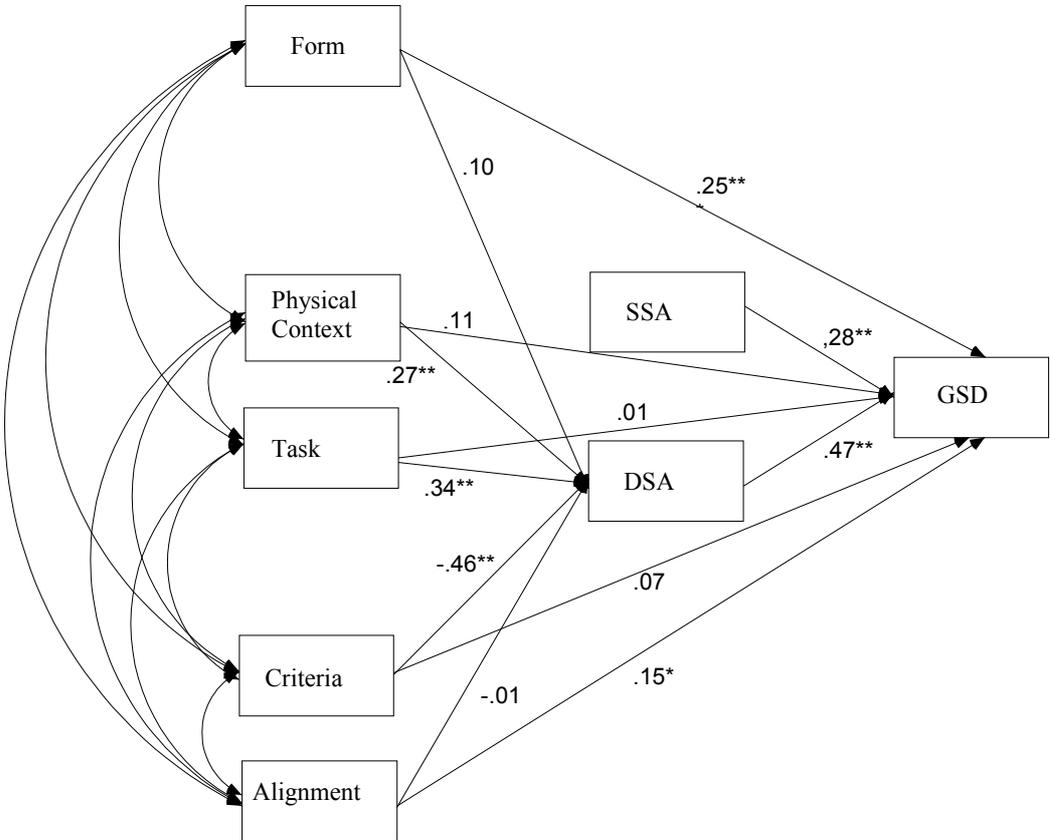
Figure 1.

Task
Integration of knowledge, skills and attitudes Meaningfulness, typicality and relevance in the students' eyes Degree of ownership of problem and solution space Degree of complexity * Solutionsspace (one / multiple) * Structure (well-defined / ill-defined) * Domains (monodisciplinary / multidischiplinary)

Physical context
Similarity to professional work space (fidelity) Availability of professional resources (methods / tools) Similarity to professional time frame (thinking / acting)

Social context
Similarity to social context of professional practice * Individual work / decision making * Groups or collaborative work / decision making

Form
Demonstration of competence Observation of performance / presentation to others Multiple indicators of learning

Criteria
Based on criteria used in professional practice Related to realistic products/processes Transparent and explicit Criterion-referenced leading to profile score

Figure 2.

* p < .05. ** p < .01