

The effects of performance-based assessment criteria on student performance and self-assessment skills

Greet Mia Jos Fastré · Marcel R. van der Klink ·
Jeroen J. G. van Merriënboer

Received: 16 June 2009 / Accepted: 16 December 2009 / Published online: 7 January 2010
© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract This study investigated the effect of performance-based versus competence-based assessment criteria on task performance and self-assessment skills among 39 novice secondary vocational education students in the domain of nursing and care. In a performance-based assessment group students are provided with a preset list of performance-based assessment criteria, describing what students should do, for the task at hand. The performance-based group is compared to a competence-based assessment group in which students receive a preset list of competence-based assessment criteria, describing what students should be able to do. The test phase revealed that the performance-based group outperformed the competence-based group on test task performance. In addition, higher performance of the performance-based group was reached with lower reported mental effort during training, indicating a higher instructional efficiency for novice students.

Keywords Competence-based assessment criteria · Mental effort · Performance-based assessment criteria · Self-assessment skills · Task performance

Introduction

In competence-based education, authentic learning tasks based on real-life problems are the driving force behind training, simultaneously encouraging the development of professional skills and more general competences like being self-directed. Competence-based education is a dominant trend in vocational education in many European countries (Wesseling et al. 2007). The aim is to prepare students for the workplace where people are expected to be broadly educated while stimulating lifelong learning (van Merriënboer et al. 2002, 2009). Because competences are context-bound and the aim of vocational education is preparing students for the workplace, students should always develop competences in the

G. M. J. Fastré · M. R. van der Klink
Open University of the Netherlands, Heerlen, The Netherlands

G. M. J. Fastré (✉) · J. J. G. van Merriënboer
Maastricht University, Maastricht, The Netherlands
e-mail: g.fastre@educ.unimaas.nl

context of a profession (Biemans et al. 2004). When teachers want to judge the competence development of their students, student assessments performed in a real-life context can support their findings.

Assessment criteria and standards are key clues for students to know what is essential in their study program. Fastré et al. (2009) show that drawing students' attention to the assessment criteria that are relevant for a particular learning task improves their understanding of the criteria and subsequently leads to better test task performance and better self-assessment skills. The following citation of Otter (1995) emphasizes the importance of being familiar with the relevant assessment criteria:

Describing and making clear and public what the learner is intended to achieve changes the nature of assessment from a tutor-led system with fuzzy objectives and undisclosed criteria, to a student-led system with greater emphasis on formative development and personal responsibility. (p. 45).

In the behavioural tradition of instruction and instructional design, assessment criteria were performance-based, meaning that they described the desired performance in terms of what the student has to do (e.g. Mager 1984). With the introduction of competence-based education, assessment criteria are often formulated as competences, in terms of what the student is able to do. However, no research so far has investigated the effects of this introduction of competence-based assessment criteria. The main goal of this study is to investigate the effects of competence-based versus performance-based assessment criteria on learning, test task performance and students' self-assessment skills.

The difference between performance-based and competence-based assessment criteria should be seen as a continuum, where on the one end assessment criteria are formulated as competences, which are an integration of knowledge, skills and attitudes; and on the other end assessment criteria are formulated as performance indicators. Performance-based criteria can be linked directly to competence-based criteria and vice versa as they complement each other. When discussing the continuum, the two extremes and their underlying connection will be tackled. The discussion will be coupled to the level of experience students have as it can be assumed that students with different levels of experience will have different needs concerning assessment criteria (Kalyuga 2007). In this article the focus is on the needs of novice students.

Figure 1 presents a summary of the continuum between competence-based and performance-based assessment criteria: (1) What is assessed, (2) the nature of the criteria, (3) holistic versus analytic, and (4) the level of mental effort.

First, with regard to what is assessed, when assessing with competence-based criteria, the competences underlying the performance are the focus of the assessment. What is assessed is the student's ability to perform a certain task. However, competences as a whole are not directly observable (Grégoire 1997). Certain aspects of competences are observable, like particular skills the students demonstrate, but certain aspects are hidden, like their self-concept and personal characteristics that influence their performance (Spencer and Spencer 1993).

When assessing with performance-based criteria, the observable behaviours produced by the students are the heart of the assessment. The question is not if the student is able to perform the task, but if the student shows good performance (Grégoire 1997). In order to show this good performance, students probably also know how to perform and consequently master the underlying competences necessary for performing the task (Miller 1990). For example, in the case of stoma care, the student shows he can remove the stoma in a correct way. An underlying competence is supporting the patient according to

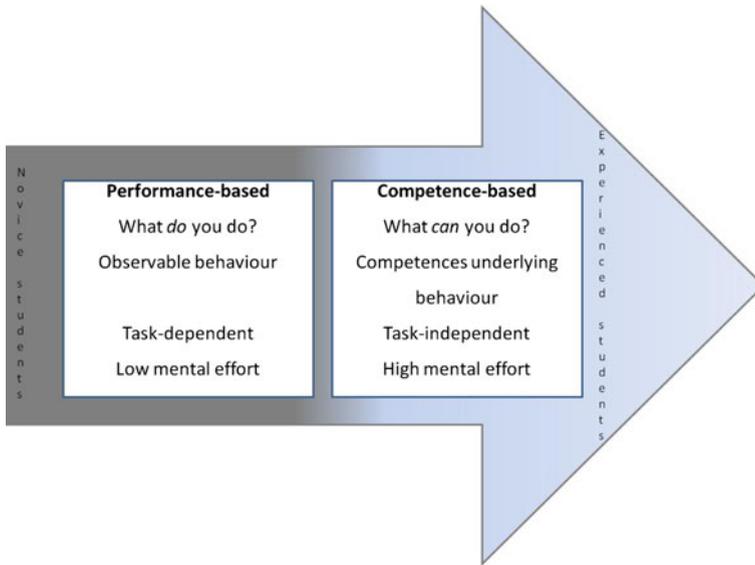


Fig. 1 Continuum of performance-based to competence-based assessment criteria

protocols, regulations and the vision of the organisation but the performance criterion is removing the stoma in a correct way. This means there is a direct link between what students show (performance) and what students are able to do (competence). Every performance shown involves one or more competences the student has to possess to perform well, and every competence can be shown in several behaviours of the student.

Because for novice students it is important in an early stage to obtain an idea of how well they are doing, the directly observable character of the performance-based criteria may be expected to be more beneficial to assess their task performance. Based on these performance-based criteria, the development of the students from the beginning on can be monitored. In order to improve novice students' self-assessment skills, it is easier to assess what they are actually doing because this is more objective than their ability to do so. Therefore, with regard to what is assessed, performance-based criteria are expected to be more beneficial for supporting novice students' learning than competence-based criteria. In later stages, it is important for students to learn to see the link with the underlying abilities they are developing.

Second, with regard to the nature of the criteria, to uncover competence development, consistency of proof of competence level across different tasks is needed (Albanese et al. 2008; Grégoire 1997). It is therefore important to formulate competence-based assessment criteria in a way that they can be used across different tasks and thus are task-independent. For example, a nurse has to be able to conduct nursing technical skills. In one situation this means replacing a stoma bag while in another situation this means washing a patient.

To judge student performance on a certain task, performance-based assessment criteria should be formulated on task-level as for each task a different set of criteria is relevant. Performance-based criteria are thus task-dependent. As is shown by Fastré et al. (2009), for novice students it is important to know the relevant criteria in every task. For example, when a nurse has to conduct stoma care, some of the relevant criteria are to remove the old stoma bag and apply a new one.

It is likely that when students know exactly what to do, their motivation, learning and performance will increase significantly (see for example Ecclestone 2001). Moreover, Miller (2003) argues that having task-specific assessment criteria leads to a better quantitative differentiation of performance levels. This more detailed view on students' performance, would argue for the use of performance-based assessment criteria. Following the results of Fastré et al. (2009), it can be concluded that the use of performance-based criteria is especially beneficial for novice students because of their task-specific character.

Third, the competence-based assessment model currently used in Europe, starts from a fixed set of competences that are categorically divided (e.g. communication skills, nursing technical skills). No further decomposition of the competences is made. The formulation of the competence-based assessment criteria is therefore holistic (Grégoire 1997). This does not mean that when working with competence-based assessment criteria only a holistic judgment on the end result is given, but the criteria are more holistically formulated than the performance-based criteria.

In a performance-based assessment model, the whole task is hierarchically analysed by developing a skills hierarchy (van Merriënboer 1997; Sadler 1985). In this way, criteria are expressed as a component of a higher-level criterion or a number of lower-level criteria. After the student performed the task, the teacher gives separate judgments on each of the preset criteria. Then, these judgments are combined to compose a final judgment which is often converted into a grade. As an example, Fig. 2 shows a part of this decomposition for performing the task of stoma care.

Gulikers et al. (2008) discuss the notions of analytic versus holistic grading from the perspective of the level of experience of students. They argue that novice students need analytic criteria as guidelines in a step-by-step process leading to the desired behaviour. In future tasks, this helps to set appropriate learning goals (Eva and Regehr 2005). For more experienced students, analytic criteria may hamper their learning process because they have to be stimulated to keep their focus on a certain outcome level and they do not need the step-by-step approach any more (Scheffer et al. 2008). Following these ideas, for novice students it would be better to receive performance-based assessment criteria.

Finally, with regard to mental effort, when designing a study program, including assessment, it is important to strive for the optimal level of using students' cognitive

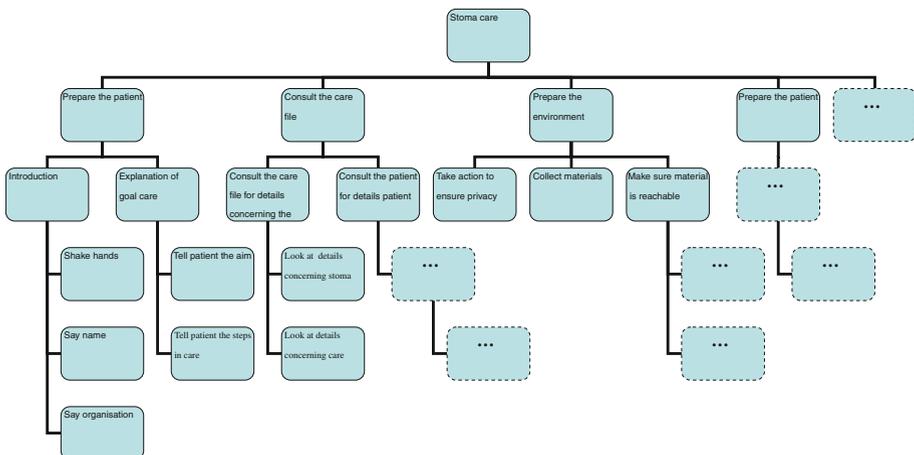


Fig. 2 Part of the skill hierarchy of stoma care

capacity (van Gog and Paas 2008). Cognitive load theory presupposes that people have a limited working memory capacity (Sweller et al. 1998; van Merriënboer and Sweller 2005). Because of this limited capacity, it is essential for learning to properly allocate the available cognitive resources (Kalyuga et al. 2003).

An important difference can be distinguished here between novice students and more experienced students. For novice students, it is important to provide sufficient guidance that compensates for the limited knowledge they have on the task at hand (e.g. stoma care) by providing them performance-based assessment criteria because this requires less cognitive capacity for the assessment and most of their working memory capacity can be devoted to the task of stoma care. For more experienced students, who already have some knowledge on the task at hand (e.g. stoma care), competence-based assessment criteria can provide them with an extra stimulus to think about the task in another way and thereby make the extra cognitive capacity beneficial for them. In addition, providing these students with performance-based assessment criteria would give them redundant information on the task which may hamper their learning. This is called the expertise reversal effect (Kalyuga 2007).

Summarising, it appears that for novice students, performance-based criteria have more advantages than competence-based criteria because: (1) They are directly observable, (2) they lead to a better quantitative differentiation of levels of performance, (3) they stimulate a step-by-step process leading to desired performance, and (4) they require less cognitive capacity for assessment leaving more capacity for learning the task at hand. The following section describes the hypotheses following this comparison.

Hypotheses

The first hypothesis is that students who receive the performance-based criteria during learning will show superior test task performance compared to students who receive the competence-based criteria because they know better what is expected from their performance. The second hypothesis is that students who receive the performance-based criteria will experience a lower mental effort during assessment than students who receive the competence-based criteria. The third hypothesis is that students who receive the performance-based criteria will be better self-assessors than students who receive the competence-based criteria because they are better able to assess their performance.

Method

Participants

Thirty-nine second-year students of a school for Secondary Vocational Education, attending a Nursing and Care program (Level 3 and 4 in the European Qualifications Framework, 2 males and 37 females) participated in this study as part of their regular training on the nursing task of stoma care. Their mean age was 18.07 years ($SD = 1.05$). Participants were randomly assigned to one of the two conditions: competence-based criteria ($n = 20$) and performance-based criteria ($n = 19$).

Materials

The whole task of stoma care, addressing the psychosocial needs of the patient, analysing the situation of the patient, changing the stoma bag, and the evaluation afterwards are

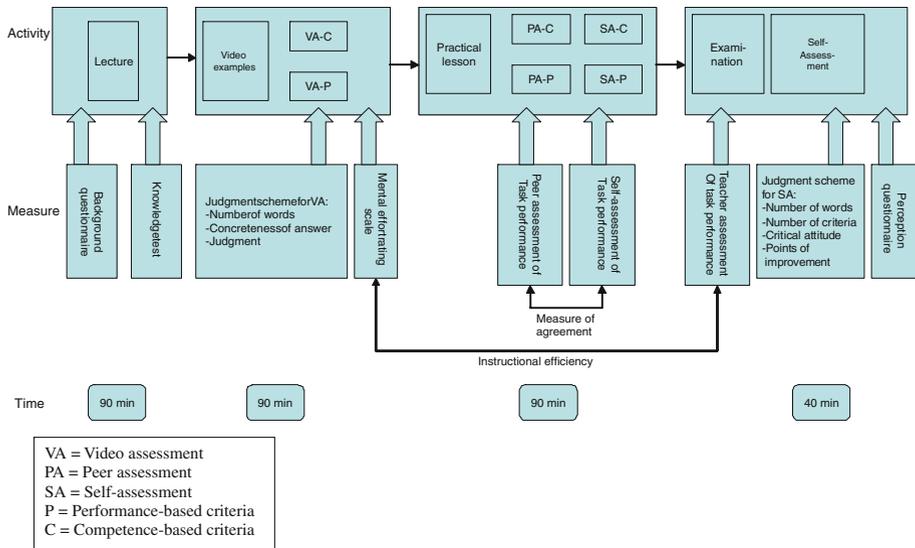


Fig. 3 Summary of materials and measures

included in the task. This means students did not only practise the technical skill of changing the stoma bag, but also needed knowledge on the stoma (e.g. possible problems with stomas), and an appropriate attitude towards the patient. The task was set up in accordance with the theory of experiential learning by Steinaker and Bell (1979) which distinguishes four important steps: (a) exposure, (b) participation, (c) identification, and (d) internalisation. Figure 3 summarises the materials described below.

Lecture

A lecture was developed that provided students with the theoretical background of stoma care. The two teachers who were responsible for this lecture set up the lecture together.

Video examples and video assessment

An electronic learning environment was developed including six video fragments (± 3 min each) in which an expert nurse shows good stoma care behavior. All fragments are subsequent parts of the whole task of stoma care: (1) Introduction, (2) preparation, (3) removing the old stoma bag, (4) applying the new stoma bag, (5) finishing off care, (6) evaluation and reporting. Students individually watched the video fragments on a computer screen. They were not allowed to put the fragment on hold, and they could watch the video a maximum of three times. On average, students watched the video 1.14 times ($SD = .29$). No differences between conditions were found.

After students watched the video, they had to assess the performance of the nurse in the video on an electronic list of preset criteria. A distinctive feature was made for the two conditions. In the competence-based condition, the assessment criteria were formulated as competences of stoma care as used previously in the study program (VA-C). Figure 4

	Criterion	How does the nurse show this criterion
Giving attention and sympathising	Shows interest, listens actively, shows empathy to the patient	
	Puts herself in position of patients, colleagues and supervisors	
Working together and consulting	Informs the patient pro-active	
	Consults patients and involved others on a regular basis and informs them	
	Makes agreements with patients and others involved in the task division	
Ethical and honest treatment	Acts consequent to own norms, professional group, organisation and legal constraints	
	Is integer, fair and acts without prejudice	
	Is discrete with sensitive topics	
	Communicates open and clear about sensitive topics	

The criteria with the text box are the relevant criteria for a particular video fragment

Fig. 4 Screenshot of competence-based assessment criteria

shows some examples, competence-based criteria as shown in the electronic learning environment.

In the performance-based condition, the assessment criteria were formulated as the underlying skills of a skill hierarchy of stoma care (VA-P). Figure 5 shows some examples of performance-based criteria as shown in the learning environment.

In order to encourage students to make the assessment criteria more concrete, students in both groups had to indicate the manner in which the nurse in the fragment showed good behaviour on the criteria by typing their answer in the text boxes.

	Criterion	How does the nurse show this criterion
Prepares the patient	Introduces herself in an appropriate way	
	Explains to the patient what her goal is in an understandable language	
Consults the care file	Consults the care file for details concerning the stoma	
	Consults the patient for details concerning the patient	
Prepares the environment for the care	Takes action to ensure there is sufficient privacy	
	Collects the right materials	

The criteria with the text box are the relevant criteria for a particular video fragment

Fig. 5 Screenshot of performance-based assessment criteria

Practical lesson, peer assessment, and self-assessment

A practical training session was developed in which students had to practice in pairs or groups of three the task of stoma care with a fellow student being the patient. After students had performed the task, they had to score their peers' task performance on the same list of criteria as in the assessment of the video examples. The students in the competence-based condition received the list with competence-based criteria (PA-C) and students in the performance-based condition received the list with performance-based criteria (PA-P). They had to indicate how well their peers mastered the criteria on a four-point scale: (1) behaviour not shown, (2) behaviour shown but insufficient, (3) behaviour shown and sufficient, (4) behaviour shown and good. In addition to this peer assessment, students had to self-assess their task performance using the identical list of competence-based criteria (SA-C) or performance based criteria (SA-P), using the same four-point scale. While practising the task, students also received oral feedback on their task performance from the instructor in the room.

Examination and self-assessment

An examination was developed in which students individually had to perform the task of stoma care with a simulation patient. Afterwards they had to assess their own performance on that particular task by filling in a blank paper with the question: assess your own performance on this task and indicate what went well and what went wrong.

Measures

Background questionnaire

A short questionnaire measured the background of the students on demographical factors such as age, sex and prior education. Student perceptions of the relevance of self-assessment and their perceptions of their ability to self-assess were measured by the self-directed learning skills questionnaire adapted from Kicken et al. (2006). This questionnaire proved reliable for the population in this study (Fastré et al. 2009). Table 1 shows the Cronbach's alpha scores of the perception scales; internal consistencies ranged from .70 to .75 and are thus quite acceptable.

Knowledge test

At the end of the lecture, a 15-item multiple choice test was taken to test the students' knowledge on this subject.

Table 1 Reliability of the self-directed learning skills questionnaire

Scale	Cronbach's alpha	# Items	Example item ^a
Self-directed learning skills questionnaire			
Relevance of self-assessment	.70	3	I think the opinion of the teacher is more important than self-assessment
Ability to self-assess	.75	12	I can assess to what extent my performance fits the assessment criteria

^a Items have been translated from Dutch

Judgment scheme for video assessment

To measure the accuracy of the video assessment, judgment schemes specified the quality of the video assessments. The overall score for quality of video assessment was the sum of the z -scores of the following aspects: how many words the students used because it is expected that performance-based criteria stimulate students more to elaborate on their answers (count of the number of words), if they gave concrete examples of the nurse's behaviour (0 = no concrete behaviour, 1 = concrete behaviour), and if they gave a judgment on the behaviour of the nurse (0 = no judgment, 1 = judgment). The higher the sum of the z -scores, the better the score for quality of video assessment as it is important that the combination of these factors is of a high quality. The quality of the video assessments was judged by two raters, with a high interrater reliability of $r = .82, p < .00$.

Mental effort rating scale

After the assessment of each video fragment, students were required to fill in the rating scale of Paas (1992) that measured their mental effort as the 'effort required to perform the assessment', ranging from a very small amount of effort (1) to a very high amount of effort (7).

Peer assessment of task performance

The peer assessments during the practical lesson indicated the task performance of the students assessed by the peers, using the competence-based criteria in one group and performance-based criteria in the other group. Peer assessed task performance was the average score on all the assessment criteria.

Self-assessment of task performance

The self-assessments during the practical lesson indicated the task performance of the students by the students' own opinion, using the competence-based criteria in one group and performance based criteria in the other group. Self-assessed task performance was the average score on all the assessment criteria.

Teacher assessment of test task performance

During the examination, the teachers observed and assessed the test task performance of the students, who took care of the stoma of a simulation patient, on the list of performance-based criteria. A second assessor co-assessed with each of the teachers to measure the reliability of the assessments. The correlation between the scores of the teacher and the second assessor, $r = .77, p < .01$, appeared to be acceptable.

Judgment scheme for self-assessment

The overall score for quality of the self-assessments during examination was the sum of the z -scores of the following aspects: how many words the students used because it is expected that performance-based criteria stimulate students more to elaborate when self-assessing (count of the number of words), how many criteria the students came up with (count of the number of criteria), if students had a critical attitude to their own performance (0 = no

critical attitude, 1 = critical attitude), and if they formulated points of improvement (0 = no points of improvement, 1 = points of improvement). The higher the sum of the z-scores, the better the score for quality of self-assessment because it is important that the combination of these factors is of a high quality. The quality of the self-assessments was judged by two raters, with an interrater reliability of $r = .82$, $p < .00$.

Perception questionnaire

The following aspects of perception were measured to evaluate the learning experience: Motivation for the study, regulation strategies, interesting course material, task orientation, pleasure and interest, pleasure and interest in reflection, and usefulness. All aspects were measured with the use of four-point Likert scales. Higher scores indicate a more positive perception of the learning experience. Two scales (interesting course material and task orientation) of the inventory of perceived study environment (IPSE; Wierstra et al. 1999) measured students' perceptions of the learning environment. Three scales (interest and pleasure, interest and pleasure in reflection, and usefulness) of the intrinsic motivation inventory by Deci et al. (1994), translated into Dutch by Martens and Kirschner (2004), were included in the questionnaire. Table 2 shows the Cronbach's alpha scores of the perception scales; internal consistencies ranged from .69 to .89 and are thus acceptable to high.

Measure of agreement

For the peer assessments and the self-assessments during the practical lesson, the agreement of the scores between the self- and peer assessments was measured by computing the Pearson's correlation.

Instructional efficiency

Instructional efficiency is calculated by relating task performance in the test task and experienced mental effort during training (Paas and van Merriënboer 1993; van Gog and Paas 2008). Performance and mental effort scores are first standardized, and then the z-scores are entered into the formula:

Table 2 Reliability of the perception measures

Scale	Cronbach's alpha	# Items	Example item ^a
Inventory of perceived study environment			
Interesting course materials	.72	8	The learning task is interesting
Task orientation	.69	3	I know what is expected of me when performing the task
Intrinsic motivation inventory			
Interest and pleasure in learning tasks	.69	7	I enjoy working on the learning task
Interest and pleasure in reflection	.89	6	I find it interesting to reflect
Usefulness	.69	4	I would like to conduct more learning tasks because they are useful

^a Items have been translated from Dutch

$$E = \frac{Z_{\text{Performance}} - Z_{\text{Mentaleffort}}}{\sqrt{2}}$$

High efficiency indicates that with a relatively low mental effort during training a relatively high task performance in the examination is accomplished, while a low efficiency indicates that with a relatively high mental effort during training a relatively low task performance is accomplished. For example, instructional efficiency is higher for an instructional condition in which participants attain a certain performance level with a minimum investment of mental effort than for an instructional condition in which participants attain the same level of performance with a maximum investment of mental effort.

Procedure

At the start of the lecture, the background questionnaire was administered. After students had filled in the questionnaire, the lecture was given and the multiple choice test was taken. This phase lasted for 90 min.

After the lecture students were instructed to assess the video examples. While doing this, students were exposed to the stoma care by watching video examples of an expert nurse showing the intended behaviour, which is the first step in the taxonomy of Steinaker and Bell (1979). Students were split up in the two experimental groups to work on the assessment of video examples. Students could work on the assessment of video examples for maximum 90 min. After the assessment of video examples, the practical lesson with peer and self-assessments took place for 90 min. In this lesson, students could participate in stoma care by practicing on a doll (second step).

One week after the practical lesson, students had to conduct the examination after which they had to assess their own performance. In this examination, they could identify with the stoma care because they were exposed to a simulation patient in performing the care (third step). Student performance was assessed by a teacher. At the end of the examination the evaluation questionnaire was filled in by the students. The examination including self-assessment lasted for 40 min. After the whole experiment, students were sufficiently prepared for further practice during internships which leads them to internalise the competence of stoma care (fourth step).

Results

This section describes the results on prior measurements, the dependent variables in the learning and test phase, and the student perceptions. Mann–whitney *U* tests were performed to test for differences between the two conditions. For all analyses, the significance level is set to .05. Table 3 presents the means and standard deviations for all dependent variables.

Prior measurements

On the background questionnaire, no significant difference between the conditions was found, indicating that students did not differ in background at the end of the lecture.

Table 3 Means and standard deviations for dependent variables

	Competence-based ($n = 20$)		Performance-based ($n = 19$)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Pretest				
Score on prior knowledge test	7.72	1.27	7.85	1.35
Learning phase				
Score quality of video assessment	.08	1.50	.22	1.65
Number of words	889	143	855	248
Concreteness of answer	.50	.52	.80	.41
Judgment	.56	.51	.43	.51
Mental effort	3.33	.75	2.36	.65
Task performance scored by peer	2.95	.34	3.14	.43
Task performance scored by self	3.12	.31	3.25	.48
Test phase				
Test task performance scored by teacher	2.89	.44	3.18	.45
Score quality of self-assessment	-.53	2.73	.56	2.92
Number of words	93	45	111	67
Number of criteria	6.35	3.80	6.32	3.86
Critical attitude	.60	.50	.74	.45
Points of improvement	.15	.37	.37	.50
Instructional efficiency	-.65	.81	.68	.91

On the knowledge test, no significant difference between the conditions was found, indicating that all students had the same level of knowledge at the end of the lecture. Thus, students had the same background and prior knowledge before they started to study the video examples.

Learning phase

On the overall score for quality of video assessment, a significant difference between the conditions was found, $z = -1.964$, $p < .05$. Students in the performance-based condition had an average rank of 18.21, while students in the competence-based condition had an average rank of 12.00. More specifically, on number of words no difference was found. In concreteness of answers, a significant difference was found, $z = -1.716$, $p < .05$. Students in the performance-based condition had an average rank of 18.40, while students in the competence-based condition had an average rank of 13.75. No significant difference in judgment was found. A further qualitative analysis of the data reveals that students in the competence-based condition often decoded the competence-based assessment criteria into the performance-based criteria as an answer but were not able to describe the concrete behaviour.

Mental effort during assessment of the video examples is an average score of the scores during assessment of the six film fragments. On mental effort, a significant difference between conditions was found, $z = -3.964$, $p < .001$, indicating that students in the performance-based condition had an average rank of 12.61, while students in the competence-based condition had an average rank of 27.03.

On peer assessment and self-assessment of task performance in the practical lesson, no significant differences between conditions was found. Yet, a moderate agreement between peer and self-assessment was found, $r = .65$, $p < .00$, indicating that students' self-assessment scores corresponded with the scores of their peers. For the performance-based condition $r = .66$, $p < .01$, and for the competence-based condition $r = .63$, $p < .01$.

Test phase

On test task performance, a significant difference between conditions was found, $z = -2.037$, $p < .05$. Students in the performance-based condition had an average rank of 23.82, while students in the competence-based condition had an average rank of 16.38. On the overall score for quality of self-assessment, no significant differences between both conditions were found. Although not significant, the direction of the differences was in line with the expectations. On instructional efficiency, a significant difference between conditions was found, $z = -3.962$, $p < .001$, indicating that students in the performance-based condition had an average rank of 27.42, while students in the competence-based condition had an average rank of 12.95.

Evaluation questionnaire

Overall, students perceived the learning environment as interesting and useful. Table 4 shows the means and standard deviations for all scales.

No significant differences were found between conditions. Being in the performance-based or competence-based condition did not influence students' perceptions of the learning task.

Discussion

The goal of this study was to investigate the effects of competence-based versus performance-based assessment criteria on students' test task performance and self-assessment skills. The first hypothesis, stating that students who receive the performance-based criteria will be better task performers than students who receive the competence-based criteria is confirmed by the data. It seems that novice students who receive the performance-based criteria during training know better what is expected from their task performance and are better able to show desired performance than students who receive the competence-based criteria. A possible explanation is the finding that students who receive the performance-based criteria had a higher

Table 4 Means and standard deviations for evaluation questionnaire

	Competence-based ($n = 20$)		Performance-based ($n = 19$)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Interesting course material	3.13	.43	3.07	.39
Task orientation	2.83	.67	2.51	.70
Interest and pleasure	3.47	.39	3.35	.37
Interest and pleasure in reflection	2.59	.43	2.51	.58
Usefulness	3.45	.52	3.50	.33

quality of video assessments in the learning phase. They were especially better in being concrete on the desired behaviour, which may have led to better task performance in the test phase. This is in line with the ideas of Eva and Regehr (2005), who state that performance-based criteria make it easier to distinguish levels of performance, enabling a step-by-step process of performance improvement.

The second hypothesis, stating that students who receive the performance-based criteria experience a lower mental effort during assessment than students who receive the competence-based criteria is also confirmed by the data. It appears that by providing novice students with performance-based assessment criteria, they have to invest less mental effort to assess their task performance. This effect is positive when it leads to a better test task performance because this would mean that during training the reduced load of assessment permits more cognitive capacity for learning to perform the task of stoma care.

Indeed, the findings concerning the first and second hypotheses together allow to conclude that the performance-based assessment criteria result into a higher instructional efficiency, since students in the performance-based condition experience a lower cognitive load during the learning phase, followed by a higher performance on the test task (Paas and van Merriënboer 1993; van Gog and Paas 2008). Providing novice students with performance-based assessment criteria thus leads to more efficient learning.

The third hypothesis, stating that students who receive the performance-based criteria become better self-assessors than students who receive the competence-based criteria, is not confirmed by the results. This finding is, however, in line with the findings of Dunning et al. (2004), who also found that for novice students knowing the assessment criteria does not necessarily imply the ability to assess their own performance on those criteria. As self-assessment can be seen as a complex cognitive skill, one of the key words in developing this skill is sufficient practice (van Merriënboer and Kirschner 2007). It is likely that students need considerably more practice than provided in the current study to improve their self-assessment skills.

Finally, students did not differ in their perceptions of the learning environment. Receiving competence-based or performance-based criteria thus did not influence their appreciation of the learning task. The findings indicate that both groups were positive about the learning task as a whole and especially valued the provided video examples.

The results of this study show that for novice students performance-based assessment criteria do lead to a lower mental effort during learning and a higher test task performance, which is in line with our theoretical assumption that for novice learners it is better to use performance-based criteria than competence-based criteria. The question remains, however, what causes the observed effects. The relative importance of the separate dimensions of Fig. 1 was not investigated in this study and further research is required to determine the contribution of the various dimensions to the reported effects on mental effort during learning and test task performance. Is it because these criteria refer to directly observable behaviour? Or is it because the criteria are more task-dependent? Maybe the analytic character of the criteria is the driving force behind these effects? These insights could serve as a guideline for teachers in the development of performance-based assessment criteria and should be further examined.

Furthermore, the effects of providing students with performance-based assessment criteria should be examined with students in later years of the educational program to explore differences between novice and more experienced students as it is expected that students in later phases of their educational program have to learn to think on a higher level and thus work more efficient with competence-based criteria.

A shortcoming of this study is the limited duration of the intervention. Because this intervention was restricted to only one learning task (i.e. stoma care), students did not get the opportunity to practice extensively on their skill development. This was most visible for the complex cognitive skill of self-assessment. According to van Merriënboer and Kirschner (2007), more training is needed to develop this kind of skill. Furthermore, only a small sample was used in the study. The question remains if the results are transferable to larger groups of students or students in other domains. Nevertheless, the fact that this intervention yielded some important results concerning mental effort expenditure during learning and test task performance is a sound basis for further research on this topic.

The findings yield the clear guideline that novice students should be provided with performance-based assessment criteria in order to improve their learning process, and reach higher test task performance. For instructing young nurses in the beginning of their study, performance-based assessment criteria are a necessity to guide their learning process. It should be noted, however, that formulating such performance-based criteria is a demanding task. To assure a sound implementation, training should be provided to teachers to increase their skills in formulating performance-based assessment criteria, based on a systematic process of drawing up a skills hierarchy with related criteria. When students progress in the study program, explicit attention should be paid to training students to interpret their own behaviours in terms of the underlying competences. In this way, students learn to see the link between performance and competence development. If this is not explicitly in the program, students remain on a lower level of thinking.

To conclude, the introduction of competence-based education primarily consisting of authentic learning tasks based on real-life problems, leads educators to solve the issue of how to redesign their assessment programs. Our results show that stating that competence-based assessment criteria are the answer to this problem is a step too far. Whereas competences seem to be a good starting point to develop professional education, they do not always serve this purpose for assessment. At least for novice students, providing them with performance-based assessment criteria is more beneficial than providing them with competence-based criteria. This study shows that novice students need less mental effort to assess their task performance and show higher test task performance, that is, they learn more efficiently when being provided with performance-based assessment criteria.

Acknowledgments We would like to thank the participants in this study and the staff of ROC A12 for all their help in conducting this research. Participants were offered confidentiality. Ethical approval was not necessary as this study was part of the normal education program.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Albanese, M. A., Mejjicano, G., & Anderson, W. M. (2008, March 15). Building a competency-based curriculum: The agony and the ecstasy. *Advances in Health Sciences Education*. Retrieved 26 May 2009, from <http://www.springerlink.com/content/g1w11258v173u6r6/fulltext.pdf>.
- Biemans, H., Nieuwenhuis, L., Poell, R., Mulder, M., & Wesseling, R. (2004). Competence-based VET in the Netherlands: Background and pitfalls. *Journal of Vocational Education and Training*, 36(4), 523–538.
- Deci, E. L., Eghrari, H., Patrick, B. C., & Leone, D. (1994). Facilitating internalization: The self-determination theory perspective. *Journal of Personality*, 62, 119–142.

- Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological science in the public interest*, 5(3), 69–106.
- Ecclestone, K. (2001). 'I know a 2:1 when I see it': Understanding criteria for degree classifications in franchised university programmes. *Journal of Further and Higher Education*, 25(3), 301–313.
- Eva, K. W., & Regehr, G. (2005). Self-assessment in the health professions: A reformulation and research agenda. *Academic Medicine*, 80(10), 46–54.
- Fastré, G. M. J., Van der Klink, M., van Merriënboer, J. J. G. & Sluijsmans, D. (2009). Drawing students' attention to relevant assessment criteria: Effects on self-assessment skills and performance. Manuscript submitted for publication.
- Grégoire, J. (1997). Diagnostic assessment of learning disabilities. From assessment of performance to assessment of competence. *European Journal of Psychological Assessment*, 13(1), 10–20.
- Gulikers, J. T. M., Kester, L., Kirschner, P. A., & Bastiaens, Th. J. (2008). The influence of practical experience on perceptions, study approach and learning outcomes in authentic assessment. *Learning and Instruction*, 18(2), 172–186.
- Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review*, 19, 509–539.
- Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, 38(1), 23–31.
- Kicken, W., Brand-Gruwel, S., & van Merriënboer, J. J. G. (2006). *Self-directed learning skills questionnaire*. Heerlen: Open Universiteit Nederland.
- Mager, R. F. (1984). *Preparing instructional objectives* (2nd ed.). Belmont: David S. Lake.
- Martens, R. L., & Kirschner, P. A. (2004). What predicts intrinsic motivation? Paper presented at the 2004 AECT convention, Chicago.
- Miller, G. E. (1990). The assessment of clinical skills/competence/performance. *Academic Medicine*, 65(Suppl.), S63–S67.
- Miller, P. J. (2003). The effect of scoring criteria specificity on peer and self-assessment. *Assessment and Evaluation in Higher Education*, 28(4), 383–394.
- Otter, S. (1995). Assessing competence: The experience of the enterprise in higher education initiative. In A. Edwards & P. Knight (Eds.), *Assessing competence in higher education*. London: Kogan Page.
- Paas, F. (1992). Training strategies for attaining transfer of problem-solving skills in statistics: A cognitive-load approach. *Journal of Educational Psychology*, 84, 429–434.
- Paas, F., & Van Merriënboer, J. J. G. (1993). The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human Factors*, 35, 737–743.
- Sadler, D. R. (1985). The origins and functions of evaluative criteria. *Educational Theory*, 35(3), 285–297.
- Scheffer, S., Muehlinghaus, I., Froehmel, A., & Ortwein, H. (2008). Assessing students' communication skills: Validation of a global rating. *Advances in Health Sciences Education*, 13, 583–592.
- Spencer, L., & Spencer, S. (1993). *Competence at work*. (New York: John Wiley&Sons, Inc).
- Steinaker, N. W., & Bell, M. R. (1979). *The experiential taxonomy: A new approach to teaching and learning*. New York: Academic Press.
- Sweller, J., van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10, 251–294.
- Van Gog, T., & Paas, F. (2008). Instructional efficiency: Revisiting the original construct in educational research. *Educational Psychologist*, 43, 16–26.
- Van Merriënboer, J. J. G. (1997). *Training complex cognitive skills*. Englewood Cliffs: Educational Technology Publications.
- Van Merriënboer, J. J. G., & Kirschner, P. A. (2007). *Ten steps to complex learning*. Mahwah: Erlbaum/Taylor and Francis.
- Van Merriënboer, J. J. G., Kirschner, P. A., Paas, F., Sloep, P. B., & Caniëls, M. C. J. (2009). Towards an integrated approach for research on lifelong learning. *Educational Technology*, 49(3), 3–14.
- Van Merriënboer, J. J. G., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review*, 17, 147–177.
- Van Merriënboer, J. J. G., Van der Klink, M. R., & Hendriks, M. (2002). *Competenties: Van complicaties tot compromis [Competencies: From complications to compromise]*. The Hague: Educational council.
- Wesselink, R., Biemans, H. J. A., Mulder, M., & van den Elsen, E. R. (2007). Competence-based VET as seen by Dutch researchers. *European Journal of Vocational Training*, 40, 38–50.
- Wierstra, R. F. A., Kanselaar, G., van der Linden, J. L., & Lodewijks, H. G. L. C. (1999). Learning environment perceptions of European university students. *Learning Environments Research*, 2, 79–98.