OpenAccess

Students' assessment preferences, perceptions of assessment and their relationships to study results

Gerard van de Watering · David Gijbels · Filip Dochy · Janine van der Rijt

Abstract The purposes of this study are to gain more insight into students' actual preferences and perceptions of assessment, into the effects of these on their performances when different assessment formats are used, and into the different cognitive process levels assessed. Data were obtained from two sources. The first was the scores on the assessment of learning outcomes, consisting of open ended and multiple choice questions measuring the students' abilities to recall information, to understand concepts and principles, and to apply knowledge in new situations. The second was the adapted Assessment Preferences Inventory (API) which measured students' preferences as a pre-test and perceptions as a post-test. Results show that, when participating in a New Learning Environment (NLE), students prefer traditional written assessment and questions which are as closed as possible, assessing a mix of cognitive processes. Some relationships, but not all the expected ones, were found between students' preferences and their assessment scores. No relationships were found between students' perceptions of assessment and their assessment scores. Additionally, only forty percent of the students had perceptions of the levels of the cognitive processes assessed that matched those measured by the assessments. Several explanations are discussed.

G. van de Watering (✉)
Faculty of Technology Management, Eindhoven University of Technology,
Den Dolech 2, De Lismortel: Pav. A.43, Eindhoven 5600 MB, The Netherlands
e-mail: g.a.v.d.watering@tm.tue.nl

D. Gijbels
Institute for Education and Information Sciences, University of Antwerp,
Gratiekapelstraat 10, 2000 Antwerpen, Belgium

F. Dochy
Centrum voor Opleidingsdidactiek, Katholieke Universiteit Leuven,
Dekenstraat 2, 3000 Leuven, Belgium

J. van der Rijt
Department of Educational Research and Development, University of Maastricht,
Tongersestraat 53, 6211, LM, Maastricht, The Netherlands

Introduction

Assessment is an umbrella term. Understanding of it varies, depending on how one sees the role of the assessment itself in the educational process, as well as the role of the partici-pants (the assessors and the assessees) in the education and assessment processes. The main difference is described in terms of an 'assessment culture' and a 'testing culture' (Birenbaum 1994, 1996, 2000). The traditional testing culture is heavily influenced by old paradigms, such as the behaviourist learning theory, the belief in objective and standard-ized testing (Shepard 2000), and testing being separated from instruction. Multiple choice and open ended assessments are typical test formats of a testing culture. In the last few decades, developments in society (National Research Council 2001) and a shift towards a constructivist learning paradigm, combined with the implementation of new learning environments (NLEs), have changed the role of assessment in education. NLEs claim to have the potential to improve the educational outcomes for students in higher education which are necessary to function successfully in today's society (Simons, van der Linden and Duffy 2000). New learning environments are rooted in constructivist theory and intend to develop an educational setting to meet the challenge for today's higher education, making the students' learning the core issue and defining instruction as enhancing the learning process.

The most fundamental change in the view of assessment is represented by the notion of 'assessment as a tool for learning' (Dochy and McDowell 1997). In the past, assessment was primarily seen as a means to determine grades; to find out to what extent students had reached the intended objectives. Today, there is a realisation that the potential benefits of assessing are much wider and impinge on all stages of the learning process. Therefore, the new assessment culture strongly emphasises the integration of instruction and assessment, in order to align learning and instruction more with assessment (Segers et al. 2003). The integration of assessment, learning and instruction, however, remains a challenge for most teachers (Struyf et al. 2001). In the UK, Glasner (1999) concludes that a number of factors, like the massification of higher education, the declining levels of resources, and concerns about the ability to inhibit plagiarism, are responsible for the persistence of traditional methods of assessment and the absence of widespread innovation. A recent report on final exams in secondary education in the Netherlands indicated that most of the exams con-sisted primarily of multiple choice questions and open ended or essay questions, in spite of the effort that had been put into the implementation of new teaching and assessment methods (Kuhlemeier et al. 2004). The situation in which NLEs are accompanied by traditional assessment methods are still very common. However theoretical underpinned, empirical research on this combination is rather scarce.

It is generally acknowledged that assessment plays a crucial role in the learning process and, accordingly, on the impact of new teaching methods (Brown et al. 1994; Gibbs 1999; Scouller 1998). The way students prepare themselves for an assessment depends on how they perceive the assessment (before, during and after the assessment), and these effects can have either positive or negative influences on learning (Boud 1990; Gielen et al. 2003; Nevo 1995). There also can be a discrepancy between what is actually asked, what students prefer and what students expect to be asked (Broekkamp et al. 2004). NLEs have been developed in which schools have a balance between a test culture and an assessment culture. The effects of such environments, however, do not always demonstrate the

expected outcomes (Segers 1996). Research results show that educational change only becomes effective if the students' perceptions are also changed accordingly (Lawness and Richardson 2002; Segers and Dochy 2001).

As mentioned before, NLEs are not always accompanied by new methods of assessment. In this article, we want to explore in such a situation which assessment formats are preferred, how students perceive rather traditional assessment formats, and what relationships exist between students' preferences, perceptions and their assessment results. Before presenting the results, we will first describe some research into students' assessment preferences and perceptions of assessment.

Assessment preferences

In our study, assessment preference is defined as imagined choice between alternatives in assessment and the possibility of the rank ordering of these alternatives. Several studies have investigate such assessment preferences earlier. According to the studies of Ben-Chaim and Zoller (1997), Birenbaum and Feldman (1998), Traub and McRury (1990) and Zeidner (1987) students, especially the males (Beller and Gafni 2000), generally prefer multiple choice formats, or simple and de-contextualised questions, over essay type assessments or constructed-response types of questions (complex and authentic).

Traub and McRury (1990), for example, report that students have more positive attitudes towards multiple choice tests in comparison to free response tests because they think that these tests are easier to prepare for, easier to take, and thus will bring in relatively higher scores. In the study by Ben-Chaim and Zoller (1997), the examination format preferences of secondary school students were assessed by a questionnaire (Type of Preferred Examinations questionnaire) and structured interviews. Their findings suggest that students prefer written, unlimited time examinations and those in which the use of supporting material is permitted. Time limits are seen to be stressful and to result in agitation and pressure. Assessment formats which reduce stress will according to these authors increase the chance of success and students vastly prefer examinations which emphasize understanding rather than rote learning. This might well be explained by the fact that students ofter perceive exams that emphasize understanding as essay type or open question exams.

Birenbaum (1994) introduced a questionnaire to determine students' assessment preferences (Assessment Preference Inventory) for various facets of assessment. This questionnaire was designed to measure three areas of assessment. The first is assessment-form related dimensions such as assessment type, item format/task type and pre assessment preparation. The second was examinee-related dimensions such as cognitive processes, students' role/responsibilities and conative aspects. The final area was a grading and reporting dimension. Using the questionnaire, Birenbaum (1997) found that differences in assessment preferences correlated with differences in learning strategies. Moreover, Birenbaum and Feldman (1998) discovered that students with a deep study approach tended to prefer essay type questions, while students with a surface study approach tended to prefer multiple choice formats. This has been recently confirmed by Baeten et al. (2008). In the Birenbaum and Feldman study, questionnaires about attitudes towards multiple choice and essay questions, about learning related characteristics, and measuring test anxiety, were administered to university students. Test anxiety seems to be another variable that can lead to specific attitudes towards assessment formats: students with high test anxiety have more favourable attitudes towards multiple choice questions whilst those with low test anxiety tend to prefer open ended formats. Clearly, students with a high level of

test anxiety strive towards more certainty within the assessment situation. Birenbaum and Feldman assumed that if students are provided with the type of assessment format they prefer, they will be motivated to perform at their best.

Scouller (1998) investigated the relationships between students' learning approaches, preferences, perceptions and performance outcomes in two assessment contexts: a multiple choice question examination requiring knowledge across the whole course and assignment essays requiring in-depth study of a limited area of knowledge. The results indicated that if students prefer essays this is more likely to result in positive outcomes in their essays than if they prefer multiple choice question examinations. Finally, Beller and Gafni (2000) gave an overview of several studies which analysed the students' preferences for assessment formats, their scores on the different formats, and the influence of gender differences. In a range of studies they found some consistent conclusions suggesting that, if gender differences are found (which was not always the case), female students prefer essay formats, and male students show a slight preference for multiple choice formats (e.g. Gellman and Berkowitz 1993). Furthermore, male students score better on multiple choice questions than female students and female students score better than male students on open ended questions than on multiple choice questions as could be expected (e.g. Ben-Shakhar and Sinai 1991).

Overall, from the studies regarding students' assessment preferences, it seems that students prefer assessment formats which reduce stress and anxiety. It is assumed, despite the fact that there are no studies that directly analyze the preferences of students and their scores on different item or assessment formats, that students will perform better on their preferred assessment formats. Students with a deep study approach tend to prefer more the essay type of questions, as do female students.

Perceptions of assessment

Perception of assessment is defined as the students' act of perceiving the assessment in the course under investigation (Van de Watering et al. 2006). Recently, interesting studies have investigated the role of perceptions of assessment in learning processes. For example, Scouller and Prosser (1994) investigated students' perceptions of a multiple choice question examination, consisting mostly of reproduction-oriented questions, to investigate the students' abilities to recall information, their general orientation towards their studies and their study strategies. The students' perceptions do not always seem to be correct: on one hand they found that some students wrongly perceived the examination to be assessing higher order thinking skills. As a consequence, these students used deep study strategies to learn for their examination. On the other hand, the researchers concluded that students with a surface orientation may have an incorrect perception of the concept of understanding, cannot make a proper distinction between understanding and reproduction, and therefore have an incorrect perception of what is being assessed. In our research, this certainly accounts for the younger, inexperienced students (Baeten et al. 2008). Nevertheless, in this study no correlations were found between perceptions of the multiple choice questions and the resulting grades. In the earlier mentioned study by Scouller (1998), relationships were found between students' preferences, perceptions and performance outcomes. Students who prefer multiple choice question examinations perceive these assessments (actually assessing lower levels of cognitive processing) to be more likely to assess higher levels of cognitive processing than students who prefer essays. Poorer performance, either on the multiple choice questions or on the essays, was related to the use of an unsuitable study

approach due to an incorrect perception of the assessment. Better performance on the essays (actually assessing higher levels of cognitive processing) was positively related to a perception of essays as assessing higher levels of cognitive processing and to the use of a suitable study approach (i.e. deep approach).

In the above mentioned studies, the multiple choice examinations intentionally assessed the lower levels, and the assignment essays the higher levels of cognitive processing. It is certainly not always evident that students perceive assessments in the ways that were intended by the staff. This general feeling that is experienced by many teachers, can also be underpinned by research results: MacLellan (2001), for example, used a questionnaire asking students and teaching staff about the purposes of their assessment, the nature and difficulty level of the tasks which were assessed, the timing of the assessment and the procedures for marking and reporting. The results showed that there are differences in perceptions between students and staff of the use and purposes of the assessment and the cognitive level measured by the assessment. For example, the students perceived the reproduction of knowledge to be more frequently assessed, and the application, analysis, synthesis and evaluation of knowledge less frequently assessed, than the staff believed.

Empirical study

In the present study we investigate the students' actual assessment preferences in a NLE and their perceptions of the traditional assessment within the NLE aimed to measure both low and high cognitive process levels. In order to gain more insight into the effect of these on students' performances, four research questions were formulated:

1. Which assessment preferences do students have in a NLE? In more detail: (a) which assessment type is preferred; and (b) assessments of which cognitive processes are preferred?
2. How did students actually perceive the 'traditional' assessment in the NLE (i.e. the cognitive processes assessed in the traditional assessment)?
3. In what ways are students' assessment preferences related to their assessment results? In more detail, what were: (a) the relationships between assessment type preferences and the scores on the assessment format; and (b) the relationships between cognitive process preferences and the scores on the different cognitive levels that were assessed?
4. In what way are students' perceptions of assessment related to their assessment results?

The focus in this study is on end point summative assessment. Other forms of formative assessment were not subject of our investigation. The NLE in this study is highly consistent with learning environments described under the label of problem based learning (Gijbels et al. 2005), as will be outlined below.

Method

Subjects

A total of 210 students, in the first year at a Dutch university, participated in the inventory of assessment preferences (pre-test). 392 students underwent the assessment of learning outcomes and 163 students participated in the inventory of perceptions of the questions after the assessment (post-test). In total, 83 students participated on all three occasions.

Learning environment

The current study took place in a NLE and was structured as follows. For a period of 7 weeks, students worked on a specific course theme (i.e. 'legal acts'). During these 7 weeks the students worked twice a week for 2 h in small groups (maximum 19 students) on different tasks, guided by a tutor. In conjunction with these tutorial groups, they were enrolled in somewhat bigger practical classes (38 students) for 2 h a week and another 2 h a week in large class lectures. Assessment took place immediately after the course, by means of a written exam (a combination of multiple-choice questions and essay questions).

Procedures and instruments

Assessment preferences (pre-test) were measured by means of the Assessment Preferences Inventory (API) (Birenbaum 1994). This was originally a 67 item Likert-type questionnaire designed to measure seven dimensions of assessment. For our purposes we selected only two dimensions of the questionnaire using a 5 point Likert-scale (from 1 = not at all to 5 = to a great extent) and translated these items. Minor adjustments were also made to fit the translated questionnaire into the current educational and assessment environment. Two dimensions were relevant in order to answer the research questions. The used dimensions are: Assessment types (12 questions about students' preferences for different modes of oral, written and alternative tests; see Table 1) and Cognitive processes (15 questions about the preferences for assessing the cognitive processes remembering, understanding, applying, and analysing, evaluating and creating; see Table 2). Students were asked to complete the API questionnaire during one of the tutorial sessions near the end of a first

Table 1 Descriptives of students' preferences of assessment type

Rank	Assessment type	Mean	SD
To what extent would you want your achievements in the course be assessed by each of the following methods?			
	Oral tests	*2.47*	*.92*
12	Individual oral tests, without supporting material (notes, books).	2.10	1.13
11	Individual oral tests wherein the questions are given half an hour prior the test, and answers can be prepared without supporting materials.	2.26	1.17
7	Individual oral tests wherein the questions are given half an hour prior the test, and answers can be prepared with supporting materials.	3.01	1.30
10	Oral tests, in the form of a group discussion where the instructor observes and assesses the contribution of each of the participants.	2.48	1.36
	Written tests	*3.42*	*.73*
5	Written tests without supporting materials.	3.39	1.15
6	Written tests, without a time limit and without supporting material.	3.16	1.24
1	Written test, with supporting materials.	3.73	1.08
3	Written tests without a time limit, with supporting materials.	3.47	1.31
4	Take-home exams.	3.41	1.32
	Alternative tests	*3.20*	*1.09*
2	Papers/projects.	3.50	1.17
8	Portfolio (your collected work, finished and in progress).	2.93	1.25
9	Computerised tests.	2.72	1.22

Table 2 Descriptives of students' preferences of cognitive processes to be measured in assessment

Rank	Cognitive process	Mean	SD
	To what extent would you want your achievements in the course be assessed by each of the following methods?		
	Remember	*3.72*	*.77*
1	Knowledge questions related to the reading of assignments.	3.84	.916
4	Questions making an appeal to the reproduction of facts.	3.60	.924
	Understand (exemplifying, comparing, inferring)	*3.33*	*.61*
2	Comprehension questions related to the material taught by the instructor.	3.76	.886
4	Questions that require the drawing of conclusions.	3.60	.834
12	Questions that require the providing of examples.	2.99	.956
13	Questions that require comparing different concepts/ideas.	2.93	.904
	Apply (implementing), problem solving	*3.61*	*.72*
3	Questions that require problem solving.	3.70	.847
8	Questions requiring the application of material learnt during the course to new situations.	3.53	.939
	Analyse (organizing), evaluate (critiquing, checking), create (generate)	*3.10*	*.62*
4	Questions that require a personal explanation or opinion.	3.60	1.027
7	Questions that require critical thinking.	3.54	1.009
9	Questions that require analysis and interpretation.	3.29	.871
10	Questions that require an overall view of the relationships between all topics learnt.	3.14	.973
10	Questions that require creativity and imagination.	3.14	1.042
14	Questions that require scientific investigation.	2.70	1.013
15	Questions in which you are asked to evaluate others' solutions or opinions.	2.27	1.001

year law course. Our translation of the questionnaire resulted in an acceptable Cronbach's alpha value for the whole questionnaire (*Cronbach's alpha* = .76). The reliability of the first part of the questionnaire (Assessment type) is moderate (*Cronbach's alpha* = .61) and that of second part (Cognitive process) is good (*Cronbach's alpha* = .82).

The summative traditional assessment was a combination of 6 open ended questions, requiring a variety of short and long answers, and 40 multiple choice questions (assessing learning outcomes). The objectives of the assessment were threefold and derived from Bloom's taxonomy (Bloom 1956; Anderson and Krathwohl 2001). The first objective was to investigate the student's ability to recall information (in terms of Bloom, knowledge; defined as the remembering of previous learned material). The second was to investigate understanding of basic concepts and principles (comprehension; defined as the ability to grasp the meaning of material). The third was to examine the application of information, concepts, and principles, in new situations (application; refers to the ability to use learned material in new and concrete situations). Cognitive processes such as analysing, evaluating and creating were also put into this last category, so the emphasis of this category is to investigate if students are able to use all their knowledge in concrete situations to solve an underlying problem in the presented cases or problem scenarios. About 25% of the assessment consisted of reproduction/knowledge based questions (37.5% of all multiple choice questions), 20% of comprehension based questions (30% of all multiple choice questions), and 55% of application based questions (32.5% of all multiple choice questions and all open ended questions).

The item construction and the assessment composition was carried out by the so-called assessment construction group. This group, consisting of four expert teachers on the subject and one assessment expert, worked according to the faculty assessment construction procedures, using several tools such as a planning meeting, a specification table, a number of item construction meetings, and a pool of reviewers, to compose a valid and reliable assessment. During the item construction meetings, different aspects of the items were discussed: the purpose of the question; the construction of the question (in case of a multiple choice question the construction of the stem, the construction and usefulness of the correct choice and the distracters); and the objectivity of the right answer. A specification table was used to ensure the assessment was a sound reflection of the content (subject domain) and the cognitive level (cognitive process dimension) of the course. The difficulty of the items was also discussed. Classifications of the items were made during the meetings. After composing the assessment, it was send to four reviewers (in this case two tutors and two professors who were all responsible for some teaching in the course). The reviewers judged the assessment in terms of the usefulness of the questions and the difficulty of the assessment as a whole. The assessment was validated by the course supervisor. On average, students scored 34.8 out of 60 on the total assessment ($SD = 9.75$); the average score on the open ended part was 12 out of 20 ($SD = 4.40$); and the average score on the multiple choice part was 22.8 out of 40 ($SD = 6.16$). The internal consistency reliability coefficient of the multiple choice part was also measured and found to be appropriate (*Cronbach's alpha* = .80). For these students it was the fifth course in the curriculum and the fourth assessment of this kind.

To measure the students' perceptions of the assessment, the 15 questions from the dimension Cognitive process of the API were used (post-test). Basically, this questionnaire asked students which cognitive processes (remembering, understanding, applying, and analysing, evaluating and creating) they thought were assessed by the combination of open ended and multiple choice questions they had just taken. Students were asked to complete the questionnaire directly after finishing the assessment. This questionnaire also has an acceptable internal consistency reliability (*Cronbach's alpha* = .79).

Analysis

Results were analysed by means of descriptive statistics for the measures used in the study and multivariate analysis of variances (MANOVAs) were conducted to probe into the relationships between students' assessment preferences, perceptions and their study results (Green and Salkind 2003). The research questions will be reported on one by one.

Results

Which assessment preferences do students have?

For the first research question, students were firstly asked about their preferences for assessment types and item format/task types in the current NLE. Students preferred written tests, including take-home exams and papers, in which they are allowed to use supporting materials such as notes and books, as well as papers or projects. Oral tests and the other modes of alternative assessment mentioned in the questionnaire, i.e. computerised tests and portfolios, are not amongst the students' preferences (see Table 1).

Secondly, the students were asked about their preferences in relation to the cognitive processes which were to be assessed. According to the students' preferences, a mix of

cognitive processes should be assessed, such as (in order of preference): reproducing; comprehending; problem solving; explaining; drawing conclusions; critical thinking; and applying. Evaluating others' solutions or opinions, scientific investigation, providing of examples and comparing different concepts, were not preferred (see Table 2).

How did students perceive the traditional assessment?

Concerning the second research question, how students actually perceived the assessment they took, students considered it to be primarily a measurement consisting of comprehension- and application-based questions that required the drawing of conclusions, problem solving, analysis, interpretation and critical thinking (see Table 3). The measurement was also considered, secondarily, as a measurement of reproduction based questions.

Additionally, correlations between student preferences (see Table 2) and perceptions of the assessment (see Table 3) turned out not to be significant, suggesting that there is a distinction between students' preferences and their perception of the assessment.

How are students' preferences related to assessment results?

The third research question concerned the relationships between students' assessment preferences and their assessment results. A multivariate analysis of variance (MANOVA) was conducted to evaluate this relationship between the preference for oral assessments

Table 3 Descriptions of students' perceptions of cognitive processes measured by the assessment

Rank	Cognitive process	Mean	SD
	To what extent do you think the exam was consisting of ...		
	Remember	*3.37*	*.68*
6	Knowledge questions related to the reading of assignments.	3.38	.807
8	Questions making an appeal to the reproduction of facts.	3.36	.820
	Understand (exemplifying, comparing, inferring)	*3.26*	*.58*
1	Comprehension questions related to the material taught by the instructor.	3.66	.820
3	Questions that require the drawing of conclusions.	3.59	.795
14	Questions that require the providing of examples.	2.81	.920
10	Questions that require comparing different concepts/ideas.	3.00	.858
	Apply (implementing), problem solving	*3.59*	*.70*
3	Questions that require problem solving.	3.59	.802
2	Questions requiring the application of material learnt during the course to new situations.	3.60	.792
	Analyse (organizing), evaluate (critiquing, checking), create (generate)	*3.00*	*.56*
13	Questions that require a personal explanation or opinion.	2.84	.972
6	Questions that require critical thinking.	3.38	.843
5	Questions that require analysis and interpretation.	3.49	.740
8	Questions that require an overall view of the relationships between all topics learnt.	3.37	.913
11	Questions that require creativity and imagination.	2.97	.936
12	Questions that require scientific investigation.	2.95	.987
15	Questions in which you are asked to evaluate others' solutions or opinions.	2.23	1.038

Table 4 Means and standard deviations on the dependent variables for the three levels of preference for written assessment	Preference level	Multiple choice questions		Open ended questions	
		Mean	SD	Mean	SD
	Not preferring (N = 17)	25.72	6.69	14.58	4.93
	Neutral (N = 116)	25.57	4.94	14.12	4.15
	Preferring (N = 77)	24.01	5.04	12.28	3.93

and the two assessment scores, between the preference for written assessments and the assessment scores, and the preference for alternative assessments and the assessment scores. The independent variable, the preference for the assessment type (oral, written and alternative), included three levels: students who prefer that assessment type, students who are neutral about that assessment type, and students who do not prefer that assessment type. The two dependent variables were the scores in the multiple choice questions and the scores in the open ended questions. Results only showed significant differences among the three levels of preferences for written assessments on the assessment scores, Wilks's $\Lambda = .95$, $F(4, 414) = 2,614$, $p < .05$, though the multivariate effect size η^2 based on Wilks's Λ was low, at .03, suggesting the relationship between the preferences and the assessment scores are weak. Table 4 contains the means and the standard deviations on the dependent variables for the three levels.

Analysis of variances (ANOVA) on each dependent variable were conducted as follow-up tests to the MANOVA. Using the Bonferroni method, each ANOVA was tested at the .025 level. The ANOVA on the open ended questions scores was significant, $F(2, 208) = 5.25$, $p < .01$, $\eta^2 = .05$, while the ANOVA on the multiple choice questions scores was not significant, $F(2, 208) = 2.31$, $p = .10$, $\eta^2 = .02$.

Post hoc analyses to the univariate ANOVA for the open ended questions scores consisted of conducting pairwise comparisons, each tested at the .008 level (.025 divided by 3), to find which level of preference for written assessments influences the outcome of the open ended questions most strongly. The students who preferred written assessments obtained lower marks on this part of the assessment when compared with those students who are neutral towards them.

Secondly, MANOVAs were conducted to evaluate the relationships between the students' preferences for cognitive processes measured by the assessment (the independent variables: remembering, understanding, applying, and analysing, evaluating and creating), and the scores on the different cognitive levels actually measured by the assessment of learning outcomes (the dependent variables: total scores on the reproduction based questions, comprehension based questions and application based questions). For this purpose, students were also divided into three groups for each cognitive process: students who prefer that cognitive process, students who are neutral to that cognitive process, and students who do not prefer that cognitive process. No significant differences were revealed by means of the MANOVAs, suggesting that no relationship exists between preferences for cognitive processes and the actual outcomes on the different cognitive levels measured with a combination of multiple choice and open ended questions.

How are students' perceptions and assessment results related?

In order to answer this question, it was assumed that students with a matching perception of the level of the cognitive processes measured by the assessment of learning outcomes

(being a clear correspondence between the perceived levels of cognitive processes and the intended levels of cognitive processes by the assessment construction group), will have better results than students with a misperception of the cognitive processes (being a mismatch between the perceived level of cognitive processes and the intended levels of cognitive processes by the assessment construction group). To investigate this assumption, students were divided into three groups: a matching group, made up of students who perceived the assessment more as applying than remembering ($N = 65$; 40%); a mismatching group, consisting of students who perceived the assessment more as remembering than applying ($N = 36$; 22%); and a second mismatching group, of students who perceived the assessment as equally remembering and applying ($N = 62$; 38%). The dependent variable was the outcome on the total assessment.

Though MANOVAs indicated that students with a matching perception scored slightly better on the assessment of outcomes compared to students with a misperception, the differences were marginal and not significant. So students with a misconception of the level of cognitive processes assessed (i.e. those who perceived the assessment as being more remembering than applying or equally remembering and applying), did not perform significantly worse than students with a matching perception (i.e. those who correctly perceived the assessment to be more applying than remembering).

Discussion

This study was designed to examine which assessment formats students prefer in a NLE, and how students perceive more traditional assessment formats in the NLE used. In addition, the relationships between students' assessment preferences, perceptions of assessment and their assessment results were examined.

With regard to students' assessment preferences we can conclude the following. Students who were participating in the described NLE preferred traditional written assessment, as well as alternative assessment such as papers or projects. According to the students, the use of supporting material should be allowed and the questions or tasks should assess a mix of cognitive processes. The preference for assessment formats with the use of supporting material is in line with the studies of Ben-Chaim and Zoller (1997) and Traub and McRury (1990) in which students prefer easy-to-take and stress reducing assessment formats. Additionally, papers and projects can be considered as assessment formats in which, generally speaking, support (by means of materials and fellow students) is allowed. This could be one of the reasons why students prefer these assessment formats. Despite the fact that oral assessments, group discussions and peer evaluation play a more important role in a NLE, these formats or modes are not preferred by our students. This is possibly also because these assessment formats are yet not very common in the curriculum of the described NLE. This would implicate that students should have the opportunity to show their competence on different assessment methods in order to build clear assessment preferences.

Our findings regarding the relationship between assessment type preferences and the resulting scores on the assessment formats showed some significant differences. Strangely, students who preferred written assessments obtained lower marks on both parts of the assessment. Written assessments, especially the multiple choice format, are often preferred because students think they reduce stress and test anxiety and are easy to prepare for and to take (Traub and McRury 1990). So it is possible that some students prefer written assessment formats because they are used to it, but not because they are good at them. No

significant relationships were found between students' preferences for the cognitive processes measured by the assessment and the scores on the different cognitive levels actually measured by the assessment of learning outcomes. As mentioned before, according to Birenbaum and Feldman (1998), students will be motivated to perform at their best if they are provided with the assessment format they prefer. Our outcomes do not support that finding. Preferences assume a choice. As in most cases, however, the students who were studied could not choose between assessment formats. They had to take the exam as it was presented. It is thus possible that their preferences only reflect what they think is a suitable assessment format to measure their abilities and to give results which are fair enough.

With regard to the students' perceptions of assessment we can draw the following conclusions. The 'traditional' assessment, a combination of multiple choice and essay questions in the NLE, was generally, as intended by the item constructors and the assessment composers, perceived more as assessing the application of knowledge, problem solving, the drawing of conclusions, and analysing and interpreting, than assessing the reproduction of knowledge. Despite this, there was a clear correspondence between the intended level of cognitive processes and the perceived level of processes in the assessment in only 40% of the cases. In 38% of the cases, students made no distinction between a more reproduction based and a more application based assessment, and 22% of the students perceived the assessment to be more reproduction based. These figures show that it is possible for students to have a clear picture of the demands of assessments. On the other hand, the presence of multiple choice questions in the assessment (and the apparently persistent perception of these questions created by previous experiences, that this format mainly assesses the reproduction of knowledge), could have caused the overestimating of the reproduction of knowledge. The standard interviews with students about the course and the assessment revealed some insights into how they perceived the assessment. For some students it was unimaginable to use cases or problem scenarios in a multiple choice assessment. So, when preparing for the assessment, they did not prepare for applying knowledge, meaning that they did not practice their problem solving skills. Students also seem to identify certain questions as being reproduction based, because the questions or the cases used in the assessment resembled the tasks or cases used in the preceding educational programme. These results imply that a lot of students need help in building up a matching perception of what is assessed by means of the assessment formats that are used. Just giving help using examples of assessment items and discussing their answers seems not to be enough for these students. The purpose of the assessment questions and the cognitive processes to be used for answering the question correctly must also made clear and preferably be practiced. This is especially true in cases where a multiple choice format is used to measure cognitive processes, rather than reproducing knowledge.

It has been argued by Scouller (1998) that a mismatch in the perception of assessment leads to poorer assessment results. The outcomes of that study were found to be strongly associated with students' general orientations towards study. Students with an orientation towards deep learning approaches will continue to rely on deep strategies when preparing for their examinations. As discussed by Scouller and Prosser (1994), students' perceptions, based on previous experiences with multiple choice questions, lead to a strong association between multiple choice examinations and the employment of surface learning approaches, leading to successful outcomes. In both studies (as in other studies of this kind), assessment formats were used in which the multiple choice assessment was reproduction based and the essay assessment was more application based. In contrast to the Scouller and Scouller and Prosser study, we used an assessment format containing multiple choice questions and essay questions with a heavy emphasis on application, using problem solving in both parts

of the assessment in a NLE. We did not find any direct relationships between students' perceptions of assessment and their assessment results. It is possible that students' approaches to learning moderate the relationships between students' perceptions and their assessment results. From previous research, we know that a substantial proportion of the student population in the described NLE do not use appropriate approaches to learning (Gijbels et al. 2005). We also now know from our study that a lot of students have a mismatching perception of the cognitive processes measured by the assessment. The relationships between study approaches and perceptions of the assessments used in this study should be further explored, in combination with the findings of Birenbaum and Feldman (1998) and Scouller (1998) about the relationships between study approaches and perceptions of assessment.

Certainly statements about internal validity must be interpreted cautiously as a result of the fact that only 39.9% of the students following the course completed all three research instruments. In view of the specific context of the institution in which the study was conducted and the limited range of subject matters studied, caution must also be exercised where external validity is concerned. Nevertheless we feel that some interesting conclusions can be drawn from this study. It shows that the use of traditional assessments in a NLE, even though a lot of attention is paid to the validity of the traditional assessment, remains problematic.

References

Anderson, L., & Krathwohl, D. (2001). *A taxonomy for learning, teaching and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.

Baeten, M., Struyven, K., & Dochy, F. (2008). *Students' assessment preferences and approaches to learning in new learning environments: A replica study*. Paper to be presented at the annual conference of the American Educational Research Association, March 2008, New York.

Beller, M., & Gafni, N. (2000). Can item format (multiple choice vs. open-ended) account for gender differences in mathematics achievement? *Sex Roles: A Journal of Research, 42*, 1–21.

Ben-Chaim, D., & Zoller, U. (1997). Examination-type preferences of secondary school students and their teachers in the science disciplines. *Instructional Science, 25*(5), 347–367.

Ben-Shakhar, G., & Sinai, Y. (1991). Gender differences in multiple-choice tests: The role of differential guessing. *Journal of Educational Measurement, 28*, 23–35.

Birenbaum, M. (1994). Toward adaptive assessment—the student's angle. *Studies in Educational Evaluation, 20*, 239–255.

Birenbaum, M. (1996). Assessment 2000: Towards a pluralistic approach to assessment. In M. Birenbaum & F. Dochy (Eds.), *Alternatives in assessment of achievement, learning processes and prior knowledge* (pp. 3–30). Boston: Kluwer Academic.

Birenbaum, M. (1997). Assessment preferences and their relationship to learning strategies and orientations. *Higher Education, 33*, 71–84.

Birenbaum, M. (2000). New insights into learning and teaching and the implications for assessment. *Keynote address at the 2000 conference of the EARLI SIG on assessment and evaluation, September 13*, Maastricht, The Netherlands.

Birenbaum, M., & Feldman, R. A. (1998). Relationships between learning patterns and attitudes towards two assessment formats. *Educational Research, 40*(1), 90–97.

Bloom, B. S. (Ed.). (1956). *Taxonomy of educational objectives, handbook I: The cognitive domain*. New York: McKay.

Boud, D. (1990). Assessment and the promotion of academic values. *Studies in Higher Education, 15*(1), 101–111.

Broekkamp, H., van Hout-Wolters, B. H. A. M., van den Bergh, H., & Rijlaarsdam, G. (2004). 'Teachers' task demands, students' test expectation, and actual test content. *British Journal of Educational Psychology, 74*, 205–220.

Brown, S., Rust, C., & Gibbs, G. (1994). *Strategies for diversifying assessment in higher education*. Oxford: The Oxford Centre for Staff and Learning Development, Oxford Brookes University.

Dochy, F., & McDowell, L. (1997). Assessment as a tool for learning. *Studies in Educational Evaluation, 23*(4), 279–298.

Gellman, E., & Berkowitz, M. (1993). Test-item type: What students prefer and why. *College Student Journal, 27*(1), 17–26.

Gibbs, G. (1999). Using assessment strategically to change the way students learn. In S. Brown & A. Glasner (Eds.), *Assessment matters in higher education: Choosing and using diverse approaches* (pp. 41–53). Buckingham: SRHE and 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 pre-, post-, and true assessment effects. In M. Segers, F. Dochy, & E. Cascallar (Eds.), *Optimising new modes of assessment: In search of qualities and standards* (pp. 37–54). Dordrecht: Kluwer Academic Publishers.

Gijbels, D., Dochy, F., Van den Bossche, P., & Segers, M. (2005). Effects of problem based learning: A meta-analysis from the angle of assessment. *Review of Educational Research, 75*(1), 27–61.

Gijbels, D., van de Watering, G., Dochy, F., & Van den Bossche, P. (2005). 'The relationship between students' approaches to learning and the assessment of learning outcomes. *European Journal of Psychology of Education, XX*(4), 327–341.

Glasner, A. (1999). Innovations in student assessment: A system-wide perspective. In S. Brown & A. Glasner (Eds.), *Assessment matters in higher education* (pp. 14–27). Buckingham: SRHE and Open University Press.

Kuhlemeier, H., de Jonge, A., & Kremers, E. (2004). *Flexibilisering van centrale examens*. Cito: Arnhem.

Lawness, C. J., & Richardson, J. T. E. (2002). Approaches to studying and perceptions of academic quality in distance education. *Higher Education, 44*, 257–282.

MacLellan, E. (2001). Assessment for learning: The differing perceptions of tutors and students. *Assessment & Evaluation in Higher Education, 26*, 307–318.

National Research Council. (2001). In J. Pelligrino, N. Chudowski, & R. Glaser (Eds.), *Knowing what students know: The science and design of educational assessment*. Committee on the foundation of assessment. Board on Testing and Assessment, Center for Education. Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

Nevo, D. (1995). *School-based evaluation: A dialogue for school improvement*. London: Pergamon.

Scouller, K. (1998). 'The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay. *Higher Education, 35*, 453–472.

Scouller, K. M., & Prosser, M. (1994). 'Students' experiences in studying for multiple choice question examinations. *Studies in Higher Education, 19*(3), 267–279.

Segers, M. (1996). Assessment in a problem-based economics curriculum. In M. Birenbaum & F. Dochy (Eds.), *Alternatives in assessment of achievements, learning processes and prior learning* (pp. 201–226). Boston: Kluwer Academic Press.

Segers, M., & Dochy, F. (2001). New assessment forms in problem-based learning: The value-added of the students' perspective. *Studies in Higher Education, 26*(3), 327–343.

Segers, M., Dochy, F., & Cascallar, E. (2003). The era of assessment engineering: Changing perspectives on teaching and learning and the role of new modes of assessment. In M. Segers, F. Dochy, & E. Cascallar (Eds.), *Optimising new modes of assessment: In search of qualities and standards* (pp. 1–12). Dordrecht: Kluwer Academic Publishers.

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher, 29*(7), 4–14.

Simons, R. J., van der Linden, J., & Duffy, T. (2000). New learning: Three ways to learn in a new balance. In R. J. Simons, J. van der Linden, & T. Duffy (Eds.), *New learning* (pp. 1–20). Dordrecht: Kluwer Academic Publishers.

Struyf, E., Vandenberghe, R., & Lens, W. (2001). The evaluation practice of teachers as a learning opportunity for students. *Studies in Educational Evaluation, 27*(3), 215–238.

Traub, R. E., & MacRury, K. (1990). Multiple choice vs. free response in the testing of scholastic achievement. In K. Ingenkamp & R. S. Jager (Eds.), *Tests und Trends 8: Jahrbuch der Pädagogischen Diagnostik* (pp. 128–159). Weinheim und Basel: Beltz.

Van de Watering, G., & Van der Rijt, J. (2006). Teachers' and students' perceptions of assessments: A review and a study into the ability and accuracy of estimating the difficulty levels of assessment intems. *Educational Research Review, 1*(2), 133–147.

Zeidner, M. (1987). Essay versus multiple choice type classroom exams: The students' perspective. *Journal of Educational Research, 80*(6), 352–358.