EUROPEAN JOURNAL OF ENGINEERING EDUCATION

EJEE

SEFI

EUROPEAN JOURNAL OF
ENGINEERING EDUCATION

Volume 29 Number 2 June 2004

Special Issue: ***Assessment of learning Results in Engineering Education***

European Journal of Engineering Education
Vol. 29, No. 2, June 2004, 291–298

Taylor & Francis
Taylor & Francis Group

The accuracy of student grading in first-year engineering courses

NATASCHA VAN HATTUM-JANSSEN[†]*,
JOSÉ AUGUSTO PACHECO[‡] and
ROSA MARIA VASCONCELOS[†]

Assessment has become a powerful tool to change student learning. In a project of the Council of Engineering Courses of the University of Minho, Portugal, students of textile engineering, apparel engineering and industrial electronics increased their participation in every aspect of their assessment process. The traditional exam was changed to three assessment moments with increasing student involvement. The goal of the project was to try to change the learning styles of the students to more profound ones and to look at the grading ability of the students, which means their ability to mark their peers' or their own work. It was supposed that engineering students would have major difficulties in grading themselves and their peers, because they were inexperienced first-year students and because it is not a common practice on the engineering courses of the University of Minho. It was found that students were able to grade their colleagues and themselves and that the level of accuracy depended on their final grade.

1. Introduction

Assessment—or evaluation—of students has a rather big impact on their learning. In the literature, assessment is often mentioned as an instrument to influence student learning (Barnett-Foster and Nagy 1996, Hager and Butler 1996, Baillie and Toohey 1997, Dochy and Moerkerke 1997, Biggs 1999). Assessment of learning is a key to change learning. Dochy and Moerkerke (1997) pointed out that as assessment changes, learning and teaching will change as well. As Hargreaves (1997: 403) indicates:

> Assessment is vitally important to students and exerts a major influence on their approach to learning. Assessment procedures should therefore promote and reward the achievement of desired learning outcomes. Teaching, learning and assessment are inextricably linked.

A good test result is normally a strong incentive for a student not only to learn, but also to learn in a certain way. The strong impact of assessment on learning makes it a powerful instrument for institutions, course directors, individual teachers and others responsible for teaching and learning to improve the quality of education. To use assessment as a tool for student learning, students need to be responsible for their own assessment process. Dochy and Moerkerke (1997) present a continuum with the teacher on one end and the student on the other end of the continuum. In traditional teacher-controlled education, the teacher carries out each assessment task. He determines

[‡] Council of Engineering Courses of the University of Minho, Centre of Educational Quality, University of Minho, 4810-058 Guimarães, Portugal.
[†] Institute of Education and Psychology, University of Minho, Minho, Portugal.
* To whom correspondence should be addressed. e-mail: natascha@eng.uminho.pt

European Journal of Engineering Education
ISSN 0304-3797 print/ISSN 1469-5898 online
http://www.tandf.co.uk/journals
DOI: 10.1080/03043790320001572S9

criteria, evaluates the performances of students and gives feedback. Dochy also explains student-controlled assessment, where students are involved in stating criteria, evaluating their own performance and giving feedback. In between these extremes, there are a number of possibilities. In co-assessment, a more teacher-controlled assessment method, students and teachers co-operate in the assessment process. They discuss assessment criteria and students can evaluate each other, but the teacher takes the final decisions. Students learn to see what it is to evaluate and make judgements. Furthermore, students learn to evaluate themselves and reflect on their own performances. Peer assessment implies less teacher involvement than co-assessment. Students evaluate each other. It is used to encourage deeper learning. Self-assessment is the opposite of traditional assessment. In this case, students are very much involved in the assessment process and they have a central role in their evaluation. The teacher is not the only person responsible for the assessment process any more. Students take part in the development of criteria and the evaluation of their own and their colleagues' performance (Houston and Lazenbatt 1996, McIlveen *et al.* 1997, Sheppard *et al.* 1998, Sluijsmans *et al.* 2001).

Apart from advantages of student involvement in assessment, there are some difficulties as well, especially when introducing peer assessment, self-assessment and co-assessment for the first time. Students may not be used to criticizing themselves and others and may feel inexperienced. They are afraid of getting an unfair judgement from their colleagues. Explicit marking procedures and guidance of students is necessary for them to gain confidence in assessing each other and to assure an adequate and fair assessment. Studies in, for example, civil engineering, describe possible peer assessment methods that help students to be fair and accurate to each other (Rafiq and Fullerton 1996, Lejk *et al.* 1996). A study at the Hong Kong Polytechnic University showed significant differences in the grading of tutors and peer groups (Kwan and Leung 1996), who assessed a simulation training exercise of third-year students. They argue that self-assessment is a skill that becomes more reliable with practice. Other studies indicated, however, that students were able to assess themselves accurately and showed positive correlations between student and teacher grades (McDowell 1995, Longhurst and Norton 1997).

In this study, students are supposed to assess their peers and themselves. They have an increasing responsibility and participation in the assessment process, going from co-assessment and peer assessment at the first and second moment, to self-assessment at the third moment. The main goal of the change in assessment method was to deepen the learning of the student, although this article is not focusing on learning styles of the students, but on the accuracy of grading. At the engineering courses of the University of Minho, there was little or no tradition of student participation in assessment of learning. It was supposed that peer and self-assessment was too complicated for most of the students, who are not very motivated in their course and who did not seem prepared for a large increase of responsibility for their own learning and assessment.

2. Sample

The sample of this study consisted of 96 students on three different courses: 36 first-year students of textile engineering and apparel engineering and 61 first-year industrial electronics students. The students of textile and apparel engineering were considered one group, because they participated in the same subject together: introduction to textile engineering. The subject of the industrial electronics group that was

involved in this project was introduction to industrial electronics. Both subjects are first-year engineering subjects and the only course-specific subject for both courses in the first year.

3. Process

Students faced three evaluation moments, instead of the traditional exam at the end of the semester. Each evaluation moment differed in the degree of student participation. The first moment can be described as co-assessment. The student and the teacher worked together closely, but the students already had an important role to play in their assessment process. They helped to define evaluation criteria and the weight of each criterion. They also graded their colleagues. At the second moment, the role of the teacher became less important. The assessment could be identified as peer assessment and students were responsible for most of the assessment activities. They graded the work of their colleagues. At the third assessment moment, the students assessed their own work. In this self-assessment process, they were responsible for the selection of criteria, weights of criteria, grading, justification of grading and recommendations.

At the first assessment moment, the teacher played an important role in the assessment activities. He/she guided the students through the process of defining assessment criteria, in the case of industrial electronics, and the correction of tests of peers, in the case of textile and apparel engineering. At the second moment, students made recommendations to their peers. At the third moment, the teachers were far less involved in the determination of assessment criteria and the students had to write a reflection about their assessment process.

The assessment tasks at the three moments were different for each group (see table 1). As it was not intended to change the curriculum of the two subjects involved in the project, the teachers of these subjects had the freedom to select the most appropriate assessment tasks for each assessment moment.

Assessment moment	Introduction to industrial electronics (industrial electronics)	Introduction to textile engineering (textile and apparel engineering)
First	Group work/report • Defining assessment criteria • Correcting report of other group • Justifying classifications	Test • Making test questions • Correcting one other student • Justifying classification
Second	Take-home assignment • Defining assessment criteria • Correcting work of colleagues • Justifying classifications • Making recommendations	Group work/report • Defining assessment criteria • Correcting report of other group in group • Justifying classifications • Making recommendations
Third	Test • Defining assessment criteria • Correcting own work • Justifying classifications • Making recommendation • Writing reflection	Individual assignment • Defining assessment criteria • Correcting own work • Justifying classifications • Making recommendation • Writing reflection

Table 1. Assessment tasks and related student activities for each assessment moment.

4. Instrument

For the information collection, an evaluation form was used at each assessment moment. The form included the assessment criteria, the teacher grade, the student grade, justifications and, for the second and third moments, recommendations. The students completed the form in a group in the case of a group assignment and by themselves in the case of an individual assignment.

5. Results

For each assessment moment, students identified criteria. Two types of criteria were used. For reports, students identified criteria such as identification of the objectives, structure, clarity, content, literature search, conclusions and references. If the assessment tasks did not include a report, the criteria included the correct answer of a test question or exercise and the calculations that would lead to the answer. Tables 2 and 3 show the results of paired t-tests that were used to identify significant differences between the teacher grades and the student grades. Comparisons of means for the first assessment moment of industrial electronics and the second assessment moment of textile and apparel engineering were not included, because of the small number of observations, due to group work. The groups of 61 and 42 students, respectively, were divided into groups of four students, who graded the work of their peers as a group. Therefore, the number of grades to compare was much lower than at the other, individual assessment moments.

Tables 2 and 3 give the results for industrial electronics at the second and third assessment moments. At the second moment, teachers and students differ significantly on two criteria and on the total grade. At the self-assessment task, students and teachers disagree significantly on five criteria and the total grade. The situation at textile and apparel engineering demonstrates more agreement between the teacher and the students. Only two criteria are assessed as significantly different at the third assessment moment.

To have a more detailed look at the differences in scoring accuracy of students and their final grades, both groups of students were divided into three equal size subgroups, based on their final grade, as given by their teachers. Tables 4 and 5 show the differences in mean grades between teachers and students for each level group. The total scores reflect the grades of the assessed students, not of the assessing students. Therefore, the highest level group does not necessarily have to have the highest grades.

	Students		Teacher		
Criterion	M	SD	M	SD	t (df = 61)
IE2.1	13.05	3.658	7.56	4.414	10.267***
IE2.2	17.41	4.511	16.70	5.614	1.247 (n.s.)
IE2.3	15.73	5.624	16.38	5.748	−1.335 (n.s.)
IE2.4	7.20	4.207	4.39	3.174	6.475***
IE2.5	10.89	4.218	11.74	4.501	−1.418 (n.s.)
IE2.6	7.41	3.303	7.62	3.527	−0.597 (n.s.)
IE2.7	3.60	1.798	3.98	1.554	−1.768 (n.s.)
IE2.Total	75.29	16.369	68.38	19.436	4.403***

Table 2. Differences between mean scores of teacher grades and student grades at the second evaluation moment of industrial electronics.

Significance level: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$. n.s., not significant.

	Students		Teacher		
Criterion	M	SD	M	SD	t (df = 49)
IE3.1	3.93	1.641	4.26	1.291	−1.733 (n.s.)
IE3.2	12.32	7.333	11.50	7.492	1.111 (n.s.)
IE3.3	2.08	2.098	1.34	1.780	3.015**
IE3.4	5.36	4.144	4.08	4.351	2.649 (n.s.)
IE3.5	4.56	4.496	4.14	4.760	1.130 (n.s.)
IE3.6	7.02	3.878	6.20	4.238	2.669*
IE3.7	3.38	2.256	3.35	2.127	0.116 (n.s.)
IE3.8	11.90	7.792	9.20	7.680	3.388**
IE3.9	2.11	1.952	1.40	1.884	3.002**
IE3.10	4.37	4.147	3.39	3.970	2.491*
IE3.Total	54.67	26.314	45.24	27.731	4.215***

Table 3. Differences between mean scores of teacher grades and student grades at the third evaluation moment of industrial electronics.

Significance level: *$p < 0.05$: **$p < 0.01$; ***$p < 0.001$.

		Students		Teacher			
Group	Criterion	M	SD	M	SD	df	t
Highest level	IE2.1	12.85	2.889	6.95	4.174	19	5.871***
	IE2.2	16.65	5.422	16.00	6.521	19	0.503 (n.s.)
	IE2.3	14.75	4.587	16.50	5.568	19	−1.437 (n.s.)
	IE2.4	6.90	4.191	3.85	2.758	19	3.612**
	IE2.5	11.35	3.027	12.55	2.892	19	−1.508 (n.s.)
	IE2.6	7.60	3.202	7.85	2.739	19	−0.383 (n.s.)
	IE2.7	3.65	1.237	3.85	1.785	19	−0.597 (n.s.)
	IE2.Total	73.75	13.796	67.55	17.769	19	1.643 (n.s.)
Intermediate level	IE2.1	14.15	2.245	12.00	3.784	19	5.904***
	IE2.2	18.05	2.964	16.85	5.204	19	0.513 (n.s.)
	IE2.3	16.95	3.734	15.10	5.447	19	0.414 (n.s.)
	IE2.4	6.63	3.579	7.98	4.681	19	4.279***
	IE2.5	10.83	3.668	11.90	3.892	19	−0.324 (n.s.)
	IE2.6	8.00	2.772	7.10	3.177	19	−1.552 (n.s.)
	IE2.7	3.38	2.032	3.60	1.465	19	−2.138*
	IE2.Total	77.97	12.116	74.53	16.291	19	2.935**
Lowest level	IE2.1	13.67	3.665	8.76	4.571	20	5.040***
	IE2.2	18.52	2.358	17.95	3.943	20	0.611 (n.s.)
	IE2.3	16.86	4.993	17.67	4.820	20	−1.110 (n.s.)
	IE2.4	7.48	3.060	3.76	2.343	20	5.766***
	IE2.5	9.81	5.026	12.10	4.381	20	−2.510*
	IE2.6	7.48	3.558	7.57	3.668	20	−0.133 (n.s.)
	IE2.7	3.12	2.061	3.81	1.436	20	−1.823 (n.s.)
	IE2.Total	76.88	12.566	71.62	14.975	20	2.225*

Table 4. Differences between mean scores of teacher grades and student grades of the students per level of grade at the second evaluation moment of industrial electronics.

Significance level: *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

Group	Criterion	Students M	Students SD	Teacher M	Teacher SD	df	t
Highest level	IE3.1	4.56	0.85	4.67	0.970	17	−0.461 (n.s.)
	IE3.2	15.59	6.23	17.11	4.575	17	−1.080 (n.s.)
	IE3.3	3.00	2.27	2.11	1.967	17	1.978 (n.s.)
	IE3.4	7.67	3.49	7.56	3.502	17	0.178 (n.s.)
	IE3.5	5.89	4.86	6.00	4.947	17	−0.356 (n.s.)
	IE3.6	9.17	1.91	8.72	2.927	17	0.889 (n.s.)
	IE3.7	4.76	0.56	4.88	0.485	17	−0.808 (n.s.)
	IE3.8	16.81	5.02	16.00	5.770	17	0.669 (n.s.)
	IE3.9	2.94	2.12	2.72	2.052	17	0.455 (n.s.)
	IE3.10	6.72	3.95	6.17	4.062	17	0.617 (n.s.)
	IE3.Total	76.84	16.10	75.67	11.37	17	0.403 (n.s.)
Intermediate level	IE3.1	3.92	1.92	4.50	0.798	17	−1.465 (n.s.)
	IE3.2	12.00	7.24	10.92	7.440	17	1.132 (n.s.)
	IE3.3	1.58	1.92	0.75	1.603	17	1.758 (n.s.)
	IE3.4	3.42	3.72	1.83	2.406	17	1.682 (n.s.)
	IE3.5	4.17	4.85	4.17	5.149	17	0.000 (n.s.)
	IE3.6	5.83	4.17	4.33	3.916	17	2.105 (n.s.)
	IE3.7	3.08	2.27	3.00	2.174	17	0.146 (n.s.)
	IE3.8	9.00	8.19	7.00	5.360	17	1.188 (n.s.)
	IE3.9	1.58	1.97	0.50	1.446	17	2.493*
	IE3.10	4.25	4.07	3.25	3.519	17	1.459 (n.s.)
	IE3.Total	48.83	14.92	40.25	3.519	17	2.872*
Lowest level	IE3.1	2.82	1.95	3.36	1.781	13	−1.075 (n.s.)
	IE3.2	8.21	7.61	4.86	5.216	13	2.396*
	IE3.3	1.36	1.55	0.57	1.158	13	1.808 (n.s.)
	IE3.4	3.93	3.79	0.43	1.158	13	3.524**
	IE3.5	2.93	3.51	1.43	2.901	13	1.916 (n.s.)
	IE3.6	4.36	4.25	3.29	4.084	13	1.670 (n.s.)
	IE3.7	2.07	2.84	1.64	2.134	13	0.598 (n.s.)
	IE3.8	6.18	7.09	1.50	2.955	13	2.763*
	IE3.9	1.39	1.52	0.29	0.611	13	3.419**
	IE3.10	1.50	2.41	0.36	0.929	13	2.332*
	IE3.Total	38.21	21.16	16.53	9.841	13	3.893**

Table 5. Differences between mean scores of teacher grades and student grades of the students per level of grade at the third evaluation moment of industrial electronics.
Significance level: $*p < 0.05$; $**p < 0.01$; $***p < 0.001$.

Table 4 concerns the second assessment moment of industrial electronics. Two means for the best students are significantly different: for the first criterion ($p < 0.001$) and for criterion IE2.4 ($p < 0.01$). The total grade mean is not different. For the intermediate group, four grade means are significantly different, although the total grade is not significantly different. The group with the lowest grades shows four significantly different grade means. Table 5 shows the differences for the mean grades of the third evaluation moment of industrial electronics. At this moment, the best students appear to be accurate in their scoring. They grade themselves about the same way the teacher does, not revealing any significant differences. The intermediate students are less accurate, disagreeing with the teacher grade on one criterion. The total score is also different. The students with the lowest grades seem to have major

difficulties with the self-assessment, as conducted at the third assessment moment. They differ significantly from their teacher on five out of ten criteria and on the total grade.

The results of textile and apparel engineering per level are similar to those of industrial electronics. There is little disagreement between teacher and students at the highest and intermediate-level group: only one significant difference between student and teacher grade ($p < 0.05$), whereas there is no significant difference in the group of students with the lowest grades. At the third moment, the self-assessment moment, students seem to have many more difficulties with accurate grading. In particular, the intermediate students and the students with the lowest grades face difficulties at nearly every criterion.

6. Discussion

The general results as presented above do not indicate clearly whether or not students are able to assess themselves. Both courses show different results. At industrial electronics, mean grades of students and teacher are significantly different in almost half of the cases. At textile and apparel engineering, there are only four significant differences. This can be explained by the nature of the criteria that were identified. At the first assessment moment of textile and apparel engineering, students were insecure about the corrections of their colleagues and they looked for very clear criteria. Combined with the short answer nature of most of the test questions, the grading process became rather structured, making it relatively easy for students to grade their colleagues. At industrial electronics, the nature of the test questions and take-home assignment was more complex, so the actual grading process was less structured and therefore more complicated for the students. From the industrial electronics results, self-assessment seems to be harder than peer assessment. Looking at the results per course per student final level, it becomes clear that the final grade of a student is a strong indicator of his or her accuracy in grading. The best industrial engineering students only differ with their teacher on two criteria at the peer assessment moment, whereas the intermediate students and the students with the lowest grades differ on, respectively, three and four criteria. At the self-assessment moment, the differences between the groups are even more obvious. The best students agreed with the teacher on each criterion, while the intermediate students differed on two criteria. The students with the lowest grades had major problems with self-assessment. More than half of the mean scores were significantly different. A similar pattern occurs in textile and apparel engineering. Although in this course the differences between levels of students do not appear at the peer assessment moment, the results of the self-assessment moment indicate more difficulties for students with lower grades. With hardly any support of students in their first-year engineering courses, many of them demonstrated the capacity of accurate grading of peers and themselves. A study of the effect of better guidance of weaker students could give valuable information about how to improve the peer and self-assessment method and how to make them benefit as well from the advantages of these forms of assessment.

Acknowledgement

This project has been supported by the FCT under POCTI 1.1 by a scholarship with number BD/21536/99.

References

BAILLIE, C. and TOOHEY, S., 1997, The 'power test': its impact on student learning in a material science course for engineering students. *Assessment & Evaluation in Higher Education*, **21**, 33–48.

BARNETT-FOSTER, D. and NAGY, P., 1996, Undergraduate student response strategies to test questions of varying format. *Higher Education*, **32**, 177–198.

BIGGS, J. B., 1999, *Teaching for Quality Learning at University* (Buckingham: SRHE and Open University Press).

DOCHY, F. J. R. C. and MOERKERKE, G., 1997, Assessment as a major influence on learning and instruction. *International Journal of Educational Research*, **27**, 415–432.

HAGER, P. and BUTLER, J., 1996, Two models of educational assessment. *Assessment & Evaluation in Higher Education*, **21**, 367–378.

HARGREAVES, D. J., 1997, Student learning and assessment are inextricably linked. *European Journal of Engineering Education*, **22**, 401–409.

HOUSTON, K. and LAZENBATT, A., 1996, A peer-tutoring scheme to support independent learning and group project work in mathematics. *Assessment & Evaluation in Higher Education*, **21**, 251–265.

KWAN, K. P. and LEUNG, R., 1996, Tutor versus peer group assessment of student performance in a simulation training exercise. *Assessment & Evaluation in Higher Education*, **21**, 205–214.

LEJK, M., WYVILL, M. and FARROW, S., 1996, A survey method of deriving individual grades from group assessment. *Assessment & Evaluation in Higher Education*, **21**, 267–280.

LONGHURST, N. and NORTON, L. S., 1997, Self-assessment in coursework essays. *Studies in Educational Evaluation*, **23**, 319–330.

MCDOWELL, L., 1995, The impact of innovative assessment of student learning. *Innovations in Education and Training International*, **32**, 302–313.

MCILVEEN, H., GREENAN, K. and HUMPHREYS, P., 1997, Involving students in teaching and learning: a necessary evil? *Quality Assurance in Education*, **5**, 231–238.

RAFIQ, Y. and FULLERTON, H., 1996, Peer assessment of group projects in civil engineering. *Assessment & Evaluation in Higher Education*, **21**, 69–81.

SHEPPARD, S., JOHNSON, M. and LEIFER, L., 1998, A model for peer assessment and student involvement in formative course assessment. *Journal of Engineering Education*, **87**, 349–354.

SLUIJSMANS, D. M. A., MOERKERKE, G., VAN MERRIËNBOER, J. J. G. and DOCHY, F. J. R. C., 2001, Peer assessment in problem based learning. *Studies in Educational Evaluation*, **27**, 153–173.

About the authors

Natascha van Hattum-Janssen works at the Council of Engineering Courses of the University of Minho, at the Centre of Educational Quality. She is involved in projects on learning and assessment, quality of higher education and staff development and has recently finished her PhD on the influence of assessment on the learning approach of students in first-year engineering courses.

Rosa M. Vasconçelos received her degree in textile engineering in 1984 at Minho University, Portugal. In 1993 she obtained her PhD in Engineering-Textile Technology and Chemistry in the speciality of textile technology, at Minho University. Since 1993 she has been working as an auxiliary professor, in the Textile Engineering Department of Minho University. Her fields of interest are textile processes and industrial automation. Furthermore, she is President of the Council of Engineering Courses and works on a university-wide approach to quality of teaching and learning. She is involved in various projects on staff development, evaluation of teaching and learning, integration of first-year students and assessment of student learning.

José Augusto Pacheco is an associate professor at the Institute of Education and Psychology at the University of Minho and publishes and teaches in the areas of evaluation, curriculum theory and teacher training. He is co-ordinator for postgraduate education in curriculum theory.

VIEW

TOMORROW'S RESEARCH TODAY

0304-3797(2004)29:2;1-M