Using Test Results for ASSESSMENT OF TEACHING AND LEARNING

H. Henning Winter
University of Massachusetts • Amherst, MA 01003

Examination time can be filled with anxiety. Teachers design a mid-term or final exam to cover the most important subjects of their courses and expect the student to apply the learned material successfully. Most gratifying for teacher and student alike is an exam in which the student answers all questions and receives a top grade. Incomplete or wrong answers generate dissatisfaction with both the student and the teacher. Reality is somewhere between these extremes, depending on the degree of success of the teaching and student committment. The exam results often suggest that the teaching needs to be improved, but the questions are where it can be improved and how. Direction can come from an assessment of exams. They contain a wealth of information, much more than just a grade for the student.[1]

Methods have been developed for assessing entire engineering programs, curricula as well as individual courses, and educational research projects.[2,3] Student portfolios[2,3] allow quantitative assessment of the students' work during the year with feedback to the campus community. This report describes a teaching tool that works on the assumption that the educational program as a whole has already been assessed and that a plan exists for individual courses. Instead of the large-scale approach, this paper will focus on methods of analyzing a single exam and generating direct feedback for the teaching of a course with well-defined objectives.

I have introduced the concept of a "grading matrix" for analyzing the results of tests in chemical engineering. The grading matrix has the purpose of detecting academic strengths and weaknesses of individual students as well as strengths and weaknesses of teaching. Most important is the identification of weaknesses so that they can be corrected in the classroom (or outside) and possibly re-assessed. The increased interest in teaching assessment has motivated me to describe the grading matrix in this report. Until now, I have used it by myself in all undergraduate and graduate teaching for over a decade and have gradually refined it. The matrix method is somewhat related to the Primary Trait Analysis of Loyd-Jones,[5] which was recently pointed out to me. But, in addition to student performance, the grading matrix also assesses teaching success. This paper briefly describes the grading matrix together with suggestions for its use in teaching and curriculum development.

THE GRADING MATRIX

The definition and use of the grading matrix can be seen in Figure 1. The example is deliberately kept simple: a typical written test is broken down into N individual subtopics ($task_1$ to $task_{16}$ since N=16 was chosen for this test) shown across the top of the matrix. Student names appear on the left side. Separately for each of the subtopics, the student's exam is evaluated on a scale from 0% to 100%. Grades are finely varied between 0% and 100% or, in yes/no fashion of a quiz, with either 1 or 0 in the matrix. This choice depends on the nature of the test or quiz. A row of grades across the matrix shows the strengths and weaknesses of that individual student. The average over the row constitutes

***H. Henning Winter** is Distinguished University Professor of Chemical Engineering at the University of Massachusetts at Amherst. He has degrees from Stanford University (MS) and the University of Stuttgart (Dr. Ing). His research includes experimental rheology, polymer gelation, and crystallization.*

his or her final grade:

$$\text{grade}\,[\%] = \frac{100}{N}\left(\text{task}_1 + \text{task}_2 + \text{task}_3 \ldots + \text{task}_N\right) \qquad (1)$$

where N is the number of tasks (=number of columns in the matrix). The actual grading process is complete at this point.

When returning the graded test, each student receives two items: their own exam booklet and the grading matrix (without names) of the entire class. No grades are written in the booklet except for the final grade on the booklet cover. Instead of grades, I write occasional comments into the exam booklet with the purpose of helping the student to understand the course material. For identification on the matrix, students need to find the row with their final grade on the right side. By knowing the row, students obtain an analysis of their personal performance in each of the subtopics of the test. This allows them not only to assess their personal knowledge but also to compare it with the rest of the class. Students told me that they especially like this comparison to others. Note that, different from Figure 1, no student names are listed on the students' copy of the matrix; privacy is maintained. Students can reveal their grade to fellow students, but their performance remains otherwise unknown. I have not had any problems arising from this procedure.

The most critical part of the entire assessment process is the design of the grading matrix itself; *e.g.* the selection of test questions (called "task" in Figure 1), which the student will be asked on the test. These tasks need to be representative for the course objectives according to an overall plan.[2,3,6] Consider the example of a Fluid Mechanics course, which has the objective that students learn to solve certain flow problems. This can be tested in an exam where one such flow problem is broken down into: ($task_1$) schematic drawing of the expected velocity field, choice of coordinate system, and definition of boundary conditions; ($task_2$) equation for conservation of mass; ($task_3$) equation for conservation of linear momentum; ($task_4$) solution for obtaining the velocity field; ($task_5$) statement of all simplifying assumptions and limitations of the solution; ($task_6$) discussion of properties of calculated flow field; and ($task_7$) prediction of pressure and stress. Most written tests are easily structured in this way.

TEACHING ASSESSMENT AND CORRECTIONS

Until this point, the exam grading has followed conventional paths, except that the data is filed in a spreadsheet,

	task 1	task 2	task 3	task 4	task 5	task 6	task 7	task 8	task 9	task 10	task 11	task 12	task 13	task 14	task 15	task 16	bonus	Σ weights	student grade
weight =	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		16	
1. student	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	2		100 %
2. student	1	1	1	1	1	1	1	0.3	1	0	1	1	1	1	1	1	1		96 %
3. student	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	2		94 %
4. student	1	0.9	0.9	1	1	1	1	1	1	1	1	1	0	1	0.9	0	1		92 %
5. student	1	0.9	0.8	1	1	0	1	0.2	0	0	0.9	1	1	1	1	0.9	1		79 %
6. student	1	0.8	0.6	1	1	1	1	0	1	0	0.9	1	1	1	1	0			77 %
21. student	1	1	0.9	1	1	1	1	0	0	0	1	0	0	0	0.5	0			53 %
22. student	1	1	1	1	0	1	0.8	0	0.2	0	0.6	0.8	0	1	0	0			53 %
23. student	1	0.8	0.5	1	0.9	1	1	0.2	0	0	1	0	0	1	0	0			53 %
24. student	1	0.5	1	1	1	1	0	0	0	0	1	0	0	0.8	0	0	1		52 %
25. student	1	0.8	1	1	1	1	1	0	0	0	0.8	0	0.7	0	0	0			52 %
26. student	1	1	0	1	1	1	0.8	0	0	0	0.8	0	0	0.8	0	0			46 %
27. student	1	0.3	0.8	1	1	0	1	0	0	0	1	1	0	0.2	0	0			46 %
28. student	1	0.8	1	1	1	1	0	0	0	0	0.5	0	0	0.8	0	0			44 %
29. student	1	0.8	0.8	0	0	0	1	1	1	0	1	0	0	0	0	0			41 %
30. student	1	0	0.4	1	1	1	1	0	0	0	0.7	0	0	0	0	0			38 %
teaching	100	84	78	96	92	86	89	27	47	16	85	42	22	81	22	9 %			
assessment								*		*		*	*		*	*			

Figure 1: *Example of the grading matrix of a test. Grades are filed in a spreadsheet. $Task_1$, $task_2$, $task_3$, etc. stand for test questions. Number codes for grades are 1=100%, 0.9=90%, 0.8=80%, ...and 0=0%. Different weights can be assigned to each of the tasks, though here all weights are set to the same value of 1. Teaching is assessed by taking an average over entire columns, top to bottom; the result shows in the bottom row. An asterisk marks topics which are not understood by the majority of the class and need to be addressed. In real application, the left column of names will be removed. All data in this example are fictitious.*

ready for further assessment. Some of the most important information is contained in the columns of the grading matrix of Figure 1. A column with mostly high marks (1 = highest mark) top to bottom shows that all students know the subject, at least at the level of the exam question. If a column, however, has mostly "0" marks, something went wrong. Reasons can be deep-rooted or only superficial (*i.e.*, the question was confusing or the students ran out of time). Discussions between teacher and students often bring clarification, and plans for further action are easily devised. Technical deficiencies and/or misunderstandings are recognized and can be addressed, for instance, in a special help session or in the next homework assignment. Experiments can be added or computer animation can be used to help visualize abstract concepts. Teachers have an opportunity to become very creative as soon as the problem is defined. This definition of the problem is the main purpose of the grading matrix.

Correction of weaknesses can then be re-assessed in the next test. This is typically done by including appropriate questions in the next exam, preferably within the same course and/or in the next homework assignment. Teaching should be corrected further if necessary. Often it is too late to introduce corrections in the same semester or quarter. If changes cannot be made in time, the weakness in one course will be passed on to the teacher of the following course. This teacher should be alerted to the problem so that corrections can be made there.

The grading matrix provides a record, which can be used even if another teacher teaches the course the following year. Adjustments can be made then and can be re-assessed until teaching weaknesses are resolved. I can imagine, however, a problem with the existence of such records, since they have a potential for misuse in the form of over-coaching of teachers. This would interfere with the learning environment and impair the matrix method. Access to the grading matrix should be restricted to the teachers and students who are directly involved.

FEEDBACK TO STUDENTS

Advising individual students is enhanced by the diagnostic property of a grading matrix. The teacher sees individual weaknesses of students and can suggest corrective measures. (*e.g.*, specific reading material or exercises). This does not require further preparation on the teacher's part. Information is available instantly when a student comes to the office for consultation. The matrix row of grades, in combination with other observations (attendance, participation during class, etc.), provides a quantitative basis for a discussion.

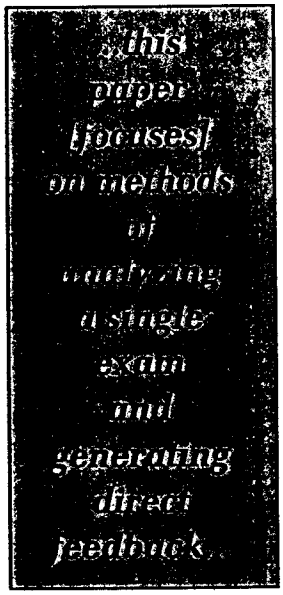
Figure 2:

This is the same grading matrix as in Figure 1, but specific weights are assigned to each of the tasks. This affects the calculation of the grade as defined in Equation 2. Everything else, including the teaching assignment, remains unchanged by the weighting system. Weights have little effect on the grade of top students but can make a large difference for a weaker student.

	task 1	task 2	task 3	task 4	task 5	task 6	task 7	task 8	task 9	task 10	task 11	task 12	task 13	task 14	task 15	task 16	bonus	Σ weights	student grade
weight =	0.5	1	3	1	2	1	5	1	2	0.5	2	1	4	1	1	1		27	
1. student	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	0	2		100 %
2. student	1	1	1	1	1	1	1	0.3	1	0	1	1	1	1	1	1	1		99 %
3. student	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	0	2		85 %
4. student	1	0.9	0.9	1	1	1	1	1	1	1	1	1	0	1	0.9	0	1		83 %
5. student	1	0.9	0.8	1	1	0	1	0.2	0	0	0.9	1	1	1	1	0.9	1		84 %
6. student	1	0.8	0.6	1	1	1	1	0	1	0	0.9	1	1	1	1	0			85 %
.....																			
.....																			
22. student	1	1	1	1	0	1	0.8	0	0.2	0	0.6	0.8	0	1	0	0			51 %
23. student	1	0.8	0.5	1	0.9	1	1	0.2	0	0	1	0	0	1	0	0			55 %
24. student	1	0.5	1	1	1	1	0	0	0	0	1	0	0	0.8	0	0	1		44 %
25. student	1	0.8	1	1	1	1	1	0	0	0	0.8	0	0.7	0	0	0			66 %
26. student	1	1	0	1	1	1	0.8	0	0	0	0.8	0	0	0.8	0	0			44 %
27. student	1	0.3	0.8	1	1	0	1	0	0	0	1	1	0	0.2	0	0			53 %
28. student	1	0.8	1	1	1	1	0	0	0	0	0.5	0	0	0.8	0	0			37 %
29. student	1	0.8	0.8	0	0	0	1	1	1	0	1	0	0	0	0	0			51 %
30. student	1	0	0.4	1	1	1	1	0	0	0	0.7	0	0	0	0	0			45 %
teaching	100	84	78	96	92	86	89	27	47	16	85	42	22	81	22	9	%		
assessment								*		*		*	*		*	*			

CURRICULUM DEVELOPMENT

Weaknesses in student learning, as detected in the grading matrices of a course (two midterms and a final, for example) should be assessed in the context of the entire curriculum. There is a possibility that students may not be sufficiently prepared for a specific class. Prevailing weaknesses should, in this case, be addressed by changing the course content of the responsible preceding course. Relevant results from the grading matrix can be integrated into the systematic curriculum development.[3] Discussions along these lines are in progress in our department.



ADAPTATION OF THE MATRIX METHOD

There are many ways of integrating the information from the grading matrix into personal approaches to teaching and student advising. It goes without saying that assessment of test performance as reported here needs to be integrated with classroom assessment. This is a dynamic process, which differs from year to year, since each group of students interacts differently and varies in its needs. As the learning process evolves, teachers adapt in their classroom assessment and in their creative teaching approaches. The integration of the grading matrix in day-to-day teaching works well for me, but a general discussion of this topic would exceed the scope of this report.

Obviously, the matrix itself can be tailored in many different ways, and adaptations are straightforward. A few will be mentioned here. It is possible, for instance, to emphasize selected parts of an exam by adding weight to some of the tasks. While I normally give uniform weight to all questions (see top row of the matrix in Figure 1), more important questions can be given an increased weight, as shown in Figure 2. The row of grades across the matrix needs to be rescaled accordingly when calculating the final grade:

$$\text{grade}\,[\%] = 100\,\frac{\sum_{i=1}^{N} \text{weight}_i \cdot \text{task}_i}{\sum_{i=1}^{N} \text{weight}_i} \qquad (2)$$

where N is the number of columns. Additional bonus points can be added wherever appropriate. The overall scale of the test will not be affected by assigning bonus points to individual students.

The concept of a grading matrix is introduced here with a chemical engineering example and on the most straightforward type of test. The proposed method for assessment of teaching is applicable at many levels, however. It is equally useful for students and teachers outside of engineering. Similar questions arise in high school teaching and even in elementary schools where standardization of tests is considered.[7] The matrix method can also be adapted to examinations of much wider scope, such as oral presentations or essay-type exams. Oral exams or essays tend to be less uniform in their structure than the written tests discussed above. This, however, does not make their grading less amenable to matrix format. New categories need to be added to the list of tasks, such as style and expression, logic of argument, depth of discussion, format of graphs, validity of conclusions, and more. The choice of categories needs to be explained to the students well in advance of the exam.

SUMMARY

The three main functions of the grading matrix are providing a grade for the student, labeling areas of weakness in the student's knowledge, and labeling areas of weakness in the teaching. For me personally, the grading matrix helped to fairly assess the abilities of students since my grading became more uniform, something I tried with less success with other grading methods. The grading matrix also alerted me to problems that students encountered with course material. It labeled weaknesses in my teaching so that I could devise different teaching methods when needed. I feel that, during office hours, my advice became better directed to the needs of individual students. The design of test content with the matrix structure in mind and the feedback from tests have positively affected my teaching and my continued search for ways to motivate students. While still being a stressful experience for the students, examinations have turned into an effective instrument for improved teaching.

ACKNOWLEDGMENTS

Support from the von Humboldt Foundation, many lively discussions with colleagues and students, and helpful suggestions from the reviewers are gratefully acknowledged.

REFERENCES

1. Walvoord, G. and V.J. Anderson, *Effective Grading: A Tool for Learning and Assessment*, Jossey-Bass, San Francisco, CA (1998)
2. Olds, B.M. and R.L. Miller, "An Assessment Matrix for Evaluating Engineering Programs," *J. Eng. Ed.*, **87**, p. 173 (1998)
3. McNeill B. and L. Bellamy, "The Articulation Matrix, a Tool for Defining and Assessing a Course," *Chem. Eng. Ed.*, **33**, p. 122 (1999)
4. Taylor, R. *Basic Principles of Curriculum and Instruction*, University of Chicago Press. Chicago, IL (1949)
5. Loyd-Jones, R. "Primary Trait Analysis" in Cooper C. and L. Odell (eds.) *Evaluating Writing: Describing, Measuring, Judging*. Urbana, IL Council of Teachers of English, Urbana (1977)
6. Olds, B.M. and R.L. Miller, "Using Portfolios to Assess a Chemical Engineering Program," *Chem. Eng. Ed.*, **33**, p. 110 (1999)
7. Saltet, J.K. "How is my Child Doing?" *J. Waldof Education*, **10**(2), p. 5 (2001) ❐