Criterion-Based Grading for Learning and Assessment in Unit Operations Laboratory

V.L. Young, D. Ridgway, M.E. Prudich, D.J. Goetz, and B.J. Stuart
Department of Chemical Engineering, Ohio University

Abstract

Goals for student learning in Unit Operations Laboratory are diverse. Students are to develop skills in the planning and execution of experiments, the interpretation of experimental data, and the communication of technical information. Additional goals include gaining familiarity with process equipment, improving teamwork skills, and developing professional habits in safety, documentation, and planning. Each laboratory report grade represents an amalgamated assessment of students’ progress towards all goals, so it cannot be used to assess progress towards any of them. Assessing student progress in our course is additionally complicated because students rotate through experiments supervised and graded by different professors.

A criterion-based grading system, inspired by Walvoord and Anderson’s Effective Grading: A Tool for Learning and Assessment, was introduced at Ohio University. A standard grading sheet for all experiments specifies “traits” indicative of progress towards each course goal (e.g., the student’s proposed test matrix indicates ability to design an experiment) and criteria for grading each trait. The purpose of the grading sheet is threefold:

1. Guide student learning by specifying the traits of a good laboratory report.
2. Improve consistency in grading and feedback to students by specifying the key traits to be considered.
3. Collect data to assess student progress towards course goals.

Success in achieving the three-fold purpose is being evaluated via student surveys, conversations with faculty, and the assessment data provided by the grading instrument itself. The results of this evaluation are the subject of this presentation.

Introduction

At Ohio University, Unit Operations Laboratory (UO Lab) is a two-quarter sequence for seniors in chemical engineering. The laboratory is equipped with pilot-scale versions of process equipment and with experimental setups for exploring particular concepts in chemical engineering. Standard laboratory and “plant floor” equipment (graduated cylinders, analytical balances, pipettes, buckets, scales, stopwatches) is also available. UO Lab is a “capstone” experience in which students must integrate and apply previous coursework. Extensive written and oral technical communication is required. Each experiment is motivated by a “design objective”. Students must identify the parameters to be determined experimentally, plan and execute the required experiments, analyze the data and interpret the results, and complete the
design objective. Data collection is completed in two five-hour lab periods for each experiment. Four teams work on different experiments during each lab period, and the teams cycle through all four experiments during the quarter. Outside of class, students complete a prelab report, a postlab report, an oral design presentation, and a design memo for each experiment. Students request frequent, individualized attention from the instructor outside scheduled class time.

Several years ago, our department changed from a model in which a group of students were assigned one faculty member to assist them in all experiments, to one in which each faculty member acts as an expert for one or two experiments and assists all students needing that expertise. The instructor for each experiment grades all work related to the experiment, with a lead instructor collecting all grades and using them to calculate course grades. The faculty believe this change improved the quality of instruction on each experiment, and made grading more fair by extending unequal faculty expectations equally over all students.

Still, a persistent complaint from students has been that expectations differ so much among faculty. Reports grew thicker and thicker, as students tried not to miss anything that anyone might think was important. “No matter what you put in, you miss something that the next guy wants.” Ask the faculty; all are looking for thorough planning and careful execution of experiments, the logical interpretation of experimental data via mathematical and statistical analysis, and clear and concise communication of technical information. Of course, each experiment lends itself to a different emphasis, and each individual has some pet peeves. As a result, the students missed completely the common elements sought by the faculty. To the students, “This guy wants a ton of statistics, that guy wants to see every design equation you can imagine, and the other guy goes nuts if you don’t write down every detail of your procedure.”

Worse, during the course students didn’t seem to improve substantially at planning and execution of experiments, interpretation of experimental data via mathematical and statistical analysis, or communication of technical information. It couldn’t be proven, because there was no straightforward way to compare student performance on different experiments. But if they were learning, wouldn’t most students working on one’s experiment at the end of the course perform better than those at the beginning did? Anecdotal comparisons of student work on different experiments suggested they were repeating their mistakes, not learning from them. Carefully-written comments made during grading were ignored, seen as peculiar to the instructor and of no value for the next experiment. In short, grades were assigned, but the extent to which course goals were achieved remained unknown, and student learning was suspected to be falling short.

Educational Technique

In their book *Effective Grading: A Tool for Learning and Assessment*, Walvoord and Anderson propose criterion-based grading as a method for assisting and assessing student learning. They call their technique “Primary Trait Analysis”. The instructor identifies “Primary Traits”, features of the student’s work that indicate whether goals for learning have been achieved. If the goal is the ability to design an experiment, a useful “Primary Trait” is the test matrix, the table of experimental trials to be conducted. Next, the instructor defines a scale for analyzing each Primary Trait. For example, an excellent (“A”) test matrix specifies all the experimental conditions with reasonable values, covers a sufficient range of values to meet the experimental
objectives, specifies a reasonable number of runs for the time available, and includes repeat runs to establish reproducibility. The instructor must similarly describe lower levels of performance. Finally, the instructor converts the scores for Primary Traits to an overall score for the assignment. The score on each trait is retained for assessment purposes.

Superficially, this appears identical to the common practice of assigning a grade to the “experimental procedure” section of the report. However, a grade for that section likely considers the logical order of the procedure, the quality of the equipment diagram, the inclusion of safety concerns, and the clarity of the prose in addition to the contents of the test matrix. Isolating the test matrix provides a more meaningful assessment of the student’s the ability to design an experiment.

A key feature of Primary Trait Analysis is that the same criteria the instructor uses to assess the work are provided to the students before they complete the assignment. Primary Trait Analysis is a learning tool because students can identify the important features of the assignment and assess their own work before submitting it. Responsibility for performance is transferred to the student; the excuse that “knowing what’s wanted is just a guessing game” is eroded.

Another useful feature of Primary Trait Analysis is the concept of “Gateway Criteria”. These are standards that must be met before the work will be graded. For example, in our class, a report with more than about one proofreading error per page is unacceptable, regardless of content. It receives an “F” and a required rewrite with a penalty. If the introduction is poorly proofread, the instructor reads no further. The responsibility to meet basic requirements of format and style is the student’s. The instructor can then provide feedback about content without distraction.

Using standard Primary Trait Analysis grading sheets for all experiments in our UO Lab had several perceived advantages. (Current grading sheets for the UO Lab may be obtained from the course archives on the departmental webpage, or by contacting the authors.) Once all instructors agreed to the Primary Traits to be assessed, they could still weight the traits differently to arrive at report grades. Students could see that features such as statistical analysis were essential to all laboratory reports, though the depth of analysis required might differ among projects. Grading sheets would identify areas of weakness, allowing students to direct effort where it was needed, and flagging topics for increased attention in the curriculum. Subjectivity and instructor discretion would not be eliminated, but consistency in grading should improve, and it should be possible to track student performance for specific traits. Students would be encouraged to take more responsibility for the quality of their work. The technique seemed to fit neatly into the requirements of EC 2000. Primary Trait Analysis should:

1. Guide student learning by specifying the traits of a good laboratory report.
2. Improve consistency in grading and feedback to students by specifying the key traits to be considered.
3. Collect data to assess student progress towards course goals.

Implementation of Primary Trait Analysis

V.L. Young tested Primary Trait Analysis for grading UO Lab prelab and postlab reports in Winter quarter 2000. Her motivations were
(1) to reduce the amount of time spent trying to ensure consistent grading
(2) to prevent good prose from overpowering poor content in her grading.
(3) to test Primary Trait Analysis as an assessment technique for meeting the EC2000
 requirements of the Accreditation Board for Engineering and Technology programs.\(^3\)

Note that Winter quarter is the second quarter in the two-quarter UO Lab sequence. This second quarter is also offered during Spring quarter.

Developing the grading sheets was time-consuming because the task was unfamiliar and the work to be graded was complex. Anyone undertaking such a task for the first time should allow several weeks for revision before the grading sheets are used. To define the Primary Traits, V.L. wrote complete reports for her experiment, then questioned why she considered them to be good reports. She selected about ten Primary Traits for each report (prelab and postlab). These traits were drawn from course and departmental goals for student learning and seemed critical to report quality. She felt that selecting a much larger number of traits would make grading more (not less) time-consuming and be too daunting to the students, while a much smaller number would omit too much. Traits such as proofreading and correct formatting became Gateway Criteria, because their lack destroys a report’s quality but their presence should not atone for poor content. For each trait, she described five grade levels (A, B, C, D, F). She reviewed and revised the descriptions to be specific to each trait, but not specific to the experiment. Her goal was that any instructor be able to use the same sheets for any experiment in the course, thus highlighting the characteristics common to all good reports and allowing comparisons of student performance among experiments. She tested the sheets by grading a report from a previous year and revised again. Finally, she offered use of the grading sheets to the other three faculty teaching UO Lab in Winter 2000. One, D.J. Goetz accepted. Goetz wrote and graded a report on his experiment, leading to a final revision before the quarter began. Grading sheets were available to students as PDF files linked to the course web page.

Both Young and Goetz are tenure-track faculty. This was the fourth time for Young to teach the course, the first for Goetz. The other two faculty who taught UO Lab in Winter 2000 are tenured, with several years of experience in the course. During Winter 2000, Young and Goetz graded reports using Primary Trait analysis, while the other faculty continued with their accustomed grading methods. All four faculty agreed to return unacceptable reports for rewriting to pressure students to learn from their mistakes. Young and Goetz defined “unacceptable” as failure to meet Gateway Criteria or grades of F on two or more Primary Traits. Goetz also used the grading sheets when he taught the course in Spring 2000, the only instructor to do so.

Young and Goetz recorded grades for each Primary Trait for each report in an Excel spreadsheet. Using the “countif” function of Excel, a table of the number of A’s, B’s, etc. earned for each Primary Trait was automatically generated. Because each grade was connected directly with a description of student performance and a goal for student learning, the tabulated results fed directly into assessment of student achievement of course goals.

During Spring 2000, V.L. Young presented the grading sheets and Winter Quarter assessment report at a departmental faculty meeting. The faculty agreed to use the sheets to grade all UO Lab reports during the 2000-2001 academic year. After each quarter, use of the sheets would be reviewed and revisions made. Of the faculty teaching UO Lab in 2000-2001, only V.L. Young had previous experience with the sheets. D. Ridgway and M.E. Prudich took an active role in
evaluating and revising them, and using them for assessment. The paper submission deadline being January 2001, results from Winter and Spring 2001 cannot be included here.

A Useful Tool for Student Learning?

Student acceptance of the grading sheets was very low at first, as evidenced both by the griping in the halls and the overall teaching evaluation scores given to V.L. Young and D.J. Goetz for Winter 2000. (See Table 1.) Both instructors received scores well below their usual for this course or all courses. While both scored at their usual above-3.0 levels in categories such as helpfulness in answering questions, they received their all-time low scores in quality of course material provided (about 2.5/4.0) and fairness of grading (about 1.2/4.0). Students that quarter did not perceive useful information or more fair grading from the sheets.

<table>
<thead>
<tr>
<th>Quarter</th>
<th>V.L. Young</th>
<th>D.J. Goetz</th>
</tr>
</thead>
<tbody>
<tr>
<td>Winter 1997</td>
<td>3.24</td>
<td></td>
</tr>
<tr>
<td>Winter 1998</td>
<td>3.23</td>
<td></td>
</tr>
<tr>
<td>Winter 1999</td>
<td>3.41</td>
<td></td>
</tr>
<tr>
<td>Winter 2000</td>
<td>2.61</td>
<td>2.77</td>
</tr>
<tr>
<td>Spring 2000</td>
<td></td>
<td>3.44</td>
</tr>
<tr>
<td><strong>Maximum possible</strong></td>
<td><strong>4.00</strong></td>
<td><strong>4.00</strong></td>
</tr>
</tbody>
</table>

However, the aversive response of these students may be a poor indicator of the intrinsic value of Primary Trait analysis. Because this was the second quarter in a sequence, students may have seen the grading sheets as an unfair change in the rules in midstream. This perception was probably exacerbated by the fact that V.L. Young and D.J. Goetz returned 80% of submitted reports for rewriting. These unacceptable reports each failed at least one Gateway Criterion (proofreading, formatting) and/or failed on two or more Primary Traits. In contrast, the two faculty not using the grading sheets apparently found no unacceptable reports, as they returned none for rewriting. Students seem to accept the grading sheets more readily when they know they are aware of them in advance. Although D.J. Goetz was the only instructor using the grading sheets in Spring 2000, his teaching evaluations were at normal levels, possibly because students were forewarned by friends in the Winter class. Grading sheets were used by all instructors from the beginning of Fall 2000, and have not generated dissatisfaction with fairness in grading, in spite of the high number of rewrites assigned compared to previous years.

Undoubtedly, the number of rewrites assigned increases with the implementation of grading sheets. In Fall 2000, one instructor assigned seven report rewrites. Most years he assigns none on the same two experiments. D. Ridgeway assigned seven rewrites, compared to his usual one or two for one experiment. M.E. Prudich assigned eleven rewrites, compared to the normal two or three for that experiment. Rewrites in Fall 2000 were assigned for failing a Gateway Criterion or failing one or more Primary Traits. This reflects a philosophy that each Primary Trait reflects a goal deserving at least some attention, and that Traits which may be comfortably ignored should just be deleted, along with their associated goals.
Interestingly, most rewrites were not due to poor performance on one of the Primary Traits, but rather to complete absence of the Trait. Traits such as uncertainties quoted for values, explanation of deviations from the procedure in the prelab report, and discussion of safety concerns were completely missing in some reports, even though they are on the grading sheet. This suggests that without the grading sheets, instructors often issue acceptable grades on reports that are missing elements listed in the course and departmental goals. It also suggests that students do not use the grading sheets effectively for self-evaluation. Indeed, one student admitted that his team didn’t look at the grading sheets, and just used their first report as a template because they had gotten a pretty good grade on it.

Based on student surveys and instructor observation, students used the grading sheets primarily to identify “categories” that should be in the report. Some even took the Traits as an outline of the report. This led to interesting phenomena, such as a few reports with a section entitled “Quantitative vs. Qualitative”, when this trait was intended to assess the nature of the discussion in the entire report. Such problems are being addressed by discussing the sheets in more detail on the first day of class and changing their format somewhat. Based on student surveys and instructor observation, students did not effectively use the descriptions associated with each Trait to assess their own performance. In some cases, after using the Primary Trait as a section title, they placed “D” or “F”-level content in the section. When surveyed, students said they found the wording vague. Many coupled this with a request for experiment-specific sheets, suggesting they want to be told exactly what to do and graded on their execution rather than having to determine what to do themselves. It appears our students need more practice both in self-evaluation and in higher-level thinking.

Both students and faculty agreed that the grading sheets provided a basis for specific discussion of weaknesses to be addressed. Faculty said that rewritten reports were certainly higher in quality than first submissions. However, faculty were not convinced that first submissions were improved over previous years. Also, faculty offered the general opinion that the current class is the weakest in recent years.

A Useful Tool for Grading?

As of January 2001, grading sheets for UO Lab have been used by five instructors. All used the sheets to grade reports; none graded reports and then filled in the grading sheets afterwards. One found grading with the sheets to be slower because it took time to become familiar with the sheets and the criteria being emphasized. One found that the sheets made little difference in time spent grading. Three, including one experienced in the class and one teaching it for the first time, found the sheets made grading much faster. All three said it was because they were more confident that they were grading consistently, and did not need to refer frequently to previous reports to make sure they had assigned similar grades for similar work.

All instructors agreed that report grades were different than they would have been without the grading sheets, with some grades higher and some lower. All agreed that using the sheets led to assigning more rewrites. Two felt that in the past, they may have given too much weight to the quality of writing, forgiving poor or lacking content for a well-written report. The grading sheets give heavy and consistent emphasis to content. After discussion of whether writing quality was
now receiving too little weight, the sheets were revised to make adjustments for writing quality more obvious.

Faculty felt that too often, student work did not quite fit the descriptions of any of the five levels (A,B,C,D,F), making it difficult to assign a score. The sheets were revised to describe only “A”, “C”, and “F” performance, making the scale more flexible. The notations “B” and “D” remained, but without descriptions. (Only “A”, “C”, and “F” were described for some traits in Winter 2000, and the “B” and “D” omitted. Students assumed that they couldn’t get a “B”, for example, and were upset.)

Some students complained that grading was not consistent from professor to professor. All instructors felt that the grading sheets assisted them in grading consistently during the quarter, and ensured more consistency between professors by defining the traits to be considered.

**A Useful Tool for Assessment?**

Primary Traits Analysis grading sheets make it easy to assess student achievement of many course goals. For example, in Winter 2000 only 50% of prelab reports scored at the C level or better in statistical analysis of results, and only 50% of postlab reports scored at the C level or better in determination and discussion of experimental uncertainty, suggesting that our students lack adequate skills in applying statistical methods or propagation of error. However, 100% of postlab reports scored at the C level or better, suggesting that after individual consultation with the instructor, students are able to correctly execute the statistical analysis. Results for fall 2001 were similar. Students in fall 2000 were surveyed to learn whether they felt adequately prepared for UO Lab by previous coursework. Students rated the Engineering Experimental Design course (which includes statistics) poorly. Their written comments stressed poor preparation in statistics and stated that they were unsure of what statistical methods to use. Instead of anecdotal reports from faculty that students may have weak preparation for statistical analysis, Primary Trait Analysis provides evidence that students are generally able to perform a specific statistical test, but that they show widespread inability to select an appropriate statistical method for use in data interpretation. This not only indicates that revising our Experimental Design course would be useful. It also suggests that this revision should be problem-based, giving student practice in answering for themselves the questions, “What do I want to learn about these numbers?” and “What statistical methods can I use to learn it?”

The assessment function of the grading sheets is facilitated by the use of spreadsheet software. In particular, the “countif” function of Excel is useful to determine how many reports receive particular grades on particular traits, and the plotting functions are useful to visually assess whether there is progress during the quarter.

**Conclusions**

Criterion-based grading using “Primary Trait Analysis” was adopted for grading laboratory reports with three goals in mind:

1. Guide student learning by specifying the traits of a good laboratory report.
(2) Improve consistency in grading and feedback to students by specifying the key traits to be considered.
(3) Collect data to assess student progress towards course goals.

Success in guiding student learning was mixed. The first hurdle was acceptance of the grading sheets by the students. It is important that the sheets be introduced at the very beginning of the course, remembering that students often think of the second course in a sequence as part of the same course. So far, students have not used the grading sheets effectively to evaluate and improve their work before submission. However, they have used the sheets to identify features that should be included in their reports (with some lapses). The sheets have also been valuable in focusing instructor-student discussions on the specific traits of their work that are lacking. With guidance from the sheets and the instructor, rewritten reports were superior to first submissions, though it remains uncertain (as of January 2001) whether the students have truly learned from this guidance and will improve in their first submissions. It should be noted that these students have rarely been asked previously to evaluate their own work. They are probably in the habit of thinking that they just do the work; the grade is the instructor’s responsibility and they have no control over that.

Based on student evaluations, students saw no clear advantage to the grading sheets in promoting grading consistency. Faculty, however, felt the sheets led to clear improvement in grading consistency, and in many cases made grading less burdensome. Sheets did require several revisions to balance specificity and flexibility. The grading sheets alone did not seem adequate as feedback to the students, at least partly because they were unable or unwilling to apply the descriptions to evaluate their work. However, the grading sheets did help the students focus their questions on the Primary Traits, leading to productive instructor-student discussion that eventually resulted in adequate performance on all the Primary Traits. The faculty perception was that the time devoted to individualized instruction was not reduced by using the grading sheets, but that the time was used more effectively.

Primary Trait Analysis proved very useful for assessing student learning. It produced data that could easily be converted to show progress towards the course and curriculum goals, and clearly showed areas requiring improvement and suggested the manner in which they might be improved. It is particularly valuable because the process of grading results automatically in the data for the assessment report; it is not necessary to grade the work and then later ask, “So, what did that mean?” It seems most useful for courses such as UO Lab, where mastery of important skills cannot be tested with an exam question.

Bibliography

2. Current course materials for Unit Operations Laboratory (ChE 415 and ChE 416) available from the course archives link from the Ohio University chemical engineering website: http://webche.ent.ohiou.edu/