Paper ID #25274

# Validation of an Instrument to Measure Student Engagement with a Standards-Based Grading System

**Prof. Heidi A. Diefes-Dux, University of Nebraska, Lincoln**

Heidi A. Diefes-Dux is a Professor in Biological Systems Engineering at the University of Nebraska - Lincoln. She received her B.S. and M.S. in Food Science from Cornell University and her Ph.D. in Food Process Engineering from the Department of Agricultural and Biological Engineering at Purdue University. She was an inaugural faculty member of the School of Engineering Education at Purdue University. Her research focuses on the development, implementation, and assessment of modeling and design activities with authentic engineering contexts. She also focuses on the implementation of learning objective-based grading and teaching assistant training.

**Ms. Laura Melissa Cruz Castro, Purdue University, West Lafayette**

Laura Cruz is native to Bogotá, Colombia, where she obtained her Bachelor degree in statistics from Universidad Nacional de Colombia. Her passion has been always education. Laura Cruz is currently an Engineering Education Ph.D. student at Purdue University in West Lafayette, where she is working in Education Assessment, specifically in high-quality assessment practice in engineering education.

# Validation of an Instrument to Measure Student Engagement with a Standards-Based Grading System

## Introduction

This research paper presents the development and validation of an instrument intended to measure the engagement of students with standards-based grading (SBG) systems. Such systems can complement the use of backwards design [1], [2], a curriculum development strategy intended to improve student learning which is taking hold in engineering education. Increasingly, engineering instructors are working towards more clear identification of intended learning objectives, alignment of curriculum, and adoption of transparent, informative, and feedback rich assessment strategies. Instructors are in essence creating complex systems with multiple interconnected parts. Well-framed and delivered systems have the potential, among other things, to increase students' active involvement in their learning and enable metacognitive development [3]. Criterion-referenced assessment strategies, like SBG, measure a student's achievement based on clear descriptions of standards for behavior (e.g., learning objectives) irrespective of other students' performance [4]. These grading strategies are becoming common in engineering, as they are a means to pull the system parts into tighter and more meaningful alignment.

When designing such systems there are a number of elements that are created by instructors for sharing with students: the learning objectives, an articulation or demonstration of performance expectations with regards to the learning objectives (e.g., rubrics, exemplars), and students' feedback with respect to the learning objectives. The true potential of these systems to improve student learning can only be realized through the engagement of each student with these elements, where engagement entails active awareness of course learning objectives (the standards) and expected performance, planning to learn, accessing of feedback, and subsequent actions. Prior work has shown that instructors find the initial workload to create an SBG system considerable [5], and students, unfamiliar with such systems, do not take as much advantage of the learning opportunities afforded by an SBG system as instructors would hope [6]. The workload to create a criterion-referenced system, coupled with students' lack of engagement with the system and thus stunted improvement in learning gains, can be a deterrent to curricular transformation. A myriad of strategies can be employed to engage students in a grading system; for instance, checks-for-understanding of the rubric, peer and self-assessment, and reflection. The question is: How effective are these engagement strategies and in what contexts?

An ability to measure students' engagement with a grading system is necessary to gauge the success of the grading system and inform improvements. However, there is no instrument in the literature that helps one understand the degree to which students are engaged with a grading system. The purpose of this study was to design and then collect and assess validity evidence for an instrument that measures students' engagement with an SBG system.

## Background

This work draws on literature from a number of areas: standards-based grading, self-regulation, and student engagement. These areas are briefly described below with an indication of their relevance to this work.

***Standards-Based Grading.*** Standards-based grading (SBG) is a grading strategy whereby students' achievement is measured relative to well-defined learning objectives [2]. While Marzano [7] points to the mis-identification of standards-referenced grading as SBG, with the former being the better label for the description above, SBG has become the recognized name of the strategy. This grading strategy, along with other criterion-referenced grading strategies (e.g., competency-based), has its roots in K-12 education, where it has been promoted as providing greater transparency of expectations and more meaningful feedback to students and more actionable information about what students know and can do to instructors [8].

SBG is being implemented in a variety of engineering education courses for those same reasons (e.g. [5], [9], [10], [11]). SBG is seen as a means to: (1) achieve tighter alignment between learning objectives, curricula, and assessment, (2) increase the validity and reliability of grading data [7], [12], and provide formative feedback to students that is actionable. SBG was the type of grading strategy employed in this study. The development of this system for the context of this study was previously described in detail [12].

***Self-Regulated Learning & Engagement.*** The degree to which a student actively monitors and regulates processes related to aspects of learning is referred to as self-regulation. The processes include goal setting, planning and use of strategies for achieving the goal, use of resources, reactions to feedback, and the production of products. In a conceptual model of formative assessment and feedback, Nicol and Macfarlane-Dick [13] implied a number of ways in which students can actively engage with an assessment system; these include (1) goal setting based on expectations laid out in learning objectives, (2) accessing high-quality feedback, (3) interpreting feedback, and (4) taking advantage of opportunities to close the gap between expected and current performance.

Engagement is a variable that researchers have studied for decades. In the educational context, it has been defined broadly to refer to the extent to which students are actively involved in meaningful educational experiences and activities [14]. The time and energy students devote to educationally sound activities inside or outside the classroom is one accurate measure of engagement [15]. Engagement is a construct with both behavioral and affective components [16], and it has been tied to outcomes that are desired in educational contexts, such as increased learning, persistence in college, and graduation [17]. Two approaches to measuring the engagement of students are popular in the literature. The first uses observational studies and the second uses instruments or surveys that attempt to measure student engagement. There are a few examples of instruments that have undergone development and validation for higher educational contexts including one that measures the engagement of students in systems thinking [18] and one that focused on academic course engagement [16].

In this study, engagement is defined as students' active involvement with elements of the SBG system. The aspects of engagement with an assessment system, as described by Nicol and Macfarlane-Dick [13], were the focus of item development for the instrument for this study.

## Methods

***Setting and Participants.*** The instrument designed and assessed in this study was administered in a large (N=1600) first-year engineering course taught at a Mid-western university. This course began using an SBG system in 2013. By Spring 2017, the semester of this study, the SBG system was well developed [12]. The system included a complete set of learning objectives (LOs) for the course that were made available on a need-to-know basis to the students with each assignment and prior to exams. LOs were fully articulated with evidence for proficiency (see the proficient column in Table 1 for an example). Assignments were graded using rubrics in which the rubric items were mapped explicitly to those LOs (see Table 1). Each learning objective was assessed on a scale of proficient (100%), developing (80%), emerging (50%), insufficient evidence (0%), or no attempt (0%); where the level of achievement of a LO was based on the number of applicable pieces of evidence sufficiently demonstrated in student work (see Table 1). Such rubrics were used to assess student work on 12 problem sets, three midterm exams, and five project milestones. Grading of problem sets and project milestones was completed predominantly through the course learning management system, Blackboard Learn™. Students could access their overall grade through Blackboard's gradebook and their individual LO assessments and written feedback by drilling into the rubrics associated with assignments on Blackboard. Grading of exams was done both on students' written exam papers which were returned to students and in Blackboard rubrics. So written feedback on exams was provided on students' papers but the individual LO assessments were posted to rubrics on Blackboard.

**Table 1. Example of a learning objective with evidence of proficiency and assessment scale**

| **Learning Objective** | 07.05 Format plots for technical presentation [using MATLAB] | | | |
|---|---|---|---|---|
| **Proficient** | **Developing** | **Emerging** | **Insufficient Evidence** | **No Attempt** |
| Evidence items for proficiency:<br>1. Correct syntax for title<br>2. Correct syntax for xlabel<br>3. Correct syntax for ylabel<br>4. A descriptive title that references the problem context, the independent (x) variable, and the dependent (y) variable<br>5. Clear x-axis label with units<br>6. Clear y-axis label with units<br>7. Gridlines | 1 (of 7) missing or incorrect item from the proficient list | 2-3 (of 7) missing or incorrect items from the proficient list | 4 or more (of 7) missing or incorrect items from the proficient list | Did not attempt the graded item |

***Instrument Design.*** To measure students' engagement with the SBG system, a 26-item instrument (the *Engagement with SBG System Questionnaire*) was developed to consider three aspects of engagement with an SBG system (21 items) and students' beliefs about these aspects of engagement and learning-objective based grading more generally (5 items) (see Appendix). Overall, the instrument was envisioned to have a three factor structure based on self-regulation and engagement theory. These aspects entailed: use of the LO lists (9 items), timely review of feedback (7 items), and use of feedback as a guide for subsequent learning (10 items). Table 2 shows how the items for these three theoretical factors mapped to the aspects of engagement with an assessment system described by Nicol and Macfarlane-Dick [13]. The LO lists mapped to goal setting, timely review of feedback mapped to accessing feedback, and using feedback as a guide mapped to interpreting feedback and taking advantage of opportunities to close the gap between current and expected performance.

**Table 2. Proposed theoretical structure of the *Engagement with SBG System Questionnaire***

| Factor | Factor Description | Self-Regulation & Engagement Theory [13] | Question IDs Original Instrument | Question IDs After Item Elimination |
|---|---|---|---|---|
| 1 | Use of LOs lists | Goal setting | Q1-Q9 | Q1-Q3 ,Q5-Q7 |
| 2 | Timely review of feedback | Accessing feedback | Q10-Q16 | Q10-Q12, Q14-Q16 |
| 3 | Use of feedback as a guide for subsequent learning | Interpreting feedback and taking advantage of opportunities | Q17-Q26 | Q17-Q20, Q22-Q26 |

Items that explicitly pointed to an aspect of engagement with the SBG system made reference to particular graded assignments (i.e., problem sets (3 items), exams (3 items), and project milestones (1 item)) and, in the case of problem sets and exams, particular time frames associated with the three exam periods of the semester (i.e., early, mid, and late semester). A four-point Likert-scale (i.e. strongly agree to strongly disagree) with additional options of 'I do not recall' or 'No opinion' was used to rate the degree of agreement with each item. The Likert-scale was coded on a scale of 1 (strongly disagree) to 5 (strongly agree) with the additional options being coded as the mid-point response of 3 [19]. Less than 5% of responses were mid-point responses, meaning that the additional options were rarely used by students.

***Data Collection and Analysis.*** The instrument was administered in class via Blackboard during the last week of class. Of the 1600 students that completed the course with a grade, 1312 responded to the instrument. The students selected the same item response option for all 26 items ($n$ = 92) were deleted to ensure data quality. This problem of careless responses has been addressed by many sources, and the decision to eliminate these students' responses was to ensure that the items capture the most representative information possible by reducing the noise that these observations can generate [20]. After the deletion of these respondents, a missing value analysis was performed on the 1220 remaining observations. There were 131 students that did not answer at least one item; seven of these did not answer more than two items. These seven observations were deleted from the sample. The remaining 124 student responses that were missing one or two items were imputed using a multiple imputation procedure by chained equations in R [21]. The demographics for those that completed the instrument ($n$ = 1312) and those that were maintained in the study ($n$ = 1213) are shown Table 3. The demographics of those maintained in the study are comparable to those that originally responded to the instrument.

Of the 26 items, the three items that corresponded to the project milestones were deleted (Q4, Q13, and Q21). The project milestones were completed and monitored by teams of students, rather than individuals. These three items were eliminated because they showed a different behavior from those that referred to assignments and actions completed by individual students, and these items did not have major loadings in any of the factors. Two more items (Q8 and Q9) were deleted from the instrument since they showed loadings that were higher than 0.30 for two factors [22]. The theoretical three-factor structure then had 6 items associated with use of LO lists, 6 items associated with timely review of feedback, and 9 items associated with use of feedback as a guide for subsequent learning (Table 2).

**Table 3. First-year engineering student participant demographics**

| Demographic | Completed Instrument ($n$ = 1312) | | Maintained in Study ($n$ = 1213) | |
|---|---|---|---|---|
| | Number | Percent | Number | Percent |
| **Gender** | | | | |
| Male | 965 | 73.55% | 885 | 72.96% |
| Female | 329 | 25.08% | 313 | 25.80% |
| Other | 3 | 0.23% | 2 | 0.16% |
| No Indication | 15 | 1.14% | 13 | 1.07% |
| **Race** | | | | |
| White | 806 | 61.43% | 770 | 63.48% |
| Asian | 280 | 21.34% | 241 | 19.87% |
| Hispanic | 117 | 8.92% | 103 | 8.49% |
| Other | 38 | 2.90% | 34 | 2.80% |
| Black | 28 | 2.13% | 26 | 2.14% |
| No Indication | 43 | 3.28% | 39 | 3.22% |
| **International** | | | | |
| No | 1053 | 80.26% | 995 | 82.03% |
| Yes | 244 | 18.60% | 205 | 16.90% |
| No Indication | 15 | 1.14% | 13 | 1.07% |

Two of the five sources of evidence of validity [23], [24] were examined. The first was internal structure which refers to the extent to which the instrument items relate to each other and with the overarching constructs (theoretical factors). An Exploratory Factor Analysis (EFA) was performed to understand the internal structure of the instrument. Before implementing the procedure, a random sample of 606 students was selected to use in the EFA, the remaining observations were used to confirm the model. The EFA was performed using an oblique rotation as suggested by Fabrigar [25]. An oblique rotation was used because it could not be assumed that the factors were perpendicular/independent; in fact, the EFA showed that there was correlation between factors. The number of the factors to be retained was based on the parallel analysis results which included multiple criteria as mentioned by Schmitt [26]. The result of the parallel analysis was a seven factor structure. A Confirmatory Factor Analysis (CFA) was performed to verify the best factor structure ($n$ = 607). Since the theoretical structure contained three factors and the EFA suggested a seven factor structure, the global (one factor), three, and seven factor structures were compared using the CFA. The seven factor structure was found to be a good fit for the instrument. After assessing the structure of the instrument and identifying the factors, the Cronbach alpha was computed.

The second source of validity evidence was an examination of the relationship between student responses to the instrument and another variable [23], [24]. Prior work has shown that students who engaged in structured reflection on their learning accessed their feedback more often than those who do not [6]. So it is hypothesized that students who do reflections will score higher on proposed factor 2 (timely review of feedback) than those that do not. Of the 1213 students, 226 were enrolled in sections of the course that used structure-reflection [6]; 874 were enrolled in sections that did no or minimal (e.g., minute paper) reflection. To test the hypothesis that the scores obtained for factors associated with timely review of feedback were higher for sections that did reflection than those that did no or minimal reflection, a Kruskal-Wallis test and a

Wilcoxon Rank test were used. These two non-parametric tests were chosen over the parametric *t*-test due to the Likert scale data and the skewed distribution of the responses to the factors [27].

A demonstration of the use of the instrument was then completed and focused on students' responses for those students who were in sections of the first-year engineering course that did not implement any strategies to strengthen students' engagement with the SBG system. There were 874 student responses from these sections retained for this analysis. This demonstration provides a baseline measure of students' engagement with the SBG system used in this course. For each student, an average score was computed for each of the seven factors. The average value for each factor was obtained by using the student's responses to each item that belonged to a given factor.

## Results

The results of the EFA, the item loadings for each factor, are shown in Table 4. The amount of variance retained by the seven factors was 67.7%. The EFA suggested seven factors in the instrument as opposed to the three suggested by the theoretical structure. The general Cronbach alpha for the instrument without the items that were deleted due to cross loading was 0.94.

**Table 4. Summary of Exploratory Factor Analysis**

| Question ID | Factor | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Q1 | | | | 0.126 | | | **0.686** |
| Q2 | | | | | | | **0.891** |
| Q3 | | | | 0.114 | | | **0.590** |
| Q5 | | **0.788** | | | | | |
| Q6 | | **0.980** | | | | | |
| Q7 | | **0.866** | | | | | |
| Q10 | | | **0.826** | | | | |
| Q11 | | | **1.007** | | | | |
| Q12 | 0.186 | | **0.647** | | | | |
| Q14 | **0.914** | | | | | | |
| Q15 | **0.930** | | | | | | |
| Q16 | **0.863** | | | | | | |
| Q17 | | | | | | **0.719** | |
| Q18 | | | 0.108 | | **0.675** | 0.138 | |
| Q19 | | | | 0.146 | **0.833** | | |
| Q20 | | | | | **0.846** | | |
| Q22 | | | | **0.854** | | | |
| Q23 | | | | **0.683** | | 0.123 | |
| Q24 | | | | **0.852** | | | |
| Q25 | | | | 0.156 | | **0.689** | |
| Q26 | | | | | | **0.778** | |
| **Eigenvalues** | 2.506 | 2.393 | 2.145 | 2.018 | 1.897 | 1.645 | 1.639 |
| **% of Variance** | 0.119 | 0.232 | 0.334 | 0.430 | 0.520 | 0.599 | 0.677 |
| **Alpha coefficient** | 0.95 | 0.93 | 0.93 | 0.91 | 0.89 | 0.84 | 0.81 |

Using the factor structure generated by the EFA, the CFA confirmed the seven factor structure as being the most suitable (Table 5). For the seven factor model, the values for the CFI (Comparative Fit Index) and the TLI (Tucker Lewis Index), which compare a model with the null (one factor) model, are higher than 0.90, which indicates a good fit for the model [28]. The TLI also indicates that there is an improvement of fit when compared to the null model [29]. The RMSEA (Root Mean Square Error of Approximation) for the seven factor model is 0.085; a value closer to 0 indicates a better fit and values less than 0.08 are indicators of a good fit [29]. The RMSEA is slightly higher than 0.08, indicating an acceptable fit. The AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) values are used to compare the fit of different models to the same data; a lower value indicates a better fit. Here the values are lowest for the seven factor model. The seven factor structure differs from the theoretical structure in that it parses engagement by assignment type (problem sets versus exams) and beliefs. The confirmed best structure available for the instrument is shown in Table 6.

**Table 5. Confirmatory Factor Analysis (CFA) comparison of models**

| Model | Global Factor | Three Factors | Seven Factors |
|---|---|---|---|
| df | 210 | 186 | 167 |
| p | 0.000 | 0.000 | 0.000 |
| CFI | 0.523 | 0.709 | 0.935 |
| TLI | 0.470 | 0.672 | 0.919 |
| AIC | 37605.540 | 35516.054 | 32993.811 |
| BIC | 37790.698 | 35714.438 | 33275.957 |
| RMSEA | 0.216 | 0.17 | 0.085 |

**Table 6. Structure of *Engagement with SBG System Questionnaire***

| Factor | Factor description | Question IDs |
|---|---|---|
| 1 | LOs access for problem sets | Q1,Q2,Q3 |
| 2 | LOs access for exams | Q5,Q6,Q7 |
| 3 | Timely review of feedback on problem sets | Q10,Q11,Q12 |
| 4 | Timely review of feedback on exams | Q14,Q15,Q16 |
| 5 | Use of feedback as a guide for subsequent learning on problem sets | Q18,Q19,Q20 |
| 6 | Use of feedback as a guide for subsequent learning on exams | Q22,Q23,Q24 |
| 7 | Students' general beliefs about learning-objective based grading | Q17,Q25,Q26 |

The results of the Kruskal-Wallis and Wilcoxon Rank Sum tests are shown in Table 7. Students' scores for timely review of feedback on problem sets (factor 3) and exams (factor 4) were significantly higher for those students in structured reflection sections.

Figures 1 and 2 show histograms of the average scores for the seven factors for the 874 students that belonged to the sections that did not implement strategies to engage students with the SBG system. Each histogram shows the mean (red dashed line) and median (blue dashed line) of the average factor scores. Figure 1 provides a side-by-side comparison of students' engagement with the SBG system average scores as that engagement related to problem sets (factors 1, 3, and 5) versus exams (factors 2, 4, and 6). Figure 2 shows the average scores for students' beliefs about the SBG system (factor 7).

**Table 7. Timely review of feedback scores (non-reflection $n$ = 874; reflection $n$ = 226)**

| Factor | Reflection | Mean | Median | St. Dev. | Kruskal-Wallis | Wilcoxon Rank |
|---|---|---|---|---|---|---|
| 3 | No | 2.60 | 2.33 | 1.08 | 115.13* | 53792* |
| | Yes | 3.61 | 4.00 | 1.18 | | |
| 4 | No | 2.66 | 2.00 | 1.25 | 56.306* | 67664* |
| | Yes | 3.37 | 4.00 | 1.25 | | |

* $p < 0.0001$

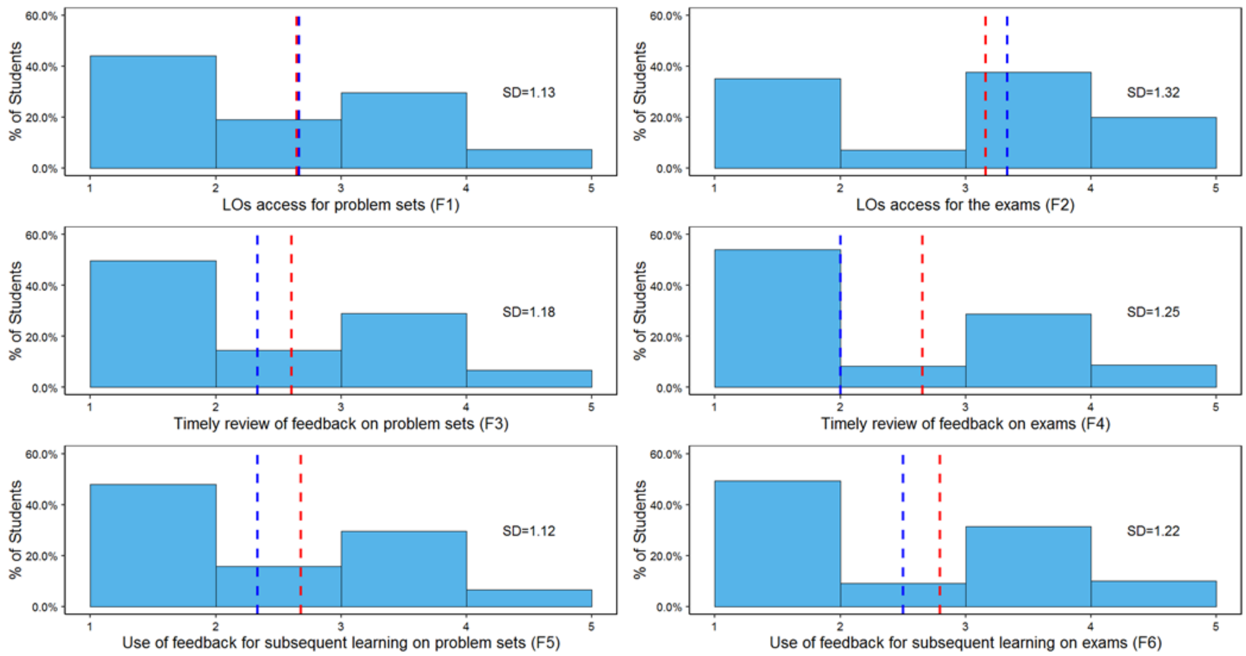


**Figure 1. Average scores for factors 1-6 for students in sections with no engagement interventions ($n$ = 874); the mean is a red dashed line and the median is a blue dashed line.**

## Discussion

***Design and Instrument Validation.*** The *Engagement with SBG System Questionnaire* was designed, based on theory, to measure three aspects of students' engagement with an SBG system: use of LO lists, timely review of feedback, and the use of feedback to guide learning on future assignments. In general, this internal structure bore out in the analysis of these first-year engineering students' responses to the questionnaire. However, those aspects needed to be parsed by assignment type (e.g., problems sets and exams) and general beliefs to improve the model fit. The distinct responses for problem sets and exams may be interpreted as differential student behavior when approaching these assignment types. This notion is better explored by looking at the results of the demonstration (see below).

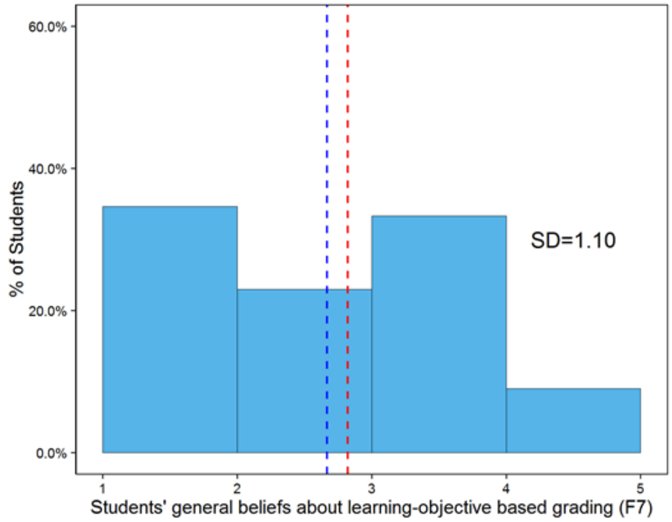


**Figure 2. Average scores for factor 7 for students in sections with no engagement interventions (*n* = 874); the mean is a red dashed line and the median is a blue dashed line.**

In addition, students' scores on the factors related to the timely review of feedback did show a difference based on whether students were enrolled in sections performing feedback structure reflection or not. As hypothesized, students in reflection sections scored higher. This is just one piece of evidence that the students' self-report of engagement with the SBG system does reflect their actual behavior. Further evidence could be gathered in the future for the other factors. Students' click-stream access the website on which the learning objectives are posted could be used to assess factors 1 and 2. Students' reflections on their feedback might be useful in assessing factors 5 and 6.

***Demonstration of the Instrument.*** The demonstration of the instrument was intended to provide a baseline of self-reported first-year engineering students' engagement with a well-defined SBG system under the condition that no interventions were employed to encourage students to engage with the system. Across factors 1 through 7, students' average engagement scores are dichotomous and always tending towards the lower ends of disagreement and agreement with statements about their use the LO lists, accessing their feedback in a timely manner, and using their feedback to guide their learning.

Students' average scores for factors related to problems sets (Figure 1: factors 1, 3, and 5) indicate that in excess of 40% of students strongly disagree that they use the LO lists, access their feedback in a timely manner, and use their feedback to guide their learning. The low use of the LO lists may be attributed to the LO lists being distributed to students in a separate document from the homework assignment each week. Further, only the most recent LOs were distributed with each homework. If students wanted to see LOs introduced earlier in the semester, they had to locate them with the associated assignment. When asked in an open-ended question at the end of this instrument what could be improved about the way the LO lists were provided, students mentioned the LO language being too vague or wordy, having too many documents to manage when doing assignments and a preference for embedding the lists into the homework document, and wishing the LO lists more resembled a rubric in format. These issues may have overwhelmed students upon their first access, deterring them from future use of the LO lists.

Students' strong disagreement with regards to the timely access and use of feedback to guide their learning was confirmed in a prior study of their click-stream Blackboard access to their grades and their feedback [6]. That prior study showed that approximately 65% of students accessed the gradebook 20 or more times during the semester (a number commensurate with the number of assignment with LO feedback) to view their assignment grades. However, less than 10% of students drilled into the rubrics that often to view their LO feedback. One issue may be that need to drill into each rubric via the gradebook. Students may either not know how to access their feedback or they are not inclined to make that effort based on their grade on a given assignment. If an assignment grade meets their expectations, students do not tend to attend to feedback [30].

Students' average scores for factors related to exams indicate that students did agree that they used the LO lists to prepare for exams more so than for problem sets (Figure 1: factor 2). There was anecdotal evidence to suggest that students used the LO lists as checklists when preparing for exams. The ease with which students could access these LOs was greater than on problem sets because the LOs for each exam period (about a third of the semester) were gathered into one document and the list was referred to in a number of course documents as being a resource for exam preparation. One could see students printing these lists and bring them to class with items highlighted for discussion during exam review sessions. The impetus for using the LO lists for exam preparation over problem sets may also be related to final course grade weighting of exams (36% total or 12% each) versus problems sets (12% total or 1% each). A student may question the time spent considering the LO lists for an assignment worth only one percent of their grade.

Students' strong disagreement to statements about timely access to their feedback on exams and use of their feedback on exams as a guide for learning (Figure 1: factors 4 and 6) may be attributable to the way in which feedback was provided to students. Unlike problem sets, where feedback was only given through Blackboard rubrics, exam feedback was provided on both students' exam papers and via Blackboard rubrics. Students could see their exam grade and some amount of written feedback physically on their written exams. Via Blackboard, they could only access their LO proficiency assessments (no written feedback). Students may have felt that they had all the feedback they needed on the exam papers.

Students' general beliefs about the SBG system are more negative than positive (Figure 2). This is not surprising given the low use of the LO lists and feedback. At least one other implementation issue may have contributed to this result in addition to those shared above. In reaction to an evaluation of grading quality [31], a grader-training program was being implemented for the first time in Spring 2017 to improve the reliability of grading and the quality of written feedback. While there were significant improvements, particularly with regards to the quantity of written feedback on problem sets, grading quality still varied considerably across the over 70 undergraduate graders for the entire course.

## Implications and Future Work

Clearly, these baseline results from the *Engagement with SBG System Questionnaire* show that these first-year engineering students were not all naturally predisposed to a high level of engagement with the SBG system. This means that they are not all engaging in a high level of planning or self-reflection on their performance. SBG system improvements and interventions intended to bolster student engagement with the system are warranted, and the *Engagement with SBG System Questionnaire* could provide a means to monitor the impact of any changes made. To start, reducing logistical issues may lower the barrier to student engagement with the system elements. While some logistical SBG system issues can be addressed, like finding less cumbersome ways to link the LO lists to the problem sets, others are not entirely within the instructors' power, like reducing the number of mouse clicks to access Blackboard rubrics. A second approach is to raise students' awareness of the value of engaging with the SBG system. This can be done through continued improvements to the grader training so that students feel they are receiving high quality feedback. Raising the profile of the feedback in the course would also be helpful. Meaning, instructors could discuss what students are and are not having difficulties with in class by showing summaries of the rubric-generated data. This encourages students to compare their LO performance and feedback to that of the class. Further, implementing structured and periodic student self-reflection on LO performance can heighten students' awareness of the SBG system elements available to help them succeed.

Further development of the instrument might include reworking the Likert scale so that the middle option is more directly interpretable from the responses and negatively phrasing some items would provide a cleaner check for careless responses. Future research with the instrument will look at how different forms of structured reflection impact students' engagement with the SBG system. Prior work with one form of structured reflection showed that students accessed their rubric feedback to a greater extent than those who did no reflection [6], but the click-stream data analysis was cumbersome to perform on a regular basis. The validity evidence for the *Engagement with SBG System Questionnaire* suggests that it holds potential for evaluating student engagement with the SBG system more easily.

## Conclusion

Evaluating the engagement of students with the different elements of a standards-based grading system is necessary because it provides instructors with important information about what materials students are using and how they are using them. The validity evidence for the *Engagement with SBG System Questionnaire* suggests that this measure of student engagement with an SBG system may be sufficient to help instructors make decisions about their SBG systems and the design of strategies for improving student engagement with their systems, while providing a way to evaluate the impact of those system changes and engagement strategies.

## Acknowledgement

This work was made possible by a grant from the National Science Foundation (NSF DUE 1503794 and NSF IIS 1552288). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

## References


[1] G. P. Wiggins and J. McTighe, *Understanding by design*. Second Edition. Alexandria, VA: Association for Supervision and Curriculum Development, 2005.

[2] C. A. Tomlinson and J. McTighe, *Integrating differentiated instruction & understanding by design: Connecting content and kids*. Alexandria, VA: Association for Supervision and Curriculum Development, 2006.

[3] T. A. Angelo, and K. P. Cross, *Classroom Assessment Techniques: A Handbook for College Teachers*. San Francisco, CA: Jossey-Bass, 1993.

[4] R. Glaser, (1963). "Instructional technology and the measurement of learning outcomes: Some questions," *American Psychologist*, vol. 8, no. 8, 519–521.

[5] E. Lee, A., Carberry, H. A., Diefes-Dux, and S. Atwood, "Faculty perception before, during, and after implementation of standards-based grading," *Australian Journal of Engineering Education*, vol. 23, no. 2, pp. 53-61, 2018.

[6] H. A. Diefes-Dux and L. M. Cruz Castro, "Student reflection to improve access to standards-based grading feedback," *in Proc. 48th ASEE/IEEE Frontiers in Education Conference, San Jose, CA, 2018.*

[7] R. J. Marzano, *Formative Assessment and Standards-Based Grading*. Bloomington, IN: Marzano Research Laboratory, 2010.

[8] P. L. Scriffiny, "Seven reasons for Standards-Based Grading," *Educational Leadership*, vol. 6, no. 2, pp. 70-74, 2008.

[9] S. L. Post, "Standards-based grading in a fluid mechanics course," in *Proc. American Society for Engineering Education Conference & Exposition, Indianapolis, IN, 2014*.

[10] S. Atwood, M. Siniawski, and A. Carberry, "Using standards-based grading to effectively assess project-based design courses," in *Proc. American Society for Engineering Education Conference & Exposition, Indianapolis, IN, 2014*.

[11] H. A. Diefes-Dux, "Student self-reported use of standards-based grading feedback and resources," *European Journal of Engineering Education*, 2018, [Online]. Available: https://doi.org/10.1080/03043797.2018.1483896 [Accessed Feb. 3, 2019].

[12] H. A. Diefes-Dux and H. Ebrahiminejad, "Standards-Based Grading derived data to monitor grading and student learning," *in Proc. American Society for Engineering Education Conference & Exposition, Salt Lake City, UT, 2018*.

[13] D. J. Nicol and D. Macfarlane-Dick, "Formative assessment and self-regulated learning: A model and seven principles of good feedback practice," *Studies in Higher Education*, vol. 31, pp. 199-218, 2006.

[14] C. N. Marti, "Dimensions of student engagement in American community colleges: using the community college student report in research and practice," *Community College Journal of Research and Practice*, vol. 33, no. 1, pp. 1–24, 2008.

[15] G. Kuh, "What we're learning about student engagement from NSSE: Benchmarks for effective educational practices," *Change Magazine of Higher Learning*, vol. 35, no. 2, pp. 24–32, 2003.

[16] S. Brown, A. Bowmar, S. White, and N. Power, "Evaluation of an instrument to measure undergraduate nursing student engagement in an introductory human anatomy and physiology course," *Collegian*, vol. 24, no. 5, pp. 491–497, 2017.

[17] E. T. Pascarella, "How college affects students: Ten directions for future research," *Journal of College Student Development*, vol. 47, no. 5, pp. 505–520, 2006.

[18] F. Camelia, and T. L. J. Ferris, “Undergraduate students’ engagement with systems thinking: Results of a survey study,” *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, no. 12, pp. 3165–3176, 2017.
[19] K. K. Tsang, “The use of midpoint on Likert scale: The implication for educational research,” *Hong Kong Teachers’ Centre Journal*, vol. 11, pp. 121–130, 2012.
[20] M. Revilla, “Impact of raising awareness of respondents on the measurement quality in a web survey,” *Quality and Quantity*, vol. 50, no. 4, pp. 1469–1486, 2016.
[21] S. V. Buuren and K. Groothuis-Oudshoorn, “mice: multivariate imputation by chained equations in R,” *Journal of Statistical Software*, vol. 45, no. 3, pp. 1-67, 2011.
[22] A. B. Costello, and J. Osborne, “Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis,” *Practical Assessment, Research & Evaluation*, vol. 10, no. 7, pp. 1–9, 2005.
[23] American Education Research Association, American Psychological Association, National Council on Measurement in Education, *Standards for Educational and Psychological Testing*, Washington, DC: American Education Research Association, 2014.
[24] D. A. Cook and R. Hatala, “Validation of educational assessments: a primer for simulation and beyond,” *Advances in Simulation*, vol. 1, no. 31, 2016.
[25] L. R. Fabrigar, D. T Wegener, R. C. Maccallum, and E. J, Strahan, “Evaluating the use of exploratory factor analysis in psychological research,” *Psychological Methods*, vol. 4, no. 3, pp. 272–299, 1999.
[26] T. A. Schmitt, “Current methodological considerations in exploratory and confirmatory factor analysis,” *Journal of Psychoeducational Assessment*, vol. 29, no. 4, pp. 304–321, 2011.
[27] J. C. F. De Winter and D. Dodou, “Five-point likert items: t test versus Mann-Whitney-Wilcoxon,” *Practical Assessment, Research & Evaluation*, vol. 15, no. 11, pp. 1–16, 2010.
[28] T. Brown, *Confirmatory Factor Analysis for Applied Research*, New York, NY: The Guilford Press, 2006.
[29] R. B. Kline, *Principles and practice of structural equation modeling*, 4th ed. New York, NY: The Guilford Press, 2016.
[30] B.R. Crisp, “Is it worth the effort? How feedback influences students’ subsequent submission of assessable work,” *Assessment & Evaluation in Higher Education*, vol. 32, no. 5, pp. 571-581, 2007.
[31] N. M. Hicks and H. A. Diefes-Dux, “Grader consistency in using standards-based rubrics,” *in Proc. American Society for Engineering Education Conference & Exposition, Columbus, OH, 2017.*

## Appendix

**Table A-1. Engagement with SBG System Questionnaire**

| No. | Factor | Question text |
|---|---|---|
| Q1 | 1 | While completing problem sets early in the semester (PS 1-3), I referred to the LO lists posted on Bb in the problem set zip folder. |
| Q2 | 1 | While completing problem sets in the middle of the semester (PS 4-7), I referred to the LO lists posted on Bb in the problem set zip folder. |
| Q3 | 1 | While completing problem sets later in the semester (PS 8-11b), I referred to the LO lists posted on Bb in the problem set zip folder. |
| Q4 | X | While completing the project milestones, I referred to the LO lists posted on Bb and/or embedded in the milestone instructions. |
| Q5 | 2 | While preparing for Exam 1, I referred to the LO lists posted on Bb. |
| Q6 | 2 | While preparing for Exam 2, I referred to the LO lists posted on Bb. |
| Q7 | 2 | While preparing for Exam 3, I referred to the LO lists posted on Bb. |
| Q8 | X | I believe that having access to a list of LOs associated with an assignment makes me aware of what I need to learn. |
| Q9 | X | I believe that having access to a list of LOs associated with an assignment guides my study habits. |
| Q10 | 3 | Early in the semester, I reviewed my performance on the LOs soon after feedback on a problem set was released (PS 1-3 Bb rubric feedback). |
| Q11 | 3 | In the middle of the semester, I reviewed my performance on the LOs soon after feedback on a problem set was released (PS 4-7 Bb rubric feedback). |
| Q12 | 3 | Later in the semester, I reviewed my performance on the LOs soon after feedback on a problem set was released (PS 8-11b Bb rubric feedback). |
| Q13 | X | I reviewed my team's performance on the LOs soon after feedback on a project milestone was released (M1-3 Bb rubric feedback). |
| Q14 | 4 | I reviewed my performance on the LOs associated with Exam 1 soon after feedback was released (Exam 1 Bb rubric feedback). |
| Q15 | 4 | I reviewed my performance on the LOs associated with Exam 2 soon after feedback was released (Exam 2 Bb rubric feedback). |
| Q16 | 4 | I reviewed my performance on the LOs associated with Exam 3 soon after feedback was released (Exam 3 Bb rubric feedback). |
| Q17 | 7 | I believe that having my work assessed based on LOs keeps me aware of what I have and have not learned. |
| Q18 | 5 | Early in the semester, I used the assessment of my performance on one problem set LOs to guide my work on the next problem set (PS 1-3). |
| Q19 | 5 | In the middle of the semester, I used the assessment of my performance on one problem set’s LOs to guide my work on the next problem set (PS 4-7). |
| Q20 | 5 | Later in the semester, I used the assessment of my performance on one problem set’s LOs to guide my work on the next problem set (PS 8-11b). |
| Q21 | X | I used the assessment of my team's performance on one project milestone’s LOs to guide our work on the next milestone. |
| Q22 | 6 | I used the assessment of my performance on the LOs to guide my preparations for Exam 1. |
| Q23 | 6 | I used the assessment of my performance on the LOs to guide my preparations for Exam 2. |
| Q24 | 6 | I used the assessment of my performance on the LOs to guide my preparations for Exam 3. |
| Q25 | 7 | I believe that having my work assessed based on LOs guides my study habits. |
| Q26 | 7 | I believe that all courses should use LOs as the basis for assessment of my work. |