International Journal of Teaching and Learning in Higher Education 2014, Volume 26, Number 2, 276-286
http://www.isetl.org/ijtlhe/ ISSN 1812-9129

Formative Plus Summative Assessment in Large Undergraduate Courses: Why Both?

Nirit Glazer
University of Michigan

One of the main challenges in large undergraduate courses in higher education, especially those with multiple-sections, is to monitor what is going on at the section level and to track the consistency across sections in both instruction and grading. In this paper, it can be argued that a combination of both formative and summative assessment is necessary in order to cope with the aforementioned challenge. A combination of the two types of assessment is necessary so instructors can provide formative assessment for learning and summative assessment for assuring that the formative assessment is done appropriately. In addition, the combination of the two also aids in other instructional challenges such as time management, instructor training, and balancing coursework overload. The proposed instructional perspective is illustrated by the Assessment Clock model that shows when to conduct the various assessment tasks, their frequency, and by whom, along with supplementary explanations and clarifications.

This paper focuses on the use of assessment to enhance consistency, particularly grading and instruction efficiency, in large post-secondary courses. Typical large post-secondary (i.e., higher education) courses include a head instructor, usually a faculty member, and several teaching assistants (TAs), typically graduate students. The head instructor is responsible for designing the course and delivering conceptual lectures, while the TAs often teach the hands-on labs and/or discussions, called sections. Large introductory level courses of 800 to 1,000 students might have as many as 40-50 sections, each with 20-30 students. In addition to the lectures, the head instructor is also responsible for coordinating the multiple sections and mentoring the TAs. One of the main challenges in such large courses is monitoring the individual section activities and tracking consistency across sections in both instruction and grading. It is important that all students be graded on the same basis regardless of the section to which they have been assigned.

The desire and call for consistency in teaching and grading across sections in a multiple-section course is mandatory, but unfortunately it has received little attention in the research literature. Nevertheless, it is a practical problem that has been observed and reported in the practical literature, such as in Mckeachie's (2002) *Teaching Tips* book. There is often a lack of consistency in teaching and grading practices as well as diversity in leniency/strictness even when all sections follow the same curriculum and the same grading guidelines.

The call for consistency is not limited to the course level. Head instructors need to ensure consistency, across sections within a given semester and across semesters, by comparing course grade distribution with that of the course sections of previous years. In addition, the head instructor has to "keep the distribution of grades consistent with that of other courses offered in the same department or school" (Ozaktas, 1994, para. 26). Arbitrariness in grading can result in unfairness and distortionary effects, such as students preferring courses by instructors issuing easier grades rather than courses for their educational content or instructors for their teaching ability. Some institutions have guidelines at the department level, such as a distribution policy of 40% A, 50% B, and 10% C.

Assessment should be equitable and fair. In higher education, especially in the case of large courses and multiple instructors, "whether it is in grading 1200 examinations or in assessing as many lab reports, first and foremost criterion in the grading rubrics is the desire and call for consistency" (B. P. Coppola, personal communication, March 20, 2006). Monitoring consistency in grading across sections throughout the semester, and between semesters and courses, is mandatory. In addition, grading issues should be one of the top priority topics to be elaborated in any TA training program, in course staff orientation, and in interactions between TAs and faculty instructors during ongoing staff meetings. Therefore, course coordinators, associated authorities such as department policy makers and the research community should focus more on the problem of a consistent grading system. To promote fairness and equality in an attempt to improve instruction in undergraduate education, it is necessary to have a combination of both formative and summative assessments, especially in large courses with multiple sections.

Assessment—Review of Relevant Literature

Assessment serves many purposes and can be implemented in many forms. Policy makers and administrators use it, among others, to track progress

and to make statistical comparisons across groups of students for budgetary decisions. In the classroom, teachers use assessment activities to monitor achievement and learning by students. In addition, teachers can use assessment tools to identify student misconceptions and also to identify strengths and weaknesses in the curriculum.

Beyond its role in student learning, assessment affects student lives. Performance on assessment activities often determines which students get into college and which colleges they attend. Assessment activities can result in achieving a degree from a first class college or from a lower class college. A fair and reliable assessment could better indicate who is really at the top. Particularly at the college level, assessment has a high value since it serves also for certification purposes. Educators should therefore pay extra attention to assure that assessment practices are not only meaningful for learning, but are also fair and consistent with respect to instructors, courses, years, and institutions, and that a student, regardless of the section/semester he/she is enrolled in, would receive the same course letter grade. Assessment is a key component in the learning cycle and should be valid, reliable, and transparent. Validity and reliability are the heart of assessment discussions especially in large-scale assessment activities (Atkin, Black, & Coffey, 2001). In an equitable and just grading system, students ideally will achieve the same final letter or numerical grade regardless of the section or semester in which they are enrolled.

Two key strategies for classroom assessment have emerged and have been debated among education scholars: formative and summative. Formative assessment uses feedback to improve teaching and learning, while summative assessment measures what students have learned to certify a grade.

Formative assessment is any task that provides feedback to students on their learning achievements during the learning process. It includes, for example, open-ended response questions, essays, and performance tasks, such as posters, presentations or projects. It may also include closed-ended questions, such as multiple-choice questions, when used for providing feedback to guide the learner's growth. Race (2009) emphasized the importance of having qualified feedback by first restating an analogy he credits to John Cowan, "Assessment is the engine that drives learning" (p. 47), and then extending it to add that, "feedback is the oil that lubricates the cogs of understanding" (p. 47). Thus, the ways feedback is produced are important for achieving maximum efficiency of the learning process (Black & Wiliam, 2003, 2006; Nicol & Macfarlane-Dick, 2006; National Research Council [NRC], 2001; Race, 2009; Weurlander, Söderberg, Scheja, Hult, & Wernerson, 2012).

Formative assessment activities are ongoing and part of the learning process in the classroom; it features activities that provide feedback to the students and teachers during the learning process, rather than after a period of instruction. The main purpose of formative assessment is to contribute to student learning through the provision of by providing information about performance (Yorke, 2003). Formative assessment may also serve as a learning tool by students (Heady, Coppola, & Titterington, 2001). It brings up opportunities to integrate activities that encourage students to think critically and to practice lifelong skills, such as presentation, communication, analytical, and problem-solving skills, as well as to practice teamwork. The exposure to such lifelong skills could also help students who are not performing well on traditional assessment tasks to demonstrate their knowledge in alternative ways (Cerny, 2005; National Center for Fair and Open Testing, 1999).

Summative assessment is used for evaluation, in which there is limited or no feedback beyond the achievement report, and is usually a numerical or letter grade score. Summative assessment is an activity, typically a written test given at the end of a term, chapter, semester, year, or the like, for grading, evaluation, or certification purposes. Summative assessment includes, for example, closed-ended questions, such as multiple-choice, true/false, and fill-in-the-blank questions. It may also include open-ended response questions when used for evaluating achievements; high-stake tests, such as ACT, GRE, and SAT. Summative assessment may further include state-standardized tests which are designed for policy and budgetary decisions. The same questions could be originally designed and used for one purpose (e.g., a summative purpose) and may later be used for another purpose (e.g., a formative purpose). Glazer, Hofstein, and Bar-Dov (2002), for example, analyzed student responses to the questions on the national matriculation exam, which questions were originally used for high-school certification and which are to be used later on for formative purposes, specifically, for providing feedback to students about common difficulties, such as misunderstandings and misconceptions, to prepare them better for their matriculation exam.

Feedback

The usefulness and effectiveness of assessment depends on the quality of the feedback. Educators and policy makers recognize such feedback as an essential factor in student learning, and therefore they strongly recommend that such feedback be prioritized in the curriculum practice (Atkin et al., 2001; Black & Wiliam, 1998a, 2003; Nicol & Macfarlane-Dick, 2006; Quality Assurance Agency for Higher Education,

2000). However, in practice, this area is still in its infancy, and many instructors still struggle with providing productive and timely feedback.

Assessment is effective only if students or instructors use the information generated from an activity to help decide on the next learning activity (Atkin et al., 2001; Biggs, 1998; Black & Wiliam, 1998a; Cowan, 2003; Sadler, 1998).

Feedback should be targeted to enhance learning and motivate students to study. Therefore, feedback should be realistic with respect to expectations and should include, not only areas for improvement, but positive feedback as well (Race, 2009; Weaver, 2006). The literature provides several suggestions as to make the feedback more useful and how to encourage students to use the feedback appropriately.

One suggestion is to have clear criteria and to share the criteria with the students before the assessment assignment. It is also suggested to use descriptive criteria and detailed comments, rather than numerical scoring, to improve feedback (Butler, 1987). Frederiksen and Collins (1989) used the term "transparency" to express the idea that students must have a clear understanding of the criteria for grading their work before they start working on the assessment task. Ideally, it should be so transparent that students will be able to evaluate their own work in the same way that their instructors do.

Another suggestion is to engage students in the feedback process in order to enable them to take control of their own learning and thereby to enhance their learning (Black & William, 1998a; Boud & Molloy, 2013; Nicol & Macfarlane-Dick, 2006; Race, 2010; Yorke, 2003).

Still another suggestion is to avoid too much feedback. Instructor should set priorities and highlight the most useful comments. Similar to other disciplines, such as usability and computer user interface, feedback should comply with the three-click rule (Zeldman & Marcotte, 2009), in which, to avoid frustration, users should click no more than three times to find the desired content. Similarly, students should have to address no more than three major feedback items at a time.

A still further suggestion is to avoid generic comments, such as "excellent," "poor," or "try again." For example, when assessing a graph, rather than commenting to the student that "the x-axis and y-axis are bad," it would be preferable that the student receive specific guidelines of how to improve the axes. These guidelines could include how to label the axes correctly, how to scale them, or how to decide on their range of values in order to eliminate wide open spaces (i.e., dead areas).

Yet another suggestion is that an appropriate feedback should be timely and frequently made in order to avoid repeating mistakes and to practice acquired skills effectively and efficiently (Black & William, 1998b; Boston, 2002; Cowan, 2003; NRC, 2001; Weaver, 2006).

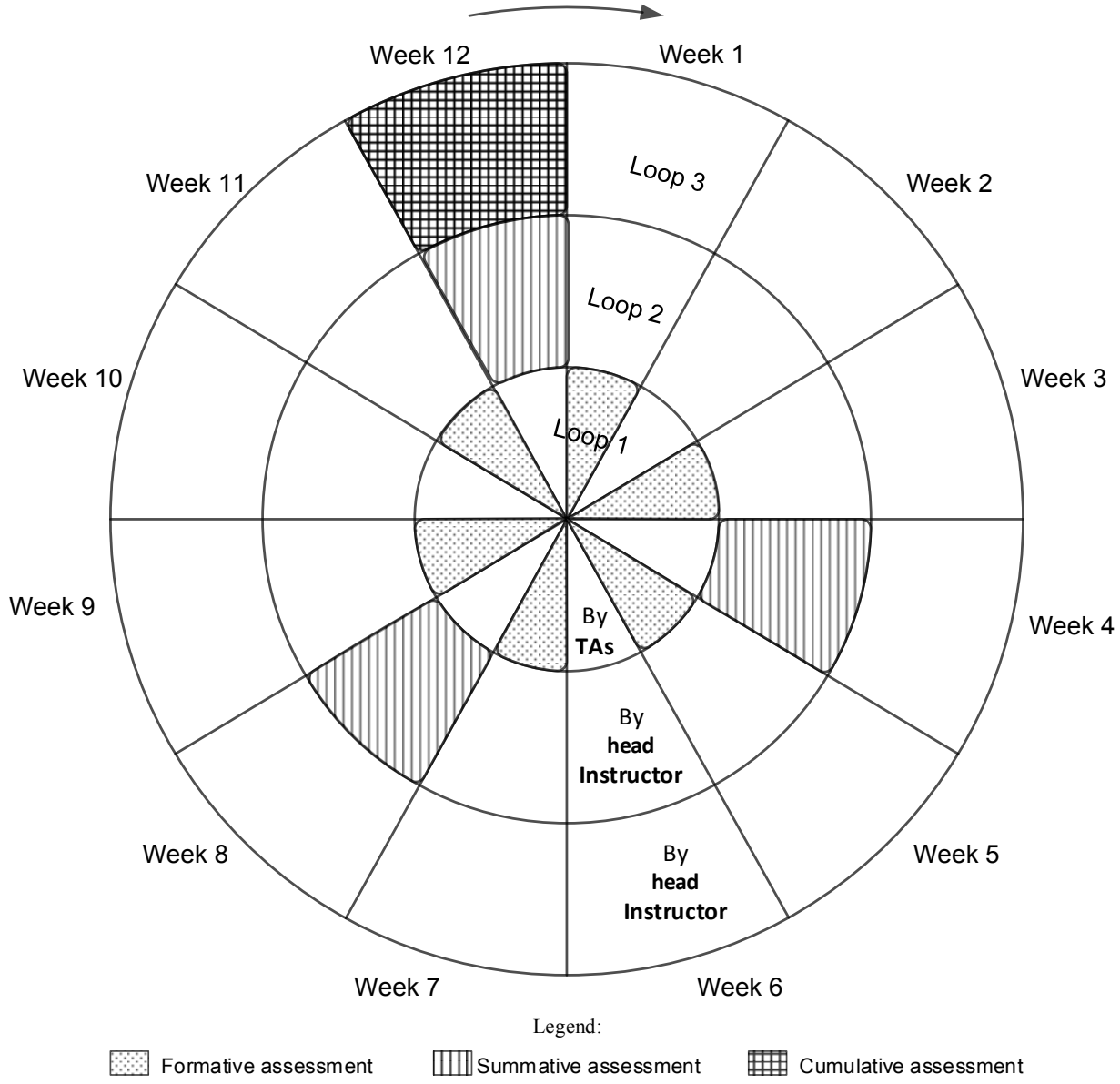
The Need for a More Consistent and Reliable Classroom Assessment

This paper focuses on the use of assessments to enhance consistency in grading across sections, and to inform instructors regarding diversity in leniency/harshness and when following grading guidelines. It deals only with classroom assessments that are part of the ongoing classroom life (e.g., assignments, exams, projects, and graded homework) involved in formal situations undertaken by the instructor of the course (Atkin et al., 2001). Such undertaken situations suggest the necessity of having both formative and summative assessment activities integrated together into multiple-sectioned courses, particularly in introductory courses at the college level.

Many papers describe the pros and cons of each assessment, formative and summative, and discuss which is more useful in various situations. Some have argued that formative and summative assessments are so different in purpose that they have to be kept apart (Black, Harrison, Lee, Marshall, & Wiliam, 2004). It is submitted, however, that in large courses at the college level, both forms are necessary, and one cannot be used effectively without the other, particularly those with multiple sections. It is also submitted that the assessment cannot be only summative because then students will not receive the sufficient feedback critical for learning. Nor can it be only formative because without outside and more objective tracking, the immediate classroom instructors might inflate grades and/or fail to cover essential material. By combining formative assessment with summative assessment (in an outside objective test that is run by the course coordinator), summative assessment will serve as a standardized test to compare the achievements of students from different sections, thereby reducing bias from subjective grading. Using the formative and summative combination method also provides more perspectives than a separate assessment and brings different forms of evidence together, which thereby increases the degree to which each assessment measures what it is intended to measure; thus using the forgoing method contributes to the validity (accuracy) of each assessment. Assessment validity is particularly important in higher education, since assessment plays a significant role in student life (Secolsky & Denison, 2012).

This instructional perspective is illustrated further by the Assessment Clock model below (Figure 1). In

Figure 1
The Assessment Clock Model



addition to the combination of formative and summative assessments, the Assessment Clock model also shows when to conduct the various assessment tasks throughout the semester, their frequency, and by whom, along with supplementary explanations and clarifications.

Instructional Implementation—The "Assessment Clock" Model

Figure 1 represents time, similar to a clock, by using patterns to allow clear observation of the frequency of the assessments and their types. As indicated by the legends, the "dotted" pattern corresponds to the formative assessment, the vertical lines pattern corresponds to the summative assessment, and the grid pattern corresponds to the final processing of all assessments. Such final processing includes the assignment of credit for non-graded aspects in the course, such as effort, safety in the lab, etc. The model is circular to show that assessments have a continuous effect on course instructions beyond a respective semester. The small arrow at the top of the Assessment Clock model

represents the continuation of the assessment process and its evolution from one semester to the next.

The balancing of formative and summative assessments is a key consideration in the model. This balancing can occur by strategically designing assessment tasks that use feedback procedures to enhance learning, and also objective baselines that allow comparisons across groups, for example, across multiple sections within the same course. To maximize consistency and to eliminate variations, it is recommended that both summative and formative assessment tasks be the same for all sections, regardless of their format.

Assessment Clock Model Structure

The proposed model entails three loops of assessment activities. The first loop (loop 1, Figure 1) is an assessment conducted by the immediate instructor (i.e., TA). The second loop (loop 2, Figure 1) is conducted by the course coordinator (i.e., head instructor). The third loop (loop 3, Figure 1) indicates the final grading process, where the head instructor determines the final grades and reports to an upper level authority (e.g., department, college, provost office).

The non-patterned areas in the model stand for the absence of any particular assessment task. During such times, the instructors should work with the students on the feedback they have previously received so that the students will be able to use that information effectively in their next assessment task.

Value of the Assessment Clock Model

Loop 1. In this loop, it is important to have frequent formative assessment as part of the ongoing instructional process. This can be done either individually, in pairs, or in teams, and can be planned as frequently as every two or three weeks. Tasks should be of the formative type, such as writing lab reports or doing poster presentations, where students have a chance to actively engage in the learning process and to benefit by being exposed to various learning skills.

Tasks, such as the lab reports or poster presentations, should be repeated as the semester progresses in order for the students to gain experience and to develop expertise in a specific skill. It is possible to have more than one type of formative assessment task, for example, writing lab reports and doing poster presentations. Each type of task should be repeated a few times every semester in order for the students to develop adequate important skills. To maximize consistency and eliminate variations, formative assessments, similar to summative assessments, are recommended to be the same for all sections. For the sake of uniformity, the assessments should be designed by the lead instructor and then could be followed and graded by the local instructors (TAs).

Loop 2. In this loop, it is important to have a summative assessment carried out two or three times in each semester. The tasks should include objective items, such as true/false, multiple-choice, and matching questions. The objective items will assure the equitability and consistency of the formative assessment guidelines with respect to multiple instructors and multiple sections of the same course. All students will do exactly the same summative tasks, ideally at the same time. In this way, the summative assessment will serve as the baseline for comparison with respect to groups of students and groups of instructors.

While the formative tasks in the immediate loop (loop 1) can be done to test either individuals or teams, the summative assessment (loop 2) should test the individual. Thus, the performance comparison between the formative and the summative assessment activities can highlight differences between an achievement of an individual and an achievement of a team; the latter does not necessarily reflect the understandings or skills of individuals in the respective team.

Loop 3. This loop occurs at the end of each semester, when the head instructor takes into account the performances of the students in loops 1 and 2, and assigns final letter or numerical grades. Decision, such as cut-offs, can then be used for normalizing grades. At this point, the immediate instructor will assign credit for non-graded aspects of the course, for example, efforts by students in the course, observations of safety procedures in the lab and contributions by individuals to the team efforts.

Example of an Assessment Clock Model Implementation

Table 1 illustrates the use of both types of assessment for providing constructive feedback to the instructor in an effort to improve grading and to maximize consistency in a large multiple-sectioned introductory chemistry course (of over 1,200 students, taught by 28 different TAs in 56 sections). All tasks, regardless of their format (summative or formative), were the same for all sections. The given example is from a large science class, but the assessment clock model is in fact useful across many disciplines.

The formative assessment (loop 1, Figure 1) in the example below was constructed of six sets, each set being constituted of lab reports and oral presentations (student-centered discussions), and the summative assessment (loop 2, Figure 1) was constituted of two written exams (mid-term and final), including mostly multiple-choice questions.

Table 1
An Implementation of the Assessment Clock Model

TA ID	Section #	Formative assessment (%)*	Summative assessment (%)
C	29	85.1	83.9
C	34	84.3	84.0
D	27	85.2	78.6
D	47	82.7	78.3
H	37	84.0	82.8
H	51	82.4	80.3
J	53	80.2	79.5
J	55	79.3	78.5
T	10	86.8	78.0
T	18	86.2	75.5
U	11	86.3	83.7
U	16	86.2	83.7
V	2	88.7	80.9
V	3	88.0	78.3
X	6	87.1	78.7
X	14	86.2	80.9
Course *M*	--	84.6	79.9
STD	--	2.14	2.9

Note. Formative assessment includes lab reports and presentations. Summative assessment includes the midterm test. *% of success up to the midterm test.

By comparing student performances in the various formative tasks, and their achievements in the summative (more objective) tasks, in a manner similar to the comparison in Table 1 above, instructors can receive feedback regarding the quality of instruction and assessment across the course sections. One common example was the case where the classroom instructor did not provide specific feedback and grades as generously as other instructors did. Typically, those sections performed poorly on the summative assessment tasks. The differences in the performances of the students in their formative assignments, and their achievements in the summative tasks showed up immediately.

If a section is performing exceptionally poorly or exceptionally well in the summative tasks, it is expected that the average performances in the formative tasks will be lower or higher, respectively, than the overall course average. If not, this would provide an alert to look for grading exceptions within the section, or to determine if the instructor grades too harshly or too leniently. This will also provide an indication whether there are one or more students who shift the average by underperforming or excelling. The comparison in Table 1 may prompt one to assume that the average performance trends would be similar in various types of tasks. However, such an assumption would be wrong because the comparison was made between the average of groups, rather than of individuals.

In case a problem can be identified, preventative actions can be taken during the semester, such as working closely and providing more guidelines for the instructor to teach and grade more appropriately. Thus, the combination of formative and summative assessment activities is necessary to create a mechanism for independent feedback in order to identify weaknesses in teaching quality. After the first summative task, the instructors may identify and correct problems of which they were not aware.

For example, instructors V, T, and X in Table 1 above appear to be too lenient. Their students received relatively higher scores in the classroom tasks; however, the exam scores were at about the average of the course, specifically, within the standard deviation (STD). In another example, instructor J appears to grade too strictly: the students in the respective sections received relatively low scores on the class work below STD, but average scores at the exam. Instructors C, D, and U, are examples of TAs that grade “just right”; their student scores for both the class work and the exam are within the course STD.

By combining the two types of assessments, such as in the latter examples in Table 1 the summative assessment can serve both as a preventative action and as a corrective action. Implementing an independent assessment as a comparison mechanism motivates instructors to follow guidelines more carefully and to provide better feedback, since any shortcomings in teaching quality will appear in the independent

summative assessment. In addition, after the first summative task, the instructors will be able to identify problems of which they were not aware, and to make appropriate corrections. The combination of formative and summative assessment is very useful also for comparing individual work to work done in a team when applicable. This combination has also other advantages, such as balancing the course workloads of students and instructors, and the management of time. However, the combination is deemed necessary mainly because it enables having assessment for learning, and creates a tracking system better assuring that the learning activities are done comparatively and are graded properly.

Model Implementation Challenges

Designing both formative and summative assessment activities, which can be integrated into the curriculum as part of the learning process, is both challenging and worthwhile. Time and training of instructors are the main challenges that are associated with implementing the Assessment Clock model in Figure 1.

Time. Time is one of the major barriers in implementing good assessment practices in the classroom; it is particularly challenging when employing both formative and summative assessment. Balancing time for instructions and time for assessment, during the instructional timeframe, becomes particularly challenging in the case of large class sizes in which the instructors are faced with large numbers of students and other constrains. Thus, the combination of summative and formative assessments helps with balancing time and course overload for both students and instructors.

Training. In an attempt to improve teaching quality at the college level, many departments now offer pedagogical training to new TAs. One implementation of such training is to incorporate assessment-related case study sessions of real-life TA situations, followed by teaching dilemmas (Coppola, 1996; Kerner, Black, Monson, & Meeuwenberg, 2002). The case study strategy exposes the new TAs to critical aspects of assessment, such as the need for quality feedback as well as a consistent grading system. The new instructors can thus better understand their roles and responsibilities and the importance of having both assessment procedures, one that provides feedback to the students, and one that allows comparisons, which increase objectivity and drive consistency with respect to sections and instructors.

A situation that frequently arises involves assessment practice and the issue of fairness in large multi-section courses. This is illustrated by a case study, developed by the author and schematically shown in Figure 2. All tasks in the case study,

Figure 2

A Sample Case Study for TA Training: Unfair Grading

Case Study for TA Training: Unfair Grading

>From: Student xxx@xxx.edu
>To: Head Instructor
>Subject: Grades

First off, I don't want you to think this e-mail is attacking you in any way. I just feel it is necessary to inform you of how the grading in chem125 is very unfair. My roommate and I both have CHEM125, but we have different GSI's [TAs]. We do many of our lab reports together and most of the time she ends up with a better grade. On top of that, she told me that her GSI informed her that her section had the highest lab report scores, but the lowest tests grades. Shouldn't this tell you something?

In addition, she had her last lab today. When I asked her how it went she said well; her GSI helped her out when they had trouble. Isn't that nice! Mine would not even give me a straight answer when I asked if we needed to include the net ion equation. This does not seem fair to me!!!!!

Discussion Dilemmas and Guided Questions:

What are the key issues presented in this case study? Why those key issues are so important?

If you (as a GSI, either the strict grader or the lenient grader) witnessed such an event, how would you respond to this particular situation?

Facts: The last lab is a "hands-on test" during lab time. The GSI served as a safety person and was not to answer any question regarding lab procedures.

Table 2
Results of Student Learning in Each Section of the Case Study of Figure 2

	Exam 1 average (%)	Exam 2 average (%)	Total average (%)
"Lenient" TA	76.6	74.5	82.2
"Strict" TA	83.0	78.9	84.6
Course Average	82.5	80.3	85.0

regardless of their format (summative or formative), were the same for all sections. The case study was developed from an email sent by a student to the head instructor and was followed by a discussion of dilemmas and guided questions, as set forth in Figure 2.

The results of student learning in the above case study, using the formative and summative combination method, are given in Table 2. Clearly, the students of the "lenient" instructor performed poorly on the summative assessment tasks (i.e., exams), in comparison to the course average; their instructor graded generously and did not provide as much specific feedback as the other instructors. In contrast, the students of the "strict" instructor performed at the course average on both the summative and the formative tasks. Overall, the students of the strict instructor finished the course with better final grades than the grades of the students of the lenient instructor.

Training sessions for TAs provide an opportunity for including practical sessions and for addressing issues in the grading of formative assessment tasks that are challenging. One such example is the grading of a student presentation in student center learning. Having students present their results in the form of oral presentations is a worthwhile learning experience. It allows students to form a greater understanding through the act of organizing their thoughts during an active verbal discourse (Kenny et al., 2002). It also provides invaluable opportunities for students to practice essential skills that are useful in their continuous learning and in everyday life such as data analysis (Glazer, 2011) and public speaking skills (Association of American Colleges and Universities, 2007; Schreiber, Paul, & Shibley, 2012). However, the grading of such an activity is very challenging, particularly when the instructor is required to evaluate the quality of the presentation and to provide appropriate feedback for making necessary corrections, all within a specified short time period. The complexity of such grading often causes a large diversity in the quality and quantity of the feedback given by instructors. This suggests a strong need for a simple grading rubric that is easy to interpret for aiding the TA to quickly grade the presentation. For example, a grading rubric that includes a list of criteria, which the instructor can evaluate quickly each criterion on a Likert scale while listening to the presentation, and finalize the total score later. A sample grading rubric for student oral presentation is provided in the Appendix.

Conclusion

Assessment has a critical impact on student life, both in providing appropriate feedback for enhancing learning, and in providing a grade, which can determine the career and academic opportunities of a student. Instructors should be concern about that impact and should adjust their teaching and grading, by using formative assessment for enhancing feedback and learning, and by using summative assessment for comparison purposes. The above argument shows the necessity of the combination of both. It also suggests a model for such combination in higher education courses, namely the Assessment Clock model in Figure 1. This model of assessment tasks represents just one of many options that an instructor should use. In the proposed model, determination of the frequency and the types of summative activities, in combination with the formative activities, are necessary for an effective assessment plan. In the proposed model, the summative activities are given no more than two to three times during a semester. The formative tasks are given more frequently, even as frequently as every other week; they are repeated in the same format as the semester progresses so that students will gain experience and develop expertise in a specific skill.

The literature clearly shows that formative assessment has a central role in enhancing learning. It is important, however, to consider real constraints since the implementation of quality assessment is time consuming for both students and instructors, and requires appropriate training of the instructors. Summative assessment is simpler to implement, especially in large courses, where technology assisted exams are commonly used. Therefore, the combination of formative and summative assessments helps with balancing work overload of instructors.

Similar to standardized tests that allow comparisons with respect to different schools and/or different teachers, the summative assessment tasks in a large college course allow comparisons with respect to different sections and/or different instructors of the same course. Results from such summative tests provide immediate feedback to the instructor regarding

the mastery of a subject area or of a specific skill by students in the instructor's section, in comparison to students from other sections. In addition, educators may use the summative results to improve instruction by providing information on how to better follow, more consistently, grading guidelines.

Consistency in grading is very important and is often neglected. Summative comparison across sections is critical to reduce differences among instructors, especially in the case of a multiple-section course. Arbitrariness in grading can result in unfairness as well as in distortionary effects, such as a preference by students for instructors in their grading (lenient, strict) or in their teaching ability, rather than for courses educational content. Using both summative and formative assessments is an important mechanism for identifying potential weaknesses regarding the instructions. It is also important for comparing the average achievements of groups of students in both assessments. Yet, before taking any further steps, the instructor should identify any exceptional students within the group that may shift the average significantly.

The literature shows that formative assessment with quality feedback enhances learning and achievement (Atkin & Coffey, 2003; Black & William, 1998a, 1998b; Boston, 2002; Bransford, Brown, Cocking, Donovan, & Cocking, 2000; Cowan, 2003; Yorke, 2003). It also shows that without informative feedback, students will exhibit relatively little progress their development. In addition, summative assessment increases objectivity and consistency with respect to various groups of students. Imagine a situation where students receive no feedback or instructors have no outside tracking system in place on their teaching quality and their grading. If students have only summative assessment, they will miss all the educational opportunities of feedback, and if they have only formative assessment, the grades may be inflated. The combination of the assessments is necessary so that there will be formative assessment for learning and summative assessment for assuring that the formative assessment is done appropriately.

References

Association of American Colleges and Universities. (2007). *College learning for a new global century: A report from the National Leadership Council for Liberal Education and America's Promise*. Washington, DC: Author. Retrieved from http://www.aacu.org/leap/documents/GlobalCentury_final.pdf

Atkin, J. M., Black, P., & Coffey, J. (Eds.). (2001). *Classroom assessment and the national science education standards*. Washington, DC: National Academy Press.

Atkin, J. M., & Coffey, J. E. (2003). *Everyday assessment in the science classroom*. Arlington, VA: NSTA Press.

Biggs, J. (1998). Assessment and classroom learning: A role of summative assessment? *Assessment in Education, 5*(1), 103-110. doi:10.1080/0969595980050106

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). *The nature and value of formative assessment for learning*. Unpublished manuscript, King's College London, London, UK.

Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education, 5*(1), 7-73. doi:10.1080/0969595980050102

Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, *80*(2), 139-148.

Black, P., & Wiliam, D. (2003). In praise of educational research: Formative assessment. *British Educational Research Journal, 29*(5), 623-637. doi:10.1080/0141192032000133721

Black, P., & Wiliam, D. (2006). Developing a theory of formative assessment. In J. Gardner (Ed.), *Assessment and learning* (pp. 81-100). London, UK: Sage.

Boston, C. (2002). The concept of formative assessment. *Practical Assessment, Research and Evaluation, 8*(9). Retrieved from http://PAREonline.net/getvn.asp?v=8&n=9

Boud, D., & Molloy, E. (2013). Rethinking models of feedback for learning: The challenge of design. *Assessment & Evaluation in Higher Education, 38*(6), 698-712. doi:10.1080/02602938.2012.691462

Bransford, J. D., Brown, A. L. Cocking, R. R., Donovan, M. S., & Cocking, R. R. (Eds.). (2000). *How people learn: Brain, mind, experience, and school* (2nd ed.). Washington, DC: National Academy Press.

Butler, R. (1987). Task-involving and ego-involving properties of evaluation: The effects of different feedback conditions on motivational perceptions, interest and performance. *Journal of Educational Psychology, 79*(4), 474-482. doi:10.1037/0022-0663.79.4.474

Cerny, C. (2005). Formative assessment. In B. Hoffman (Ed.), *Encyclopedia of educational technology*. Retrieved from http://eet.sdsu.edu/eetwiki/index.php/Main_Page

Coppola, B. P. (1996). Progress in practice: Teaching and learning with case studies. *The Chemical Educator, 1*(4), 1-13. doi:10.1007/s00897960050a

Cowan, J. (2003). Assessment for learning: Giving timely effective feedback. *Exchange*, *4*, 21-22. Retrieved from http://www.exchange.ac.uk/files/eissue4.pdf

Frederiksen, J. R., & Collins, A. (1989). A systems approach to educational testing. *Educational Researcher, 18*(9), 27-32. doi:10.2307/1176716

Glazer, N. (2011). Challenges with graph interpretation: A review of the literature. *Studies in Science Education, 47*(2), 183-210. doi:10.1080/03057267.2011.605307

Glazer, N., Hofstein, A., & Bar-Dov, Z. (2002, April). *New resources for educational enrichment: How a non-conventional research source revealed crucial learning difficulties in the Israeli chemistry curriculum*. Paper presented at the annual meeting of the National Association for Research in Science Teaching, New Orleans, LA

Heady, J. E., Coppola, B. P., & Titterington, L. C. (2001). Assessment standards. In E. D. Siebert & W. J. McIntosh (Eds.), *College pathways to the science education standards* (pp. 57-94). Arlington, VA: NSTA press.

Kenny, S. S., Thomas, E., Katkin, W., Lemming, M., Smith, P., Glaser, M., & Gross, W. (2002). *Reinventing undergraduate education: Three years after the Boyer Report.* Stony Brook, NY: Boyer Commission on Educating Undergraduates in the Research University.

Kerner, N. K., Black, B., Monson, E., & Meeuwenberg, L. (2002). Training instructors to facilitate collaborative inquiry. *Journal of Student Centered Learning*, *1*(1), 29-36.

Mckeachie, W. J. (2002). *Teaching tips: Strategies, research, and theory for college and university teachers* (11th ed.). Boston, MA: Houghton Mifflin.

National Center for Fair and Open Testing. (1999). *The value of formative assessment.* Cambridge, MA: Fairtest. Retrieved from http://www.fairtest.org/sites/default/files/FormativeAssessment_0.pdf

National Research Council (NRC). (2001). *Knowing what students know: The science and design of educational assessment*. Arlington, VA: NSTA Press.

Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education, 31*(2), 199-218. doi:10.1080/03075070600572090

Ozaktas, H. M. (1994). *A teaching manifesto: A personal view on undergraduate university education*. Retrieved from http://cgi.stanford.edu/~dept-ctl/tomprof/posting.php?ID=711

Quality Assurance Agency for Higher Education. (2000). *Code of practice for the assurance of academic quality and standards in higher education. Section 6: Assessment of students*. Gloucester, UK: Author.

Race, P. (2009). *Designing assessment to improve physical sciences learning*. Hull, UK: Higher Education Academy. Retrieved from http://www.heacademy.ac.uk/assets/ps/documents/practice_guides/practice_guides/ps0069_designing_assessment_to_improve_physical_sciences_learning_march_2009.pdf

Race, P. (2010). *Making learning happen: A guide for post-compulsory education*. London, UK: Sage.

Sadler, D. R. (1998). Formative assessment: Revisiting the territory. *Assessment in Education, 5*(1), 77-84. doi:10.1080/0969595980050104

Schreiber, L. M., Paul, G. D., & Shibley, L. R. (2012). The development and test of the public speaking competence rubric. *Communication Education*, *61*(3), 205-233. doi:10.1080/03634523.2012.670709

Secolsky, C., & Denison, D. B. (2012). *Handbook on measurement, assessment, and evaluation in higher education.* New York, NY: Routledge.

Weaver, M. R. (2006). Do students value feedback? Student perceptions of tutors' written responses. *Assessment & Evaluation in Higher Education*, *31*(3), 379-394. doi:10.1080/02602930500353061

Weurlander, M., Söderberg, M., Scheja, M., Hult, H., & Wernerson, A. (2012). Exploring formative assessment as a tool for learning: Students' experiences of different methods of formative assessment. *Assessment & Evaluation in Higher Education*, *37*(6), 747-760. doi:10.1080/02602938.2011.572153

Yorke, M. (2003). Formative in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher Education, 45*(4), 477-501. doi:10.1023/A:1023967026413

Zeldman, J., & Marcotte, E. (2009). *Designing with web standards* (3rd ed.). Berkeley, CA: New Riders.

NIRIT GLAZER holds a PhD in Educational Studies from the University of Michigan. Her research interests focus on the use of technology to enhance teaching, learning, assessment, and curricular activities, and on the study of usability and perception to promote emerging technologies in education at both secondary and college level education. Correspondence should be addressed to Nirit Glazer (nirit.glazer@gmail.com).

Acknowledgments

I would like to thank Dr. Joseph Krajcik, Christine Feak, and Benjamin Barish for their feedback, and Nancy Kerner for allowing me to collect data in her class.

Appendix
A Sample Grading Rubric for Assessment of Oral Presentations

Date: _____ Section#: _____ Team#: _____ Question#: _____				
%	**Criteria**	**Scale (circle one) Weak(1)…strong(10)**	**Comments**	**Total % (%*scale)**
10%	**Organization** Presentation includes introduction, main section, conclusions Each part is clearly defined	0…1…2…3…4…5 …6…7…8…9…10		
10%	**Introduction: the question/problem** is addressed & presented clearly	0…1…2…3…4…5 …6…7…8…9…10		
10%	**Conclusion** of the question/problem is addressed & presented clearly	0…1…2…3…4…5 …6…7…8…9…10		
25%	**Overall accuracy of content** (e.g., clear, scientifically correct, trend/relationship addressed correctly)	0…1…2…3…4…5 …6…7…8…9…10		
10%	**Appropriate use of evidences** The main points are made clearly and supported by evidence	0…1…2…3…4…5 …6…7…8…9…10		
20%	**Visuals** (clear fonts, appropriate titles, labeling, a reasonable choice for the types of visuals such as the type of the chart/tables)	0…1…2…3…4…5 …6…7…8…9…10		
5%	**General impression:** confidence, familiar with the material, a suitable pace for comprehension, appropriately loud, eye contact, and clear	0…1…2…3…4…5 …6…7…8…9…10		
5%	**Handling of Questions** Provides accurate and appropriate (length and depth) responses when answering questions to classmates or to the TA	0…1…2…3…4…5 …6…7…8…9…10		
5%	**Overall Effort**	Zero or 5%		
	Total (%)			
	Total (points)			
General comment:				