Supplemental, ELO-based Campus-wide Course Evaluations

A White Paper

[Mar. 3rd 2016]

Teaching Academy ad hoc committee on course evaluations (David Baum, Chair)
1. **Purpose of this Report**

At a time of increasing academic accountability and demands for transparency and data-driven educational program design, UW-Madison retains a fully decentralized system of course evaluations. The current system neither allows for neither comparative assessment across programs, centralized dissemination of data, nor consistent evaluation of our stated essential learning outcomes (ELOs; Appendix A). The UW-Teaching Academy Executive Committee sought to assess whether the current system of course and instructor evaluations could be modified to generate useful, campus-wide course evaluation data or if, instead, a new, supplemental course evaluation system should be developed.

2. **Abstract**

As a campus we need effective ways to evaluate how individual courses contribute to student progress in achieving our essential learning outcomes (ELOs). This is especially important at a time when budgetary allocations to units are made, in part, on credit hours, which is to say teaching *quantity*, with no mechanism to look at educational *quality*. Currently, the implementation of course evaluations is delegated to departments. Departmental course evaluation serve important needs in relation to improving course content and delivery and are also used to compare performance among instructors within a unit. However, as currently implemented, departmental evaluations have limited values for academic planning and analysis or for guiding students on course choice. There is variation across departments in the content of their surveys and no consistent mechanism for making data available to stakeholders. Furthermore, departmental evaluations typically focus on qualities of the instructor and pedagogical techniques rather than trying to assess the educational impact of a given course. To solve these problems, we considered the possibility of modifying how departments implement course evaluation mechanism but concluded that this would not be feasible and would undermine the ability of departmental evaluations to serve other important goals (e.g., obtaining course-specific feedback, helping in teacher assignments). Instead, we propose that central administration implement a supplemental course evaluation system, focused on ELOs, with summary data made widely available.

3. **Challenge Statement**

It has long been appreciated that all universities, but especially public institutions like UW, have an obligation to evaluate the quality of their courses. Course evaluations have typically been designed to serve two primary needs: (1) provide feedback to instructors on how to improve the course, and (2) provide departments, tenure review committees, award committee, etc. with data on relative teaching effectiveness of different instructors. However, this does not exhaust the potential uses of course evaluations. Evaluation data can be used by departments, colleges, and other administrative units to assess program quality (e.g., course evaluation data are sometimes included in program reviews).
Furthermore, course evaluations can be used by students to help them in course selection. Indeed, in this data rich age, students have an understandable expectation to be able to access information that might influence their personal educational choices.

The implementation of course evaluations at UW has historically been handled by the academic department responsible for the course and/or the departmental home of the instructor (these are not always the same). For a time, the Associated Students of Madison (ASM) collected and disseminated data on course evaluations, but this has not been institutionalized and ceased at least ten years ago.

Concerned at the lack of consistent course evaluation mechanisms, in 1996 the University Committee charged a Teaching Academy Task Force on the Student Assessment of Learning to work with representatives from the Associated Students of Madison and the Office of the Provost to address the following questions:

- What are the mechanisms for soliciting student evaluations of courses in different settings and class formats?
- What should be the balance between departmental autonomy and university-wide course evaluation guidelines, and to what extent should common elements be included in departmentally administered evaluations?
- How might departments and students most appropriately and effectively use student course evaluation information?
- What is the most effective mechanism to allow students to publish evaluations in a format that makes it easy to ascertain information about courses?

This task force conducted a thorough analysis, resulting in a detailed report (Appendix B). This report was presented to the faculty senate and yielded a motion that was presented to the senate at the April 7th 1997 meeting. This resolution contained four specific recommendations:

1. Whereas student course evaluation results are a useful mechanism to provide information to students about their course options, these results should be made available to students at the UW-Madison.
2. Whereas student course evaluation results are only one measure of the effectiveness of teaching at UW-Madison, assessment of faculty and instructional staff should not be based primarily on student course evaluation results.
3. Whereas the following areas are those identified by ASM and the Ad Hoc Committee on Student Course Evaluations as important to student course decisions, departments should develop questions that will address these areas. The areas identified are: (1) course organization, (2) clarity of instruction, (3) instructor accessibility, (4) grading fairness, (5) an overall course rating, and (6) an overall instructor rating.
4. Whereas student course evaluation results have most value when administered and disseminated fairly and uniformly, the recommendations and advice enclosed in the attached report should serve as guidelines for departments as they construct their own questions.

As a result of the report and resolution greater consistency was achieved across departments in the scale used and the questions included on course evaluation. However, there was a
never any system for sharing data with students, and departmental course evaluations have continued to differ in terms of number of questions, wording of questions, and whether a four- or five-point scale is used.

In the time since the 1997 task force a number of further changes have impacted the course evaluation landscape, of which notable examples are listed below.

- A number of departments have adopted online surveys as well as or in place of in-class scantron evaluations. In many cases the switch to online surveys has been accompanied by a drastic shortening of the survey in an attempt to achieve a reasonable response rate.
- The culture of radio-button rating has now permeated the internet, from Yelp to TripAdvisor. In the educational sphere, the market leader is RateMyProfessors.com. This site collects information from a small number of students who voluntarily score faculty in four categories (Helpfulness, Clarity, Easiness, Hotness) and provide a written review. It is well known that UW undergraduates often use ratemyprofessor.com when selecting classes.
- UW Madison adopted its own version of the AACU LEAP Essential Learning Outcomes (ELOs). These are organized in four umbrella areas: Knowledge; Practical and Intellectual Skills; Social Responsibility; and Integrative Learning. The ELOs have been used to structure the graduation survey and questions were added recently to a departmental course evaluation instrument developed by Testing and Evaluation.
- The UW has adopted a budget model that assigns resources, in part, based on credit hours, a measure of teaching quantity. This has highlighted the need to develop complementary measures of educational quality.

4. Proposed Solution(s)

a. Introduction to a Solution

In Spring 2012 the Teaching Academy Executive Committee formed an ad hoc subcommittee (Janet Batzli; David Baum; Jamie Henke, with later contributions from Brian Yandell) on course evaluations to revisit the question of whether a consistent, centralized mechanism could be adopted. This subcommittee obtained funding from the University Assessment Council to work with the UW Survey Center (UWSC) to try out an ELO-based survey and evaluate its effectiveness using cognitively-based group interviews with students. Additionally, the subcommittee held several meetings with administrators (Associate Vice Provost, the Associate Provost and Director of Academic Planning and Institutional Research, and the Vice Provost for Teaching & Learning) and with the University Committee.

The committee identified its goal as being to develop a campus wide course evaluation system that would allow uses such as the following:

- An educational analyst can query course data to better understand the educational development of students over their time at UW.
• A legislator or accreditor can assess the role of UW courses in generating knowledgeable, capable, and well-rounded graduates.
• A departmental Executive Committee can compare the educational impact of their courses with those of other departments.
• A divisional committee can compare instructors in terms of the learning that their courses engender.
• An advisor can look at historical data to guide a student on the sorts of educational impact to expect from a given course.
• A student can gain reliable data on the learning outcomes to expect from different courses.

The committee noted that the current system of course evaluations can yield valuable data for instructors and individual programs, but does not achieve the goals we had identified for campus-wide course evaluation, because (a) survey mechanisms vary across departments making them difficult to compare, (b) survey tools generally (and appropriately) focus on qualities of the instructor and pedagogical techniques, rather than on educational impact of the course on learning outcomes, and (c) data collected at the department level are not widely available.

One possible approach considered by the committee was to adjust existing department-based course evaluation mechanism to achieve the desired goals. This would require the following changes:

a. At least a subset of questions on all departmental surveys should be required to be the same
b. Data, or at least those for the shared questions, collected by departmental surveys should be required to be uploaded to a central database for access by students and other stakeholders
c. Requirements would have to be established to standardize survey implementation to improve comparability (e.g., online versus paper)
d. One survey would be required per course (rather than one per instructor).

We believe that such an approach would be impractical. First, we do not believe there is an effective mechanism to impose particular survey mechanisms on all units in all colleges. Second, even if there were, the administration of this mechanism, and particularly the uploading of data to a central database, would represent an undue burden on departmental administrations. Third, mandating a prescribed survey delivery mechanism is impractical. Finally, because departmental evaluations usually focus on individual instructors and TAs, and it is commonly the case that multiple instructors teach in a single course, it would not be practical to obtain useful summary data at the course level.

Given the many problems of imposing a consistent survey mechanism on departments we propose instead developing a supplementary online survey system that would be built to gather student self-reported of ELO impacts for all courses, with data collated into a centralized database.
b. Proposed Solution

The vision that emerged is that we should develop a short, supplemental course evaluation system that would be layered on top of departmental instructor evaluation to achieve the goals we have identified. This supplemental evaluation should be electronic, quantitative (non-textual), and implemented centrally. We also proposed that the supplemental survey be structured around the ELOs, while also asking at least one summative question on the overall educational impact of the course and one on how effectively the course was taught so as to maximize its educational impact.

An important component of the proposed solution is that aggregate, summary data for each course be made publicly available in a timely manner. This fact will, we believe, enhance student response rate and provide critical information to students at an opportune moment (help in selecting courses for the subsequent semester). Summary data could include response rate broken down by whether the student is or is not a major in the same division and could be enhanced by graphical summaries of the quality questions by ELOs, though the details of data presentation remain to be worked out.

Using such a mechanism, data collected by the supplemental course evaluation system should be available to department chairs and other administrators to be used in curriculum design. Additionally, when complemented with more targeted department course and instructor evaluation data it may be hoped that the results would be useful for calibrating campus-wide assessments of teaching quality in promotion, merit, and teaching–award decisions. In this regard, we would propose that the expectation should that a course impact at least one ELO (not that it necessarily yields gains in all ELOs).

A potential side benefit of our proposed solution is that it would force students to reflect on ELOs at the end of each semester. This should enhance knowledge of the ELOs in the student body, hopefully helping students better communicate what they are gaining from a UW education. In addition, the emerging data will foster discussion and normalize the use of ELOs among administrators, faculty and students.

c. Proposed Supplemental Course Survey

The ad hoc committee developed a draft ELO-based supplemental evaluation, which was assessed using cognitively based group interviews conducted by the UW Survey Center (UWSC), as reported in Appendix C. Among the students surveyed, the supplemental evaluation and the idea of making the data available were very positively received. Students seemed to appreciate the idea of a course evaluation that asks about various benefits a course may provide. Respondents did mention that there is great variability in how one might rate a humanities course versus a science course in regards to different learning goals, as one might expect. After having taken the survey and looking at the results, students seemed to have increased their understand of the ELO’s.

The ad hoc committee worked with the UWSC through three rounds of question-testing using focus groups. During the process, wording for individual questions was modified. Also, it came to be recommended that the second ELO umbrella, “skills,” should be divided into intellectual skills (“such as critical or creative thinking, quantitative reasoning, and
problem solving”) and practical skills (“written and oral communication, computer literacy, and working in teams”).

The following eight survey questions, while still in need of editing, define the starting point for further work.

1. At the time you enrolled, did you take this course primarily to fulfill a requirement for your major? [yes/no] Helpful for data interpretation This was assessed in round 3 and seemed fine. However, the committee now thinks that it might be better to use demographic data available for the student instead: for example whether they have declared a major in the same division as the course.

2. In general, how much did this course enhance your knowledge of the world, such as knowledge of human cultures, society, or science? [Not at all, A little, Somewhat, Quite a bit, A great deal] Relates to ELO #1 (Knowledge) This was assessed in round 3 and seemed fine.

3. How much did this course help you develop intellectual skills, such as critical or creative thinking, quantitative reasoning, and problem solving? [Not at all, …] Relates to ELO #2 (Skills), in part. This was assessed in round 3 and seemed fine.

4. How much did this course help you develop practical professional skills, such as written and oral communication, computer literacy, and working in teams? [Not at all, …] Relates to ELO #2 (Skills), in part. This was assessed in round 3 and seemed fine. The term “professional” might give the erroneous impression that the other learning outcomes (knowledge etc.) are not professional skills. Given this we might just call these “practical skills.”

5. How much did this course increase your sense of social responsibility, for example by increasing your knowledge of cultures or providing you with opportunities for civic or community involvement?” [Not at all, …] Relates to ELO #3 (Social responsibility) This question still caused problems and the Survey Center suggested we modify it, perhaps by providing more of a definition of social responsibility. Also, we were advised that the term “civic engagement” might be a useful addition.

6. How much did this course improve your ability to combine knowledge or skills from different fields of study? [Not at all, …] Relates to ELO #4 (Integrated learning). This was assessed in round 3 and seemed acceptable (though we still wonder if further improvements are possible).

7. How would you rate the overall educational value of this course, that is the extent to which the course improved your all-around education or prepared you for the future?” [Very poor, Poor, Fair, Good, Very good]. Based on the focus group, the recommendation was to cut either “all-around education” or “prepared you for the future” from the question. We would like to evaluate these two options in the future.

8. How would you rate the overall quality of this course, that is the extent to which it was structured and taught in order to maximize its educational value?” [Very poor, …] This was assessed in round 3 and seemed fine.

The focus groups led the ad hoc committee to believe that a survey composed of 7-8 radio-button questions per course would not be too onerous and that a reasonable response rate could be hoped for. If it is later felt necessary to increase the number of students who respond,
it might be possible to allow students to access the database of past scores in the period of course selection only if they have completed the requested course evaluations from the preceding semester.

**d. Suggestions for implementation**

There are many details will need to be worked-out by the body charged with implementing the proposed supplemental survey system, but we would like to highlight some principles and suggestions.

We believe that the optimal survey mechanism would entail an electronic survey being generated for each student soon after the end of each semester that would be pre-populated with the list of courses that the student had completed. Data entered by the student should automatically populate a database, suitably linked to relevant student demographic information drawn in from other sources.

While the raw data should be accessible only to a limited number of individuals, we imagine that a subset of data would be made publicly accessible. For example, we can imagine course means and standard deviation being available for each offering of each course section, plus perhaps means and standard deviations of each course across the last (say) five offerings. Further analysis is needed to decide on the minimum class size needed for data to be made available in the publicly accessible course evaluation database.

Ideally, the publicly accessible database will be linked directly to Course Guide so that students can readily obtain quantitative data for classes they are considering enrolling in. There should also be broad discussion of whom to grant access to the database: should access require a NetID or should it be open to the public?

There is much further work to be done on the presentation of data. One suggestion is that an initial view of a course should show: name the ELO that scored highest; give the score for that ELO; give the score for overall educational impact, and; give the score for teaching effectiveness. Then a user would be able to drill down to break the data down by year and section or to obtain means and standard deviations of all ELO scores.

Further discussion involving shared governance bodies would be needed to decide if overall metrics of course “quality,” perhaps a composite score of selected data (e.g., scores of the highest-scoring ELO, overall educational impact, and teaching effectiveness) should be extracted from these data.

**5. Next Steps**

Based on conversations with the Associate Vice Provost, the Associate Provost and Director of Academic Planning and Institutional Research, and the Vice Provost for Teaching & Learning it has become clear that the decision to invest in a supplemental course evaluation system, which would represent a considerable expense, would depend upon strong support from all three shared governance constituencies: faculty, academic staff, and students. Given this, we propose the following process.
1. This document, with an suggested motion (draft included as Appendix D), should be 
circulated to the University Committee, ASEC, and ASM and selected members of 
central administration.
2. If the proposed motion is passed, a joint implementation committee would be convened 
to (a) finalize the initial survey and design and interface for presenting the data, (b) guide 
the Provost in the design and eventual roll-out of the supplemental course evaluation 

6. Conclusion

The ad hoc subcommittee on course evaluations suggests that the Teaching Academy 
Executive Committee requests that steps be taken to develop a campus-wide 
supplemental course evaluation system and make summary, quantitative data available to 
campus users—namely students, instructors, and administrators. This system should not 
replace departmental course/instructor evaluations, which will continue to play an 
important role in guiding course improvement and in evaluating teaching. It is hoped that 
the resulting data will provide new insights into educational impact and that the 
assessment process will foster a broader campus discussion of the ELOs and encourage 
teaching practices that enhance the educational experiences of UW students.

Appendices

Appendix A – ELOs

Appendix B – 1997 Task Force report

Appendix C – USC Report

Appendix D – Draft Motion
Essential Learning Outcomes for UW–Madison Students

These learning outcomes were adapted from those developed through extensive national surveys and interviews done by the Association of American Colleges & Universities with employers, faculty, staff, and alumni, asking the basic question, “What qualities and skills do you want in college graduates?”

Beginning in their first year, and continuing at successively higher levels across their college studies, students should prepare for twenty-first-century challenges.

Knowledge of Human Cultures and the Physical and Natural World
- Through study in the sciences and mathematics, social sciences, humanities, histories, languages, and the arts
  Focused by engagement with big questions, both contemporary and enduring

Personal and Social Responsibility
- Civic knowledge and engagement—local and global
- Intercultural knowledge and competence
- Ethical reasoning and action
- Foundations and skills for lifelong learning
  Anchored through active involvement with diverse communities and real-world challenges

Intellectual and Practical Skills
- Inquiry and analysis
- Critical and creative thinking
- Written and oral communication
- Quantitative literacy
- Information, media, and technology literacy
- Teamwork and problem solving
  Practiced extensively across the curriculum in the context of progressively more challenging problems, projects, and standards for performance

Integrative Learning
- Synthesis and advanced accomplishment across general and specialized studies
  Demonstrated through the application of knowledge, skills, and responsibilities to new settings and complex problems

We urge you to explore these opportunities at www.learning.wisc.edu or contact the Division of Student Life at 608-263-5700, or the Office of the Vice Provost for Teaching and Learning at 608-262-5246.

Division of Student Life • Office of the Provost for Teaching and Learning

All photos by University Communications.
The Wisconsin Experience at UW–Madison

UW–Madison graduates become extraordinary citizens, community members and national and global leaders. We have produced more Peace Corps and Teach for America volunteers than almost any other university in the country. More leaders of major corporations have graduated from UW–Madison than any other university in the country. We are among the top producers of faculty members who teach at research-intensive institutions around the world. Something about the UW–Madison experience prepares our students to become outstanding leaders who are engaged locally, nationally and globally.

That “something” is the Wisconsin Experience. Grounded in the 100-year old Wisconsin Idea and our progressive history, our historical mission has evolved to create an expectation for all of us—faculty, staff, and students—to apply in and out of classroom learning in ways that have significant and positive impacts on the world. What we do matters, and together we can solve any problem. It is this unique Wisconsin Experience that produces graduates who think beyond the conventional wisdom, who are creative problem-solvers who know how to integrate passion with empirical analysis, who know how to seek out, evaluate, and create new knowledge and technologies, who can adapt to new situations, and who are engaged citizens of the world.

The Wisconsin Experience comprises the following inquiry-based, high-impact practices:

- Substantial research experiences that generate knowledge and analytical skills
- Global and cultural competencies and engagement
- Leadership and activism opportunities
- Application of knowledge in the “real world”

The nature of these opportunities and how we offer them makes UW–Madison unique in higher education. The resulting Wisconsin Experience is characterized by intentional integration of in- and out-of-class learning, by creative and entrepreneurial engagement in real world problems, and through active student leadership.

Here are a few examples of these opportunities:

- Service-learning and community-based research
- Research apprenticeships on competitively funded projects
- Summer internships in for-profit and non-profit organizations
- First Wave and other multicultural arts initiatives
- Peer mentoring in the classroom
- The Undergraduate Research Scholars Program
- First-Year Interest Groups (FIGs)
- Entrepreneurship opportunities for graduate and undergraduate students
- Intercultural dialogues through small group learning communities
- Training for health professions in rural, central city, and other underserved areas
- Leadership certificates in schools, colleges and majors
- Activism and leadership through student organizations and campus-community partnerships
- Comprehensive honors and honors theses in the major
- Year-long and semester-long study and research abroad
- National and international internships
- Residential Learning Communities
- Research communities for graduate students
- Graduate and professional student development
FINAL REPORT
AD HOC COMMITTEE ON THE USE OF STUDENT COURSE EVALUATIONS
(The appended resolutions were adopted by the Faculty Senate at its meeting on 6 October 1997.)

Committee Charge and Composition
The ad hoc committee was appointed by the University Committee to continue the debate on the use of student course evaluations. This debate was initiated by a report presented to the Senate in Spring 1996 prepared by a Teaching Academy Task Force on the Student Assessment of Learning, working with representatives from the Associated Students of Madison and the Office of the Provost. That report called for the potential institution of a standardized university-wide approach to student course evaluations—a proposal that raised many questions and concerns.

After seeking applications from across the campus, the University Committee appointed the ad hoc committee at the beginning of the fall 1996 semester. The committee includes 9 faculty members, 1 student, and 2 members of the academic staff.

The charge to the Ad Hoc Committee on the Use of Student Course Evaluations requested that it seek answers to the following questions:

• What are the mechanisms for soliciting student evaluations of courses in different settings and class formats?

• What should be the balance between departmental autonomy and university-wide course evaluation guidelines, and to what extent should common elements be included in departmentally administered evaluations?

• How might departments and students most appropriately and effectively use student course evaluation information?

• What is the most effective mechanism to allow students to publish evaluations in a format that makes it easy to ascertain information about courses?

The ad hoc committee interpreted its charge broadly, to make recommendations not just on the above questions, but also to examine the purposes and contexts under which student course evaluation are and are not useful, and to make some cautionary remarks on the use of student course evaluations.

We have taken the stance that our role is not to advocate for a position, per se, but to serve as a "thoughtful filter" for comments, concerns, and recommendations that we have gathered from the university community. We have also examined these comments, concerns, and recommendations in the context of the literature and institutional examples we could collect.

One final note must be made to set the context for this report: Our meetings with university attorneys assure us that course evaluations produced under administrative direction (either centrally or by a department) are considered open records. Therefore, even though student groups are only interested in publishing portions of teaching evaluations, all questions asked on an institutional evaluation form should be considered accessible upon request. We do not have to publish all questions unilaterally, of course, but the university community should be aware that they are subject to requests based on the open-records law.

(continued)
Summary of Meeting Chronology

It was important for this committee to act in as inclusive a manner as possible. Two public hearings were held to elicit feedback from the university community on the committee's recommendations, and other opportunities were offered to provide feedback to the committee at each step in its development of this report.

The committee held its first meeting on October 18, 1996 and held 3 subsequent meetings on November 1, 8, and 15. The meeting on November 8 was an interview with Professor John Centra, a nationally-known expert on course evaluations from Syracuse University, and the meeting on November 15 was a public hearing held to solicit input from the university community at large. Twenty-one individuals provided oral and/or written testimony at the hearing.

An interim report was presented to the faculty senate in its December meeting. The committee then met through the month of February to draft its recommendations, which were presented at a second public hearing on February 21st. The committee met on February 28th and March 10th to incorporate the comments from the public hearing, and then submitted this report on March 17th, for discussion at the April Faculty Senate meeting.

Introduction

To the best of the ability of the committee to synthesize the literature\(^1\), a modest and complex relationship exists between student course evaluations and student learning—that is, student evaluations do more than simply measure student satisfaction. The research literature suggests that carefully crafted questions can be helpful in evaluating the effectiveness of a course, in helping students learn more effectively and efficiently than they could on their own, can be used to measure more than student satisfaction, and the responses can be useful in student advising.

For example, student ratings correlate with student achievement, with independent ratings of instructors (those done both by other instructors and by alumni of the course), and with difficulty of course material (the harder the course, the higher the ratings). However, these relationships exist primarily when student course evaluations are administered in the right way and under the right contexts. Many of our recommendations, then, concern these right "ways" and "contexts."

It is the committee's opinion that student course evaluations serve a valid function within the enterprise of evaluating teaching and learning. They provide one way to evaluate what goes on in the classroom. Furthermore, students deserve to hear other students' opinions about courses they are considering. Because of this, and since Regents have mandated course evaluations, we did not pursue options to abandon course evaluations altogether.

The committee reviewed the course booklets published by student groups at several peer institutions. We also collected information from those we knew at some of these institutions to help us determine the range of issues that they faced. A sample page from the student booklet used at the University of Michigan is included in the Appendix to illustrate how one set of student course evaluations were successfully published, with no ill effects.

We identified three separate purposes for the use of student course evaluations: (1) for students to make course decisions, (2) for instructors' use to improve their teaching, and (3) for faculty and administrators to use in making faculty/staff merit and promotion decisions.

(continued)
Using students to rate courses is only one way to meet these three purposes, and in fact, it is the committee's opinion that student ratings only best accomplish purpose #1. The committee strongly recommends that the university institute additional mechanisms to evaluate teaching and courses for purposes #2 and #3. Several recommendations and advice given below identify what some of these mechanisms could be.

Organization of This Report

We begin our presentation of recommendations by listing those that are primary to the committee's discussion of the use of student course evaluations. These are followed by a set of recommendations concerning the publication of student course evaluations by student groups and then by a set of recommendations meant to assist colleges and departments in developing their own procedures for course and teaching evaluation.

In all cases, our recommendations and advice are followed by justification and amplification.

Central Recommendations

1. THE COMMITTEE RECOMMENDS THAT STUDENT EVALUATIONS ARE CONDUCTED, BUT THAT WE DO NOT ADOPT A UNIVERSAL FORM FOR ALL DEPARTMENTS AND COURSES.

Courses and departments differ too greatly for the same items to apply equally well to each. Although common formats (large lecture courses, small seminars, etc.) share similarities, there are enough differences in approach within different disciplines that even requiring common questions for similar formats can obscure cross-disciplinary variance.

FURTHERMORE, THE FOLLOWING AREAS ARE THOSE THAT THE COMMITTEE IDENTIFIED AS IMPORTANT FOR STUDENTS IN THEIR COURSE SELECTION. THE COMMITTEE RECOMMENDS THAT DEPARTMENTS DEVELOP QUESTIONS THAT WILL ADDRESS THESE AREAS:

A. Course organization, planning, or structure, including an assessment of the level of student/instructor interaction (whether or not the instructor uses a variety of learning strategies in the course) and a description of the course format.

B. Clarity of instruction, communication skills of the instructor.

C. Instructor accessibility outside of class (holds regular office hours, responds to e-mails, etc.).

D. Grading procedures (i.e., How well grading procedures are explained and followed).

E. An overall rating of the course.

F. An overall rating of the instructor.

2. THE COMMITTEE RECOMMENDS THAT THE UNIVERSITY CAMPUS ADMINISTRATION SHOULD NOT ITSELF PUBLISH COURSE EVALUATIONS--ELECTRONICALLY OR OTHERWISE. IF INDIVIDUAL DEPARTMENTS OR STUDENT GROUPS DESIRE TO DO SO, THAT SHOULD BE THEIR DECISION.
The committee is concerned that if campus administration publishes course evaluations, they will take on an undue weight relative to all other forms of evaluation. "Officially" published evaluations receive a "stamp of approval" that is not commensurate with their role in the evaluation process, nor with their merit. We are further concerned that a campus-sponsored publication may encourage faculty and instructional staff to "dumb-down" instruction in an effort to achieve higher ratings. The evidence is mixed as to whether there is a correlation between grading leniency and student ratings. There are, however, findings showing that first-year teachers have lower ratings than more experienced teachers. Therefore, we are additionally concerned that new instructors/probationary faculty might suffer if ratings from their first instructional efforts were made "officially" public by the campus.

3. THE COMMITTEE RECOMMENDS THAT QUANTIFIED QUESTIONS SHOULD ALL USE A COMMON SCALE FORMAT CONSISTING OF 5 POINTS, WITH 1 BEING "LOW" AND 5 BEING "HIGH." FURTHERMORE, ITEMS SHOULD BE ORDERED SO THAT THE "1" IS ON THE LEFT AND THE "5" IS ON THE RIGHT. SCALES SHOULD BE ANCHORED AT EACH POINT, USING "LOW" AND "HIGH" ANCHORS THAT ARE THEMSELVES EXTREME, AND WITH REMAINING ANCHORS THAT EQUALLY SPAN THE RANGE FROM "LOW" TO "HIGH."

Using a standard fixed-point scale will alleviate many of the problems that student groups encounter when the burden is on them to convert differently numbered scales to a common format. Since 5-point scales are currently the most commonly used format at the university, the committee thought it most efficient to make this format the standard.

Anchoring attaches a word or phrase with a point on a scale. Relating a word at each scale point increases the reliability of the ratings. Using extreme low and high anchors increases the usefulness of the scale. The specific anchors to use are to be left up to the discretion of the department, to best fit the questions asked.

4. THE COMMITTEE RECOMMENDS THAT FREQUENCY DISTRIBUTIONS SHOULD BE REPORTED IN COURSE EVALUATION SUMMARIES IN ADDITION TO REPORTING MEANS.

Means can be misleading. Frequency distributions (for example, percentages, bar graphs, histograms, etc.) can help differentiate between two courses, both of which have received an average rating of 3 on a 5-point scale, but one of which has received "mixed" reviews—both high and low responses, while the other receives uniformly average responses.

The committee discussed whether means, medians, or modes were the most appropriate measure of central tendency to report. While the committee was not unanimous in its opinion, it was felt that since means represent the most easily recognizable measure of central tendency, and since we were recommending that frequency distributions also be reported, those interested in obtaining medians and modes could do so based on the data that would be reported.

5. THE COMMITTEE STRONGLY RECOMMENDS THAT DEPARTMENTS DO NOT USE STUDENT EVALUATIONS AS THEIR SOLE SOURCE OF EVALUATION INFORMATION ABOUT COURSES.

As mentioned previously, student course evaluations are only one way to evaluate courses and instructors, and are not the best way to meet all three purposes listed at the beginning of this report—that is, in addition to helping students make course decisions, using evaluations for instructors' own

(continued)
feedback on their courses, and for merit/promotion decisions. Multiple evaluation tools insure that instructors maintain a high quality of instruction, and that student questionnaires do not become a mechanism for the "dumbing down" of instruction to manipulate student course ratings.

The two other mechanisms that have been identified most often are peer evaluations—having other instructors sit in on their peers' classes; and alumni evaluations—asking former students to evaluate the course well after they took it.

Lastly, it has been found that open-ended questions provide information that is very useful to instructors interested in evaluating their own teaching efforts.

6. THE COMMITTEE RECOMMENDS THAT FACULTY AND INSTRUCTIONAL STAFF ENGAGE IN THEIR OWN FORMATIVE EVALUATION AT MID-SEMESTER.

These evaluations, termed formative for their helpful nature, seek constructive feedback on the course while it is in progress. They differ from those termed summative, that are used for making decisions and are normally given at the end of the semester. The formative use provides a chance to make changes in response to the student comments, whereas the summative one provides only a chance to influence how the particular course is taught in a subsequent semester.

7. THE COMMITTEE RECOMMENDS THAT A "SUNSET CLAUSE" OF THREE YEARS SHOULD BE BUILT INTO THE PRESENT RECOMMENDATIONS ON THE USE OF STUDENT COURSE EVALUATIONS SO THAT THEIR ADMINISTRATION AND USE CAN BE EVALUATED AT REGULAR INTERVALS.

We need systematically to determine how evaluations are used by the university community, how evaluations are administered, and how they affect classroom teaching and student learning. This implies that the data generated from the use of student course evaluations must be stored in a manner that lends itself to longitudinal compilation, and a mechanism for longitudinal analysis must be developed and put into place.

8. THE COMMITTEE RECOMMENDS THAT THE SUBSEQUENT WORK RELATED TO STUDENT COURSE EVALUATIONS—EVALUATING ITS USE AND IMPLICATIONS, ASSISTING DEPARTMENTS IN THE DEVELOPMENT OF THEIR QUESTIONS, ASSISTING STUDENT GROUPS IN THE PUBLICATION OF THE COURSE BOOKLET, ETC.—SHOULD BE TAKEN UP BY AN ON-GOING COMMITTEE WITH BROAD-BASED INTERESTS IN TEACHING AND LEARNING.

The committee felt that rather than assign these tasks to another ad hoc committee, they should fall to one that is on-going and already committed to the larger teaching and learning enterprise. Our committee considered, but then rejected, the Undergraduate Education Committee absorbing these responsibilities since graduate courses also administer student course evaluations. The Teaching Academy might fit this description.

Recommendations for the Publishing of Course Evaluations by Student Groups

1. THE COMMITTEE RECOMMENDS THAT THE PUBLISHED BOOKLET SHOULD BE ENTITLED "COURSE EVALUATIONS" RATHER THAN "PROFESSOR EVALUATIONS".

(continued)
It is more accurate to recognize that most of the questions asked of students describe the course rather than the professor. Alternatively, the name of the book could be similar to the title used by the University of Michigan booklet: "Advice."

2. **THE COMMITTEE RECOMMENDS THAT THE STUDENT GROUP PUBLISHING THE BOOKLET WORK CLOSELY WITH DEPARTMENTS TO DEVELOP A SCHEDULE THAT WILL ALLOW THEM TO DELIVER THEIR SUMMARIZED RATINGS (I.E., MEANS AND FREQUENCY DISTRIBUTIONS) IN A TIMELY AND ACCURATE MANNER.**

   The schedule to collect information from departments must fit into the department's work activities, into the student group's production schedule, and into the end-user's time line so that the booklet can actually be used for course decisions. Working out a cooperative schedule assures that the course evaluation booklet does not inadvertently misrepresent data.

3. **THE COMMITTEE RECOMMENDS THAT THE COURSE EVALUATION BOOKLET SHOULD ONLY CONTAIN RATINGS THAT HAVE BEEN COMPLETED BY 15 OR MORE STUDENTS, AND/OR BY AT LEAST 50% OF THE COURSE ENROLLMENT.**

   Evaluations based on small sample sizes are often unreliable and unrepresentative. The literature suggests that 15 ratings provide a reasonable benchmark for reliability purposes. **It is the responsibility of the department** to alert the student group publishing the booklet when the responses do not meet these criteria of representativeness.

4. **THE COMMITTEE RECOMMENDS THAT THE SEMESTER AND YEAR THAT THE EVALUATION WAS COMPLETED SHOULD BE INCLUDED ALONG WITH THE DATA FOR EACH COURSE.**

   Evaluations should be current. Instructors, course material, or both often change from year to year, making old evaluations obsolete. Dating evaluations allows students to assess their validity.

5. **THE COMMITTEE RECOMMENDS THAT THE STUDENT GROUP THAT PUBLISHES COURSE EVALUATIONS SHOULD EXPLAIN CLEARLY TO READERS HOW TO USE, AND HOW TO READ, THE INFORMATION CONTAINED IN THE BOOKLET.**

   The course evaluation booklet currently produced by ASM does include instructions for students—alerting readers to the fact that there are limits to comparing ratings across departments, across levels, and across course formats. The committee recommends that these instructions continue to be used, and that they are highlighted further.

**Recommendations and Advice for Departments to Develop Student Course Evaluations**

1. **IN ADDITION TO THE AREAS IDENTIFIED IN "CENTRAL RECOMMENDATION #1" THAT ARE IMPORTANT FOR STUDENT COURSE DECISION-MAKING, THE COMMITTEE RECOMMENDS THAT DEPARTMENTS DEVELOP QUESTIONS THAT COVER THE FOLLOWING ADDITIONAL AREAS OF TEACHING AND LEARNING:**

   a. Work load, course difficulty;

   b. Student learning, student self-rating of accomplishments or progress, students' own level of effort;

   (continued)
c. Descriptive information about the students and the course: is the course required or an elective; is the course team-taught; is the course introductory or advanced; is the student a major in the area or taking the course as an elective;

d. An assessment of classroom and other instructional facilities.

Areas A-C have been found to influence student ratings of courses. Area D assesses information about the teaching environment, which has been shown to affect how well instructors can teach and students can learn—factors which, in turn, have an influence on course ratings. We recommend that this question about facilities be asked with an open-ended format rather than with a fixed-point format. The information collected about facilities can be provided to campus administrators for use in setting budget and renovation priorities—with the objective that the overall emphasis on teaching and learning will lead to constructive change.

The committee recommends that evaluations not include questions on students' GPA, ability, age, and the time the course is offered—questions in these areas have been found to not significantly influence student ratings of courses.

2. THE COMMITTEE RECOMMENDS THAT FACULTY DEVELOP THEIR OWN OPEN-ENDED QUESTIONS TO BE USED TO COMPLEMENT FIXED-POINT RATING SCALES.

These written comments can be especially useful to improve instruction when elicited in response to questions such as "How do you think the course could be improved?" and "What do you consider the major strengths and weaknesses of the course?" These questions can also be asked in mid-semester to assist in instructors' course planning and development.

3. THE COMMITTEE RECOMMENDS THAT, WHEN POSSIBLE, DEPARTMENTS SHOULD DEVELOP MULTIPLE ITEMS TO MEASURE EACH AREA OF INSTRUCTION.

The literature on internal consistency suggests that a minimum of four items be used to create good measures of teaching areas. Recognizing that it is often impractical for students to complete lengthy questionnaires that contain multiple measures on all areas that we suggest should be evaluated, departments will have to make compromises among reliability and validity issues, inclusiveness issues, and the use of students' time.

4. FOR PROMOTION DECISIONS, THE COMMITTEE RECOMMENDS THAT THE INSTRUCTOR BE EVALUATED ON THEIR TEACHING ACROSS SEVERAL COURSES, AND IN SEVERAL SEMESTERS.

Multiple assessments greatly increase the validity of student ratings. And since almost no instructor is equally as well suited to teach all course formats, departments need to use evaluation patterns (i.e., not ratings from only one course) to determine how and in what context an instructor is and is not effective.

5. THE COMMITTEE RECOMMENDS THAT COURSE EVALUATIONS SHOULD BE ADMINISTERED AT THE BEGINNING OF THE SECOND TO LAST CLASS SESSION, AND WHEN POSSIBLE, WITHOUT THE INSTRUCTOR PRESENT.

Administration in the last class does not assure representation of students; administration at the end of class or as a take-home assignment undervalues their importance; administration in a class close
to the end of the semester allows students to provide summative information about their experience; and administration with the instructor present biases student ratings.

6. THE COMMITTEE RECOMMENDS THAT STUDENT NAMES SHOULD NOT APPEAR ON THEIR COURSE EVALUATIONS. COURSE EVALUATIONS SHOULD NOT BE DELIVERED TO INSTRUCTORS UNTIL AFTER GRADES ARE POSTED.

While the literature identifies only a weak relationship between fear of reprisal and course ratings, it is the committee's recommendation that the problems with fears of reprisal far outweigh the possible benefits of being able to identify students' responses. For similar reasons, students should be confident that their evaluations will not be read until after grades have been reported.

FOOTNOTES

1. One of the difficulties that this committee struggled with concerned the relationship between student course evaluations and their effects on teaching and learning—everyone had a story to tell and an anecdote to convey. To develop a broader perspective on this relationship, the committee reviewed several meta-analyses and secondary reviews of the research literature on student course evaluations. We followed standard guidelines to evaluate the findings from these reviews and to make recommendations as to their application. We did not, however, re-review the primary studies that made up these reviews. We refer interested readers to the following meta-analyses and review studies that formed the basis for our opinions about the validity and reliability of student course evaluations:


2. The literature does provide modest support showing that students rate the "hard" sciences less highly than the humanities, with the social sciences in between.

3. ASM has identified specific questions that they have found particularly useful for student course decisions. They are provided here to serve as examples for departments' use. Note that these are meant to offer different ways to collect information on similar areas. These are not meant to represent a set of questions that a department should use.

Course Organization
Were the classes clearly organized and well prepared?
Were the course objectives/expectations/requirements clearly defined?
Did the overall structure and organization of the course contribute to a better understanding of the course objectives?

(continued)
Clarity of Instruction
Were ideas and concepts clearly presented in the class periods?
Did the instructor clearly explain relationships among course topics?
How would you rate the instructor's ability to communicate course material?

Instructor Availability
Was the instructor available outside of class to answer questions?
If you contacted the instructor during office hours or outside of class, how accessible and helpful was he or she in these meetings?
Was the instructor available when needed and requested?

Grading Fairness
Were grading procedures presented clearly?
Were grading procedures followed in a predictable way?
How well have the examinations and other assignments measured your knowledge of the course material?

Overall Instructor Rating
Overall, how do you rate the quality of the instructor's teaching in this course?
How would you rate this instructor compared to others you have taken in this subject area?
This instructor is an effective teacher (rate the level of agreement with this statement).

Overall Course Rating
How would you rate this course compared to others you have taken in this subject area?
How strongly would you recommend this course to other students?
Overall, how would you rate this course?

4. The committee offers these models as examples of scales and anchors that can be adapted for various purposes:

The first scale and set of anchors can be used for questions that ask about the frequency of an event happening (such as how often lectures were well organized):
1 = almost none
2 = few
3 = about half
4 = most
5 = almost all

The second scale and set of anchors can be used for "how much"-type questions (such as how much a student felt they learned from reading assignments):
1 = little or nothing
2 = a slight amount
3 = a moderate amount
4 = a great deal
5 = a very great deal

The third scale and set of anchors can be used for questions that measure "intensity" of response (such as how well assignments and examinations helped a student learn course material):
1 = slightly or not at all
2 = moderately

(continued)
3 = a good amount
4 = very
5 = extremely

The fourth model and set of anchors can be used for general quality ratings (such as how students rate the quality of the instructor’s teaching):

1 = poor
2 = fair
3 = good
4 = very good
5 = excellent

We would like to thank Nora Cate Schaeffer (Sociology) for her assistance in developing these models of scales and anchors.

5. Departments may face another compromise: obtaining information each semester vs. the “fatigue effect” that results from completing repetitive evaluation forms. The literature suggests that course evaluations may be less valid when the exact same items are used repetitively in every class, in every semester. A fatigue factor takes over when students face the exact same questions for each and every course. In this case, students stop taking them as seriously as they once did. Departments can vary the wording of their questions to help students continue to complete course evaluations as seriously as possible.

Committee Members:
Aaron Brower, School of Social Work (Chair)
Mark Albanese, Medical School, Preventive Medicine
Joyce Anderson, School of Nursing
George Cramer, School of Education, Art
Gery Essenmacher, Chemistry
Stacy Hartman, Student & ASM representative
Jacqueline Hitchon, School of Journalism & Mass Communications
Evelyn Howell, Landscape Architecture & University Committee
Rolf Reitz, Mechanical Engineering
Joel Robbin, Mathematics
Bob Skloot, Associate Vice Chancellor (Ex officio)
James Taylor, Chemistry

(continued)
RESOLUTIONS FROM THE AD HOC COMMITTEE ON THE USE OF STUDENT COURSE EVALUATIONS.

1. Whereas student course evaluation results are a useful mechanism to provide information to students about their course options, these results should be made available to students at the UW-Madison.

2. Whereas student course evaluation results are only one measure of the effectiveness of teaching at UW-Madison, assessment of faculty and instructional staff should not be based primarily on student course evaluation results.

3. Whereas the following areas are those identified by ASM and the Ad Hoc Committee on Student Course Evaluations as important to student course decisions, departments should develop questions that will address these areas. The areas identified are: (1) course organization, (2) clarity of instruction, (3) instructor accessibility, (4) grading fairness, (5) an overall course rating, and (6) an overall instructor rating.

4. Whereas student course evaluation results have most value when administered and disseminated fairly and uniformly, the recommendations and advice enclosed in the attached report should serve as guidelines for departments as they construct their own questions.
### Comparative Literature Evaluations

<table>
<thead>
<tr>
<th>Course</th>
<th>Intro Compar Lit</th>
<th>Students Responding</th>
<th>Enrolled</th>
</tr>
</thead>
<tbody>
<tr>
<td>240</td>
<td>Miranda, Patricia</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>240</td>
<td>Silbey, Jessica</td>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td>240</td>
<td>Lenkos, Frauke</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td>240</td>
<td>Liu, Maichen</td>
<td>15</td>
<td>18</td>
</tr>
</tbody>
</table>

### Economics Evaluations

<table>
<thead>
<tr>
<th>Course</th>
<th>Principles Econ 1</th>
<th>Students Responding</th>
<th>Enrolled</th>
</tr>
</thead>
<tbody>
<tr>
<td>201</td>
<td>Jin, Ying-Fang</td>
<td>17</td>
<td>31</td>
</tr>
<tr>
<td>201</td>
<td>Debasco, Peter</td>
<td>20</td>
<td>31</td>
</tr>
<tr>
<td>201</td>
<td>Wang, Yue</td>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td>201</td>
<td>Singh, Karan</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td>201</td>
<td>Hogan, Chad</td>
<td>13</td>
<td>14</td>
</tr>
</tbody>
</table>
1. **Overview of the Current Project**

The overarching goals of the current project were to review, test, and provide recommendations for an original set of five items to be included in the Essential Learning Outcomes survey (see Appendix B for the wording of the items programmed into Qualtrics for the round 3 group interview).
This project involved four steps. First, UWSC provided instrument review of the items noting any potential problems with the layout, wording, or ordering of the questions. As part of their review, UWSC also drafted follow-up questions or probes to include in the group interviews. Second, UWSC consulted with the client and reviewed the items. Based on these discussions, some items were revised and a final set of probes was created (see Appendix C for the protocol used by the moderator in round 3). Third, UWSC conducted three cognitively-based group interviews with undergraduates to evaluate the wording and layout of the items. The goals of the group interviews were to provide comments and feedback on:

- Overall layout of the questions
- Choice of response options and answer categories
- Problematic items and suggestions for alternative question wording
- Ordering and flow of the survey items
- Other ways to increase the validity and reliability of items

Fourth, UWSC analyzed the audio-recorded group interviews and compiled a list of recommendations, which are presented in Appendix A (round 3).

2. Cognitively-Based Group Interviewing: An Overview

A focus group study is a carefully planned series of discussions designed to obtain perceptions on a defined area of interest in a permissive, non-threatening environment. Groups are typically conducted by a skilled moderator with six to eight people. Participants are selected with the goal of matching their characteristics to the topic under study. The discussions are relaxed, and participants usually enjoy sharing their ideas and perceptions.\(^1\) In contrast to a traditional focus group in which the goals of inquiry are to uncover themes, topics, and the native ways in which participants refer to concepts, the group discussions held by UWSC also drew on principles and techniques from cognitive interviewing.\(^2\) In cognitive interviewing a trained interviewer makes use of structured probes to test respondents’ understanding of the survey questions. Structured probes are asked in reference to specific items and are designed: 1) to determine whether participants understood the questions as intended by the objectives of the question writer; 2) to test whether respondents’ definitions of terms are consistent with the question writers; 3) to uncover areas in which questions might be problematic for participants; and 4) to explore potential problems with cognitive processing and judgment making.

In the current study, group dynamics and discussion were used to test and evaluate the items drafted for inclusion in the Essential Learning Outcomes survey. Thus, the moderator in the group interviews focused the participants’ attention by asking one or more scripted, open-ended probes about each item. These probes were written in advance. The probes were designed to uncover problems with comprehension, retrieval, judgment, and mapping answers onto the response categories provided.

---

3. **Methods**

3.1 **Recruiting**

UWSC telephone interviewers were recruited to participate in the group interviews. All participants were undergraduates including sophomores, juniors, seniors and fifth-year seniors in a variety of majors. These students were specifically targeted for recruitment because they are part of the population who will ultimately complete the web-based ELO survey. The first two groups consisted of eight participants and there were seven in the third group.

3.2. **Data Collection – Mode/Study Instruments**

The group interviews were conducted face-to-face. To test cognitive processing with different administrations of the survey, each of the group interviews were conducted using a slightly different format. In Round 1, the moderator began by introducing the topic and explained the goals of the group. Then, the moderator had each participant list four courses they had taken the previous semester. The moderator then asked the participants to complete Question 1 in the self-administered questionnaire for each of their four courses. After they completed Question 1, the moderator asked a set of structured probes to facilitate discussion of Question 1. Once discussion of Question 1 was completed, the moderator asked participants to complete Question 2 for each of their four courses; and immediately following their completion of Question 2, the moderator briefed the participants on Question 2. Administration of the remaining questions, Questions 3 through 5, followed in a similar manner until all five questions from the survey were reviewed.

In Round 2, the moderator began the interview by introducing the topic and explained the goals of the group meeting. The moderator then asked the participants to complete the eight questions in the self-administered questionnaire for each of their four courses. After they completed the questionnaire for all four of their courses, the moderator asked a set of structured probes to facilitate discussion of the questions one-by-one.

In Round 3, the revised survey items were programmed into web survey software (Qualtrics). Participants were asked to access the web survey on a laptop, read the introduction, and list three courses they had taken last semester. They were instructed to answer the eight survey questions for their first course only, and were then asked several structured probes before completing the survey for their other two courses. After they had completed the survey for all three courses, they were asked for their thoughts on the survey overall and for suggestions on how the data might be presented.

The scripted probes for round 3 were prepared in advance (see Appendix C). Probes were written to focus the discussion on potential problems with the items, such as with the respondent’s understanding of individual terms. The group interviews each lasted about an hour and a half.

3.3. **Moderator**
The group moderator was a project director at the UW Survey Center. Focus group and cognitive interviewing techniques were reviewed and incorporated into the protocol script.

4. **Analysis and Final Recommendations**

The group interviews were audio-recorded in order to facilitate analysis and review. Recommendations are outlined in Appendix A. Appendix A is organized in the following manner. First, we provide the question that had been tested in round 2. Next, we give a brief “background and history” that includes any recommendations or changes we made during review of the items with the client via email or meetings after round 2. Third, under “Candidate revision tested in round 3,” we present the wording of the item tested in the most recent group interview. Fourth, the probe or probes administered by the moderator are listed under “Summary of the group interviews.” Immediately following the probes we provide a summary of what respondents reported during the group discussion. Finally, our recommendations and comments appear under “UWSC recommendation.”

5. **The University of Wisconsin Survey Center (UWSC)**

The UW Survey Center (UWSC) is a department of the College of Letters and Science at the University of Wisconsin-Madison. UWSC is supported by the College and revenue generated from contractual work. UWSC serves the survey research needs of University of Wisconsin faculty, staff, and administration; faculty at other universities; federal, state, and local governmental agencies; and not-for-profit organizations. The mission of UWSC is to assist researchers by providing the highest quality survey research services and as such, the Survey Center provides the complete range of survey research capabilities. Professor Nora Cate Schaeffer is the Faculty Director of the UW Survey Center. Jaime Faus and Tara Piché served as Project Directors on this project with assistance from Survey Methodologist Jennifer Dykema.
Appendix A:
Summary of Results and Recommendations from the Group Interviews

Question-by-Question Comments

Introduction to web survey in round 3:

“The UW-Madison is committed to ensuring that students like you achieve a set of Essential Learning Outcomes throughout your college experience (see here). By focusing on these topics, the hope is that students will leave college with a wide range of knowledge about the world and skills that will help them succeed in the future.

Today, we are asking you to evaluate how well the courses you took last semester contributed to each of the essential learning outcomes. We will also ask you to evaluate the overall impact and quality of each of your courses. Please note that this survey is not intended to replace course or instructor evaluations that are conducted each semester by individual departments.”

Summary of round 3 group interview:

PROBE: After having read this paragraph, what do the Essential Learning Outcomes (ELOs) mean to you?

- All respondents had read through all of the questions before this probe was asked. It likely would have yielded more feedback if it had been asked immediately after they read the paragraph. One respondent said ELO’s were intellectual skills, professional skills, social responsibility – what each question had asked. Another respondent said they thought the main point was “success after college.”

PROBE: Did you click the link where it said “see here?” How thoroughly did you read the descriptions of the ELOs?

- No one reported having clicked on the link. When asked if there was any particular reason why they did not use the link, no one responded. We typically suggest against using optional links (such as hover-over text) for this very reason – some will use it and most won’t.

UWSC recommendation:

- When introducing the survey, we suggest you avoid links and include all the text that you’d like respondents to read on the first screen of the web survey (including any brief definitions of ELO’s). This way, all respondents will read the same explanations before answering the questions.
**Question used in round 2:**

“What is the main reason you took this course?”

[To fulfill a course requirement; As an elective]

**Background and history:**

- Respondents did not seem to understand what was meant by an “elective,” as an elective course can be taken to fulfill a breadth or a major requirement.
- Suggested specifying that they should answer according to their intent at the time they enrolled for the course.

**Candidate revision use in round 3:**

“At the time you enrolled, did you take this course primarily for your major?”

[Yes/No]

**Summary from round 3 group interview:**

**PROBE:** In your own words, what does it mean to take a course “primarily for your major?”

- Respondents gave the following responses: “That it’s a requirement for your major;” “That it fulfills the amount of credits you need;” “That you have to take it;” “That you don’t necessarily have interest in it.”

**PROBE:** Did anyone have difficulty selecting a response for this question? If so, why did you have difficulty?

- Respondents answered “no,” indicating that they didn’t have any difficulty.

**UWSC recommendation:**

- The “yes”/”no” format of the question seemed to be more clear than the response options that were used in round 2.
- Overall, it seems like respondents understood the intended meaning of the phrase “primarily for your major” and were able to give a response. You could word the item more specifically using the terms respondents actually used when responding to the probe in round 3. For instance, in the meeting we had before round 3, we talked about...
alternative wording such as, “At the time you enrolled, did you **primarily** take this course to fulfill a requirement for your major?” Even more specific might be, “At the time you enrolled, did you **primarily** take this course to fulfill any credits required for your major?” The goal is to word it so that respondents know exactly what you are asking, which makes different interpretations among respondents much less likely.

**Question used in round 2:**

“In general, how much did this course **enhance your knowledge** of the world, which might include knowledge of human cultures or the physical world?”

[Not at all, A little, Somewhat, Quite a bit, A great deal]

**Background and history:**

- Respondents had not understood what “natural” world had meant in the original version of the question, thus we decided “physical world” might work better. In round 2, respondents thought of “physical world” in terms of “biology and physics and how things work together.”
- It was decided to try “society” and “science” in place of “physical world.”

**Candidate revision tested in round 3:**

“In general, how much did this course **enhance your knowledge of the world**, such as knowledge of human cultures, society, or science?”

[Not at all, A little, Somewhat, Quite a bit, A great deal]

**Summary from round 3 group interview:**

**PROBE:** In your own words, what does it mean to “enhance your knowledge of the world?”

- One respondent said that they thought this phrase meant “to become aware of new things that you didn’t know.”
- A few respondents reported this item as being “broad,” and said that they had not understood what “world” had meant. One respondent said that the examples “kind of narrowed it down a bit.” Another respondent said he had answered the question about a Macroeconomics course, and since he learned about different things all over the world, he “jumped right to that.”
Another respondent rhetorically asked, “…like physical world or like the world figuratively speaking?” When asked which definition he had used to answer the question, he said he had thought about the world figuratively.

PROBE: When answering this question, what did “society” mean to you? What did “science” mean to you?

- Respondents said that these terms were also broad. One respondent said society could refer to “American society.” Another respondent said one might interpret it different depending on the course they took. For instance, he said he had been answering about a class about the history of Peru and had thought about society in terms of “the society of Peruvians.”

PROBE: Did the examples “human cultures, society, or science” make it easier or harder for you to answer this question?

- Several respondents said “easier.”

UWSC recommendation:

- This question seemed to work just as well as it had in round 2 when the term “physical world” was used. We did not get good feedback about what “science” had meant to respondents; however, we are wondering if “science” might be too broad. If the course being evaluated is a science course, will students automatically give it a higher rating, and if so, is that really measuring whether the course “enhanced their knowledge of the world?” It seems like that example is almost asking, “Did you take a science course?”

- Again, “the world” is interpreted vaguely, with the examples consisting of things that are almost opposite one another (i.e. “culture and society” versus “science”). If following best practices in survey research, we would recommend asking separate questions about enhancing their knowledge of “culture and society” and then enhancing their knowledge of the “physical world.” However, we think combining these ideas in this item is working for the purposes of your survey.

Question used in round 2:

“How much did this course help you develop intellectual skills, such as critical and creative thinking, quantitative reasoning, and problem-solving?”

[Not at all, A little, Somewhat, Quite a bit, A great deal]

Background and history:
• It had been clear in round 2 that respondents relied on the examples to define “intellectual skills.” One respondent interpreted developing intellectual skills as “developing intellectual skills as opposed to just learning facts,” and noted that “problem-solving” had been the key word in the question for them. Another respondent “…saw it more as if the course required you to use these skills, and then you tend to improve.”
• Some respondents seemed to interpret “quantitative reasoning” as having something to do with math or data, but others reported that this example was abstract or not helpful. Perhaps this term is more relevant to students who have had exposure to data or math in their courses, and more abstract to those who have not.

Candidate revision tested in round 3:

“How much did this course help you develop intellectual skills, such as critical or creative thinking, quantitative reasoning, and problem solving?”

[Not at all, A little, Somewhat, Quite a bit, A great deal]

Summary from round 3 group interview:

PROBE: When answering this question, what did the term “intellectual skills” mean to you?
• A couple respondents said “intellectual skills” had meant everything that came after it (i.e. the examples).

PROBE: What does it mean for a course to help you “develop intellectual skills”?
• One respondent said “teaching you new ways of thinking…like new approaches or things you haven’t done in other classes…and being able to apply it to other classes.” Another respondent said “acquiring knowledge about a certain area.”
• One respondent said if the course was “difficult and made you work really hard.”

PROBE: In this question, what did “critical or creative thinking” mean to you? What did “quantitative reasoning” mean to you?
• Respondents seemed to focus on the distinctions between and among the listed examples. One respondent said she had thought of creative thinking as “something that doesn’t have a set of rules and is up to you,” while quantitative reasoning has “rules and guidelines to follow.” Another respondent said the examples included things that have “right and wrong answers,” and if there are no right or wrong answers then “it’s more like creative thinking.”
• One respondent reported that they had not understood what creative thinking had meant.
• One respondent said “quantitative reasoning” made them think of math, a couple others nodded and said “Mmm hmm” in agreement.
UWSC recommendation:

- This question remained unchanged from round 2 to 3. The fact that respondents seemed to notice the distinctions between the examples seems fine, since you are asking whether the course developed any of those skills and they understand the differences. Again, in round 3, respondents relied heavily on the examples to define intellectual skills.
- In round 2, some respondents seemed to focus on whether the course had helped them to develop intellectual skills (i.e. “developing skills as opposed to just learning facts”), while others focused on whether the course had simply required them use the listed skills. We think this is alright, as the two ideas are correlated and there’s no way to be more specific than saying “help to develop intellectual skills.”
- We also noticed that some respondents mentioned thinking about how they would apply these skills in the future, even though the question does not explicitly state this. In round 2, one respondent said “intellectual skills may improve, but it’s hard to know for sure until you use those skills later on and apply them.” In round 3, one respondent talked about being able to apply the skills in other courses. It seems some students are able to grasp the concept that ELO’s relate to applying knowledge or skills to tasks in the future (perhaps because the introduction talks specifically about preparing them for the future).

Question used in round 2:

“How much did this course help you develop professional skills, such as written and oral communication, computer literacy, and working in teams?”

[Not at all, A little, Somewhat, Quite a bit, A great deal]

Background and history:

- In round 2, it was evident that respondents had thought of skills they may need for a job or future profession when they read “professional skills.” In fact, a few respondents said that even if they were asked to write papers or speak in class, which could be considered written and oral skills, they did not necessarily consider this as helping to develop professional skills (one respondent said especially if they don’t receive feedback in classes).
- Respondents had relied on the example to define what “professional skills” were. They seemed to be more likely to rate this course higher if the course was important for their major or related to what they wanted to do for a career in the future.

Candidate revision tested in round 3:

“How much did this course help you develop professional skills, such as written and oral communication, computer literacy, and working in teams?”
Summary from round 3 group interview:

**PROBE**: When answering this question, what did the term “professional skills” mean to you?

- Respondents gave the following responses: “I think just being able to succeed in the workplace;” “Yeah, I immediately thought of a job and being able to function at a job;” “Formal, interpersonal skills;” “I just thought of generally anything that would make you attractive to an employer.”
- When asked if anyone had thought of anything that was not related to a job or employer, no one had anything to say.

**PROBE**: Did the examples of professional skills given in this question make it easier or harder for you to answer?

- One respondent said easier and that the examples had actually been in line with what she had thought about when she first read “professional skills.”
- Another respondent said the examples were helpful because when he had first read “professional skills,” he had thought of getting a certification or something more concrete.

**PROBE**: Are there other examples of professional skills you were thinking of when you answered this question?

- Two respondents said that they would not think of “computer literacy” as a professional skill right away, but that it “made sense.”
- When asked to talk about the courses they rated and why they had rated them the way they did, one respondent said she had answered the question for an Experimental Psychology course. She said she had rated the course as helping her to develop professional skills ‘a great deal,’ thinking about oral communication, because she had to give multiple presentations. Another respondent said he had answered for an Econ 101 course and that since they didn’t have any group assignments and the discussion section had been optional, he had rated the course as helping him develop professional skills ‘not at all.’ However, he rated the course as developing intellectual skills ‘quite a bit.’ He added that it wasn’t a very interactive course, as he took notes in lecture, studied on his own, and took the test; but he had learned something. He said that if he had been graded on participation in discussion section, or had had to write a paper, he would have given the course a higher rating for developing professional skills.

**UWSC recommendation:**
This question remained unchanged from round 2 to round 3. In both rounds, respondents reported that the phrase “professional skills” made them think immediately of a job or things that would make them attractive to an employer.

In round 2, at least one respondent said that even if they had been required to do written and oral activities in the course, they wouldn’t necessarily rate the course as having helped them develop professional skills unless they had received feedback. In round 3, when asked how they had rated their courses and why, respondents reported having rated the course high if they were required to do written and oral tasks.

Respondents seemed to focus on written and oral communication, as opposed to working in teams or computer literacy (which was less salient to them). However, they reported that the examples had made it easier to answer the questions overall.

This question seems to work well. Again, we suggest ensuring the examples you use directly reflect how you want respondents to define “professional skills.” The major “fix” for the original item was splitting it up into two separate questions.

**Question used in round 2:**

“How much did this course influence your sense of social responsibility, for example by teaching you about cultures, providing opportunities for civic and community involvement, or by enhancing your ability to reason ethically?”

[Not at all, A little, Somewhat, Quite a bit, A great deal]

**Background and history:**

- In round 2, respondents defined “social responsibility” using the examples, but were not entirely clear on what it meant to “influence” one’s social responsibility. They didn’t know if this meant being more aware of social problems throughout the world, or being motivated to go out and do something about these problems.
- Respondents seemed to have broadly understood the question and were able to answer it. However, interpretations of the concept ‘social responsibility’ still varied.

**Candidate revision tested in round 3:**

“How much did this course increase your sense of social responsibility, for example by teaching you about cultures, providing opportunities for civic and community involvement, or by enhancing your ability to reason ethically?”

[Not at all, A little, Somewhat, Quite a bit, A great deal]

**Summary of the round 3 group interview:**
PROBE: In your own words, what does it mean for a course to “increase your sense of social responsibility?”

- One respondent said they rated their Spanish phonetics course as increasing their sense of social responsibility ‘a little.’ They said they learned how dialects are different in different areas of the world, but the course did not help them to “gain a sense of social responsibility, as far as civic duty or what I can do to help the work, or what things could be done.”
- A couple respondents mentioned that the examples were very specific and that one can’t expect all courses to increase one’s sense of social responsibility. One respondent said they had answered the question about a logic course and thus this question had not seemed very applicable to the content of the course.

PROBE: When answering this question, what did it mean to “reason ethically?”

- One respondent said that they hadn’t understood what this meant, so they ignored it. Another respondent said that it meant “being able to weigh pros and cons on each side of an issue,” and being able to make a decision about what should be done.

PROBE: Besides the examples listed in this question, are there any other ways a course might influence your sense of social responsibility?

- Respondents reported that social responsibility was related to the world and what is going on around you. One respondent said they are a Sociology major, so they get “a lot of this from their classes” (“…it’s all about how to better a community…”), but said that not all classes deal with stuff like that.

UWSC recommendation:

- The only change in this item from round 2 to round 3 was changing the verb from “influence” to “increase.” In round 3, at least one respondent said that the phrase “increase your sense of responsibility” meant to “gain a sense of social responsibility,” which was to her, what is going on around you or what you can do to help.
- The examples seemed to help clarify what it meant to influence one’s sense of social responsibility; however, “reason ethically” seemed to carry a lot of cognitive burden, such that respondents skim over it as opposed to really thinking about what it means. When asked specifically what it means, at least one respondent was able to give a detailed definition (i.e. weighing the pros and cons of an issue to decide what to do). Again, we suggest making sure the examples are concrete and reflect how a course might increase someone’s sense of civic duty.
- There was still a little confusion (especially in round 2) about whether “social responsibility” means feeling like one should do something, or if it just means increasing one’s awareness of what is going on in the world. You might consider revising this to include an actual definition, instead of examples (similar to the last questions about
educational value and overall impact). The examples give the impression that social responsibility might entail both increasing one’s knowledge of cultures, but also participating in civic activities and getting out and doing something (e.g. “civic duty”). For instance, you might ask, “How much did this course increase your sense of social responsibility, that is increased your knowledge of cultures or what is happening around the world, or provided you with opportunities for civic or community involvement?” This is just an example, as we do think this question, as it is, is measuring whether the course increased their social awareness. It all depends on what you want to measure here, i.e. what social responsibility entails.

---

**Question used in round 2:**

“How much did this course prepare you to solve problems in the future by improving your ability to combine knowledge and skills from different areas?”

[Not at all, A little, Somewhat, Quite a bit, A great deal]

**Background and history:**

- In round 2, several respondents ignored the part of the question asking them to think about solving problems “in the future,” and had had difficulty interpreting “areas.” One respondent said he wasn’t sure what “areas” meant – was it “fields of study, like humanities and sociology?” One respondent felt that the phrase “in the future” distorted the question a bit because students don’t always know what they’ll be doing in the future, or if the class will help them solve those future problems.
- We had recommended eliminating the portion of the question asking about solving problems and simply asking whether the course improved their ability to combine knowledge. We also recommended being more specific about what “areas” meant. We suggested using the native language of the respondent in round 2 (i.e. “fields of study”).

**Candidate revision tested in round 2:**

“How much did this course improve your ability to combine knowledge or skills from different fields of study?”

[Not at all, A little, Somewhat, Quite a bit, A great deal]

**Summary of round 3 group interview:**

**PROBE:** In your own words, what was this question asking?
Several respondents reported that they had not understood what the question was asking until discussing it with the group. When taking the web survey, they had thought the question asked whether the course itself combined knowledge from different fields. One respondent said, “I thought it just meant was the course interdisciplinary or not.” Another respondent said, “Yeah, I thought the same thing. It didn’t have to do with my ability, but more the class itself.”

**PROBE:** When answering this question, what did it mean to “combine knowledge and skills from different fields of study?” What were the “different fields of study” you thought of when answering this question?

- One respondent said “I just took it to mean interdisciplinary studies. I took an Old English course and have a background in Latin, so combining those.”
- When asked about “different fields of study,” one respondent said they had answered the question about a math class on logic, and since it was related to their other major [Philosophy], they considered it interdisciplinary.

**UWSC recommendation:**

- Before we tested this item in round 2, we knew that one of the potential problems was that it was double-barreled, meaning it is actually asking two questions: (1) did this course combine knowledge from different fields of study; and (2) did you get something from this that will help you to solve problems in the future. Thus, for round 3, we decided to eliminate part of the question asking about solving problems in the future, and being more specific by changing “areas” for “fields of study.” We also changed “and” to “or” when referring to “knowledge” and “skills.”
- Again, in round 3, respondents reported that they focused on whether the course integrated knowledge, as opposed to whether it had improved their ability. Unfortunately, there isn’t an easy fix for this except to ask two separate questions. However, we suggest underlining the portion of the question about improving their ability – we had not done that in the web survey: “How much did this course improve your ability to combine knowledge or skills from different fields of study?”
- Respondents in round 3 seemed to think about the subjects of other courses they had taken when asked what “fields of study” had meant. Several respondents simply said this made them think “interdisciplinary.” You could possibly add examples to highlight what fields of study means (using “such as”); however, we are not sure what these examples might be? Adding examples might possibly make it more confusing, depending on what they are. Perhaps it is OK for respondents to think “interdisciplinary,” as long as they are focusing on the part of the question asking it improved their ability to combine knowledge.

**Question used in round 2:**
“In this next question, educational value is the extent to which a course improves your all-around education and prepares you for future endeavors.

How would you rate this course for its overall educational value?”

[Very poor, Poor, Fair, Good, Very good]

**Background and History:**

- We recommended incorporating the definition into the text of the question to be sure that all respondents read this and to help them process how they are meant to measure “educational value.”
- Respondents in round 2 seemed to interpret “educational value” in terms of whether they learned anything from the course that they can or will use in the future. One respondent said that even if you major in a specific subject, the goal of college is to be well-rounded, so she had thought about whether the course had helped her in that way.

**Candidate revision tested in round 2:**

“How would you rate the overall educational value of this course, that is the extent to which the course improved your all-around education or prepared you for the future?”

[Very poor, Poor, Fair, Good, Very good]

**Summary of the round 3 group interview:**

**PROBE:** In your own words, what is the “educational value” of a course? Did the definition that was provided make it easier or harder to rate the course?

- When asked what “educational value” had meant, respondents gave the following responses: “How much you gain and take out of the course in general;” “How much you learned or how much you were challenged by the course and the subject;” “I also considered the practicality of it, for using it in the future and if it will benefit you in some way;” “Yeah, I was thinking just how happy you were that you took the course.”
- Several respondents reported that the definition made the question easier to answer. One respondent said the definition “kind of narrowed it” and made them think about whether the class provided any information that they will use in the future.

**PROBE:** Did the response options (Very Poor, Poor, Fair, Good, and Very Good) make it easier or harder to answer this question?

- One respondent said “I had an easy time selecting an answer;,” and several others indicated that they agreed.
UWSC recommendation:

- For round 3, we changed the tense of the verbs to reference the past and incorporated the definition of “educational value” into the text of the question itself. We also changed “prepares you for future endeavors” to “prepared you for the future,” which is more general.
- Respondents seemed to interpret this question broadly and said they thought about how much the course had contributed to their overall education or how much they had learned from it in general. Their explanations seem to mirror those from round 2.
- One thing to note is that it seemed participants either paid attention to the part of the definition about improving their “all-around education” or the part about “preparing them for the future.” Since these are two different ideas (especially preparing them for the future), you might want to pick just one by which to define educational value, so you know all respondents are using the same definition when rating the course.

Question used in round 2:

“In this next question, overall quality of a course refers to how well it was constructed and taught in order to maximize its educational value.

How would you rate this course for its overall quality?”

[Very poor, Poor, Fair, Good, Very good]

Background and history:

- We recommended incorporating the definition into text of the question to be sure that all respondents read this and to help them process how they are meant to measure “overall quality.”
- Respondents in round 2 seemed to interpret “overall quality” in terms of how well the professor organized and presented lectures, and the everyday mechanics of running and teaching the course. One respondent said that this question seemed to be more concentrated on the construction and organization of the course, not the educational value. She said if it’s an organized class, you’re going to get more out of it as far as its value.

Candidate revision tested in round 2:

“How would you rate the overall quality of this course, that is the extent to which it was structured and taught in order to maximize its educational value?”
Summary of the round 3 group interview:

PROBE: In your own words, what does the “overall quality” of a course include? Did the definition that was provided make it easier or harder to rate the course?

- Several respondents reported having thought of the instructor’s performance as in how they “structured the course to maximize its educational value.” One respondent said they thought about the “the way the course was structured and how the professor presented the materials, what type of materials they chose to present.” Another respondent added, “And if those presentations were logical, and built on one another.”
- One respondent did report having thought about the “syllabus, books, homework, and tests, but not particularly the professor or students in the class.” Other respondents said they disagree and had thought about quality as described in the statements in the bullet point above.
- One respondent said the definition made the question “tougher” to answer, and explained that without the definition she would have had an easier time rating the course just based on how well she had liked it. She said the definition made her think about the question. Another respondent agreed and said that when he thinks about the overall quality of a course he considers the amount of work he had to do, the professor’s performance and interaction with the students, and “then the definition narrows it down more, I think.”

PROBE: Did the response options (Very Poor, Poor, Fair, Good, and Very Good) make it easier or harder to answer this question?

- One respond said she was confused because she didn’t know if she was supposed to rate how good the professor was or how good the course was other than the professor.
- No one expressed having any problems with the rating scale itself.

UWSC recommendation:

- In round 3, we included the definition of overall quality in the text of the questions itself. Several respondents reported that the definition had made them think more specifically about how they were supposed to measure the quality of the course (e.g. thought more about how the course was structured as in what materials were used and how well they fit together and were presented).
- Without the definition it would be difficult to know what respondents are actually measuring. The definition narrows the scope of what “overall quality” entails. As a result, you know that most respondents will think about how the course is structured when rating quality. Without a definition it seems respondents might measure overall quality according to whether they liked the course, which doesn’t seem like what you would want.
• Adding the word “taught” in the definition seemed to make respondents think they should be rating the performance of the professor, whereas the word “structure” triggered them to think about the course itself. If you would like students to consider the professor as well as the content of the course you may want to be more explicit. For instance, you might ask, “How would you rate the overall quality of this course, that is the extent to which it was structured well and how the professor taught it in order to maximize its educational value?”

General Comments and Questionnaire Layout and Format

Summary of the round 3 group interview:

PROBE: Thinking of the questionnaire overall, did anyone have any general comments?

• One respondent reported having found the survey hard to answer for science courses versus “non-sciency” classes. The same respondent said “…you can talk about social responsibility for basically any humanities or social science class because it’s somewhat relevant, but if you’re taking O Chem or something like that it [the survey] is almost nonsensical or meaningless.”

• Another respondent said that the survey had made them reflect on their courses in ways they hadn’t before and thought the survey asked about aspects of courses where they did gain something.

PROBE: As you filled out the questionnaire, did you find yourself comparing the answers you gave from one course to another? Did you go back to previous items to change your initial response? If so, which items were these?

• A few respondents said that they had mentally compared the answers they gave from one course to another. Only one respondent said they went back to change a previous response for a different course, but did not remember which item it had been.

PROBE: Overall, was it easy or hard to answer the questions using the response categories provided – that is, not at all, a little, somewhat, quite a bit, and a great deal?

• Respondents did not have much to say in response to this probe. It seemed like the answer categories had made sense and worked well.

PROBE: In your own words, what are the Essential Learning Outcomes?

• One respondent referred to what we had talked about earlier in the group discussion and said, “…to gain skills for the future that we would use.” In response to this comment another respondent said they thought it differs from class to class and that they take certain classes for specific reasons. He said, “…Some classes are related to what I want
to do in the future, and I take other classes because I think it’s going to be easy and I’m going to get an ‘A.’” Another respondent agreed and said you can appreciate classes for different reasons and “these questions reflected that.”

UWSC recommendation:

- The overall reception of the utility of the survey was very positive. Students seemed to appreciate the idea of a course evaluation that asks about various benefits a course may provide. Respondents did mention that there is great variability in how one might rate a humanities course versus a science course; however, we believe your team expects to see these differences.
- It has yet to be determined how these surveys will be administered to students. That is, whether they will take one survey for one course in one setting, or if they will be sent links to surveys for all the courses they took in a semester. You will want to decide whether students will be able to go back to previous surveys to change answers once they read the questions in the first survey. As they take the survey for more than one course, they will reflect more on what the questions are asking and may want to adjust their ratings accordingly.
- The response options come from a standard scale used in survey research. Respondents did not seem to have any trouble classifying themselves according to the objectives of the questions using these options.
- After having taken the survey and looking at the results, students seemed to understand the main point of the ELO’s. We think the introduction to the survey helps with this as well.
- Finally, we had recommended using horizontal scales for all questions. We suggest doing this in the actual administration of the survey as well, as this will help respondents interpret the items as being ‘ratings’ and saves room as far as the layout of the survey.

Comments on Example Data Report

Summary of the round 3 group interview:

PROBE: Are these graphs easy or hard to interpret? Can you tell me what about them makes them easy or hard to interpret?

- At least three respondents said that the graphs seemed “complicated” at first or didn’t seem very “intuitive.” They said that they relied heavily on the key to help them figure out what they were looking at, especially what the tick marks had meant.
- One respondent suggested using three bars for each category, one with the overall median, one for the ‘major’ median, and another for the ‘non-major’ median. A couple additional respondents made comments suggesting that the data for ‘majors’ versus ‘non-majors’ could be presented in a clearer way (i.e. using a separate graph). One
respondent said including the number of non-majors in the label (e.g. “331 student; 105 majors”) might help. In addition, it may be useful to say “total students” (e.g. “331 total students; 105 majors and 226 non-majors”), so it’s clear the class had 331 students and not 436.

- Another respondent suggested having one graph for each skill category for every class, so one could compare just that skill [questions] among several courses.

PROBE: Please take a minute to compare the information shown in the first two graphs. Both are 100-level, introductory courses. Which course would you chose based on the graphs? Why? Please do the same for the pair of 400-level, advanced courses.

- When comparing the 100-level courses at the top of the data report, one respondent said she would choose the first one because her “first thought was to compare the overall quality of the course as the most important factor.” Another respondent said he would go with the first one because “everyone seemed to enjoy it more, but if I was looking for specific skills, I might go for the second one.”

- Another respondent said it would depend on what year in school he was. If he were a freshman, he said he’d probably take the “easy one” because he wouldn’t care as much about professional development at that point. He said if he were a senior he might look at the graphs as far as skills he might gain that he could put on a resume.

- Another respondent said it seems like the second [100-level course] was “more rigorous” and that “fewer people liked it but said they learned more from it so I guess it was harder.”

PROBE: Do you think you would use the results of this survey to help you select classes in the future? Why or why not?

- Several respondents said the results of this survey would be beneficial when selecting courses. One respondent said it would definitely help because there isn’t currently a way to know if a class is good or bad besides “anecdotes.” This respondent mentioned “RateMyProfessor.com” for information about professors, but said that those ratings don’t necessarily tell you what the class is going to be like. Another respondent added that they have sometimes read the title and description of a course before taking it, but it had not been what they expected or they hadn’t learned what they thought they would learn when they completed it.

- One respondent said he thought the survey would be beneficial, but because people learn differently and have different interpretations of their experiences, he wouldn’t put all his “trust into what the graph might say.”

PROBE: Is there any other information from the survey not shown here that would be useful to have when selecting courses?

- Several respondents talked about wanting some indication of how difficult a course had been or how much work they had had to put into it. One respondent said having this knowledge would help her balance her courses over the semester better. She said if she
knows she has a really hard course, she might look for a course that’s less “work intensive” so she knows she can get an easy ‘A.’

- One respondent said that he’d like to know how much people liked the course, and noted that this type of rating would be similar to impact or quality.
- Another respondent said it would be interesting to see a correlation between the amount of work students put in and how much they feel they actually learned from the course.
- A couple respondents talked about wanting results displayed by the professor or TA, since the same courses can sometimes be very different depending on how they are taught.
- One respondent said it might be helpful to know if students would recommend the course.

**PROBE:** What did you think about the tick marks indicating the means for majors versus non-majors? Did you notice them?

- Several respondents said they had looked at the key to see what they meant, but had dismissed them, especially if they looked similar across the different categories. One respondent said the information [major versus non-major] would be clear if in a separate graph.

**UWSC recommendation:**

- Respondents said the graph seemed crowded and suggested different parts of the graph (e.g. means for majors versus non-majors; individual ELO’s) be separated out into their own graphs. We know the data report was just an example to get feedback about what students would like to see, so hopefully their comments will prove useful when you are designing how the data will be viewed. Overall, spacing the bars of the graph out and giving them more explicit labels might help students interpret them more easily.
- When comparing the 100-level courses on the report, it was interesting to hear several respondents interpret the bars for “educational impact” and “overall quality” as meaning the students had “liked” the course more or less. Perhaps putting the ELO label right below the bar of the graph will help clear this up. As was found in all rounds of group interviews, it seemed like overall quality was the “most important factor” for many respondents. However, several respondents did note that which graph they consider important would depend on their purpose for taking the course.
- Again, several respondents indicated the survey would be beneficial in helping them to select courses. Many said they would like to have information about how difficult students had thought the course was, but such a rating does not seem to support the purpose of the ELO’s.
- Finally, we recommend using horizontal scales for all questions, as shown below. This will help respondents interpret the items as ratings, and will make the questionnaire feel shorter.
Appendix B:

Questions Tested in Round 3 (Qualtrics)

PLEASE NOTE: ALL RESPONSE OPTIONS WERE PRESENTED HORIZONTALLY AND NOT VERTICALLY AS THEY APPEAR BELOW.

>intro<

UW-Madison Essential Learning Outcomes Evaluation

The UW-Madison is committed to ensuring that students like you achieve a set of Essential Learning Outcomes throughout your college experience (see here). By focusing on these topics, the hope is that students will leave college with a wide range of knowledge about the world and skills that will help them succeed in the future.

Today, we are asking you to evaluate how well the courses you took last semester contributed to each of the Essential Learning Outcomes. We will also ask you to evaluate the overall impact and quality of each of your courses. Please note that this survey is not intended to replace course or instructor evaluations that are conducted each semester by individual departments.

>courses<

Please fill in the names of any three courses you took last semester, in the fall of 2012. You can list them in any order that you like.

Course 1:
Course 2:
Course 3:

>q1<

Next we will ask you to answer a set of questions about each of the courses you listed. For this next set of questions, please think about Course 1.

At the time you enrolled, did you take this course primarily for your major?
In general, how much did this course enhance your knowledge of the world, such as knowledge of human cultures, society, or science?

- Not at all
- A little
- Somewhat
- Quite a bit
- A great deal

How much did this course help you develop intellectual skills, such as critical or creative thinking, quantitative reasoning, and problem solving?

- Not at all
- A little
- Somewhat
- Quite a bit
- A great deal

How much did this course help you develop professional skills, such as written and oral communication, computer literacy, and working in teams?

- Not at all
- A little
- Somewhat
- Quite a bit
- A great deal

>q5<

How much did this course increase your sense of social responsibility, for example by teaching you about cultures, providing opportunities for civic and community involvement, or by enhancing your ability to reason ethically?

- Not at all
- A little
- Somewhat
- Quite a bit
- A great deal

>q6<

How much did this course improve your ability to combine knowledge or skills from different fields of study?

- Not at all
- A little
- Somewhat
- Quite a bit
- A great deal

>q7<

How would you rate the overall educational value of this course, that is the extent to which the course improved your all-around education or prepared you for the future?

- Very poor
- Poor
How would you rate the overall quality of this course, that is the extent to which it was structured and taught in order to maximize its educational value?

- Very poor
- Poor
- Fair
- Good
- Very good

Thank you. Please wait for further instruction from the moderator before moving forward.

[Same questions repeated for Course 2 and Course 3.]
Appendix C:

Moderator’s Script and Interviewing Protocol

INTRODUCTION:

Hello and welcome to our group interview. My name is Jaime Faus and I will be the moderator this evening. On behalf of the UW Survey Center, we would like to thank you all for agreeing to participate in this important project.

Before we begin, let’s go around the room and tell us your name, your major, and how long you’ve worked as an interviewer at the Survey Center? I’ll start.

Now I’d like to tell you a little more about the project and what we will be doing over the next hour.

We are working with professors in the UW-Teaching Academy here at UW-Madison to create a survey that may be used across campus to measure how effectively courses are meeting the Essential Learning Outcomes.

The Essential Learning Outcomes were adapted from those developed through national surveys and interviews done with employers, faculty, staff, and alumni, asking the basic question, “What qualities and skills do you want in college graduates?” The Essential Learning Outcomes aim to ensure that students gain knowledge about the following topics throughout their courses at the University: Knowledge of Human Cultures and the Physical and Natural World, Intellectual and Practical Skills, Personal and Social Responsibility, and Integrative Learning. The goal of this survey in particular is to create a tool that students and faculty can use to gauge the effectiveness of a course in equipping students with the Essential Learning Outcomes listed above. Ideally, data from this survey could be used by students to help them select courses, and by faculty to improve the ways that they teach their courses.

We have scheduled group interviews with undergraduates at UW-Madison, such as all of you, to get your feedback on the survey. The feedback, suggestions, and comments that you provide will help make the next administration of the survey a success and ensure that the data collected accurately reflect the perspectives of students at UW-Madison.

Today, we are going to ask you to complete a web survey about three courses you took last semester. If you took more than three courses, you can choose any three to answer the survey questions about. If you took less than three courses, you will answer questions for each course you took. The subject of your courses or the order in which you type the names into the survey does not matter. You will be asked to answer a set of seven questions for each of your classes. When you are done with the survey we will come together as a group to discuss your thoughts about the questions.
We will not focus too closely on your specific responses to the question, but more on how you interpreted certain terms or phrases in the question. For example, if the question was, “Yesterday, about how many meals did you eat?” I may ask you things like “What time period were you thinking about when you answered that question?” or “What did you consider as a ‘meal’ in that question?” Does that make sense?

As a guideline, there are obviously no right or wrong answers. We value everyone’s point of view so please listen respectfully as other’s share their opinion. We are also going to be recording our discussion, so if everyone could speak loud enough and limit our group conversation to allow one person to speak at a time, that would be great. Do I have everyone’s permission to record the interview?

PERMISSSION TO RECORD: YES NO

TURN ON RECORDERs

My role as a moderator will only be to guide the discussion, so it is very important that you feel free to tell me any concerns, comments, or suggestions you may have as we go along. Your participation of course is voluntary, so if there is a question you do not wish to answer just let me know and we can move on to the next question. I also wanted to assure everyone that your answers may be used only for educational purposes and will not be disclosed, or used, in any identifiable form.

Also, please turn off or silence all cell phones.

Before we start, are there any questions about what we are going to be doing?

We sent an email to each of you at your UW student account. You will start today by accessing this email, and will then click the embedded web link to access the survey. Please go ahead and do this now. Stop when you have the first page of the survey up on your screen.

Great. Thank you.

Please take the next 10 minutes to complete the survey questions for your first course. When you are finished, remain on that page and wait quietly. We will move on once everyone is done. Please save any questions you may have for our discussion.

[Give respondents time to fill out the questionnaire – roughly 10 to 15 minutes.]

(QUESTION BY QUESTION START BY READING EACH QUESTION ALoud):

A. The UW-Madison is committed to ensuring that students like you achieve a set of Essential Learning Outcomes throughout your college experience (see here). By focusing on these topics, the hope is that students will leave college with a wide range of knowledge
about the world and skills that will help them succeed in the future.

Today, we are asking you to evaluate how well the courses you took last semester contributed to each of the essential learning outcomes. We will also ask you to evaluate the overall impact and quality of each of your courses. Please note that this survey is not intended to replace course or instructor evaluations that are conducted each semester by individual departments.

PROBE: After having read this paragraph, what do the Essential Learning Outcomes (ELOs) mean to you?

PROBE: Did you click the link where it said “see here?” How thoroughly did you read the descriptions of the ELOs?

IF NECESSARY: Was there anything that was confusing about this paragraph? Was there anything that could be clearer?

C. At the time you enrolled, did you take this course primarily for your major?

PROBE: In your own words, what does it mean to take a course “primarily for your major?”

IF NECESSARY: Did anyone have difficulty selecting a response for this question? If so, how come?

D. In general, how much did this course enhance your knowledge of the world, such as knowledge of human cultures, society, or science?

PROBE: In your own words, what does it mean to “enhance your knowledge of the world?”

PROBE: When answering this question, what did “society” mean to you? What did “science” mean to you?

IF NECESSARY: Did the examples “human cultures, society, or science” make it easier or harder for you to answer this question?

E. How much did this course help you develop intellectual skills, such as critical or creative thinking, quantitative reasoning, and problem solving?

PROBE: When answering this question, what did the term “intellectual skills” mean to you?

PROBE: What does it mean for a course to help you “develop intellectual skills”?
PROBE: In this question, what did “critical or creative thinking” mean to you? What did “quantitative reasoning” mean to you?

D. How much did this course help you develop professional skills, such as written and oral communication, computer literacy, and working in teams?

PROBE: When answering this question, what did the term “professional skills” mean to you?

PROBE: Did the examples of professional skills given in this question make it easier or harder for you to answer?

IF NECESSARY: Are there other examples of professional skills you were thinking of when you answered this question?

E. How much did this course increase your sense of social responsibility, for example by teaching you about cultures, providing opportunities for civic and community involvement, or by enhancing your ability to reason ethically?

PROBE: In your own words, what does it mean for a course to “increase your sense of social responsibility”?

PROBE: When answering this question, what did it mean to “reason ethically”?

PROBE: Besides the examples listed in this question, are there any other ways a course might influence your sense of social responsibility?

F. How much did this course improve your ability to combine knowledge or skills from different fields of study?

PROBE: In your own words, what was this question asking? [reread question]

PROBE: When answering this question, what did it mean to “combine knowledge and skills from different fields of study”? What were the “different fields of study” you thought of when answering this question?

G. How would you rate the overall educational value of this course, that is the extent to which the course improved your all-around education or prepared you for the future?

PROBE: In your own words, what is the “educational value” of a course? Did the definition that was provided make it easier or harder to rate the course?
PROBE: Did the response options (Very Poor, Poor, Fair, Good, and Very Good) make it easier or harder to answer this question?

________________________

H. How would you rate the overall quality of this course, that is the extent to which it was structured and taught in order to maximize its educational value?

PROBE: In your own words, what does the “overall quality” of a course include? Did the definition that was provided make it easier or harder to rate the course?

PROBE: Did the response options (Very Poor, Poor, Fair, Good, and Very Good) make it easier or harder to answer this question?

________________________

General:

PROBE: Thinking of the questionnaire overall, did anyone have any general comments?

PROBE: As you filled out the questionnaire, did you find yourself comparing the answers you gave from one course to another? Did you go back to previous items to change your initial response? If so, which items were these?

PROBE: Overall, was it easy or hard to answer the questions using the response categories provided – that is, not at all, a little, somewhat, quite a bit, and a great deal?

PROBE: In your own words, what are the Essential Learning Outcomes?

________________________

MODERATOR HANDS OUT DATA REPORT

PROBE: Are these graphs easy or hard to interpret? Can you tell me what about them makes them [easy/hard] to interpret?

PROBE: Do you think you would use the results of this survey to help you select classes in the future? Why or why not? How might these graphs be changed to be more clear or useful?

IF NECESSARY: What do you think about the tick marks indicating the means for majors versus non-majors? Did you notice them?

________________________

THANK YOU FOR YOUR PARTICIPATION!
Appendix D.

Draft Motion

Whereas UW-Madison has a need for course assessments that yield comparable data on course educational impact, for use in program assessment, and to help guide students in the choice of courses, yet such a system is not in place.

Whereas it is impractical to burden departments with the obligation to collect comparable student evaluation of courses and enter it into a centralized database.

Whereas the Essential Learn Outcomes (ELOs) have been selected to define our overarching objectives for what students learn in their time at UW.

It is hereby moved that:

1) A Joint Governance, Supplemental ELO-based Course Survey Implementation Committee (SECSIC) be formed and charged with (a) designing an online survey that could be used to collect data from all students each semester and (b) guiding the development of a user interface and a code of practice for presenting the results of the survey to relevant stakeholders.

2) Central administration be asked to work with SECSIC during the design process and then to find funds to implement the system as quickly as possible.