Developing the Online Course Questionnaire: Report from the Subcommittee

2012

A summary report on the Online Course Questionnaire process and results from pilot-testing campus core items.

Online Course Instructor Evaluations at Indiana University

Introduction

On April 17, 2012, the Bloomington Faculty Council (BFC) approved Circular B40-2012, a policy that addressed 10 issues. (1) All courses will have an instructor evaluation element in the form of student-completed questionnaires. (2) The course questionnaires should be administered online. (3) All course questionnaires will have a common set of items that will be used for promotion and tenure (P&T). (4) Individual instructor results on these items cannot be made public. (5) Additional questions can be developed by schools, colleges, departments, programs, and instructors and included in the Online Course Questionnaire (OCQ). (6) The results of these questions will be released to the units that developed the items. (7) A set of "student-return" items will be on all OCQs, and these results will be available for all students using a password-protected interface. (8) The circular also addresses who has access to the common multiple-choice questions; and (9) any open-ended questions that might be included. (10) The BFC also approved the requirement that all courses offered at IUB be evaluated by students.

Before Circular B40-2012 was approved, the structure of course evaluations at Indiana University had several disadvantages. Professors have disliked taking time from class to administer the evaluations and often left too little time for students to complete them. Many students have been skeptical of the process, believe that the results have no bearing on teacher performance, and are not used towards change or improvement. Furthermore, the responses have not been shared with current or future students, limiting their usefulness to students, and encouraging them to seek information about IUB instructors on other publicly available websites. To explore the possibility of correcting these problems, the Office of the Vice Provost for Undergraduate Education (OVPUE) and the Office of the Vice Provost for Faculty and Academic Affairs (OVPFAA) created a task force in 2009 to design and pilot test items for an online questionnaire. The online questionnaire is designed to be completed outside of class to avoid diverting valuable class time. The questionnaire has also been streamlined to increase the clarity and relevance of the results. The quantitative results could be efficiently placed online for future students to learn more about a professor and a course before choosing to enroll in a course.

In 2009, the VPFAA and VPUE assembled the Task Force on Online Course Evaluation to address these issues. The first task was two-fold: 1) investigate the advantages and any issues associated with moving to an online system and 2) devise a campus-wide core set of items, in which a predetermined number of items would be used for the P&T process and others identified as student-access items. The student-access items are separate from the items used for P & T.

The Ad Hoc Item Subcommittee, hereafter referred to as Item Subcommittee, was created in September 2010 and was charged with the research and development of the OCQ. The Item Subcommittee focused on creating reliable and valid items for the campus core items and a corresponding rationale for each item (see Appendix A). The framework adopted is one that many researchers have used, including Churchill (1979), Krueger and Casey (2009), Pike and Ouimet (2009) and others, which recommends specifying the domain, generating items, obtaining subject experts' input to establish content and face validity, conducting focus groups, pilot-testing items, analyzing the data, and making changes accordingly. Also adopted in the design and implementation of the OCQ were Fink & Kosecoff's (1998) suggestions that good scale construction requires the following five steps for construct validity: 1) articulation of a clear theoretical construct statement, 2) delineation of the content domain, 3) adequately writing items that sample the domain, 4) evaluation of the item pool, and 5) obtaining evidence of construct validity.

From September 2010 to May 2011, the Item Subcommittee revised a draft instrument of seven questions designed to capture evaluative information about the quality of instructor teaching and to provide administrative information for merit and P&T decisions. This was achieved in collaboration with teaching and learning subject matter experts in the Center for Innovative Teaching and Learning (CITL) and the survey research and item design experts in the Center for Survey Research (CSR). In May 2011 the items were honed and vetted by the Item Subcommittee. Each item created was supported with a rationale and justification for its inclusion.

Summer Pilot

In summer 2011, we used quantitative and qualitative methods to capture student input on the items. The first method was to capture data from students at the end of their course. The pilot instrument (see Appendix B) was administered in 18 classes: 12 Intensive Freshman Seminars (IFS) and six graduate-level courses of which five were School of Library and Information Science (SLIS) courses. One hundred and twelve first-year and 86 graduate students completed the evaluation using a paper form and 18 graduate students completed the online form, for a total of 216 completed evaluations. These courses did not constitute a random sample of IUB summer courses but instead were a convenience sample of faculty who were asked by the Item Subcommittee to participate. The Item Subcommittee added an open-ended question to the evaluation to capture additional thoughts or comments students might have about their course, instructor, or the evaluation.

The second method was the qualitative approach**.** Two student focus groups were conducted to obtain direct feedback. Participants included seven undergraduates majoring in School of Public and Environmental Affairs SPEA, Informatics, sociology/political science, Spanish education, and

mathematics in one group and eight graduate students in SLIS and Informatics/Public Policy in the other group. The objective was primarily to gain insights into the item content, item order, and response sets; the quality and meaning of the wording; how students approached evaluations; and what incentives were needed to persuade students to complete evaluations online. Overall, both groups provided excellent feedback. The graduate students were more analytical and spent time dissecting each question and often coming full circle and agreeing with the original wording, while the undergraduate students were direct in their assessment and concerns about an online evaluation. Undergraduates had more positive responses than did graduate students.

The outcome from the summer work was insightful and produced a few key changes. First, several of the items were slightly revised[1] and the "overall" course and faculty items, sometimes known as the "global items," were removed. Students reported difficulty differentiating between the two when reporting on "overall quality" of the course and instructor. The item subcommittee took these comments under advisement and removed the two overall items. Second, students wanted an opportunity to provide additional comments, hence the inclusion of two open-ended questions.

Fall 2011 Pilot Summary Findings and Changes

In October, the Item Subcommittee convened the campus Online Course Evaluation Task Force to report on the work and finalize the items. The items were subsequently approved by this group. In November, members of the Task Force and Item Subcommittee attended a joint meeting of the Faculty Affairs Committee and Education Policy Committee. Dennis Groth, Alice Robbin and Judy Ouimet shared the items and opened the floor to discussion. The main concern focused on Items 1 and 2[2] with the issue being use of the passive voice. Two versions of the evaluation form, one with a passive voice and the other an active voice were pilot tested at the end-of-semester OCQ (called the "Fall 2011 Pilot"). An additional question was added that focused on "challenge" to pair with the "motivate" question. Appendix C displays the two versions of the Fall pilot instrument. When comparing the two versions (see Table 1), the active voice items produced a slightly higher mean than did the passive voice items. A two-tailed t-test was conducted and was not significant at the .001 level.

[1] Q4 How effectively was class time used to help you learn?; Q5 How much did assignments, readings, or activities help you learn essential material (e.g., facts, ideas, concepts and techniques)?; Q6 How useful was instructor feedback (written comments, grades, verbal communication, etc.) in helping you to understand what you did well and where you needed to improve?

[2] Q1 How clearly were learning goals or objectives communicated to you?; Q2 How effectively was class time used to help you learn?

Table 1: Fall Pilot Passive & Active Voice Means and Standard Deviations

Item Wording	Version	N	Mean	Std. Deviation	Std. Error Mean
Goals & Objectives					
How clearly were course learning goals and objectives communicated to you in <class name >.	Passive	601	3.26	.753	.031
How clearly did the instructor communicate course learning goals and objectives in <class name >.	Active	585	3.38	.761	.031
Class time					
How effectively was class time used to help you learn in <class name >.	Passive	601	3.20	.832	.034
How effectively did the instructor use class time to help you learn in <class name >.	Active	585	3.35	.797	.033

In the Fall 2011 pilot we also included items to obtain student feedback on the process and item wording. In general, students had positive statements about the process and items. Most students found the items to be clear and easy to read, but some found that the items were not applicable to their particular class. For example, one student wrote in response to the graded work item "…there should be a 'not applicable' choice…when I never had any graded homework." Other response set option comments were "…no middle category like 'undecided' or 'no opinion' or 'does not apply'." Another student noted that, "there was a VERY large gap between "somewhat effective' and "not effective at all" and wanted to say something in-between the two." Several students thought that the evaluation would benefit from short text boxes after each of the generic multiple-choice questions, which would allow them to expand on their answers, to focus on a specific aspect of the course for detailed feedback or to explain why they chose a particular response. These students noted the availability of comment boxes at the end of the survey, but they thought that additional boxes would improve the quality of the evaluations. One student suggested that "Comments on Instructor and Comments on Course boxes would be very helpful, as multiple-choice questions are not very comprehensive and provide almost no material that can lead to improvement in either the instructor's teaching skills or the course layout." It was our judgment that the cost of item-specific comment boxes would exceed the benefit, by increasing length of the OCQ, adding response burden, while complicating the reporting and analysis of results. Many students did request an opportunity to write overall comments about the instructor and the course, and this has been included among the proposed OCQ items.

Students were also asked if the evaluation did an adequate job of allowing them to express their views. Most answered in the affirmative, reporting that they were generally satisfied with the questions and responses on the evaluation. Several respondents commented on the brevity of the evaluation compared to the written evaluations in class. Most felt that this brevity

produced a clearer and more useful evaluation. As one student wrote, "the evaluation questions were relevant and condensed in comparison to the in-class form, which is tedious, repetitive, and often doesn't apply to the specific course." However, some students felt that additional questions focusing on exams, readings, and the professors' teaching styles would also be helpful.

Finally, participants were also asked to explain their understanding of Items 6 and 7 to gauge their interpretation of the terms "challenge" and "motivate." The responses indicated that the vast majority of students both shared common definitions of the terms and understood the nuanced differences between the terms. Interestingly, slightly more students seemed to feel motivated by their professors than challenged by the coursework. A cross tabulation of the two questions illustrates the difference (see Table 2).

Table 2: Crosstab of Motivate and Challenge

		Motivate				
		Not at all	Very little	Some	Very much	Total
Challenge	**Not at all**	0.9%	0.7%	0.3%	0.1%	2.0%
	Very little	1.0%	2.4%	2.3%	0.8%	6.5%
	Some	1.0%	2.6%	18.1%	14.4%	36.1%
	Very much	0.4%	1.5%	11.0%	42.6%	55.4%
	Total	3.2%	7.1%	31.7%	57.9%	

Spring item changes

Based on the results of the Fall 2011 Pilot, the Item Subcommittee made a few changes for the Spring 2012 Pilot to reflect the student feedback, even though the Spring pilot was designed to test the items on a larger scale in many different types of classes before we changed the items too drastically. First, we removed the "class name" embedded in the items as the students felt it was unnecessary. The Item Subcommittee examined other questions that faculty, departments and schools currently employ and noticed the items are primarily passive voice. The passive voice was chosen to maintain a common voice throughout the items. We maintained a 4-point Likert scale to be consistent across all items but agreed that the 4 point scale with an additional option of a "not applicable" would be reviewed once the data from the spring pilot were analyzed. We intentionally decided to omit the "undecided" or "no opinion" options because we want student to take a stand on these issues.

Spring 2012 Pilot

At the end of the 2012 spring semester, the CSR and OVPUE partnered to implement the online course evaluation with a larger number of courses. The primary purpose of the Spring 2012

pilot was to test seven quantitative and two qualitative items for consideration as the campus core section and three beta items for the student-access section (see Appendix D). The secondary purpose was to test the online process for completing course questionnaires outside of class.

The first step in the pilot test implementation was to obtain faculty participation in the pilot. All faculty who taught either a lab or lecture were eligible to participate in the Spring Pilot. To promote the study, invitation letters were sent from the Education Policy Committee (EPC) chair and the Faculty Academic Committee (FAC) chair, along with personal solicitations from various Item Subcommittee members. Mary Wennerstrom, Associate Dean for Instruction at the Jacobs School of Music, was instrumental in recruiting the Music faculty. The Registrar's Office provided the list of students in each of the classes.

Many faculty, including most of the Music faculty, had students complete two evaluations—the standard evaluation form and the OCQ pilot. In an effort to reduce the burden for students we created two versions of the OCQ (see Appendix D Spring 2012 Pilot Items). Version A consisted of the seven quantitative items originally designed for the campus core section along with three open-ended questions, two of which focused on the course and instructor and the third designed to acquire student feedback on the OCQ items and their corresponding response options. Version B differed slightly in that three additional items were added to address possible student-access questions. The additional items were a first attempt to create items geared towards areas that students might find helpful when deciding on courses and instructors.

The Center for Survey Research (CSR) created the OCQ with Qualtrics Software. The CSR sent emails to students in 434 class sections targeting 9,028 students, using the email address the students had on file. One or two reminders were sent. On April 19th, the initial invitation was sent to 7,231 students. On April 23rd, 26th, and 27th, an additional 1,797 invitations were sent after the first wave per the requests of the faculty members to mimic previous timing, or because faculty agreed to participate after our cut-off date. A first reminder was sent on April 23rd for the initial group, and for the late invitees on April 25th, 26th or May 1st. A final reminder was sent to only the students in the initial group who were first contacted on April 19th (see Table 3). After one or two reminders we obtained 4,669 completed evaluations for an overall response rate of 52%.

Table 3: Online Course Evaluation Dispositions

		Total	Percent Returned
Mailings			
	Invitation	9028	
	Reminder 1	6589	27.0%
	Reminder 2*	4634	48.6%
Total Adjusted Response		4669	52.1%
	Total Response Assessed**	4654	52.0%
	Total Students with Corrected Grade***	8954	

*Only students in the initial group received two reminders. ** 15 cases were removed from the sample in the data clean-up process.
***Students counted in the total pool are those who were in the class roster at time of corrected grade

Spring 2012 OCQ Results

Sample Descriptive Statistics

Recruiting faculty to participate was primarily limited to faculty who were either members of EPC, FAC, or were recruited by members of the Task Force or Item Subcommittee. A total of 229 faculty members teaching 434 sections, with 9,028 invitations sent to 6,671 unique students[3] of which 4654 OCQ were completed for a 52% overall response rate (see Table 3). The Jacobs School of Music had 146 faculty members participate, which represented 268 sections. Overall, our convenience sample had a larger proportion of small classes (<20 students) participate than large classes.

Table 4 shows the participation by faculty status, class size, faculty discipline and number of sections by discipline. A greater number of tenured or tenure track faculty volunteered to participate than the combined non-tenure and associate instructors, 136 and 93, respectively. The Jacobs School of Music volunteered to pilot test the seven items in as many of their sections as possible. As a result, Music faculty and classes are over-represented, but the richness of the varying sizes and types of classes provided important input from students with regard to the applicability of particular items to courses. The Schools of Informatics and Computing, Journalism, Nursing, Optometry, and Social Work did not participate.

[3] 1407 students were enrolled in more than one course see Table 6

Table 4: Participation by Faculty Status, Class Size, Discipline & Sections

	N	%
Faculty Status (N=229)		
Associate Instructors (AI's)	24	10.5%
Non-Tenure Track	69	30.1%
Tenured or Tenure Track	136	59.4%
Class Size (N=434)		
Fewer than 5	69	15.9
5 to 9	102	23.5
10 to 19	124	28.6
20 to 29	69	15.9
30 to 49	42	9.7
50 to 99	13	3.0
100 to 150	9	2.1
>150	6	1.4
Faculty by Discipline		
Business	23	10.0%
College of Arts and Sciences	43	18.8%
Education	6	2.6%
Public Health	4	1.8%
Law	1	0.4%
Music	146	63.8%
SLIS	4	1.8%
SPEA	2	0.9%
Sections by Discipline		
Business	50	11.5%
College of Arts and Sciences	71	16.4%
Education	6	1.4%
Public Health	8	1.8%
Law	2	0.5%
Music	287	66.1%
SLIS	7	1.6%
SPEA	3	0.7%

Response Rates

The overall response rate was 51.7% (see Table 5). Looking at the response rate by class section (not shown because of the anonymity agreements made with faculty), the mode is 50% (43 classes), the mean and median are 52%; however there is greater variability in the response rate by school, course level, student level, and class size. For example, the Jacobs School of Music had the lowest average response rate, 44.4%, but had the greatest number of courses participating (Note: many of their students completed both the OCQ and their required end-of-semester evaluation form). Unfortunately, 27 music sections had no respondents, representing 2% (87 students) of our sampled Music students; the 27 sections ranged in size from one to 12

students with a median class size of two. Some faculty reported not being aware that their students were being asked to complete a second evaluation, thus not informing students of the importance of completing both forms, thus compounding the low response rate. Overall, professional (73.0%) and doctoral (60.0%) students had the highest response rates with seniors having the lowest (47.5%). Students in 400-level classes, which are typically taken by seniors, had the lowest participation by course level at 41.2%. Classes with fewer than 10 students had a lower response rate than larger classes. Classes with the highest response rate (55.7%) had 50 to 99 students enrolled; however, without the Music classes, the average response rate is 69.0%. Given the unequal numbers of students for each of the subsets (school, level, etc.), only a descriptive and exploratory analysis is presented (see Table 5).

Table 5: Response Rates by School, Enrollment Level, Student Classification, Course Level & Class Size

	Sample	Response	School RR
Overall	9028	4669	51.7%
School			
Business	1560	830	53.2%
College of Arts	2768	1562	56.4%
Education	110	73	66.4%
Public Heath	389	240	61.7%
Law	150	108	72.0%
Music	3883	1723	44.4%
SLIS	128	99	77.3%
SPEA	41	34	82.9%
Enrollment Level			
Undergraduate	7297	3749	51.4%
Graduate	1657	905	54.6%
Student Classification			
Freshman	1796	970	54.0%
Sophomore	2097	1101	52.5%
Junior	1403	726	51.7%
Senior	1975	939	47.5%
Non-degree	26	13	50.0%
Masters	1196	611	51.1%
Professional	131	96	73.3%
Doctorate	330	198	60.0%

Source: Spring2012RRFacStats.xlsx

Table 5 continued: Response Rates by School, Enrollment Level, Student Classification, Course Level, & Class Size

	Sample	Response	School RR
Course Level			
100	2312	1223	52.9%
200	2776	1435	51.7%
300	1259	679	53.9%
400	1042	429	41.2%
500	894	513	57.4%
600	192	116	60.4%
700	28	20	71.4%
900	526	254	48.3%
Class Size			
Fewer than 5 (N=73)	201	82	40.8%
5 to 9 (N=101)	681	289	42.4%
10 to 19 (N=124)	1670	857	51.3%
20 to 29 (N=57)	1284	680	53.0%
30 to 49 (N=23)	866	455	52.5%
50 to 99 (N=19)	1307	728	55.7%
100 to 150 (N=14)	1678	875	52.1%
>150 (N=6)	1342	703	52.4%
Source: Spring2012RRFacStats.xlsx			

The sample contained 1,407 students who were enrolled in more than one of our participating classes, which produced 6,671 unique students (see Table 6). Of the 5,264 students who were enrolled in only one class, the response rate was 57.1% compared to 56.8% of the students in multiple classes who responded to at least one of the multiple OCQ requests. Of 1407 students who were enrolled in more than one class participating in the pilot, 53.1% responded to all OCQ requests. Students who were enrolled in more than one class and completed all the OCQs had a response rate of 38.0%.

Table 6: Response Rate Breakdown by Number of Classes a Student is Enrolled

Number Classes Students are Enrolled	Invitation Sent N	Unique Students N	Responded* to at Least One OCQ N	Responded* to at Least One OCQ %	Responded to All OCQ Requests N	Responded to All OCQ Requests %
1	5264	5264	3007	57.1%	3007	57.1%
2	1606	803	472	58.8%	351	43.7%
3	1218	406	196	48.3%	126	31.0%
4	564	141	80	56.7%	42	29.8%
5	215	43	22	51.2%	9	20.9%
6	66	11	8	72.7%	4	36.4%
7	21	3	2	66.7%	2	66.7%
Subtotal (2 to 7)	*3690*	*1407*	*780*	*55.4%*	*534*	*38.0%*
Total	**8954**	**6671**	**3787**	**56.8%**	**3541**	**53.1%**

Comparison of Responders to Non-responders

In survey research, one of the first steps in analyzing results is to determine whether responders are representative of the sample or the overall population. We know that the sample has a bias towards Music and is not a true representation of the courses taught in the Spring 2012 semester. However, what is important, and of interest, is who responds to the online course questionnaires. Are students with higher grade point averages (GPA) more likely to complete the survey? Do OCQ responders follow the same response patterns as typical survey takers in that women are more likely to respond than men? Is there a bimodal distribution by GPA? For students in multiple classes in the study, how many do they complete? The next set of tables will address these specific questions.

To address the questions above, student demographic data provided by the Registrar were linked to the student responses. In the pilot the students were told that their responses would be confidential and that their individual responses would not be reported or provided to the faculty member. We offered confidentiality because we wanted the option to analyze the data with student characteristics. When we linked the OCQ files with the demographic data we were not able to match student record data for 15 responders. This was due to various reasons, including the fact that students evaluated the same course(s) multiple times, students were missing from the end-of-term roster, the faculty responded to the survey as a trial, or there was no available identifier linking the response to the student record. From this point forward the data reported represent this linked file of 8,954 students with their demographic data. Table 7 below, compares responders to non-responders on various student demographic characteristics.

In the pilot sample, there were slightly more males than females; however, approximately 10% more females completed the OCQ than did males (χ^2 =112.77, p<.0001). When comparing responders to non-responders by the race/ethnicity category, overall the groups were similar in response. All groups except Asians had more responders than non-respondents and the non-resident alien (international) students had many more responders than non-responders, 60.6% and 39.4% respectively (χ^2 =42.3, p<.0001). In reviewing the sample by grade point average (GPA), a higher proportion of higher-performing students, as measured by GPA, responded (χ^2 =131.35, p<.0001) than did lower-performing students. When reviewing the cell chi-squares, GPAs lower than 1 and between 2.5 to 2.99 have the smallest cell chi-squares. This indicates that the differences between the responders and non-responders are in the students with GPAs between 2 to 2.49 and 3.0 and higher.

Table 7: A Comparison of Responder to Non-responder Characteristics

	Responder		Non-Responder	
Gender				
Female	2539	57.7%	1863	42.3%
Male	2115	46.5%	2437	53.4%
Ethnicity				
Asian	238	48.8%	250	51.2%
African American	161	52.8%	144	47.2%
Hispanic/Latino	183	51.7%	171	48.3%
Non-Resident Alien[4]	718	60.6%	466	39.4%
Other	115	50.7%	112	49.3%
White	3210	50.7%	3127	49.4%
GPA				
3.5 to 4.0	2412	56.1%	1885	43.9%
3.0 to 3.49	1330	52.5%	1203	47.5%
2.5 to 2.99	622	46.8%	708	53.2%
2 to 2.49	227	39.9%	342	60.1%
1 to 1.99	58	29.7%	137	70.3%
Less than 1	5	16.7%	25	83.3%

A common concern raised by faculty is that the grade distribution of responders will be bimodal—students who are either low- or high-achieving are those who typically respond to the OCQ. To test the bimodal grade distribution hypothesis, a chi-square test (χ^2 =187.732, p<.000) and a t-test (-13.28, p<.000) were conducted using responders and non-responders and their course grades (Table 8). The large chi-square suggests that there is a difference in course

[4] Race and Ethnicity categories are based on U.S. Department of Education reporting requirements (Federal Register, Volume 72, Number 202, pp. 59266-59279: http://edocket.access.gpo.gov/2007/pdf/E7-20613.pdf). http://nces.ed.gov/ipeds/news_room/ana_Changes_to_10_25_2007_169.asp

grades between responders and non-responders. Further chi-square analyses revealed that students who earned an F in the course being evaluated were least likely to respond (cell χ^2 =97.880), while those earning A's were most likely to do so (cell χ^2 =40.907). Those earning C's were also overrepresented among non-respondents (cell χ^2 =31.108), while there was no difference in response propensity for students who earned B's. This finding indicates that lower-performing students are less likely to complete the OCQ than are higher-performing students. Other evidence confirms this finding; the t-test for mean GPA indicates that the non-responders' mean GPA was 3.04 compared to 3.32 for responders (-13.28, p<.000), indicating that on average, non-responders have slightly lower grades than responders.

Table 8: Comparison of Course Grades by Responders to Non-responders by Level and Course Grade

	Undergraduate				Graduate			
	Responders		Non-responders		Responders		Non-responders	
	N	%	N	%	N	%	N	%
As	1903	57.5%	1407	42.5%	695	54.9%	572	45.1%
Bs	1131	51.2%	1077	48.8%	186	58.7%	131	41.3%
Cs	490	44.3%	616	55.7%	13	31.0%	29	69.0%
Ds	145	41.0%	209	59.0%	0	0.0%	1	100.0%
Fs	67	23.1%	223	76.9%	2	33.3%	4	66.7%

Online Course Questionnaire Summary Findings

Psychometrics and Item Analysis

Among the concerns faculty raised about administering a core set of common items to which all students in all IUB classes will be asked to respond, is applicability. More specifically, how can a set of items be created that will be appropriate to the diverse types of courses taught (content, size, level, type); the wide range of pedagogy employed; and whether the course is required or is an elective? All are appropriate concerns. This section considers these concerns; however, due to the structure of the data not all these concerns can be adequately addressed.

Validity and Reliability

When reviewing survey items, validity and reliability are often discussed together. According to Messick (1993), "validity is an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores and other modes of assessment. (p.13)" This definition is an important distinction in that validity is a property of how data are used, not the test or

assessment itself. The purpose of the campus core is to support faculty to enrich their pedagogy in ways that will enhance student learning. The item analyses and their corresponding rationales (see Appendix A) are used to demonstrate the validity of the seven core quantitative items.

One common procedure used to demonstrate reliability is to conduct a psychometric test known as a Cronbach's alpha[5]. This is simply a measure of internal consistency and is commonly used as an indication of the degree to which items measure a single construct. The higher the alpha score, the more reliable the generated scale is. The reliability procedure in the SPSS software was run on the following set of seven items:

1. How clearly were course learning goals and objectives communicated to you?
2. How effectively was class time used to help you learn?
3. How effectively did out-of-class work (assignments, readings, practice, rehearsing, etc.) help you learn?
4. How effectively did graded work allow you to demonstrate what you learned?
5. How much did the course challenge you to do your best work?
6. How much did the instructor motivate you to do your best work?
7. How available was your instructor to provide help when needed (in person, by email, etc.)?

The OCQ Cronbach's alpha[6] is .873, which indicates that the items are measuring one construct. All seven items create the strongest measurement of the construct as indicated in the item-total statistics shown in Table 9 below. The far right column shows what the Cronbach's alpha would be if an item was deleted. For example, if the first item, "How clearly were course learning goals and objectives communicated to you" was not included in the scale, the alpha would be .846, which is slightly lower than the .873 if the item remains in the scale.

[5] Alpha coefficient ranges in value from 0 to 1. The higher the score the more reliable the scale.
[6] Nunnaly (1978) states 0.7 is an acceptable reliability

Table 9: Cronbach's Alpha Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Item Deleted
How clearly were course learning goals and objectives communicated to you	19.59	13.371	.717	.544	.846
How effectively was class time used to help you learn.	19.64	13.331	.722	.545	.845
How effectively did out-of-class work (assignments, readings, practice, rehearsing, etc.) help you learn	19.84	13.171	.675	.491	.852
How effectively did graded work allow you to demonstrate what you learned	19.85	13.026	.690	.516	.850
How much did the course challenge you to do your best work	19.40	15.069	.521	.327	.871
How much did the instructor motivate you to do your best work	19.41	13.815	.710	.532	.848
How available was your instructor to provide help when needed (in person, by email, etc.)	19.39	15.012	.532	.312	.869

Item Analysis by Student Classification

Are these items appropriate for both undergraduate and graduate classes? One way to determine whether these items are applicable to each group is to examine Table 10, which presents the frequencies of student responses in each response category for each item by student classification (freshmen through seniors, professional, masters, and doctoral students). Non-degree students are also included in this table, but their numbers are too small to be considered reliable results. Looking at the item breakdown by student classification, there are similar patterns across undergraduate levels and a slightly more positive pattern when looking at graduate-level responses. Undergraduates tend to use the lower response categories more often than do graduate and professional students.

Among undergraduates, the pattern of responses across all items in Table 10 is similar: most of the responses (between 60% and 90% of all responses for each item) are in the top two

categories (clearly and very clearly, effectively and very effectively, some and very much, available and very available). For Items 1, 2, 5, 6, and 7, between 80% and 90% of respondents indicated that their instructors were in the top two categories, suggesting that faculty, in general, are doing quite well in those aspects of teaching. For Items 3 and 4, however, the responses are somewhat different. For these items, students' responses are in the top two categories only 60% to 70% of the time, suggesting that students perceive less consistency in the effectiveness of out-of-class work in helping them learn (Item 3), and in graded work allowing them to demonstrate their learning (Item 4). The lower two response categories were selected less often than the top two categories for most of the items by most of the students, but all response categories were selected by at least a few students for every question. This indicates that students are willing to use all the response categories for these questions, which is an important consideration in evaluating the validity of the questions and response options. (If students had *not* been willing to use all the response options for a particular question, it would have been necessary to change the question or the response options to ensure that they were appropriate.)

The same basic pattern of results can be seen in the professional, masters, and doctoral student responses. Graduate students show statistically significant higher responses than undergraduate students.. The majority of responses were in the top two categories for each item; this is particularly true for Items 1, 2, 5, 6, and 7. Again, however, Items 3 and 4 yielded different results, particularly for professional students. In responding to Items 3 and 4, professional students chose responses in the bottom two categories (somewhat effectively or not at all effectively) over 60% of the time. It seems unlikely that instructors of professional students are ineffective in these areas. Instead, it may be the case that these questions are not applicable for some students in the professional schools. Since a "not applicable" response option was not available for these questions, the professional students, and others, may have chosen responses in the lower categories rather than leave the items blank. Adding a "not applicable" option for these questions might address this problem.

Small trends are evident when comparing responses across undergraduate classifications (freshmen, sophomores, juniors, and seniors) for individual items. For example, for Item 1 (how clearly were course learning goals and objectives communicated to you), there is some indication that scores on this item change from freshmen to seniors, suggesting that course goals may be communicated more clearly in the upper-division courses in which juniors and seniors are likely to enroll or that this group understand them better so they seem clearer. Another trend occurs in Item 4 (how effectively graded work allowed you to show what you learned). The scores rise slightly from freshmen to seniors, suggesting that graded work in upper division courses, as reported by students, is more effective in allowing students to show what they have learned, compared to graded work in the lower division courses that freshmen

and sophomores often enroll in. In Items 5 and 6 (how much did the course challenge you to do your best work, and how much did the instructor motivate you to do your best work), there is a slight trend for scores to be higher for freshmen, and decline from freshmen to seniors. This may suggest that freshmen find their courses, in general, more challenging and motivating, although other explanations are possible, of course. When making comparisons, it is important to compare similar classes, level, etc.

The trends described here require more investigation to determine if they are consistent across more students and more courses. However, such trends, and the average overall results presented in Table 10, will be useful for faculty development because they suggest a place to start in thinking about how to improve teaching. If students in lower-division courses perceive that graded work does not allow them to show what they have learned, for example, this suggests that instructors might communicate the goals of assignments more clearly to students, or even redesign assignments to align them more closely with overall course goals and in-class activities. The results for other items can also provide a starting point for improving teaching by providing specific information about how students are responding to various aspects of a course.

There are two items, one that focuses on "out-of-class work" and the other that addresses "graded work." Responses from professional students indicate there is a difference in what students experience in the professional schools. For instance, 61.5% of professional students indicated that out-of-class work was not at all or somewhat effective compared to 17.2% and 11.8% for masters and doctorate students, respectively (Table 10). Prior to implementing the OCQ, a faculty member (EPC/FAC/BFC) voiced a concern that the class design of courses in professional schools is such that no out-of-class assignments are required nor were professional students' work graded for feedback. This makes it difficult for students to respond adequately as there was no "not applicable" option. We plan to include a not applicable response for some items where relevant.

Figures 1 through 3 display the item means for the seven campus core items. Figure 1 displays the mean differences between undergraduates and graduate students by item for both the common core. As is apparent, the items that focus on "out-of-class work" and "graded work" have a different pattern for graduate students.

When creating the campus core items, the Item Subcommittee and the Task Force focused on items that would be valid across disciplines. Figure 2 represents the item means for the seven campus core items. In general, students across the various areas have similar patterns with the exceptions of the "workload" and "graded work" items where the professional schools differ. For more detail, see radar figures (see Appendix E) for campus common-core item means.

Table 10: OCQ Responses by Item and Student Classification

		Student Classification Level							
		Freshman N=969	Sophomore N=1099	Junior N=726	Senior N=939	Non-degree N=13	Professional N=96	Masters N=611	Doctorate N=198
1. How clearly were course learning goals and objectives communicated to you	Not at all clearly	2.1%	3.8%	3.0%	5.1%	.0%	1.0%	2.8%	3.5%
	Somewhat clearly	18.7%	17.7%	16.8%	12.6%	.0%	10.4%	10.6%	4.5%
	Clearly	40.1%	34.7%	33.9%	31.7%	23.1%	39.6%	27.5%	26.3%
	Very clearly	39.1%	43.9%	46.3%	50.6%	76.9%	49.0%	59.1%	65.7%
2. How effectively was class time used to help you learn	Not at all effectively	2.8%	3.2%	3.5%	6.2%	0.0%	0.0%	3.5%	1.5%
	Somewhat effectively	14.5%	17.1%	17.5%	14.8%	15.4%	12.5%	14.2%	12.8%
	Effectively	41.4%	39.1%	34.3%	35.4%	46.2%	42.7%	32.3%	24.0%
	Very effectively	41.3%	40.6%	44.8%	43.5%	38.5%	44.8%	50.0%	61.7%
3. How effectively did out-of-class work (assignments, readings, practice, rehearsing, etc.) help you learn	Not at all effectively	4.4%	7.0%	6.3%	8.2%	7.7%	19.8%	3.0%	2.6%
	Somewhat effectively	24.0%	25.5%	25.6%	20.6%	7.7%	41.7%	14.2%	9.2%
	Effectively	41.2%	39.0%	35.7%	33.5%	53.8%	26.0%	31.2%	25.1%
	Very effectively	30.4%	28.5%	32.5%	37.7%	30.8%	12.5%	51.6%	63.1%
4. How effectively did graded work allow you to demonstrate what you learned	Not at all effectively	5.8%	6.9%	6.6%	7.0%	7.7%	36.4%	6.3%	1.5%
	Somewhat effectively	24.6%	22.6%	21.0%	20.2%	23.1%	24.7%	15.5%	16.5%
	Effectively	39.0%	40.4%	38.4%	34.6%	38.5%	26.0%	32.1%	24.2%
	Very effectively	30.6%	30.1%	34.0%	38.2%	30.8%	13.0%	46.1%	57.7%
5. How much did the course challenge you to do your best work	Not at all	0.8%	1.8%	1.4%	3.6%	7.7%	0.0%	3.0%	2.0%
	Very little	4.1%	5.3%	8.0%	9.7%	0.0%	12.9%	5.8%	2.6%
	Some	39.1%	36.6%	37.7%	36.0%	15.4%	54.8%	28.0%	28.6%
	Very much	55.9%	56.2%	52.9%	50.7%	76.9%	32.3%	63.2%	66.8%
6. How much did the instructor motivate you to do your best work	Not at all	2.4%	2.8%	2.2%	5.3%	0.0%	0.0%	4.5%	2.0%
	Very little	6.4%	7.5%	9.2%	8.2%	0.0%	10.6%	4.7%	3.1%
	Some	36.1%	34.9%	31.7%	30.4%	30.8%	38.3%	24.0%	19.4%
	Very much	55.2%	54.9%	56.9%	56.1%	69.2%	51.1%	66.8%	75.5%
7. How available was your instructor to provide help when needed (in person, by email, etc.)	Not at all available	1.3%	1.5%	1.1%	1.7%	0.0%	1.1%	1.0%	1.0%
	Somewhat available	10.6%	8.9%	8.4%	9.3%	7.7%	1.1%	7.4%	3.6%
	Available	37.4%	36.0%	30.7%	29.2%	15.4%	38.2%	25.6%	19.4%
	Very available	50.8%	53.7%	59.8%	59.8%	76.9%	59.6%	66.1%	76.0%

Figure 1: Item Mean Scores by Enrollment Level

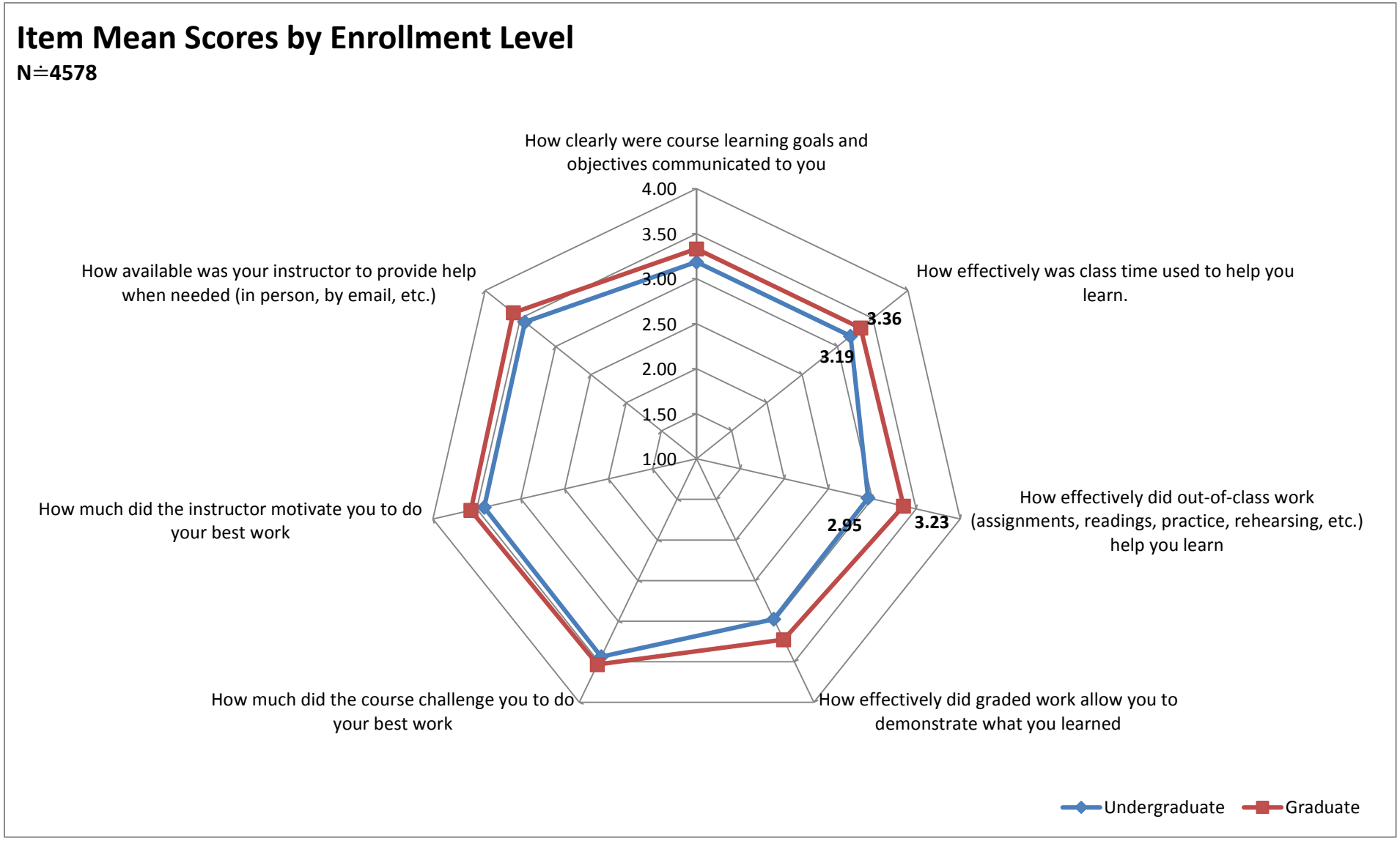


Figure 2: Item Mean Scores by College or School

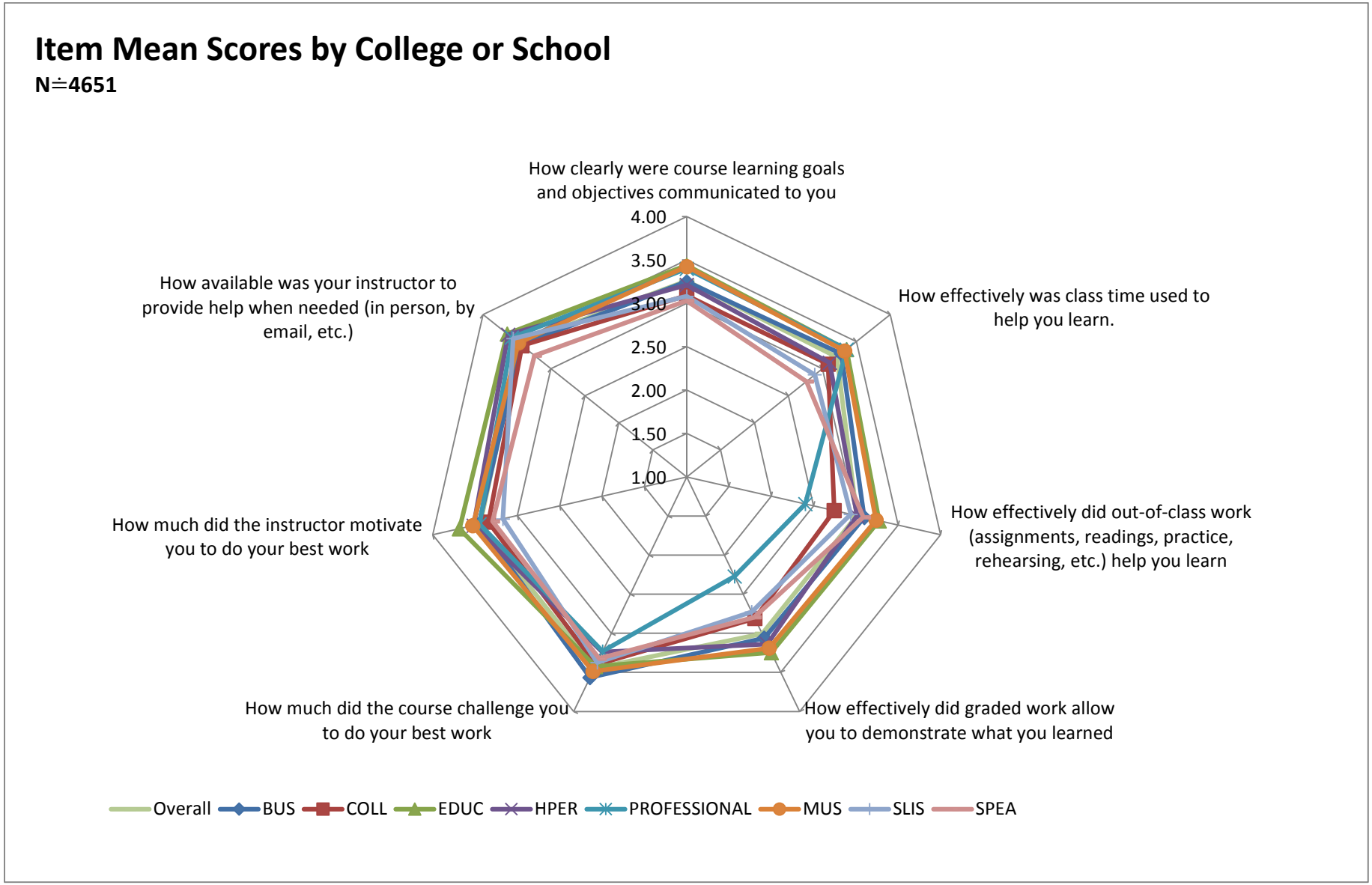
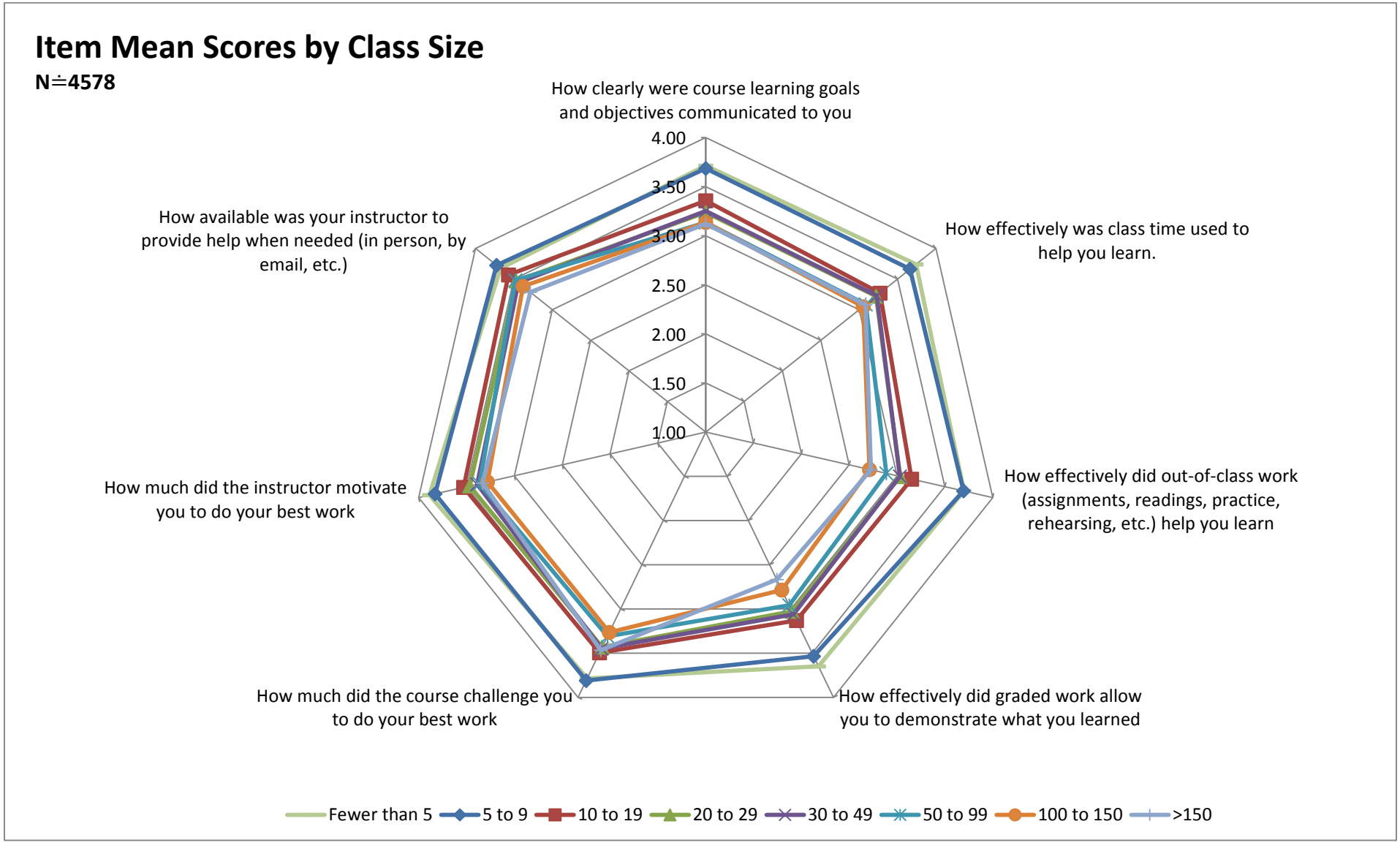


Figure 3: Item Mean Scores by Class Size



The suitability for the diversity of class size is equally important to the items being appropriate across disciplines (see Figure 3). When reviewing the seven campus core items, the pattern for very small classes (those with five or fewer or with five to nine students) is identical. For other class sizes the patterns are very similar but the level of reporting is different. Table 11 shows the means for each item by class size. Looking down the columns one can see that, in general, as the class size increases the mean score decreases. What is interesting from a teaching and learning perceptive is the data seem to be grouped into three class sizes: group one = nine or fewer, group two = 10 to 49, and group three = 50 or more.

Table 11: Campus Core Items Means by Class Size

Class Size	Campus Core						
	Q1	Q2	Q3	Q4	Q5	Q6	Q7
Fewer than 5	3.72	3.74	3.70	3.65	3.79	3.89	3.68
5 to 9	3.69	3.66	3.69	3.53	3.81	3.83	3.72
10 to 19	3.36	3.27	3.15	3.13	3.50	3.53	3.57
20 to 29	3.24	3.22	3.03	3.03	3.43	3.47	3.47
30 to 49	3.25	3.23	3.03	3.06	3.45	3.39	3.44
50 to 99	3.14	3.09	2.89	2.96	3.31	3.36	3.48
100 to 150	3.13	3.05	2.71	2.79	3.27	3.28	3.38
>150	3.12	3.08	2.73	2.67	3.47	3.33	3.29

Recommendations Part I

The Committee recognizes that students' responses to the recommended campus common items provide consistent but limited information for evaluating an instructor's teaching abilities and, more importantly, an instructor's impact on student learning and development. Teaching is a complex activity requiring subject matter expertise, effective pedagogical methods, and accommodation to diverse learning styles and learners. Additionally, students' perceptions of teaching effectiveness are influenced by some student and curricular characteristics that are beyond the control of an instructor. For these reasons, evaluating teaching requires multiple, convergent methods and measures. Moreover, student course evaluations should be judged in relation to contextual characteristics, such as class size, level, major requirement status, and other factors that systematically influence student perceptions.

After reviewing the data from the Summer focus groups, the open-ended questions from the Fall and Spring pilots, and input from faculty who participated (see Appendix F), the following changes were recommended (see Appendix G) to the items tested in the Spring 2012 pilot.

- Add a parenthetical qualifier to provide a description of what "graded work" entails to Question 4—How effectively did graded work (papers, exams, presentations, etc.) allow you to demonstrate what you learned?
- Add "material" to help delineate the course from the faculty member to Question 5—How much did this course material challenge you to do your best work?
- Include a "not applicable" response option to Question 2—How effectively did out-of-class work (assignments, readings, practice, rehearsing, etc.) help you learn?
- Include a "not applicable" response option to Question 3—How effectively did graded work allow you to demonstrate what you learned?
- Include a "not applicable" response option to Question 4—How effectively did graded work (papers, exams, presentations, etc.) allow you to demonstrate what you learned?
- Include a "not applicable" to Question 7 – How available was your instructor to provide help when needed (in person, by email, etc.)?
- Include open-ended questions for students to provide more detailed feedback on their experience.
- Change the two open-ended questions to prompt students to think about both the course and instructor versus just the course.
- Change the response options to read positive to negative from left to right (very clearly to not at all clearly), which will convey an optimistic expectation.

Recommendations Part II

The Item Subcommittee met November 26th, 2012 to discuss and identify the items that meet criteria provided by Vice Provost Gieryn and adjust to the BFC Circular B40-2012, which limits the campus core to four multiple-choice items with an unspecified number of open-ended questions. In 2011, Vice Provost Gieryn discussed the need for a set of questions that focused on effective teaching to assist the campus P&T Committee and the OVPFAA in promotion and tenure decisions. The design of the instrument was narrowed to course organization, value of the course, instructor accessibility, and the student's perception of learning, which we used to reduce the current set of seven multiple-choice items to four (see Appendix H). It was noted that Question 6 is identical to a suggested student-access item, so that one will be in the student-access section. After some discussion of which questions apply to each criterion specified by VP Gieryn, the subcommittee recommends the following:

To address the criterion of course organization:

Question 1: How clearly were course learning goals and objectives communicated to you?

To address the criterion of instructor accessibility:

Question 7: How available was the instructor to provide help when needed (in person, by email, office hours, etc.)? [7]

To address the criterion of the value of the course:

Question 8: What did you like most about this course and instructor?

Question 9: What did you like least about this course and instructor?

To address the criteria of perception of learning, value of the course, and course organization:

Question 2: How effectively was class time used to help you learn?

Question 3: How effectively did out-of-class work (assignments, readings, practice, etc.) help you learn?

[7] Add "not applicable" to the response set. This was a sidebar conversation after the meeting.

Appendix A: Rationales

The Campus Core Items

The core module is designed to measure aspects of teaching practice and student experience that are related to effective teaching practices. This questionnaire does not capture all aspects of high-impact teaching; nor is it designed to provide the only information on faculty teaching performance. Questions 1 through 7 focus on effective teaching (whether instructional responsibilities have been met), high-quality interactions with the faculty, academic challenge and effort, and high expectations for student performance. Open-ended questions 8 and 9 provide the student with an opportunity to comment more specifically about the course. Each of the nine questions is presented with a brief rationale. Also provided is a list of resources used including results from the Wabash National Study of Liberal Arts Education[8].

1. How clearly were course learning goals and objectives communicated to you?
   Very clearly Clearly Somewhat clearly Not at all clearly

   Rationale: Course goals and objectives provide students with a clear understanding of what the course is about and what is expected of them. Without such goals, teaching can be unfocused and learning haphazard. Research demonstrates that explicit learning goals improve student learning, especially when they are specific and challenging, that is, they set high expectations for student performance.

2. How effectively was class time used to help you learn?
   Very effectively Effectively Somewhat effectively Not at all effectively Not Applicable

   Rationale: The effective use of class time is critical for providing students with an optimal learning experience. The importance of effective classroom time management is a consistent theme in the research studies related to distinguishing effective and ineffective teachers, as well as those related to student learning gains. Elements of good classroom management practice include: clear explanations, good use of examples and illustrations to explain difficult points, materials presented in a well-organized way, instructor is well prepared for class, class time used effectively to allow opportunities for active learning.

3. How effectively did out-of-class work (assignments, readings, practice, etc.) help you learn?
   Very effectively Effectively Somewhat effectively Not at all effectively Not Applicable

   Rationale: Good instructional practice requires planning and reinforcement of what is learned during class time. For out-of-class work to be most useful and instructive to the

[8] The 2006 - 2009 Wabash National Study was conducted to help researchers identify highly effective experiences and practices for college students. Data were collected from 2,200 students who completed tests and surveys three times during their four years at each of 17 four-year colleges and universities. (Center of Inquiry in the Liberal Arts, 2011.)

student, these reading assignments, problem sets, etc. need to be carefully created/selected and sequenced to correspond to in-class activities. Furthermore, out-of-class experiences provide opportunities for students to reinforce and to supplement what they learn in the classroom or connect it to outside experiences; integrate ideas or information from various sources; and synthesize and organize ideas, information, or experiences into new and more complex interpretations and relationships.

4. How effectively did graded work (papers, exams, presentations, etc.) allow you to demonstrate what you learned?

   Very effectively Effectively Somewhat effectively Not at all effectively Not Applicable

Rationale: Good instructional practice requires the student to demonstrate what has been learned in the course. The primary way that students demonstrate what they have learned is through graded work such as tests, assignments, and papers; graded work also constitutes an important form of feedback provided by the instructor to the students concerning their progress in the course. This question also ties into campus-wide assessment and is consistent with P&T language for what constitutes an effective teacher, including the concept that an effective teacher designs assignments that provide students with the opportunity to demonstrate what they have learned.

5. How much did this course material challenge you to do your best work?

   Very much Quite a bit Somewhat Not at all

Rationale: Good instructional practice requires academic challenge and high expectations. This item measures the quality of effort made by students, that is, the extent to which the course encouraged them to do well and engaged them. It also measures whether students believe that they were challenged to do their best work and reflects the philosophy that the student's academic experience should reflect the institution's challenging curriculum and its standards for rigor and high quality work.

6. How much did the instructor motivate you to do your best work?

   Very much Quite a bit Somewhat Not at all

Rationale: Good instructional practice requires that instructors set and communicate high expectations with the intent that setting high expectations will motivate students to perform at a high level in the course. This item focuses on the instructor's role in challenging students and eliciting their best performance. This item measures the ability of the instructor to use strategies that motivate students to work hard in a course: for example, helping the students see value in course content, creating the expectation that hard work will enable students to succeed, and establishing an environment that supports learning. An instructor's ability to motivate students is a key component in students' eventual success in a course.

7. How available was the instructor to provide help when needed (in person, by email, office hours, etc.)?

   Very available Available Somewhat available Not at all available Not Applicable

   Rationale: Good instructional practice requires high quality interactions with faculty. Instructors demonstrate their interest in teaching and student development by being accessible and available to assist students. Accessibility is one measure of the quality of non-classroom interactions with the instructor.

8. What did you like most about this course and instructor?

9. What did you like least about this course and instructor?

   Rationale: Students want an opportunity to comment on the course, and feedback informs improvement in process and content. Instructors often find these questions very helpful for revising course content and improving their instructional practice. Using these phrasings allows students to provide both positive and negative feedback about both the course and instructor (which eliminates the need to distinguish students' feedback about the course from that about the instructor), and it will be easy to distinguish the positive from the negative comments.

Appendix B: Summer 2011 Pilot Items

Course Evaluation

Directions: Please take a few minutes to respond to the below questions regarding your course. Your responses will remain anonymous and only aggregate data will be provided to your instructor. Please circle your response.

Part One:

1. Overall, I would rate this instructor as:

   Poor Fair Good Excellent

2. Overall, I would rate this course as:

   Poor Fair Good Excellent

Part Two:

3. How clearly were learning goals or objectives communicated to you?

   Not at all clearly Somewhat clearly Clearly Very clearly

4. How effectively was class time used to help you learn?

   Not at all effectively Somewhat effectively Effectively Very Effectively

5. How much did assignments, readings, or activities help you learn essential material (e.g., facts, ideas, concepts and techniques)?

   Not at all Somewhat Quite a bit Very much

6. How useful was instructor feedback (written comments, grades, verbal communication, etc.) in helping you to understand what you did well and where you needed to improve?

   Not at all useful Slightly useful Useful Very useful

7. How much did this course challenge you to do your best work?

   Not at all Somewhat Quite a bit Very much

8. Please share other thoughts or comments you might have regarding this course, instructor, or the evaluation.

Appendix C: Fall 2011 Pilot Items

Course Instructor Evaluation

Online Directions: Please take a few minutes to respond to the questions below regarding your course. Your responses will remain confidential and only aggregate data will be provided to your instructor.

PASSIVE VOICE VERSION

1. How clearly were course learning goals and objectives communicated to you in <INSERT COURSE Name>?
   Not at all clearly Somewhat clearly Clearly Very clearly
2. How effectively was class time used to help you learn in <INSERT COURSE NAME>?
   Not at all effectively Somewhat effectively Effectively Very Effectively

ACTIVE VOICE VERSION

1. How clearly did the instructor communicate course learning goals and objectives in <INSERT COURSE NAME>?
   Not at all clearly Somewhat clearly Clearly Very clearly
2. How effectively did the instructor use class time to help you learn in <INSERT COURSE NAME>?
   Not at all effectively Somewhat effectively Effectively Very Effectively

COMMON ACROSS BOTH VERSIONS

3. How effectively did out-of-class work (assignments, readings, practice, rehearsing, etc.) help you learn in <INSERT COURSE NAME>?
   Not at all effectively Somewhat effectively Effectively Very Effectively
4. How effectively did graded work allow you to demonstrate what you learned in <INSERT COURSE NAME>?
   Not at all effectively Somewhat effectively Effectively Very Effectively
5. How much did <INSERT COURSE NAME> challenge you to do your best work?
   Not at all Very little Some Very much
6. How much did <INSERT **FACULTY** NAME> motivate you to do your best work?
   Not at all Very little Some Very much
7. How available was <INSERT **FACULTY** NAME> to provide help when needed (in person, by email, etc.)?
   Not at all available Somewhat available Available Very available
8. What did you like most about <INSERT COURSE NAME>?
9. What did you like least about <INSERT COURSE NAME>?

Appendix D: Spring 2012 Pilot Items

Online Directions: Please take a few minutes to respond to the questions below regarding your course. Your responses will remain confidential and only aggregate data will be provided to your instructor.

1. How clearly were course learning goals and objectives communicated to you?
   Not at all clearly Somewhat clearly Clearly Very clearly

2. How effectively was class time used to help you learn?
   Not at all effectively Somewhat effectively Effectively Very Effectively

3. How effectively did out-of-class work (assignments, readings, practice, rehearsing, etc.) help you learn
   Not at all effectively Somewhat effectively Effectively Very Effectively

4. How effectively did graded work allow you to demonstrate what you learned?
   Not at all effectively Somewhat effectively Effectively Very Effectively

5. How much did the course challenge you to do your best work?
   Not at all Very little Some Very much

6 How much did the instructor motivate you to do your best work?
   Not at all Very little Some Very much

7 How available was your instructor to provide help when needed (in person, by email, etc.)?
   Not at all available Somewhat available Available Very available

8 What did you like most about the course?

9 What did you like least about the course?

OPTIONAL Beta Questions:

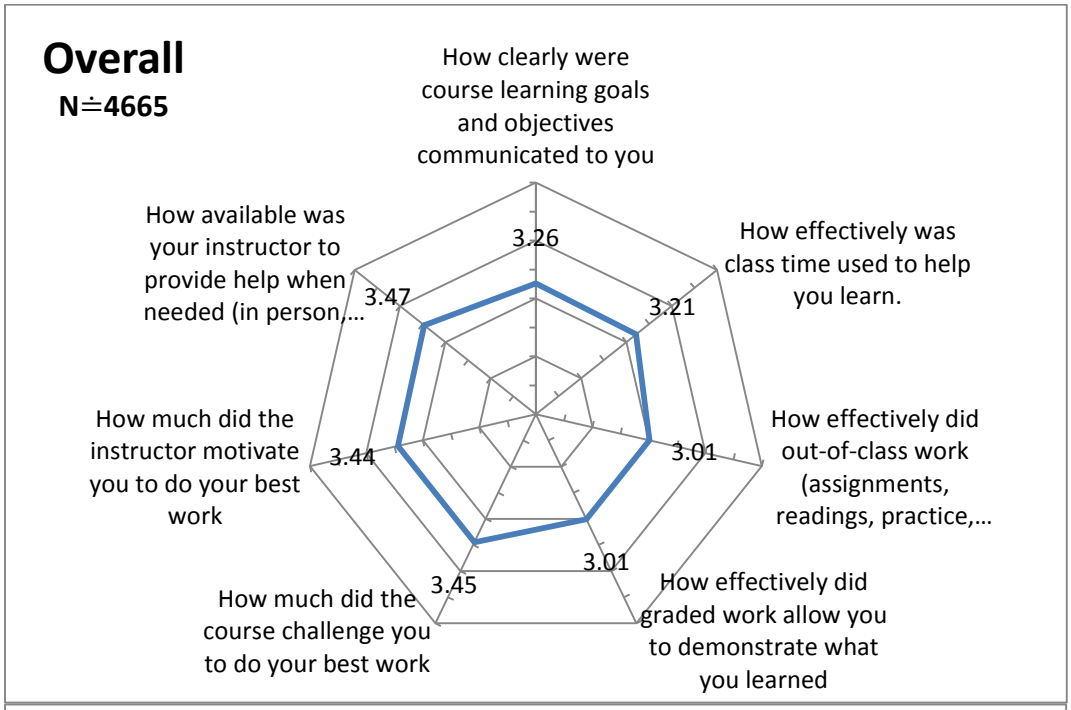
10. How likely would you be to recommend this class to a student who is considering taking it?
    Not at all likely Somewhat likely Likely Very likely

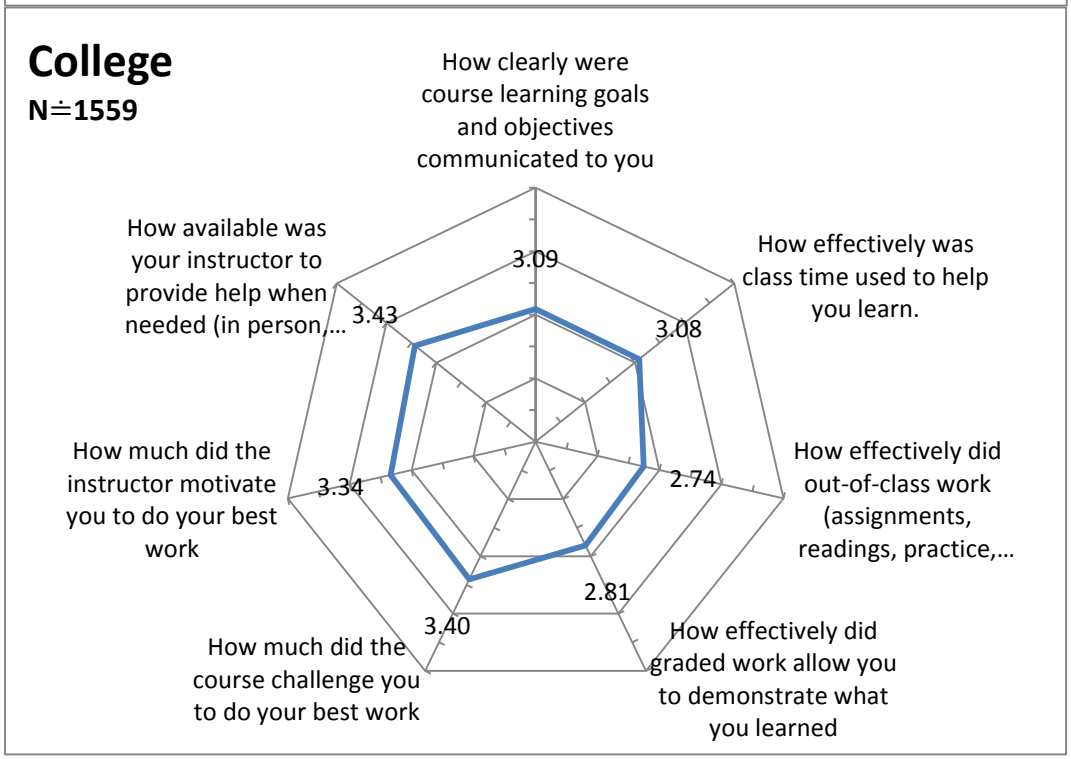
11. If you had the opportunity, how likely would you be to take another class from this instructor?
    Not at all likely Somewhat likely Likely Very likely

12. Compared to other courses at this level, how challenging was the workload?
    Much less challenging Less challenging Equally challenging More challenging Much more challenging

13. Please share other thoughts or comments you have with regard to the items and response options of this online course questionnaire.

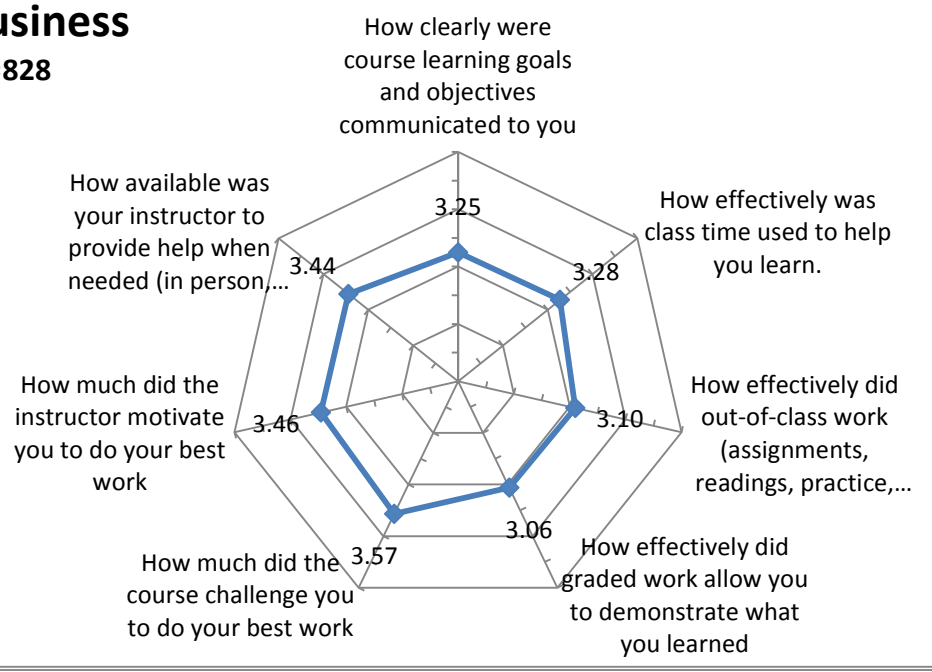
Appendix E: OCQ Campus Core Item Means by School





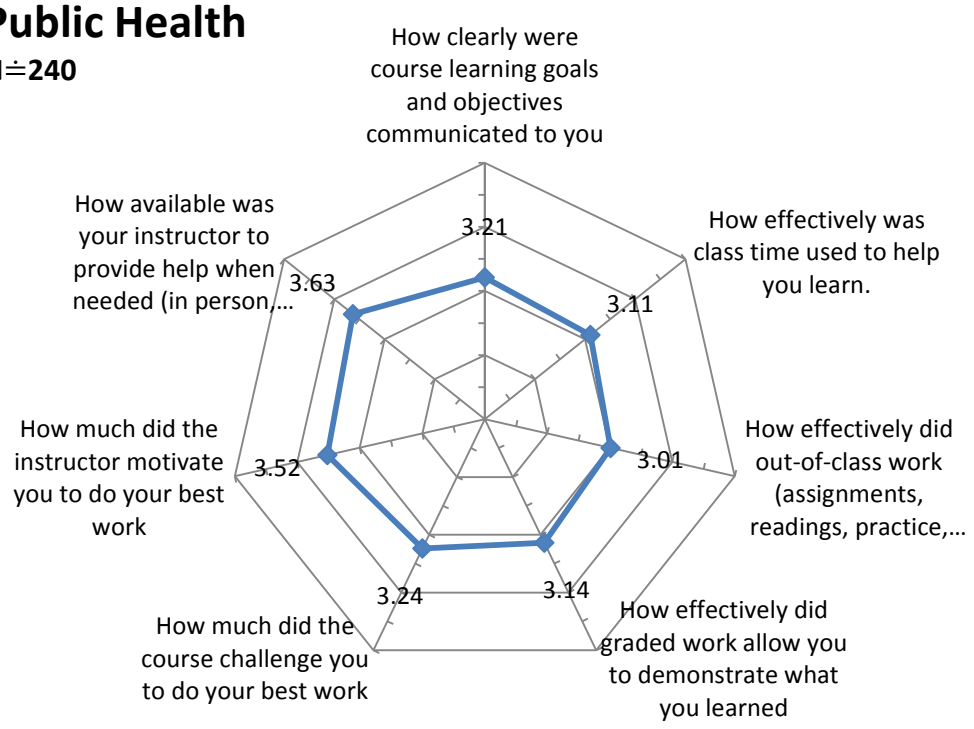
Business

N≒828



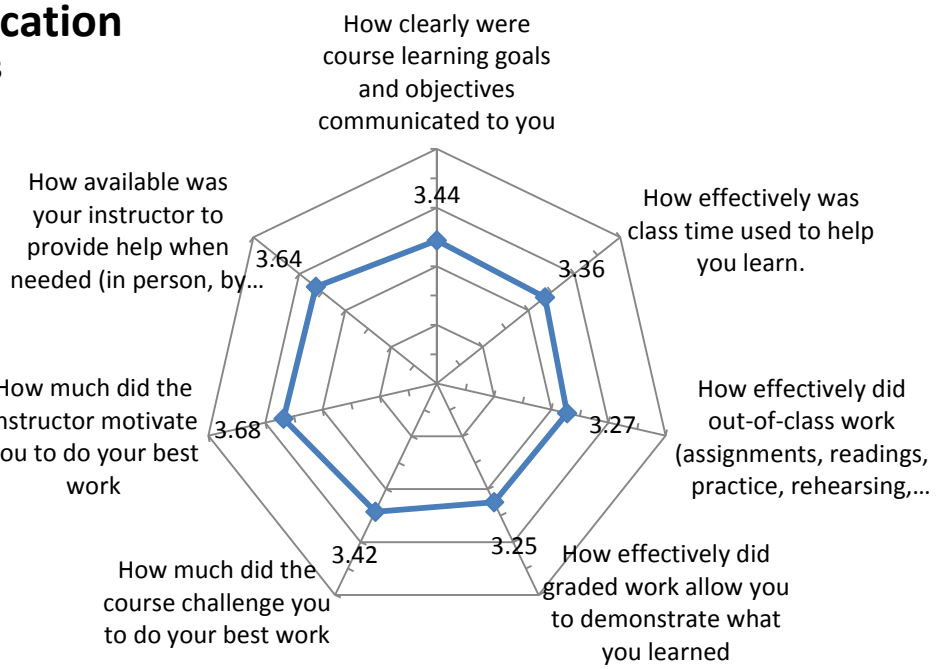
Public Health

N≒240



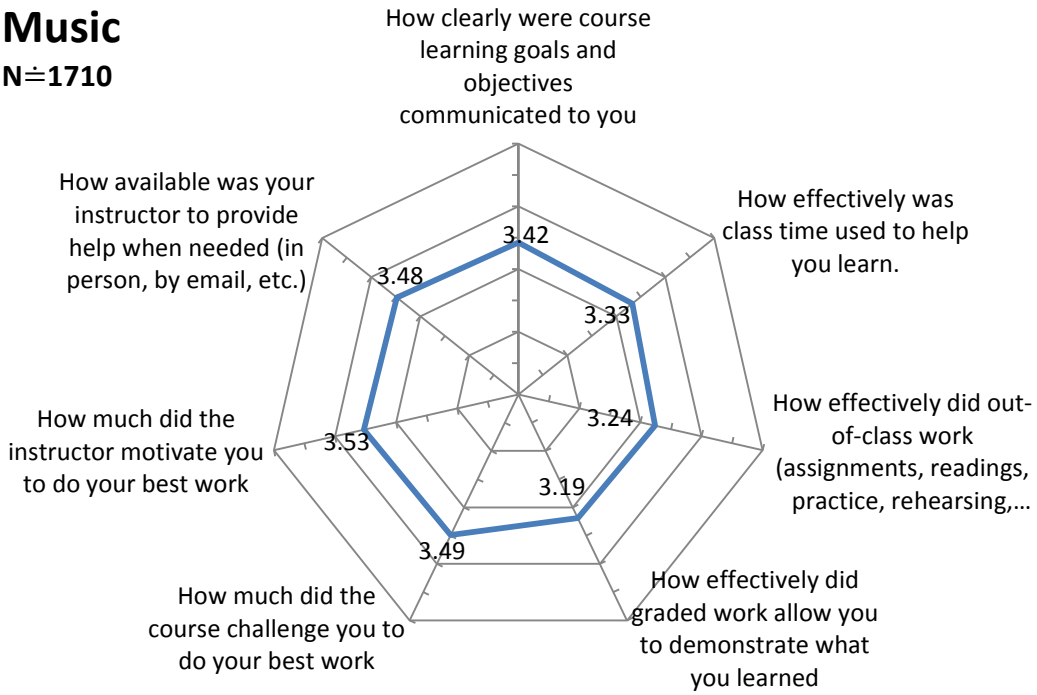
Education

N≒73



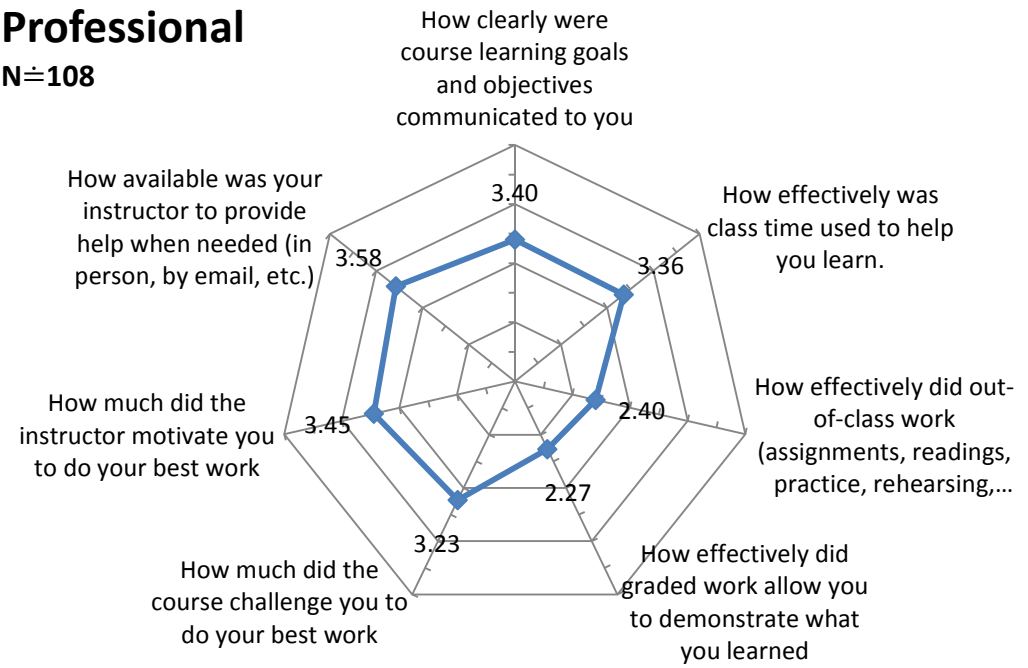
Music

N≒1710



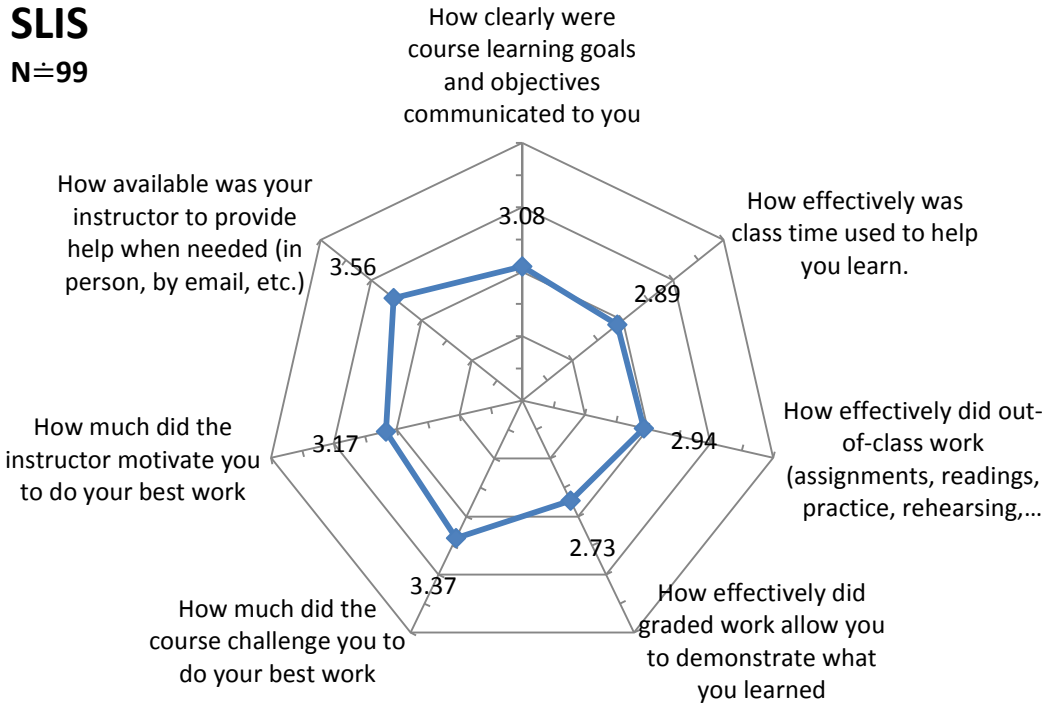
Professional

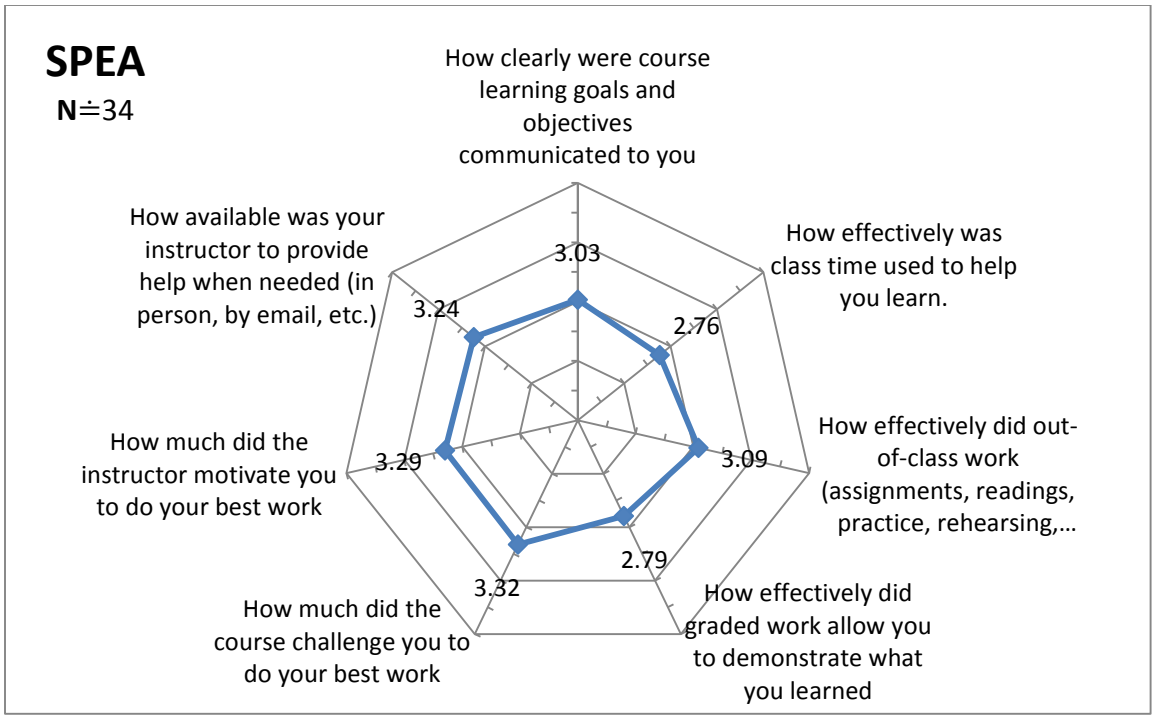
N≒108



SLIS

N≒99




SPEA
N≒34
How clearly were course learning goals and objectives communicated to you
How effectively was class time used to help you learn.
How effectively did out-of-class work (assignments, readings, practice, rehearsing,…
How effectively did graded work allow you to demonstrate what you learned
How much did the course challenge you to do your best work
How much did the instructor motivate you to do your best work
How available was your instructor to provide help when needed (in person, by email, etc.)
3.03
2.76
3.09
2.79
3.32
3.29
3.24

Appendix F: OCQ Instructor Feedback

Spring 2012 Instructor Evaluation of Online Course Questionnaire Brief

In July, a follow-up survey to all faculty who participated in the 2012 OCQ pilot was sent electronically. A total of 78 faculty completed the entire survey for a 35% response rate. The items below are just a few of the questions and some of the comments in response to the open-ended questions. This brief is an attempt to share positive and negative comments about the OCQ.

Faculty respondents:

Table 1: Cross tabulation between years taught at any college or university by years taught at IU

		How many years have you taught at any college or university?					
		less than 1 year	1 to 3 years	4 to 7 years	8 to 11 years	12 or more years	Total
How many years have you taught at IU?	First year teaching at IU	2 18.18% 100.00%	6 54.55% 35.29%	2 18.18% 11.76%	0 0.00% 0.00%	1 9.09% 2.50%	11 100.00% 12.50%
	1 to 3 years	0 0.00% 0.00%	11 61.11% 64.71%	4 22.22% 23.53%	3 16.67% 25.00%	0 0.00% 0.00%	18 100.00% 20.45%
	4 to 7 years	0 0.00% 0.00%	0 0.00% 0.00%	11 64.71% 64.71%	3 17.65% 25.00%	3 17.65% 7.50%	17 100.00% 19.32%
	8 to 11 years	0 0.00% 0.00%	0 0.00% 0.00%	0 0.00% 0.00%	6 37.50% 50.00%	10 62.50% 25.00%	16 100.00% 18.18%
	12 or more years	0 0.00% 0.00%	0 0.00% 0.00%	0 0.00% 0.00%	0 0.00% 0.00%	26 100.00% 65.00%	26 100.00% 29.55%
	Total	2 2.27% 100.00%	17 19.32% 100.00%	17 19.32% 100.00%	12 13.64% 100.00%	40 45.45% 100.00%	88 100.00% 100.00%

We asked a few questions regarding helpfulness of the OCQ and the responses to the open-ended OCQ items. Seventy-seven percent of the faculty reported that the results were very helpful or somewhat helpful and 5% reported not looking at them at all. When asked about the open-ended questions, 74% found the student responses very or somewhat informative with 6% reporting not reading them.

Table 2: How helpful were the results of the online evaluation to you?

Answer	Response	%
Very helpful	27	32%
Somewhat helpful	38	45%
Not too helpful	7	8%
Not at all helpful	8	10%
Did not look at them	4	5%
Total	84	100%

Table 3: More specifically, how informative/helpful were the open-ended questions?

Answer		%
Very informative	28	35%
Somewhat informative	31	39%
Not too informative	11	14%
Not at all informative	4	5%
I did not read them	5	6%
Total	79	100%

Q19: How do the two versions compare? (Faculty comments are presented verbatim.)

General:

- Many faculty could not compare the versions as they had not received the paper reports, even though the survey was sent in the second week of July, two months after the end of the semester. One faculty member wrote, “This is a great question, but unfortunately I still do not have access to the paper version. Which is my main reason in advocating for the quick and instant online version =).”
- Questions were different and there were more on the paper version.
- In the case of course XX, questions in the online evaluation seem to be oriented around actual course activities and work in relation to the individual student (e.g., How were course goals and objectives communicated? How was class time used? How much did the instructor

motivate you?), whereas questions in the paper evaluation seem to be about attributes of the instructor (knowledge of subject matter, ability to organize, enthusiasm, and so on). OCQ online questions also appear to cover a broader range than the limited questions in the paper evaluation. Whether due to this difference in emphasis and breadth, to the characteristics or feelings of the (smaller) population of students who completed the online evaluation, or to a combination of both, the online evaluation results are perhaps somewhat less immediately positive than the paper evaluation results. I do not necessarily view this as a bad thing; it may be that the questions address areas and issues that fall between the cracks in the paper evaluation.

Positive:

- My assumption was that the online version would result in more "troll" like comments, such as is the case with YouTube, Facebook, and other social media as the psychology seems to be different when students are online away from the classroom. However, I was surprised to see that there was more thoughtful feedback provided in the online version. I did have more students complete the paper versions but I think that I (sic) had more to do with students not wanting to provide evaluations more than once.
- They were comparable. However, students provided more comments in the online version. Also, these comments were more detailed.
- Much preferred not losing class time for paper evaluations
- I find the online version to give better results and more thorough comments
- Very similar
- I think the online versions are better - they take up no class time and students all have a chance to fill them out at their leisure (instead of only those who attend class on the day paper forms are given out).

Negative:

- The questions seemed different. Also, the students who attend the class period where the SET is administered face-to-face tend to be students with good attendance (poorly attending students forget). These attending students are the best judges of my teaching performance. In contrast, students who respond only electronically, may or may not have had good attendance; and if they didn't attend regularly, then they are poor judges of my performance.
- I feel the students took more time and put more thought into the paper version.
- I highly prefer the paper versions for multiple reasons: 1) I issue them before the final, so I always have 100% response rates; 2) I issue them right before giving the final, so the

"course" is truly finished (last days of class, study sessions, last-minute questions, feedback from presentations, submission of final papers/projects, etc.) and therefore best able to evaluate the course in its entirety; 3) when students are writing as a group they are less likely to whip through the evaluation and put more thought into free responses, and they are in a "class mindset" while they are filling it out. However, I like the wording of the new on-line questions, since they eliminate some of the bias and bad design that many of the standard paper questions exhibit.

- We ask just two multiple-mark questions and then open-ended questions. We get a much higher response rate on our paper form and the answers given are more useful and less biased by the pedagogical prejudices of the people who constructed the questions.

Q14: How did the Spring 2012 OCQ compare to your evaluations from previous semesters that you taught this course?

Neutral:

- about the same with some new information that I had never received before.
- Far lower # of participants. Averages were about the same.
- Less participants. My ratings were a bit higher than my first semester teaching the course. This was indicated on the paper version as well.

Positive :

- More students responded and the comments were more detailed than "I like him, more pizza, or less work."
- There was a little less participation from my section this time around, but overall I am happy with them and ultimately prefer them to the paper versions.
- Not as large as response because they are not captives in the room like with paper/pencil, but I think I received more sincere and useful information from those who participated. I prefer the electronic survey version
- I found that students took time to write more comments (or more words) as compared to the fall paper version.
- I generally have ratings between 3.5 and 4.0 for all questions, except "ability to Organize" for which I am rated between 3.0 and 3.5. comments are generally extremely positive and include constructive ideas, which I implement in the next semester

Negative:

- All but useless. The questions on the form tell me almost nothing I can use to improve the class. I have long used my own evaluation form: a blank sheet of paper on which I ask students to tell me, in words, what things they would like to see changed or that work well. The statistical analysis that accompanied the form was excessive; one shouldn't need SPSS to read course evaluations.
- Fewer people responded on-line as opposed to in-class paper evaluations
- Ratings were lower, again (I'd guess) because students who never came to class could fill out the online form.

Q16: Do you have any additional thoughts about your experiences with the online evaluation?

- I'd certainly prefer an online system, but perhaps a question could be included that ask students how often they attend the class. That would help add some context to the results .especially if that could be tied to the open-ended questions. In a big class like this one where attendance isn't taken, there will be some students who rarely if ever show. For the paper version that's not a problem .if they don't come to class, they wouldn't have an opportunity to evaluate it. For the online version, this becomes an issue. I'd obviously weigh the open-ended comments about lecture (both positive and negative) more heavily for those who attend versus those who didn't.
- It seems to me that students who LOVE a class will bother to fill out an evaluation. Students who hate a class will perhaps take advantage of an opportunity to vent and complain. Those in the middle will likely not fill it out at all. This seems to me as if it will result in skewed results.
- I am really glad that this is being worked on. I would like to see more university resources put toward taking the best aspects of in-class and on-line evaluations (and identifying and addressing the weaknesses of the medium that is decided upon) so that we wind up with a really good product. Just putting it on-line without proper care is not the answer.
- One of the questions asked about the something like how well the students understood the course objectives but not how well the course met the objectives. This was frustrating for those who wanted to complain about the execution, not the intent.
- I was disappointed in the response rate, but my school does its own online course evaluation so that probably hurt. For most faculty, I don't think the detailed statistical report (beyond frequencies and means) will be used. I would be interested in more comparisons, for example to other graduate level courses or to other courses of comparable enrollment

size. Not clear what to make of a comparison to all courses because graduate classes--and students, and their standards--are so different than undergrad.

- Years ago, I worked many hours to convince our faculty to change to our current form, a form that does not signal to the students what we think is important in teaching but lets them choose their own criteria for their assessment of our teaching. After that, but still years ago, I participated in a Campus committee that examined the question of whether one form is appropriate for all departments, and in some of our meetings we concluded that it was not. From what I have seen, the Campus now seems to be headed in exactly the wrong direction. Each unit should be allowed to opt out of the standard form in favor of their own form. The forms are NOT comparable across units anyway because the population sampled is different and the comparison group of teachers is different. So, there is no reason to mandate the same form for all units. Indeed, there is potential harm in that people looking at the results may think, erroneously, that the responses can be compared across units.

- The only negative I see to online evaluations is the response rate. The ideal situation would be for them to do the evaluation online, but in class (on their laptops, iPads, etc.). This is how evaluations worked at my undergraduate institution, and it worked wonderfully and gave students a sense of freedom in response, since their answers really were anonymous (no one can pretend that students' handwriting isn't recognizable after a whole semester of grading their work). However, I know this would be difficult to implement since students aren't required to have laptops at IU.

Q19: As part of the development of the online course questionnaire, there have been discussions about providing few of the closed-ended questions available to students. Making the data available would provide better information to students than RateMyProfessor, other online information, or informal evaluations.

Table 4: How much would you favor or oppose making the summaries of your online evaluations available to all students?

Answer	Response	%
Strongly oppose	21	27%
Somewhat oppose	12	15%
Don't care	22	28%
Somewhat favor	8	10%
Strongly favor	15	19%
Total	78	100%

Although 27% are strongly opposed, 58% either do not care or are in favor of making the summaries available to all students. As the question is worded, we do not define which questions students would have access to.

Appendix G: Proposed Campus Core Items (7 quantitative plus 2 open-ended)

Online Directions: Please take a few minutes to respond to the questions below regarding your course. Your responses will remain confidential and only aggregate data will be provided to your instructor.

1. How clearly were course learning goals and objectives communicated to you?
   Very clearly Clearly Somewhat clearly Not at all clearly

2. How effectively was class time used to help you learn?
   Very effectively Effectively Somewhat effectively Not at all effectively Not Applicable

3. How effectively did out-of-class work (assignments, readings, practice, etc.) help you learn?
   Very effectively Effectively Somewhat effectively Not at all effectively Not Applicable

4. How effectively did graded work (papers, exams, presentations, etc.) allow you to demonstrate what you learned?
   Very effectively Effectively Somewhat effectively Not at all effectively Not Applicable

5. How much did this course material challenge you to do your best work?
   Very much Quite a bit Somewhat Not at all

6. How much did the instructor motivate you to do your best work?
   Very much Quite a bit Somewhat Not at all

7. How available was the instructor to provide help when needed (in person, by email, office hours, etc.)?
   Very available Available Somewhat available Not at all available

8. What did you like most about this course?

9. What did you like least about this course?

Appendix H: Proposed Campus Core Items (Meets BFC Circular Requirements)

Online Directions: Please take a few minutes to respond to the questions below regarding your course. Your responses will remain confidential and only aggregate data will be provided to your instructor.

1. How clearly were course learning goals and objectives communicated to you?
   Very clearly Clearly Somewhat clearly Not at all clearly

2. How effectively was class time used to help you learn?
   Very effectively Effectively Somewhat effectively Not at all effectively Not Applicable

3. How effectively did out-of-class work (assignments, readings, practice, etc.) help you learn?
   Very effectively Effectively Somewhat effectively Not at all effectively Not Applicable

4. How available was the instructor to provide help when needed (in person, by email, office hours, etc.)?
   Very available Available Somewhat available Not at all available Not Applicable

5. What did you like most about this course and instructor?

6. What did you like least about this course and instructor?

References

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academies Press.

Center of Inquiry in the Liberal Arts at Wabash College (2011). *Wabash National Study of Liberal Arts Education. High-Impact Practices and Experiences from the Wabash National Study*. Retrieved from http://www.liberalarts.wabash.edu/study-research/

Churchill, G.A. (1979). A paradigm of developing better measures of marketing constructs. *Journal of Marketing Research (JMR), 16* (February), 64-73.

Fink, A., & Kosecoff, J.B. (1998). *How to conduct surveys: A step-by-step guide.* Thousand Oaks, California.

Krueger, R. A. & Casey, M. A. (2009). *Focus groups: A practical guide for applied research* (4th Ed.). Thousand Oaks, CA: Sage.

Mesick, S. (1993). In Robert L. Linn (Ed.), *Educational Measurement* (3rd edition). New York: American Council on Education/Macmillan Publishing.

Marton, F. & Säljö, R. (1976). On qualitative differences in learning I: Outcome and process. *British Journal of Educational Psychology,* 46, 4–11.

Nunnaly, J. (1978). *Psychometric theory*. New York: McGraw-Hill.

Ouimet, J. A., Bunnage, J. B., Carini, R. M., Kuh, G. D., & Kennedy, J. (2004). Using focus groups to establish the validity and reliability of a college student survey. *Research in Higher Education*, *45*, 233-50.

Ouimet, J. A. & Pike, G. R. (2009). Rising to the challenge: Developing a survey of workplace skills, civic engagement, and global awareness. In V. M. H. Borden and G. R. Pike (eds). *Assessing and accounting for student learning: Beyond the Spellings Commission*. New Directions for Institutional Research, No. 139. San Francisco: Jossey Bass.

Tagg, J. (2003). *The learning paradigm college.* Boston: Anker.