

# A Deliberative Poll<sup>®</sup> on Faculty Course Evaluations and the methods of delivering them



Coro Center for Civic Leadership, Pittsburgh 

**Carnegie Mellon**  
**UNIVERSITY LIBRARIES**

Southwestern Pennsylvania  
Program for Deliberative Democracy

Carnegie Mellon began the Campus Conversation in 2005 under the assumption that “when we think together we often think better.” This document initiates the process of deliberative involvement regarding the nature of Faculty Course Evaluations and the methods of delivering them.

## A Deliberative Poll® on Faculty Course Evaluations

Page	Contents
3	<b>Deliberative Polls are not Town Hall Meetings</b>
4	<b>What Carnegie Mellon Would Like to Accomplish Using Deliberative Polling</b>
5	<b>The FCEs: How they came to be and what they are meant to do at Carnegie Mellon</b>
7	<b>Many Voices: Working Toward Dissonant Understanding</b> The FCEs are a useful/problematic instrument for evaluation because... The FCEs help/prevent instructors improve their courses because... Instructors read them/pay no attention because... Students complete/fail to complete the FCEs because... Methods of delivery affect the outcome because...
13	<b>Summary</b>
14	<b>Many Voices: Using Dissonant Understanding to Improve the FCEs</b> Delivery: Voice 1: “Use an online delivery method.” Delivery: Voice 2: “Return to the paper delivery system.” Delivery: Voice 3: “Offer a hybrid delivery system.” Delivery: Voice 4: “Extend the evaluation period.” Instrument: Voice 1: “Stay with the new instrument.” Instrument: Voice 2: “Return to the old instrument.” Instrument: Voice 3: “Use a shortened version of the current instrument.” Instrument: Voice 4: “Recalibrate the scoring scale.”
18	<b>Appendices</b> <ol style="list-style-type: none"><li><b>Past FCEs</b></li><li><b>The 2004 FCE</b></li><li><b>Focus, Winter, 2004: Why the University Should Abolish FCEs</b></li><li><b>Focus, April, 2004: Letter to the Editor</b></li><li><b>The 2006 Faculty Senate Survey</b></li></ol>

Town hall meetings are one of the most common ways for individuals to discuss important community issues. Even so, attendance at these invited meetings is often low either because the format is less than optimal, or the session is dominated by “squeaky wheel” or “know nothing” opinions that do not reflect the majority. Given either of those assumptions, why bother?

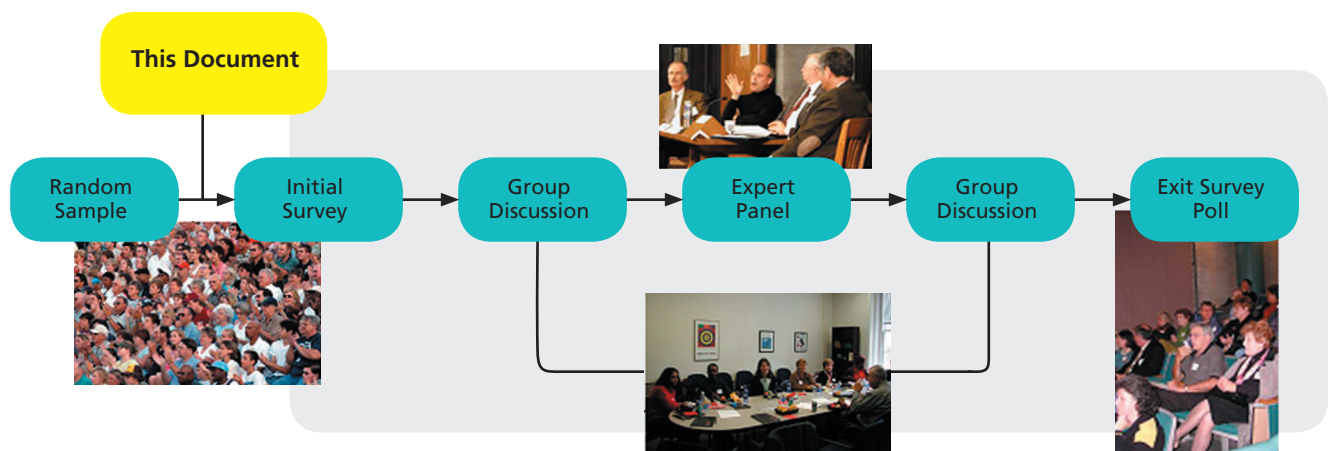
## Why Deliberative Polls are not Town Hall Meetings

Deliberative polls offer new opportunities that improve upon the town hall model. In a deliberative poll, a random sample from the community is invited specifically to increase the range and variation of opinion concerning an issue while that issue is still undecided. The success of the poll, and its potential impact on policy, depends on a substantial turnout of participants because they provide the data that policy makers will incorporate while developing plans and carrying out actions. Significant turnout results in findings that have more consulting power.

Deliberative polling is a process in which participants:

- receive balanced information about an issue,
- fill out a survey that identifies initial opinions,
- gather in small randomized groups,
- develop questions /voice opinions, within those groups,
- discuss their questions and opinions with experts,
- develop an informed understanding of the issue,
- fill out a second survey based on that informed understanding, which can provide policy makers with consulting power as they develop and finalize their recommendations.

You can see an illustration of this process below. Notice that the small group discussions sandwich the question and answer session with the expert panel.



# What Carnegie Mellon Would Like to Accomplish Using Deliberative Polling

This method of discussion was developed and trade-marked by Professor James Fishkin, now at Stanford University's Center for Deliberative Democracy. Carnegie Mellon uses Professor Fishkin's method to realize the following goals:

- Improve the quality of campus-wide discussions,
- Encourage the campus community to question, expand, and articulate their perspectives on issues of concern,
- Encourage a more diverse sample so that individual voices are not drowned out by vocal minorities,
- Enrich the campus community by valuing diversity and encouraging further collaboration across groups,
- Encourage the community to see the usefulness of deliberative polling.

This deliberative poll will consider major opinions of Carnegie Mellon's most recent version of the Faculty Course Evaluations (FCEs). **Through your participation in this deliberative process you can have a voice in future practice.** On the next page, you will find background information that describes the development of the FCEs and opinions about their usefulness. Later on, you will have the opportunity to consider major opinions concerning Carnegie Mellon's most recent version.

**In this deliberative poll, you have the opportunity to influence policy concerning the FCEs as an instrument and method of delivery at Carnegie Mellon.**

## As a Participant:

The results of this process will be best realized when you, the participant, agree to:

- Explain your own perspective,
- Listen to other views without interruption,
- Focus on relevant facts within reasoned arguments,
- Be open to challenges based on competing relevant facts and,
- Treat group members, who do not share your opinion, with respect.



Work on the instrument and its delivery continued over the next two years under the guidance of professionals in education, online information design, and survey development. Their goals included the development of:

1. A standard instrument with the potential for customization.
2. An instrument that could be used to both improve and evaluate teaching.
3. A system that would not change current policies.
4. A method of timely feedback for faculty.
5. A plan for participant involvement during the development and testing phases.

They developed the instrument and system of delivery concurrently by:

1. Studying the practice in other universities
2. Initiating a student protocol and interview study concerning the instrument and its delivery
3. Piloting a revised instrument, carried out in 23 courses with 635 students
4. Creating an outreach program for Faculty Senate, University Education Council, the department heads, and former members of the promotion committee
5. Conducting a second pilot in the spring of 2003
6. Analyzing the online FCE by validating the instrument, modeling the ratings, and analyzing the student rate of response

The developers were careful to incorporate feedback from faculty and students concerning the value of the questions, their wording, and the appropriateness and comprehensibility of the scales used. In fact, follow-up interviews revealed that the new instrument had more useful formative and summative features (2006).

However, potential challenges were also identified. While online delivery produced a more rapid turnaround for faculty (a concern that was central in the early planning stages) ensuring that student feedback could be quickly incorporated into subsequent classes, there was some concern that an online survey might reduce participation. Participants in the 2003 study believed that the new FCE would only enjoy 50% compliance. However, because that prediction was close to average in-class participation, the suggestion did not overly concern the developers who had observed participation rates in the ten years prior to online deployment.

The other potential problem is that the instrument may produce lower scores than those of the 1983 instrument for the same instructor teaching the same class. Ambrose

and Smith (2006), note a pilot study conducted on Carnegie Mellon's campus that compared the two instruments. Students were asked to fill out both evaluations. The newer one produced ratings that were significantly lower in statistical terms. And the different ratings were not merely a by-product of the order in which respondents completed the two forms.

***Instructor ratings generally decrease with the new instrument because response options contain more specific wording.***

In particular the overall course and instructor questions are now preceded by a larger number and range of specific questions and the response options for those questions are now fully labeled. As an added measure, some students were asked to complete the new survey on paper, while others were asked to complete it online. The mode of delivery (paper and pencil vs. online) did not have an effect on scores in that study. Two years later, these two concerns have not been resolved.

In November 2003, a four-year pilot was proposed because research suggested that the switch from paper to online delivery is not always an easy transition. In fact, Institutions such as Northwestern University and Georgia Institute of Technology benefited from a four-year cycle that improved response rates. But the Faculty Senate agreed to a two-year campus wide pilot period, at which time they would re-evaluate the progress of the program.

During this two-year period, the system would be deployed along with the development of:

- Improved system functions to be integrated with the SIS (Student Information System) database,
- Continued interface design and testing,
- Continued development of the FCE website.

*Now that the two year pilot program has ended, we find ourselves considering the pilot FCE, comparing that instrument with the 1983 survey, and deciding what course of action might best meet the needs of the institution as a whole. This moment has justifiably brought out many competing voices, which this deliberative poll hopes to move from noisy cacophony to informed opinion.*

## Many Voices: Working Toward Dissonant Understanding

A deliberative poll does not necessarily result in participants who agree on an issue, but it is likely to result in participants who can respect each other's perspectives. Further, it is likely to result in an informed opinion that tends toward one solution over another. That informed opinion produces consulting power – in other words, influence on the development of the final plan. A deliberative poll moves the early opinions of many voices from noisy cacophony to the sound of dissonant understanding. In a deliberative poll, not all voices will harmonize. But, like some modern symphonies, the dissonant tones will still produce a compelling sound as individual opinions become more richly informed by the presence of respectful interaction.

Where the FCEs are concerned, prominent voices emerge following contradictory perspectives:

- The FCEs are a useful/problematic instrument for evaluation because...
- The FCEs helps instructors to improve/prevents instructors from improving their courses because...
- Instructors read them/pay no attention to them because...
- Students complete/fail to complete the FCEs because...
- Methods of delivery affect/do not affect the outcome because...

Perhaps these voices reflect the particular group's use of the FCEs. **The chart below notes those functions.**

### **The FCEs are a useful/problematic instrument when used to evaluate instructors and courses because...**

Clark Glymour, Alumni University Professor of Philosophy (2004) has publicly challenged the validity of student evaluations. He bases part of his perspective on his belief that students use flawed criteria to make their assessments. For example, Glymour believes that students are more likely to downgrade a professor for teaching outside of expected formats, which undermines the development of innovative forms of teaching. Along these lines, a professor of English from an undisclosed university, writing under the pseudonym Louise Churchill in the *Chronicle of Higher Education* (2006), states that in her quest to achieve tenure, she decided to ease her grades as one part of a plan to raise her FCEs, a plan that led her to call herself "Professor Goodgrade" although research shows only a weak relationship between ratings and grades.

While these are compelling perspectives, another emerges in a survey developed by the Faculty Senate and Carnegie Mellon Professor James Hoburg (2006), which communicates a more complicated narrative. This survey enjoyed a 35.4% response rate (604 of 1705 faculty solicited). Of that group, over 80% agreed that the university should administer FCEs.

### **Students**

Gain information about a course/instructor.  
Reflect on learning outcomes and student behavior.

### **Faculty**

Gain formative feedback about course process/components.  
Receive summative evaluation.

### **Administration**

Receive summative evaluations that affect raises, promotion, & tenure.

Further, approximately half of the respondents found student feedback, concerning course content, either moderately or very valuable. About 60% found student feedback about teaching practices either moderately or very valuable. That good news stood in contrast to less encouraging findings. Less than 40% found student feedback concerning what students learned in the course either moderately or very valuable. Does this indicate that feedback is not valuable because it is not valid as Glymour would contend? Additionally, like Professor Goodgrade, almost 68% of assistant professors felt either a great deal or a moderate amount of pressure to be lenient in their grading policies. But we have no data that they actually changed their grading criteria.

Glymour argues that pre- and post-tests of covered material would yield a better method of course evaluation than the evaluative questions on the FCEs. Others say that learning gains can occur independently of the instructor when students use academic development services, tutors, etc.

While the university could initiate pre- and post-tests, it might also consider that the evaluative aspect of the FCE is fully integrated into the system. Therefore, aside from individual department initiatives, there would be little chance of changing current practice in the near future. We might even defend the need for students to have an institutional voice concerning evaluative measures. Still, which is the better instrument?

The experts who developed the 2004 pilot argue that the 1983 instrument contains no questions that invite student reflection. **As you can see in the 1983 FCE questions shown below**, only nine questions concerning instructor behaviors precede the overall course and instructor rating which were never shown to be valid.

1. Did the instructor state clearly the goals of the course?
2. Was the course well-planned?
3. Did the instructor state clearly the criteria for grading?
4. Did the instructor provide adequate feedback concerning performance?
5. Did the instructor respond to questions adequately?
6. Did the instructor show concern for the needs of individual students?
7. Did the instructor communicate clearly and effectively?
8. Did the instructor present material and activities at an appropriate level of difficulty?
9. Did the instructor stimulate critical thinking about the subject?

This lack of context shortens completion time, but also produces a student evaluation that might be more biased toward personality and style with less emphasis on learning outcomes. In the revised 2004 instrument, additional questions provide the opportunity for students to reflect on substantive issues. *The Status Report on the On-line Faculty Course Evaluations* (2006) notes that biases were removed by priming the students with specific questions before requesting overall assessments. In this way, the evaluative purpose of the instrument could more closely reflect learning issues while downplaying the instructor's personality.

Certain unique situations exist, such as Engineering and Public Policy's project courses. In these courses, evaluations invited immediately after course completion are strikingly different from the evaluations solicited five years later. While current students routinely give project courses low evaluation scores, alumni continually give them high scores because the required work tends to mimic the workplace, providing real world experience that students could not have anticipated. However, perspectives on instructors tend to remain unchanged.

Given this mixed data, what about Glymour's call for a student developed and run FCE that would meet information needs without impacting tenure and promotion? That possibility signals two potential problems. First, will students fill out these forms if their impact becomes negligible? If response rates are moderate now, will they become more so if students lose their institutional voice? Further, will students take a student run instrument seriously? *Rate Your Professor* is one example of a website devoted to providing feedback to students contemplating a course with a particular instructor. Perhaps because the stakes are small, this site includes a "hotness" rating — emphasizing the personality and style concerns that the FCE pilot works hard to downplay. Certainly, the APO could administer a useful survey. Still, we must consider the impact on student/faculty relationships if students lose an important connection to university policy.

**This information leaves you, the participant, with important questions to address.** Should the university maintain the overall course and instructor questions in the FCEs, believing that students have a right to this aspect of their institutional voice? Or does the time consuming process of priming student thinking, which seems to negatively impact response rates, suggest a reexamination?

## The FCEs help/hinder instructors in improving their courses because...

The FCEs have historically been designed to be both summative and formative. In other words, they not only evaluate teaching (summative), discussed in the earlier section, they also help faculty *improve* their teaching (formative). Given that goal, the question becomes, do the FCEs help or hinder faculty as they attempt to improve their courses. Again, the answer to that question is not an easy yes or no.

Questions that would help faculty improve their teaching were developed for the 2004 pilot so that those questions would be neither too “low” nor too “high inference.” For example, a low inference question might be “the instructor came to class on time,” which while asking that a student infer little, provides little data concerning how well that instructor taught course content. An example of a high inference question might be, “overall how would you rate the quality of the instructor?” This is a high inference question because the student has to infer what qualities and characteristics to use in their evaluation and how to rate them. Instead, the questions developed for the 2004 FCE lie somewhere in the middle, allowing students to communicate opinions of teacher performance that they are qualified to answer.

But, how can we be sure that student opinion mirrors expert opinion? Would the university be wise to substitute faculty observation for student feedback and comments? This practice would take faculty away from other responsibilities, ultimately making the program more expensive not only because of the additional time needed to complete this work, but also because the university might need more individuals to do this work. But, would it be more reliable? Not necessarily, because data from the journal *New Directions for Institutional Research* indicates that there is a high correlation between student observation and peer (faculty) observation. Further, student observation also correlates highly with Teaching Center observation and alumni observation (except for the project courses mentioned earlier). Therefore, it would seem that students are not only qualified to provide formative data, they are a cost-efficient method of gathering that data.

Still, perception and fact often produce different effects. Half of the assistant professors that took part in the 2006 Faculty Senate survey, professors who have not yet achieved tenure, felt that the FCEs created pressure to simplify or slow down their courses. Others say that those courses might have been too fast paced to begin with. As those assistants gain experience and become associate and full professors, the pressure seems to lessen. Forty percent of the professors with more experience indicated that they felt a great deal or some amount of pressure to slow down their courses. Is even that figure one with which we are comfortable?

Should students provide formative comments or does faculty response negate the value added in this effective and efficient communication? In other words, we might all agree that you don’t have to be a shoemaker to know if the shoe fits. But what do you do if the shoemaker’s work suffers under this form of observation? Well, you might say that shoemakers still need feedback in order to make shoes fit better. Even so, getting the right feedback using the present tools could be problematic. For example, concerns have been voiced over the inflexibility of the current online delivery system. An instructor cannot drop a subset of questions from the form when those questions are not relevant for a particular course. Still one must remember that, given the budget, the system as it stands now is the result of the university’s best efforts. In order to make the system more flexible, the university would have to direct substantial funds to that project.

Some ask why we did not purchase software that would have suited the university’s purposes? The developers note that in 2003, when the new system was in development, they could not find a single online evaluation system that recapitulated the features of the paper system (which was a fundamental requirement). For example, no system existed in which department heads and deans could receive information they found most necessary — a simple but important requirement. Such software exists today (essentially because one company took Carnegie Mellon’s requirements and built it with a staff of several programmers over the last three years). Using that software now would mean additional costs. However, opportunities exist to modify the present system at reasonable cost — once the requirements are known.

Finally, does the 2004 pilot study provide better formative or summative data than the 1983 survey? **If you look again at the 1983 instrument**, performance ratings are noted on a Likert Scale that has vague interpretations of the numbers. “Poor” to “excellent” leaves much room for interpretation.

Course
Instructor

Rating Scale
1 - Poor
2 - Fair
3 - Average
4 - Good
5 - Excellent

### EVALUATION QUESTIONS

1. Did the instructor state clearly the goals of the course?
2. Was the course well planned?
3. Did the instructor state clearly the criteria for grading?
4. Did the instructor provide adequate feedback concerning performance?

## **– In this course, to what degree have you...**

**KEY:** [0 - Not Relevant to Course] [1 - None/Not at all] [2 - A Little] [3 - Some] [4 - A Lot] [5 - An Exceptional Amount]

gained factual knowledge (terminology, classifications, methods, or trends).

learned fundamental principles, generalizations, or theories.

learned to apply course material (e.g., to improve thinking skills, problem solving, or decision-making).

What the points of the scale mean within the continuum is up to the student. Researchers performing a protocol study (Fall 2001) found that student participants interpreted the points on that scale in ways that were not uniform or consistent, leading to problems with the reliability of the instrument. One student's 3 was another's 4.

The 2004 pilot addresses this problem by providing a uniform interpretation of points along the scale. **On one particular series of questions shown above, you can see the labels on that continuum.**

Some believe that with this revised instrument, the university finally has a way to identify excellent performance. These revisions have removed the ceiling effect of the old instrument, whereby scores were so high that they did not allow for the differentiation between different levels of expertise. Further, analysis shows that adequate variability now exists in the ratings so that you can distinguish between "excellent" and "poor." For example, by looking at the ratings within a department or college (in relation to one standard deviation from the mean), one finds ratings that deviate that much from the mean, which supports claims that this new instrument does identify excellence (or its lack).

However, increased precision has not come without costs concerning faculty confidence in the new instrument, as can be seen in the next section. Additionally, others feel that the performance descriptions on the 1981 survey (poor, fair, average, good, and excellent) were more useful because those terms are more commonly used. They also wonder if certain questions, such as the number of hours spent working on a class, can ever be accurately addressed. For example, will the answer to that question include time spent over spring break, which could change the average time per week by as much as ten percent?

## **Instructors read/do not read them because...**

While students often believe that their instructors pay no attention to the information provided by the FCE, faculty are acutely aware of FCEs, even if they are not always convinced that the results are valid. Some additional information might help to clear up this misunderstanding. FCE scores are used in tenure and promotion as Glymour points out in an article that criticizes the FCEs for that very reason. Achieving tenure, the process by which a faculty member attains a permanent position in the university, means that the faculty member will show continued promise and progress in publications, teaching, and service.

A faculty member must produce publications, which include writing for peer-reviewed journals, submitting work to certain highly respected conferences, or producing monographs of substantial theoretical work distributed by respected publishers. Service includes developing or running university programs, participating in departmental job searches, volunteering for the Faculty Senate, or other community outreach programs. Teaching is judged by FCEs, syllabi, and sample course materials, letters and emails from students, and publications that have an educational focus.

## ***Scores that are particularly high or particularly low prompt action by the administration.***

Throughout the process of tenure and promotion, faculty members are mentored to improve their chances of success. The FCEs are an important part of that mix.

However, as Glymour points out, paying attention to the results does not always mean agreeing with them. Faculty have voiced a concern with scores because of two issues mentioned earlier. First, average FCE scores tend to be lower than they used to be, some would argue, courtesy of the new instrument. Second, while the new delivery system has resulted in higher response rates for some courses, other courses have response rates that do not reach the rate necessary to produce statistical confidence (30%). Again you, the participant, must consider.

- If the FCE instrument results in lower scores, even if those scores are more accurate, and if faculty rely on those scores to help them with cases for tenure and promotion, what should a reasonable next step include?

- The faculty survey mentioned earlier suggests that our faculty feels strongly about the importance of the FCE. They also believe that the FCE gives them valuable feedback concerning their teaching practices. However, if student participation is too low to produce confidence in the ratings, how can the administration use those scores as a decision making tool?

- Finally, how should the university communicate to students that instructors not only read the FCEs, but also often value and incorporate aspects of the data into their teaching practice? Would that information invite increased response rates?

### **Students complete/fail to complete the FCEs because...**

When the FCEs moved from paper delivery to online delivery, the common complaint went something like this. "Only the students who hate my class, or really love my class will fill these out, which is why I'll get scores I don't trust." In fact, when participation is below 30% that could be the case. Therefore, policy makers must consider ways to increase participation, discussed later. However, there is another compelling reason that Carnegie Mellon students either complete or fail to complete the FCEs. Susan Ambrose and Joel Smith (2006) note that:

***"Only the students who hate my class, or love my class will fill these out..."***

Over a third of the students (37%) completed evaluations for all their courses, regardless of how many course they were enrolled in and over a third of the students (36%) completed none. The remaining 27% completed FCEs for 10% to 90% of their courses. We are taking this as evidence that most students fall into two groups: engaged and nonengaged. This is contrary to the perception that students fill in evaluations only for the courses they are really upset or really happy about" (2006, p. 23).

Some students believe that it is their duty to speak to the faculty and administration through the FCEs. Other students do not. When students are engaged, how they feel about a particular instructor does not seem to affect their decision to complete the forms. Similarly, when students do not fill out any of the FCEs, their completion rate is the same for teachers that they love or hate.

If the love/hate factor is not the real issue, response rates are. As a participant, ask yourself this question:

- If a full 63% of students do not complete their FCEs, thereby robbing themselves of the power of their institutional voice along with the real service that voice provides to faculty, how might we increase that percentage?

That question also leads to a final concern: the online delivery method that invites students to complete their FCEs on their own time.

### **Methods of delivery affect/do not affect the outcome because...**

When the method of delivery changed from paper, filled out during class, to an online delivery, filled out at a time that the student selects, certain concerns were anticipated and others were not. One major reason to go online versus continuing the paper forms concerned timeliness. Having students fill out the forms on paper, in an online age, meant a considerable lag time between the completion of the form and the opportunity to view the results. Faculty using the paper form were unable to incorporate the requested changes until the early part of the next semester. This lag time made it difficult to initiate meaningful change.

The developers anticipated that online participation might decrease, given that the forms would no longer be completed in class. They also recognized that students who rarely attended class could now have equal access to the FCE forms. While the first problem is a real concern, the second is less so. Those non-attending students do take the exams, and they do complete the readings. They might not perceive an added value from attending the class. Therefore, they are qualified to comment on the class. Further, paper delivery can only sample the students attending that particular class, which can be as low as 30% of the class — surely not a random sample.

What the developers did not anticipate concerned the students' interest in filling out all of their forms in a single sitting — a considerable challenge since the piloted FCE contains more questions than the earlier form. Perhaps for that reason, developers also noted the students' interest in having an extended period to fill out the FCEs during finals week. Students state that they are often not as busy during that time as they are in the final two weeks of class.

Participation must be considered from a college as well as a university level. As stated earlier, this new method of delivery has increased participation in some colleges such as Computer Science, while it has led to a decrease in participation in other colleges, such as H&SS and CFA. Further, some schools in the university, such as the Heinz School, use their own system. They found online delivery problematic in terms of response rates and soon reverted to paper.

This phenomenon raises the concern that the university as a whole might benefit if colleges or departments could choose one form of delivery over another.

However, a hybrid delivery model introduces a serious problem for cross-listed courses. At Carnegie Mellon, a cross-listed course has the same title, units, meeting time, building, room and instructor, but different course numbers, and is offered by different departments within the main university, or the main university and a branch campus, or two different universities. Carnegie Mellon offers approximately 180-230 cross-listed classes. If one department or university chooses paper delivery, while another opts for online, anonymity could be compromised. For example, let's say the School of Design and the department of English offer a course. Let's further assume that English uses paper delivery while Design uses online delivery. If a class with 15 enrolled students has only two students enrolled in the course under the English number, and only one of those students attends class the day that the instructor administers the FCEs, that student's anonymity is severely compromised. Additionally, increasing the number of tasks that a department must perform to take part in FCE evaluations only increases the possibility of errors. Finally, as our branch campuses grow, the need to streamline processes such as the FCEs might grow with it.

Furthermore, students may become confused as to how the FCE is being delivered for a given course. If a course is cross-listed across colleges, some students registered under one course number, might have to complete FCEs using a delivery mode that is different from students registered under another course number. Both instructor and student will have to keep this issue in mind. Long term consequences of a hybrid model are also unclear. For example, if H&SS chose paper delivery, but other colleges chose online, might students whose home college is H&SS feel less compelled to fill out the online forms?

As CMU's branch campuses grow, the need to streamline processes such as the FCEs grow with it. Increasing the number of tasks that colleges with global connections must perform only increases the possibility of problems. Even instructors could have problems with a hybrid model. While the initial pilot study (2003) suggested that the method of delivery *would not* affect instructor scores on this campus, the results of that one study, completed in a more controlled environment, might not have produced data on which tenure track faculty want to bank their careers. While some studies outside of academia suggest that delivery does not affect outcome, others suggest that it does. Perhaps most importantly in these days of rising tuitions, the hybrid model would require substantial changes to the system, which means substantial expense.

The practicalities bring dilemmas for individuals as well as groups. For example, one individual might believe that

FCEs should be delivered on paper because that is the most likely way to ensure 100% participation. That individual also wants to see the results immediately after grades are sent out to students. However, often, the instructor who uses online delivery in order to receive feedback quickly will not realize full participation.

Would it be more practical to use scan sheets? According to Enrollment Services, scan sheets are problematic, as the administrators of the SAT found out recently. When scan sheets become damp, they are unreadable, and have at least a 10% error rate even under optimal conditions. Scan sheet technology doesn't always offer a comments section, and when it does, the OCR scanners that must interpret student writing provide hit or miss translations. In this case, additional costs might deliver unacceptable results.

Given student and faculty feedback, other possible solutions emerge. Enrollment Services finds that reminders spike response rates – even though students claim that reminders are annoying. Asking students to fill out the FCEs as a homework assignment, listed in the syllabus, might also act as an effective reminder. If FCE information became part of Freshman orientation, students might also begin to value the process.

Others have suggested that cluster time be scheduled for each class to fill out the FCEs during class time. This solution poses problems because only 18 clusters ranging from 20 to 45 seats are available while 77 classrooms ranging from 278 to 12 seats plus a few non-registrar controlled spaces that are filled daily produce something close to a 4:1 ratio between cluster and classroom. Additionally, cluster space is used during the last 2 weeks for classes when FCE evaluations occur. So those clusters would not really be available for FCE input.

Therefore, while it would be next to impossible to schedule cluster time so that classes could be scheduled to fill out the FCE during class time, it might be possible for electronic devices to be brought to the students. Using electronic devices, only certain classes would be available for evaluation during particular times. The technology, at this time, might only be able to recode standard responses, and not comments. However, participation would be more stable, and all instructors would use the same instrument and delivery.

**In this area of concern, as in others, you the participant must consider the consequences of these important**

### policy decisions.

- Should the deadline for completion of the FCEs be extended?
- What incentives should be given for students completing their FCEs? Might incentives mean that the CMU community could view only courses with a student response rate of at least 30%? Or should only students who complete all of their FCEs have access to this information? While the first possibility does not potentially affect anonymity, the second might.
- Finally, does the timeliness of feedback trump participation rates, which would mean that the FCE should remain online rather than returning to its earlier paper form?

## Summary

This section summarizes the ideas discussed in the background section.

Those questions are a good place to start, but in a deliberative poll, they are in no way, a good place to finish. Instead, they should inspire other questions and comments, from which we can develop a richer understanding of the FCEs and how they might work best for the Carnegie Mellon community. The next section will ask you to consider not only the method of delivery we should use, but also the instrument that Carnegie Mellon should adopt to best serve students, faculty, and the administration.

- **The FCEs are a useful/problematic instrument for evaluation because...**

Should the university continue to include the overall questions in the FCEs, believing that students have a right to this aspect of their institutional voice? Or does the considerable time it takes to prime thought related to instructor performance suggest that we reexamine the evaluative aspect of the FCEs?

- **The FCEs help/prevent instructors improve their courses because...**

Should students provide formative comments or does the response of some faculty to this feedback negate the value added in this effective and efficient communication?

- **Instructors read them/pay no attention to them because...**

How should the university communicate to students that their instructors not only read the FCEs, but also often value and incorporate aspects of the data into their teaching practice? Would that information help students value their institutional voice on campus thereby increasing response rates?

- **Students complete/fail to complete the FCEs because...**

From a viewpoint of engagement, how a student feels about a particular instructor does not affect his or her decision to complete the forms. Similarly, the non-engaged student's decision not to fill out the FCEs also calls into question the opinion that students complete FCEs for teachers that they love or hate. But if a full 30% of students do not complete their FCEs, thereby robbing themselves of the power of their institutional voice through this venue, along with the real service that voice provides to faculty, how might we increase that percentage?

- **Methods of delivery affect/do not affect the outcome because...**

Would a hybrid delivery in which one department stays with the paper version while another stays online, solve problems, or would it bring about an administrative nightmare? Should students have additional time to complete their FCEs? Would that additional time give students the opportunity to evaluate the course in its entirety – including the final exam (if given)? Finally, how might the administration accommodate student interest in filling out all of their FCEs at one time?

### Many Voices: Using Dissonant Understanding to Improve the FCEs

The following perspectives use the background information presented earlier, to produce five concrete points of view concerning the next step this university should take. Three of those perspectives consider the method of delivery for the FCEs, while two others argue for the instrument itself. These alternatives do not represent every possible argument or option. Instead, they attempt to show a broad range of possible courses of action. In important respects, however, the voices point in different directions. They indicate that choices must be made, and that you, the participant, have the opportunity to take part in that selection process.

Delivery: Voice 1

#### **“Use an online delivery method.”**

Carnegie Mellon should either continue to improve the online delivery system they have worked so hard to develop, or explore other technologies that not only reflect our academic mission, but also improve upon our promise to maintain student privacy and provide faculty with timely feedback.

In support: In order to attain a higher compliance rate, other universities have found that online delivery benefits from a four year introduction. Carnegie Mellon’s pilot program has only been available for two years.

In opposition: While we can’t know with any certainty, the length of this instrument coupled with online delivery might negatively affect compliance rates. If we agree that this instrument is superior to the earlier version, online delivery might not be our most effective method.

Delivery: Voice 2

#### **“Return to the paper delivery system.”**

Carnegie Mellon should return to a tried and true delivery system that has served all of its departments well for the last 45 years. This system not only seems to provide better response rates across most colleges in the university, it also is a class time activity that does not ask students to spend time outside of class on duties that do not directly benefit students during a part of the semester when they most need to conserve their time wisely.

In support: In-class participation might alleviate the problem of response rates for this new, more time-consuming instrument.

In opposition: Many responses cannot be counted when paper forms are not readable. These forms are not included in the response rates. Further, faculty members must question the data if they believe there is some error. No set way exists to know the error rate in the paper process.

Delivery: Voice 3

### **“Offer a hybrid delivery system.”**

Carnegie Mellon should offer a hybrid delivery system so that each college could choose the method of delivery that best serves their students. While the costs might increase, the importance of the student’s institutional voice, and its impact of faculty development should not be under-estimated.

In support: Faculty and student satisfaction with the delivery method seems to improve compliance. As we have seen, some colleges enjoy higher response rates with the paper form while other colleges enjoy higher response rates with the online form.

In opposition: At a time when tuition rates are rising so rapidly, the university needs to consider the most efficient method of delivery rather than the most costly. Additionally, cross-listed classes pose problems for anonymity when one department uses online delivery while another uses paper.

Alternate Option

### **“Extend the evaluation period.”**

Carnegie Mellon should extend the evaluation period to finals week when students tend to have more time.

In support: If participation rates are valued, and if the more complex instrument is preferred, the university owes it to both students and faculty to optimize student opportunities to complete the FCEs.

In opposition: When students have completed all of their classes, their memory of instructor performance is fresh. The lag time between the last class and the end of finals week dims those clear memories, and could unduly affect FCE outcomes in overly negative or positive ways.

Instrument: Voice 1

### **“Stay with the current instrument.”**

Carnegie Mellon should stay with the new instrument. Unlike the old instrument, which allowed students to confuse style and personality with teaching effectiveness, the developers of this new instrument have created an arena for reflection that helps students focus on teaching behaviors and their own learning outcomes.

In support: Helping students so that they don’t confuse or conflate style and performance gives those students a more credible voice.

In opposition: The new instrument, which students must often access outside of class, is so time-consuming that many students are not able to complete the form, decreasing their numbers, and with it their credibility.

Instrument: Voice 2

### **“Return to the old instrument.”**

Carnegie Mellon should return to the old instrument. That instrument not only gave students an adequate sense of instructor performance, it did not punish faculty on the road to tenure whose scores have been adversely affected by the new contextualized structure and content.

In support: The old instrument not only adequately addressed performance, but many departments also enjoyed higher participation rates, which allows for more confidence in the scores.

In opposition: The old instrument not only failed to identify excellence in teaching, but some departments also had lower participation rates, which lowers confidence in the scores.

Instrument: Voice 3

### **“Use a shortened version of the current instrument.”**

Carnegie Mellon should edit the current instrument so that filling out the form won't be so time consuming. While the current number of questions might prime the student to evaluate performance rather than personality, that number undermines participation. A shortened instrument could also be a customized instrument.

In support: The current instrument is an improvement over the old instrument. Therefore, it would be a mistake to simply revert to the earlier version. By limiting the number of questions that students must answer using an instrument that keeps its focus on performance, student participation might increase. As regards customization, there are in fact opportunities to somewhat modify the existing system at a reasonable cost - as long as the requirements are clearly defined.

In opposition: If the instrument could be effective using fewer questions, those unnecessary queries would have been removed earlier. Thinking about performance must be primed. The number, order, and type of questions in the current instrument cannot be altered if that priming is to reliably take place. Further, customizing the instrument could ultimately prove too costly. While the project might start out with clearly defined parameters, scope creep could soon overtake those original goals.

Instrument: Voice 4

### **“Recalibrate the scoring scale of the current instrument.”**

Carnegie Mellon should recalibrate the scoring scale to eliminate the drop in FCE scores that professors are likely to encounter even though they are teaching the same course in the same way.

In support: If the instrument did not result in almost automatically lower FCE scores, faculty might not find it so objectionable.

In opposition: The new instrument is the first to offset the ceiling effect and recognize excellence in teaching. Therefore, the scoring scale should remain unchanged.

## References

- Ambrose, S., & Smith, J. (2006). *Status report of the on-line faculty course evaluations*. Pittsburgh: Carnegie Mellon.
- Churchill, L. (2006, February, 24). Professor Goodgrade: Or how I learned to stop worrying and five lots of A's. *The Chronicle of Higher Education*, C1, C4.
- Eberly Center for Teaching Excellence (2004). *Faculty Course Evaluations: Frequently Asked Questions*. Retrieved August 22, 2006, from <a href="http://www.cmu.edu/fce/students/faqs.html">
- Glymour, C. (2004). Why the university should abolish faculty course evaluations. *Focus*, 33, 1-2.
- Hagan, S. (2006). *Interview with members of the Eberly Center, Carnegie Mellon*. Pittsburgh.
- Hagan, S. (2006). *Interview with members of Enrollment Services, Carnegie Mellon*. Pittsburgh.
- Hoberg, J. (2006). *Faculty Senate Results of Survey of CMU faculty on FCEs*. Pittsburgh: Carnegie Mellon Faculty Senate.
- Kulik, J. A. (2001). Student ratings: Validity, utility, and controversy. *New Directions For Institutional Research*, 110, 9-25.
- Theall, M., & Franklin, J. (2001). Looking for bias in all the wrong places: A search for truth or a witch hunt in student ratings of instruction? *New Directions For Institutional Research*, 110, 45-56.

## Credits

Robert Cavalier, Program Director/Principal Investigator  
Liz Style, Program Manager  
Michael Bridges, Survey Specialist  
Susan Hagan, Researcher/Writer  
James Vroom, Librarian  
Cindy Carroll, Editorial Assistant  
Neil Guzy, Coro Fellow in Local Democracy 2006-07

Special thanks to faculty and staff who contributed to this project.