ROBERT E. STAKE is emeritus professor and director of the Center for Instructional Research and Curriculum Evaluation at the College of Education, University of Illinois at Urbana-Champaign.

EDITH J. CISNEROS-COHERNOUR is a doctoral candidate in evaluation and higher education at the University of Illinois at Urbana-Champaign and an associate professor at the Autonomous University of Yucatan.

6

The University of Nebraska at Lincoln's peer review project is described and critiqued as a guide to others planning on using the process.

An Examination of the Implementation of Peer Review of Teaching

Daniel J. Bernstein, Jessica Jonson, Karen Smith

Peer review of teaching will likely play an increasing role in the summative arena as well as its current formative use. This chapter offers a description of a campuswide program to provide faculty with an opportunity to undertake substantive peer review of a course. The model of interaction was derived directly from the American Association for Higher Education (AAHE) project on peer review (Hutchings, 1996b), which has an explicitly formative flavor. We will describe what the faculty participants did and also report some of our observations of the impact of participation on those faculty and their students. Based on these observations, we will offer some suggestions for improving the implementation of peer review of teaching and also describe some likely future directions for peer interactions on teaching.

AAHE Peer Review Model

In 1994 the AAHE Teaching Initiative began a project on peer review of teaching that included representatives from twelve universities. As documented by the project leadership team (Hutchings, 1995, 1996a), faculty from the twelve campuses worked on individual versions of a general model

Note: This project was supported by grants from the Fund for the Improvement of Post Secondary Education (U.S. Department of Education), the University of Nebraska Teaching Council, and the University of Nebraska Pepsi Quasi-endowment Fund. We are especially grateful to Joan Leitzel and Rick Edwards for their support and encouragement and to Bob Brown for consistently excellent counsel. The chapter is dedicated to the memory of Al Kilgore, whose vision remains our beacon.

of peer interaction and consultation about the teaching of a particular course. Faculty working in pairs within academic disciplines exchanged written materials in three categories. The first interaction focused on the intellectual content of a course, including the statement of course goals, a rationale for the inclusion of those goals, and an account of the intellectual decisions that went into the construction of the course. The faculty members exchanged annotated syllabi and narrative accounts of the creation of the courses, offering written comments to each other on what they read.

The second interaction focused on teaching practice in the identified course. Each faculty member identified in writing specific goals for class or other contact time with students, along with a rationale for the planned activities. The peer colleague made several visits to those settings, followed by written comments about how well the planned goals were accomplished. This approach is explicitly not normative; the feedback from a peer is about the accomplishment of one's own goals, not a comparison with a hypothetical ideal class or teacher. As with the first interaction, the written comments are exchanged privately between the peers for their own use.

The third interaction focused on student learning, that is, what kind of understanding students reached about the course material and how many students achieved identified levels of understanding. Faculty members exchanged copies of examinations, written assignments, and project descriptions that were used to assess the level of student achievement in the class, along with examples of actual student performance. Whenever possible, the faculty members also provided examples of the feedback given to students on their performance—both grades and suggestions or explanations. Peers read the materials and offered comments on the quality of the understanding asked for, the depth of understanding students actually achieved, and the usefulness of the feedback given to learners.

The emphasis on the effects of teaching was the most unusual of the three components of the AAHE peer review model. Faculty are accustomed to discussions of course content in individual conversations with disciplinary peers and in the occasional documents provided to curriculum committees that oversee the nominal content of courses. In general, however, faculty hold themselves responsible for presenting coherent, appropriate, and accessible material in their discipline and students responsible for the amount they learn (as a function of their ability, motivation, prior preparation, and time available). The AAHE model pushed the limits of faculty culture by suggesting that faculty are in some way accountable for the effects of their teaching on the understanding of learners.

University of Nebraska–Lincoln Peer Review Project

The seven faculty from the University of Nebraska–Lincoln (UNL) who participated in the original AAHE project were uniformly positive about the experience. The quality of time talking with colleagues about teaching was very high, and all individuals found benefits to their own teaching from both receiving a colleague's reactions and the process of learning about the peer's teaching. Over a three-year period, thirty additional faculty members engaged in the basic three-step process. Each year's faculty recipients participated in a two-week-long seminar with all the pairs at the end of the academic year to review the project activities and to discuss teaching at UNL in general.

Monitoring the Process and Outcomes

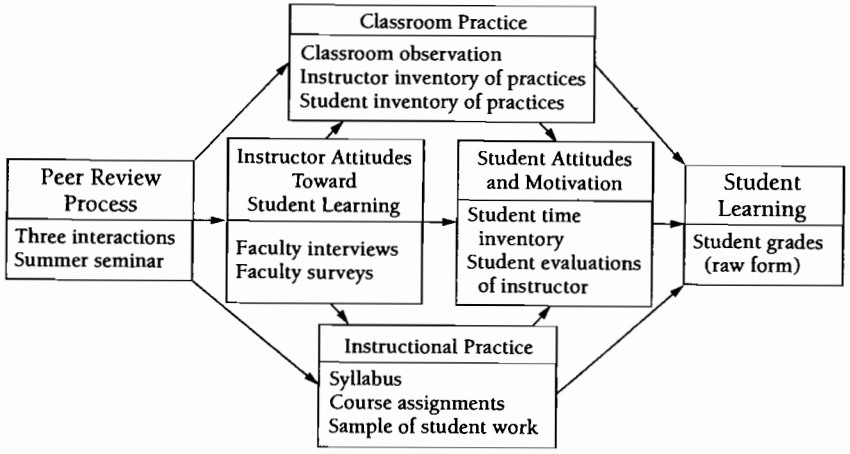
Because administrative support for the project came from the Fund for the Improvement of Post Secondary Education (FIPSE), which always expects a strong evaluation component, many layers of data were collected during the course of the project. To inform the development of measures and to guide the evaluation of the project, a general model of the possible impact of peer review was articulated by the project team. Figure 6.1 shows a schematic diagram of the possible interactions and influences that the team anticipated might occur. For each conceptual component of the process identified, an operational measure was developed that was intended to track that feature as shown in the figure.

At the left side of Figure 6.1 is the peer review process itself, consisting of three written interactions and a two-week summer seminar in teaching. Possible effects of this intervention were tracked in three areas: classroom practice, teacher and student attitudes and motivation, and course materials. These factors should all contribute to student learning, which is the final measure at the right side of Figure 6.1.

The experience of peer review was assumed to influence student learning through an indirect route. There could be direct influence on instructor attitudes, instructional practices, and classroom practices, which could influence student attitudes and motivation directly. Student attitudes and motivation are assumed to be the main direct influence on learning, while instructional and classroom practices may also have a direct effect on learning. The figure lists the measures used for each assumed component of the process.

For each participating fellow, these measures were taken at least once before and at least once after participation in the peer review interactions. Given that faculty teaching assignments change frequently and unexpectedly (even when stability was assured), it was not possible to keep the number of observations exactly equal for all faculty. Our examination of the project is actually based on pre- and postintervention data from twenty-three faculty members who taught significant numbers of undergraduate students and for whom we had sufficient data. The range of data gathered allow examination of how participation in the program influenced several features of teaching performance and student participation, as well as how effective the teaching was in changing student understanding.

Figure 6.1. Peer Review Model

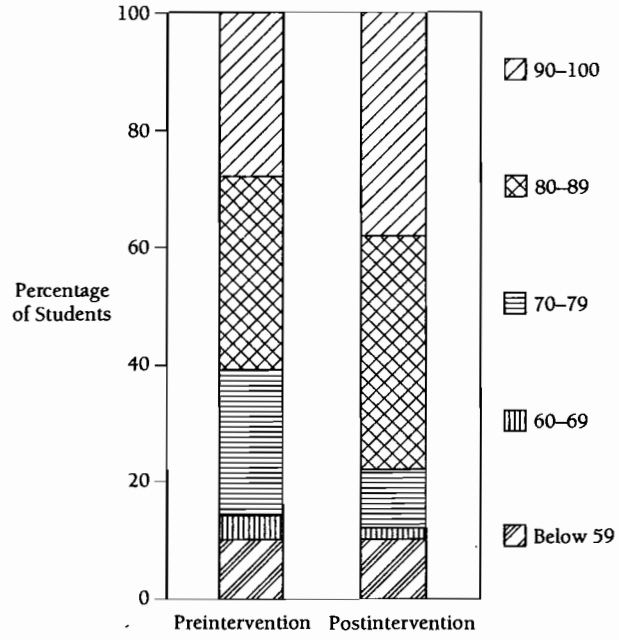


Impact on Student Achievement. When all participants are viewed as a group, there was not a systematic, consistent change in student achievement attributable to peer review. There were faculty whose students demonstrated higher achievement in their courses after participation than before, but the students of over half of the faculty sample showed no increase in learning after peer review. It is worth noting that these were not neophyte teachers who could readily improve their effectiveness every time a class is offered. Despite their level of experience, only one of the twenty-three faculty for whom we had adequate data showed a noticeable decline in student achievement after participation. Perhaps that is less regression in student performance than one might expect merely by chance.

We examined the teaching practices of those faculty whose students showed better understanding after peer review to see if there were any consistent changes in their practices that might be related to the improved work by learners. Two changed features of teaching were associated with improved learning in our sample of faculty: quality of feedback to learners and raising expectations of student performance. In one cluster of six peer review participants, we identified a consistent improvement in student achievement. As shown in Figure 6.2 there were more high-achieving students in these instructors' classes after peer review than before. Figure 6.3 shows data from students' reports of the usefulness of feedback received from the same faculty, also before and after their participation in peer review. The frequency of ratings in the two lowest categories decreased substantially, while the frequency of ratings of the two high-utility categories both increased.

Another cluster of eight faculty was identified that also had consistent improvement in student achievement. Figure 6.4 shows the achievement data for these faculty. The biggest difference is the increase in students who

Figure 6.2. Student Achievement in Course Work, Before and After Faculty Participation in Peer Review Program



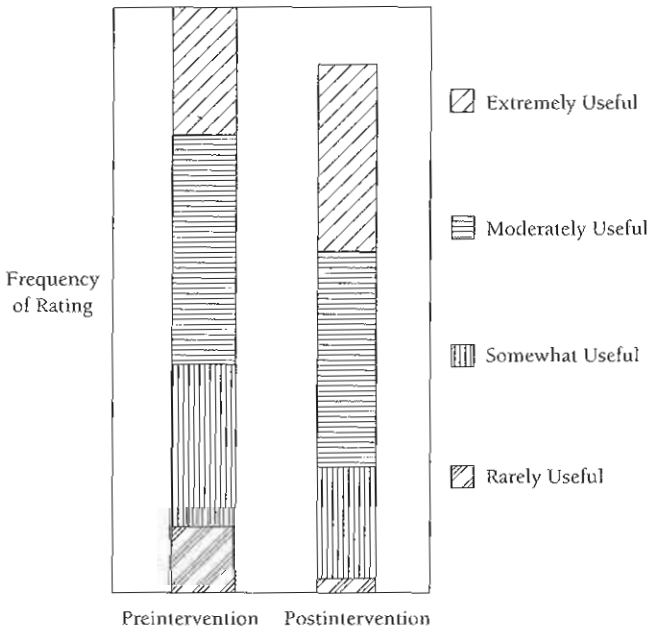
Note: Represents six faculty.

achieved in the 80–89 percent correct range. All three of the lowest categories decreased, and the top category remained about the same. Examination of this cluster of faculty revealed that they had in common a noticeable shift in the level of cognitive understanding they were requiring in their course. Figure 6.5 shows that the amount of rote or comprehension assessment used declined after participation in peer review, with substantial increases in the amount of application, analysis, synthesis, and evaluation required by assignments. These faculty reported in project meetings that peer review had led them to reexamine their course assessments, and if anything, they expected an overall decrease in achievement of these more difficult goals. In fact, their students achieved a higher percentage correct than did students who had taken the courses with the less challenging assessment.

Although these results were not universal, perhaps the observed patterns of selectively changed achievement suggest some features of peer review that could be emphasized in future seminars.

Impact on Student Attitudes and Motivation. In the model of peer review we described, the immediate source of student learning comes from the efforts and perceptions of the learners. Faculty teaching practices would have their influence on learning in part by arranging a context in which

Figure 6.3. Ratings of Usefulness of Feedback for Faculty, Before and After Participation in Peer Review Program



Note: Represents six faculty.

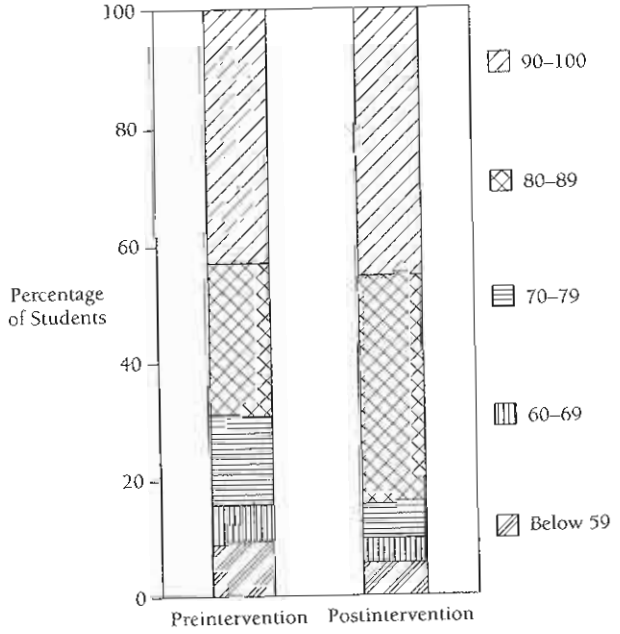
students engage in productive learning activities with greater intensity or focus than previously. Accordingly, we examined student perceptions and activity to see if the collective efforts of the faculty had resulted in changes in what students felt or did.

There was no evidence that changes in faculty performance had consistent effects on student attitudes or motivation. None of the comparisons of student survey responses showed a systematic pattern of change from before peer review began to afterward.

The most important finding from the student surveys came from the items that asked students to report how much time they spent in preparation for various aspects of the course. On a variety of specific questions, students consistently reported spending considerably less time on out-of-class preparation than even the most modest faculty expectation, and only a rare student reported spending the mythical two hours outside class for every hour in class. Like the attitude measures, these reports showed no change across the semester.

Because the project staff gave all faculty participants complete accounts of the data collected in their classes, each year's group of faculty became aware of the apparently low level of student effort, and there emerged among faculty a stream of consistently critical commentary about students.

Figure 6.4. Student Achievement in Course Work, Before and After Faculty Participation in Peer Review Project



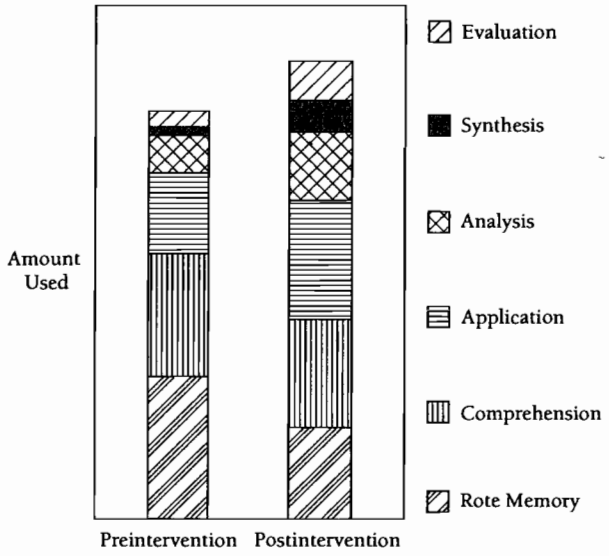
Note: Represents eight faculty.

This backlash from the reported time measure was one of the most striking effects of the process, and the data further mitigated against faculty's assuming greater responsibility for student success. We have tried to make the discussion of these data an occasion for faculty to examine teaching practices that are more and less supportive of consistently high student effort, but that was not always successful. As a result of these data and the subsequent discussions, student time on task has become one of the focal points of our collective peer conversations about teaching practices and learning.

Impact on Faculty Attitudes and Practices. The model of peer review suggests that the student attitudes and actions result in part from the instructor's attitudes toward teaching and learning and in part from the collection of teaching practices arranged by the instructor. Given that we found no systematic change in student attitudes or behavior from peer review of faculty, we asked whether the peer review process had an impact on the faculty themselves.

Attitudes. Overall, the attitude of faculty toward the peer review project itself was extraordinarily positive. All of the original participants completed the entire sequence of activities, and they reported enjoying the seminar far more than they had expected. The opportunity to discuss substantive issues around teaching with colleagues proved to be highly stimulating. Rather

Figure 6.5. Frequency of Assessment in Conceptual Level for Faculty, Before and After Participation in Peer Review Project.



Note: Represents eight faculty.

than being seen as a burden, this process was widely regarded as an opportunity to spend time on development of teaching.

Despite all the goodwill generated by and toward the peer review process, there was little or no change in faculty attitudes toward student learning or faculty responsibility for student learning. In general, the seminar discussions did not alter the basic faculty view that students are responsible for learning and that attitude remains a cornerstone of a typical teaching philosophy. There was also no change in the traditional faculty attitudes about grading; the dominant view remained that grades should primarily reflect relative standing among students so they can be differentiated based on their performance. The notion of criterion-referenced grading that puts student achievement ahead of student ranking was not widely adopted, despite extended discussion of the issue in the seminar.

In-Class Teaching Practices. Several individual faculty members dramatically changed their class practices as a result of their participation in peer review. In all of those cases, the amount of time they spent lecturing (measured by observation and by self-report) decreased substantially, and active participation (often group work) increased dramatically. For most faculty, however, there was no systematic change away from the conventional class in which the instructor does most of the talking. The amount of time needed to generate new class plans is significantly greater than the amount of time needed to refine a set of lecture notes, and it proved exceedingly difficult to motivate faculty to undertake sweeping revision of their time with learners. Faculty who did change their use of class time had mixed results on student learning, with some showing increases in student achievement and others not.

Assignments and Evaluation of Students. Our examination of the collected course materials from the faculty revealed some interesting changes in teaching practice, though not consistently. About half of the faculty required students to do lab projects or term papers in stages, but the prevalence of that practice was unaffected by participation in peer review. In the peer review seminar, the effectiveness of a range of teaching strategies was presented and discussed by each group at length, but faculty reported that in the end, their choices were driven more by limitations on their time than by potential increases in student achievement.

There was a clear increase in the number of faculty who asked students to demonstrate higher-order intellectual skills and critical thinking. Our examination of the level of understanding required by assignments showed that following peer review, about half of the faculty participants added assignments that asked for application, analysis, and evaluation of material, and many faculty decreased the percentage of their assessments based on rote learning or simple comprehension. Faculty clearly do value intellectual depth, and many used their interaction with peers to reexamine their own assessment practices.

There was also a cluster of faculty who increased the quality of the feedback they gave to students on assignments. We examined both the quantity and the quality of the feedback provided on examinations, projects, and papers, and there was a cluster of faculty who increased either the quality or the quantity of what they wrote to students when course work was returned. Although not all faculty showed an increase, none showed a decrease in the amount or quality of feedback. Overall the impact of peer review of quality feedback to students was modest but positive.

Those faculty who chose to increase the cognitive complexity of their assessments and those who improved the quality of the feedback they gave on assignments tended to be the faculty whose students increased their level of achievement. There is no support in these observations to suggest that these features of teaching were the cause of the increased achievement, but they seemed to be the two most useful markers of changed teachers in our sample of peer review participants.

Summary of Impact on Faculty. Overall there were no large, pervasive, or consistent changes in faculty attitudes toward learning or in teaching methods that followed the process of peer review. This process of formative interaction followed by a brief summer seminar did not overwhelm all the other factors that influence faculty choices about teaching. However, about a third of the participants made significant changes in some individual component of their teaching, and they typically attributed their decision to initiate change to the process of peer review. Those were also typically the

faculty whose learners showed improved achievement. It is interesting to note that the virtually universal enthusiasm for the peer review process translated into changed teaching practices for only a minority of participants, and perhaps as a result there was only selective impact on the learners in this project. Programs that report warm reception from faculty as their primary outcome may wish in the future to target changes in practice as an additional critical index of success.

Increasing the Impact of Peer Review

Given the original goals of the project, it was somewhat disappointing to find uneven results on student learning and achievement. All faculty participants embraced the program and engaged each other in meaningful interactions and consultations on teaching. Some faculty took the conversation to the next level and made substantive changes in their courses that resulted in improved levels of student understanding, although most others recognized the potential benefits of developing new teaching practices but chose not to make substantive changes in their teaching.

An understanding of this choice begins with explicit recognition of learner responsibility for education. Faculty believe that students' actions and choices are the major determinant of their success in higher education. Our background reviews of teaching literature confirmed that student time on task, ability, and previous preparation consistently account for more variance in outcomes than do most features of teaching. Faculty typically will feel comfortable in a conversation about learning only if the campus culture makes it clear that students' first priority should be spending time on their studies. Without acknowledgment of that premise, it will be very hard to engage many faculty in extensive efforts to improve student understanding by changing or developing their teaching practices.

Within the smaller portion of the variance in learning that is attributable to teaching practices, however, there is plenty of room for faculty growth, and our challenge is to identify what form of peer review program might increase the percentage of faculty who change their teaching practices in ways that would influence student learning. Programs like supplemental instruction (Martin and Arendale, 1994) show sizable increases in student achievement despite relatively poor preparation, and mastery or competence-based teaching approaches (Kulik, Kulik, and Bangert-Drowns, 1990) greatly improve achievement by increasing learner motivation and time on task. Given that our participants read about and discussed these programs, it was discouraging that virtually no faculty in our sample added these teaching practices to their courses. The changes made were much more modest refinements of existing practices rather than fundamental change in course structure or attitude toward learning.

Our conversations with participants over the four years of the project revealed that faculty have no reason to believe that they will be either

rewarded for good student achievement or held accountable for poor achievement. The most common contingency on teaching seems to be that consistently high ratings by students will result in the best feedback from the institution on the teaching portion of professional work. When global ratings of satisfaction with the course and instructor are higher than local averages, faculty are considered good teachers for purposes of annual review. Even in the materials requested for teaching awards, the majority of attention is given to testimonial evidence, either by students or from faculty who often relay student comments. Comment on the intellectual quality of the work done is not routinely required in such portfolios, and evidence of sustained high achievement by learners is rarely asked for or provided. If anything, in most faculty circles consistently high student performance is taken as evidence of low standards, usually without review of the assessments used or the work generated by learners.

It may simply be the case that peer review of teaching (or any other program intended to improve student understanding and achievement) will succeed only when community standards expect and require evidence of impact on student learning. Department and college leaders need to honor evidence of exceptional achievement in learning in all the ways that faculty recognize (annual review, tenure, promotion, and awards). Once learning outcomes are important to faculty, there is a greater likelihood that they will choose to implement the practices they learned about through peer review.

Implications for Improvement

Inevitably there is some conflict between the formative and summative roles of peer review of teaching (Cavanagh, 1996), since the first step in growing as a teacher is the frank identification of areas that could use development. Including such self-assessment in personnel materials for evaluation is unlikely to be rewarded, so having a summative function of peer review could be seen as diminishing the formative value of the process. We believe it is possible to save the baby and the bathwater in this case by using alternating periods of development and evaluation. If faculty have access to resources and time for formative peer review that is shielded from scrutiny for periods of a few years, they can generate substantial improvement in teaching that can be brought forward during periodic times of accountability. This kind of structure has been used successfully within an academic unit (Bernstein, 1996) to build a firewall between the two classes of peer interactions, allowing faculty enough unscrutinized space to explore their teaching honestly while also providing clear occasions for evaluation of the effectiveness of current practice.

It appears that a key feature of the ongoing peer review process at UNL is the voluntary nature of participation. As a formative exercise and an opportunity to generate evidence for summative purposes, faculty participate only when they have an individual goal to accomplish. In both cases, the process keeps the focus on student understanding. In the formative case,

faculty work to upgrade the level of understanding, in what they ask for and what students can demonstrate; in the summative case, faculty would be given credit for the adoption of teaching practices that have a demonstrated impact on the quality and quantity of student learning.

Future Directions for Peer Review

The uses of peer consultation and interaction about teaching and learning are in principle endless; they could be seen as parallel to the many ways that faculty interact with each other about other forms of intellectual work, such as research in laboratory and field settings. It has been said that to have great poets, one must have great audiences, and the same may be true in teaching. Only when the feedback we get on teaching comes from intellectual peers (not simply from interested novices) will the level of our work grow to its greatest potential.

At UNL we are continuing our work with peer review in two new directions. First, peer review fellowships are being given to departmental teams of faculty who teach courses in the core undergraduate curriculum of a discipline or program. These faculty engage in the usual forms of interaction focused on their individual courses, including the extensive look at student learning. Among the questions asked in that part of the conversation is whether the learning observed is appropriate for the curriculum and the general goals of the major program. This kind of conversation is especially useful in disciplines that have meaningful prerequisite orderings among their courses. All parties have a clear interest in holding specific conversations around the understanding and achievement of students in courses that are gateways to any organized curriculum. As an extension of this idea, we plan to form interdisciplinary faculty teams to examine the same student materials for evidence of performance congruent with the goals of our general education program.

Second, we are developing a community of faculty who are willing to read integrated documents that give evidence about teaching and learning; those readers could in principle provide an independent, arm's-length evaluation of the substance of the evidence provided. This development proceeds slowly but steadily. As a first step, peer review fellows are asked to integrate their three separate written peer review interactions into a coherent and concise document that summarizes the results of their sustained inquiry into their course. These documents are an evolving form of what is being called a course portfolio (Cerbin, 1994, 1996; Hutchings, 1998), and they give an account of teaching practices used to accomplish specific learning goals and what kind of understanding was achieved. The second step involves faculty's voluntarily sharing those documents with local colleagues to get feedback on them as records of their work. Once they are comfortable with their work, we circulate the documents to colleagues in the same discipline at other universities for their comments on the intellectual quality of the material presented and the understanding asked for and demonstrated by students. These independent evaluations will be crucial to the use of course portfolios as evidence of exceptional scholarly work in teaching.

These evolving processes of faculty peer review at UNL will continue to be supported. In the short term, the benefits have been clear, but mostly for the faculty participants. An improved sense of community and the chance to develop teaching in a professional way are desirable outcomes already achieved. It is also likely that the process will be influential in the way teaching is regarded in personnel decisions. Having credible peer-reviewed representations of the intellectual work in teaching could potentially increase its impact in the process. It will be especially important to include evidence of student learning as a central part of the representation of effective teaching. The individual cases of faculty in our project who improved their conceptual goals and student understanding are sufficiently encouraging to merit continued refinement and implementation of peer review. We hope to extend the effectiveness of the process to a larger percentage of our participating peer review fellows. The search for broad impact on student learning will also continue as a longer-term goal. We believe students will derive more benefit from the process when peer-reviewed evidence of student understanding is a major criterion for summative evaluation of teaching.

References

Bernstein, D. J. "A Departmental System for Balancing the Development and Evaluation of College Teaching." *Innovative Higher Education,* 1996, *20,* 241–248.

Cavanagh, R. R. "Formative and Summative Evaluation in the Faculty Peer Review of Teaching." *Innovative Higher Education,* 1996, *20,* 235–240.

Cerbin, W. "The Course Portfolio as a Tool for Continuous Improvement of Teaching and Learning." *Journal on Excellence in College Teaching,* 1994, *5,* 95–105.

Cerbin, W. "Inventing a New Genre: The Course Portfolio at the University of Wisconsin-La Crosse." In P. Hutchings (ed.), *Making Teaching Community Property.* Washington, D.C.: American Association for Higher Education, 1996.

Hutchings, P. (ed.). *From Idea to Prototype: The Peer Review of Teaching.* Washington, D.C.: American Association for Higher Education, 1995.

Hutchings, P. "The Peer Review of Teaching: Progress, Issues, and Prospects." *Innovative Higher Education,* 1996a, *20,* 221–234.

Hutchings, P. (ed.). *Making Teaching Community Property.* Washington, D.C.: American Association for Higher Education, 1996b.

Hutchings, P. (ed.). *The Course Portfolio.* Washington, D.C.: American Association for Higher Education, 1998.

Kulik, C. C., Kulik, J. A., and Bangert-Drowns, R. L. "Effectiveness of Mastery Learning Programs: A Meta-Analysis." *Review of Educational Research,* 1990, *60,* 265–299.

Martin, D. C., and Arendale, D. R. (eds.). Supplemental Instruction: Increasing Achievement and Retention. New Directions in Teaching and Learning, no. 60. San Francisco: Jossey-Bass, 1994.

DANIEL J. BERNSTEIN is professor of psychology at the University of Nebraska—Lincoln and a Carnegie Scholar in the Carnegie Academy for the Scholarship of Teaching and Learning.

JESSICA JONSON is the coordinator of University-wide Assessment of Student Outcomes at the University of Nebraska—Lincoln.

KAREN SMITH is assistant professor of psychology at Truman State University.

7

Colleagues, perhaps through special faculty committees, can play an important part in evaluating teaching portfolios, especially for summative decisions.

Evaluating the Teaching Portfolio: A Role for Colleagues

John A. Centra

Although the rewards for research are often cited as the reason that teaching is undervalued in many colleges and universities, other reasons may well be more critical. The lack of information and evidence about teaching performance is one. Another is having valid means to judge that evidence if and when it is collected.

The teaching portfolio has been promoted as a vehicle for collecting and documenting information about an individual's teaching performance. But the problem of how best to assess the information has not yet been addressed adequately. This chapter describes how one group, a teacher's colleagues, can provide valid assessments of portfolios and similar self-reported information on teaching.

The Portfolio and Teacher Self-Reports

The teaching portfolio has been adapted from such fields as art and architecture, in which professionals display samples of their work for clients or employers. In the mid-1980s a similar product was called a teaching dossier and was defined as a "summary of a professor's major teaching accomplishments and strengths" (Shore and others, 1986). A project to identify the kinds of information a faculty member might include in a teaching dossier was sponsored by the Canadian Association of University Teachers. The project report suggested three major areas containing forty-nine specific items (Shore and others, 1986):