Inside the peer review process: how academics review a colleague’s teaching portfolio

Kathleen M. Quinlan*

Office of Educational Development, S 2 0 1 4 Schurman Hall, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853-6401, USA

Received 18 May 2001; received in revised form 24 September 2001; accepted 4 December 2001

Abstract

Understanding how academics review and make judgments about teaching portfolios is important as the peer review of teaching becomes more common. This think-aloud study examines the processes seven academics used in reasoning through a colleague’s biochemistry course portfolio. Analysis of the interview transcripts revealed that participants used a normative, case-based reasoning approach, comparing the reviewee’s practices to their own experiences, their colleagues, and to prototypical or traditional practices. They considered contextual factors and their pre-existing knowledge of the teacher and the context. The readers justified their final decisions based on the appropriateness and achievement of educational goals. The teacher’s reflective commentary, the student evaluations of teaching, and the syllabus were important in their review. Implications are discussed.

© 2002 Elsevier Science Ltd. All rights reserved.

Keywords: Faculty evaluation; Evaluation methods; Cognitive processes; Problem solving; Scholarly communication; Protocol analysis; Course evaluation

1. Introduction

Much attention has been focused recently on the peer review of university teaching and learning (Cosser, 1998; Hutchings, 1996; Palmer, 1995; Ramsden, Margaton, Martin, & Clarke, 1995; Trow & Clark, 1994; Valimaa, 1994; Seldin, 1999). The peer review of teaching movement builds on the metaphor of peer review of research, viewing teaching as another form of scholarship requiring substantive intellectual reasoning and experience (Boyer, 1990). Much of the literature on peer review describes sources and methods of documenting one’s teaching accomplishments through teaching portfolios (Edgerton, Hutchings, & Quinlan, 1991; Seldin, 1991) or course portfolios (Cerbin, 1994, 1996; Hutchings, 1998).

However, while peer review of teaching is being advocated and methods are being tested, there has been little systematic attention to the ways in which documentation of teaching is read or reviewed by academics (Centra, 2000). Fully integrating teaching into the reward structure through the use of teaching portfolios—particularly in summative contexts—will require a better appreciation of how academics read and evaluate portfolios.

This study draws on theoretical and methodological traditions in cognitive psychological studies...
of problem solving to look more closely at the reading and thinking process used by academics when they are reviewing documentation of a colleague’s teaching. This investigation is modeled on studies of how academics do their work and problem-solve in their disciplinary context. For example, in medical education, there have been two decades of expert-novice studies of the reasoning processes of physicians and students, pioneered by Elstein, Shulman, and Sprafka (1978). Similar methods have been used to study the reasoning and reading processes used by professional historians, to better inform history education (Leinhardt, Beck, & Stainton, 1994; Wineburg, 1991). For example, Leinhardt and Young (1996) have constructed a detailed model of how historians read primary sources in their own specialty areas and in areas outside of their immediate specialty area. Their model includes activities associated with identifying a document, such as classifying, corroborating, sourcing and contextualizing; and interpreting a document, such as noting the structure and language and referring to their own theories about history and its purposes. Studies of problem solving in other fields have also emerged (Ochs, Gonzales, & Jacoby, 1993).

Boyer (1990) and his colleagues (Glassick, Huber, & Maeroff, 1997) have argued that teaching, like interpreting a set of unfamiliar historical documents, is a scholarly activity. It follows, then, that just as researchers have fruitfully studied academics’ thinking and reasoning processes in conducting research and practice-related work, one might also study the ways in which faculty read and reason about evidence of teaching scholarship. The current study attempts to do so, with the aim of contributing theoretically and practically to greater understanding of the process of peer review of teaching, and to our understanding of the scholarship of teaching and teachers’ thinking.

2. Conceptual framework

From the existing literature on recommended strategies for evaluating teaching (Dwyer & Stufflebeam, 1996; Centra, 2000), on assessing scholarship (Glassick et al., 1997), and on academic reasoning processes (Elstein et al., 1978; Leinhardt et al., 1994; Wineburg, 1991; Ochs et al., 1993), one can predict possible approaches to the reading and review of a teaching portfolio. Much of the literature on the evaluation of teaching (for a review of teacher evaluation literature see Dwyer & Stufflebeam, 1996) advises that criteria and standards should be established against which to review teaching. Such considerations can be developed based on the values of the academic community involved and the extensive literature on effective teaching. In this study, the academic readers who participated were given a sample portfolio to read and evaluate, but were not given a set of criteria or standards. Rather, they were given a broad, hypothetical review task and instructed that “The University has given reviewers considerable freedom to define ‘excellence in teaching’ in ways that are appropriate to each discipline. Therefore, it is important that you explain your evaluation and how you reach your decision.” If reviewers followed the advice common in the evaluation literature, they would be expected to explicitly articulate a few criteria that they used or intended to use.

Carnegie Foundation researchers on the assessment of scholarship conclude that scholarly work is typically evaluated against six standards: “(1) clear goals (2) adequate preparation (3) appropriate methods (4) significant results (5) effective presentation (6) reflective critique” (Glassick et al., 1997, p. 25). Those authors argue that successful works of scholarship are ones in which there is congruence among purposes, methods and results. Thus, authors (or teachers, in this case) are able to guide readers’ evaluations by explicitly stating their purposes. If the academics naturally treated teaching as a scholarly activity in the way that Glassick and his colleagues have defined it, they would search for the teacher’s goals and trace for appropriateness of preparation, methods and results, framing their critique against the teacher’s goals.

Previous literature on academic reasoning, conducted in the expert-novice paradigm, has identified systematic processes that academics
engage in when working in their fields (Elstein et al., 1978, 1990; Leinhardt et al., 1994; Wineburg, 1991; Ochs et al., 1993). Academic historians, for example, use their own unique heuristics, such as a “sourcing” heuristic in which they look systematically for clues about who, when and from what perspective the document was written (Wineburg, 1991). In the current study, focused on scientists, it might be expected that the academic readers would engage in a thought process that resembles or draws upon the scientific method, in which they state a hypothesis and then test it systematically against the evidence available in the portfolio.

If readers engage in any of these processes, we might expect them to search through the portfolio, pursuing particular key pieces of information at the beginning. The ability to recognize key pieces of information is a hallmark of expert thinking in the professions, whether it be the source of a document—as historians do—or whether it is particular symptoms or indications that would allow differential diagnoses to be ruled out, as in the case of physicians. Novices, on the other hand, may not yet use the same focused process of analysis that experts use. Those inexperienced in particular domains typically lack the structure of knowledge that allows them to distinguish important pieces of evidence from surface features of the problem (Chi, Feltovich, & Glaser, 1981).

The current study looks more closely at the thinking of academics engaged in the review of teaching documentation. Gaining a better understanding of how academics (who have not received a formal introduction or written guidelines about the evaluation of teaching) naturally read and make judgments about teaching based on teaching portfolios may help in the practice of writing and reviewing teaching portfolios.

3. Method

A think-aloud interview protocol (Ericsson & Simon, 1980) was used, in which participants responded to a sample teaching portfolio based on a single course. As the academics read the materials, they spoke aloud what they were thinking into an audio tape.

3.1. Sample portfolio

I worked together with a professor in biochemistry, “Dr. John Smith” (pseudonyms are used throughout) in developing a concise, sample portfolio documenting a particular course. Dr. Smith was chosen because he had already prepared a dossier of materials about that course to support a teaching award nomination and a recent application for promotion. He is a well-respected teacher and researcher in his department, with many years of experience. Following the idea of a “course portfolio” (Cerbin, 1994, 1996; Hutchings, 1998), the portfolio focused on a single semester-long course, “Biochemistry of Metabolism and its Regulation”; a second year science course in the department of Biochemistry and Molecular Biology at “Ames University”. As is common in that department, a number of other academics contribute lectures in their specialty areas to the course and thus have some familiarity with the course.

The portfolio contained the Teacher’s Self Reflection on Teaching “Biochemistry of Metabolism and its Regulation”; Syllabus; Sample of a Summary of Lecture Notes Handed out to Students; Sample Tutorial Test; Sample Instructions for Write-Up on Tutorial Practical; Guidelines for Essay Assignment; Final Examination; and Results of Student Evaluations of Teaching. Because the study design involved reading the portfolio in a face-to-face, think-aloud interview, the portfolio had to be very short to ensure an interview of reasonable length. Therefore, the portfolio items were often selected for brevity, as well as representation of a teaching approach. Examples of student work were not included nor was there information on student grades or grade distributions. The total length of the portfolio was 31 pages of large print material. More than a third of those pages contained diagrams (in the lecture notes), large spaces for student responses (test and final exam) or tables (student evaluations of teaching) rather than substantial amounts of text to be read.
3.2. Participants

Once the portfolio was compiled, seven of Dr. Smith’s colleagues from the department of Biochemistry and Molecular Biology were invited to participate in individual interviews in which they read and thought out loud about the portfolio. The interviewees all knew Dr. Smith and were familiar to some degree with how the course fit into the department’s curriculum. They all had at least 1 year of experience teaching in the department and represented a range of specialty areas and years of teaching experience. Four interviewees had less than 8 years of experience each (1, 2, 3, 7 years). Three interviewees had between 22 and 27 years of teaching experience each (22, 25, 27 years). Each participant was asked to rate the extent to which their area of expertise overlapped with the course content, on a scale of 1–10 (1 = little expertise in the course; 10 = perfect fit, indicating that they would be comfortable enough with the material to teach the course themselves, or had previously taught the course). Ratings ranged from 2 to 10, with a median rating of 8. Thus, most participants felt reasonably comfortable and familiar with the material. Two participants were female. Five were male. This gender ratio was chosen to approximate that of the department as a whole.

3.3. The interview protocol

In the interviews, typically approximately 90 min, each interviewee was briefed on the purpose of the study and asked to read the portfolio with a hypothetical task in mind. In the task description, they were told that: “the university was initiating a new award for teaching based on excellence in a particular course. To be granted an award, applicants must select one of their courses through which to demonstrate excellence in teaching…. You have been asked to be one of the reviewers for this new award program… The university has given reviewers considerable freedom to define ‘excellence in teaching’ in ways that are appropriate to each discipline. Therefore, it is important that you explain your evaluation and how you reach your decision. How would you evaluate the following unit? Would you recommend it for the award? Why or why not? On what grounds do you make your case?”

The study was designed to simulate the way in which peer review of teaching materials might take place. Generally members of the candidate’s own department are the first group who must nominate or approve a nomination for a teaching award, or promotion or tenure (Ory, 2000). Those reviewers typically read the portfolio individually before discussing their recommendations in a group setting. Thus, selecting members of the same department for an individual reading is similar to peer review in many summative contexts.

The faculty interviewees, in addition to having this task in mind while they were reading, were instructed to think out loud about their responses to the materials as they were reading. They were instructed in accordance with recommendations on the proper use of think-aloud protocols (Nisbett & Wilson, 1977; Ericsson & Simon, 1980; Rowe, 1985). Thus, they were told explicitly to verbalize everything that came into their mind, rather than recounting their thought processes retrospectively. Most interviewees appeared to be comfortable with the research setting very quickly and talked out loud articularly, spontaneously and unself-consciously. The interviews were all taped and transcribed verbatim.

After reading the portfolio, making a judgment about whether they would give the teaching of the unit an award for excellence in teaching, and explaining that judgment, each faculty interviewee completed a survey. The survey listed each of the items in the portfolio. The faculty were asked to rate the importance of each item on a scale of 1–7 (1 = unimportant and 7 = very important) in making the task judgment, and to list additional sources of evidence they wanted. While the primary focus of the study was on the processes used in reading and reasoning about the portfolio, this second part of the study served both to provide additional information on the readers’ retrospective thinking and served as a debriefing for the participants.
3.4. Data analysis

The transcripts were analyzed using a general pattern-matching analytic technique (Yin, 1989) done through an iterative process of data collection, data reduction, data display and conclusion drawing and verification (Miles & Huberman, 1984). As described in the theoretical framework above, a number of possible types of responses were anticipated, based on the literature on the evaluation of teaching (Dwyer & Stufflebeam, 1996), the assessment of scholarship (Glassick et al., 1997) and on academics’ reasoning processes (Elstein et al., 1978, 1990; Leinhardt et al., 1994; Wineburg, 1991; Ochs et al., 1993). However, as this type of study has not been conducted previously, it was important to identify and characterize emergent reading and review processes that were not expected.

The analysis involved an initial search for patterns across interviews. After several readings, distinct processes began to emerge. Each interview was read, labeling segments for the type of process that was being used and transferring those excerpts into a matrix. During this process, categories were refined and new ones created iteratively as each excerpt was compared to existing categories. To crosscheck the flavor of the processes reduced from the data, each interview was examined to identify patterns within interviews. Regular exchange with colleagues on the project and its tentative findings also took place, adding other perspectives and questions about the data and data analysis process.

4. Results

4.1. How did readers review their colleague’s portfolio?

From the literature discussed in the theoretical framework above, it was expected that academics would engage in a systematic process of review. One obvious manifestation of such a process might be that the readers would flip back and forth through the pages of the portfolio to pursue particular hypotheses and to confirm the consistency of Dr. Smith’s statements against each other. None of the seven readers did this. All of them read from the beginning straight through the portfolio in the order it was presented. Although they had a table of contents at the beginning of the portfolio that would have made it easy for them to selectively choose a topic, they did not skip ahead. They also did not revisit any previous pages, either as they read or at the end when they were making a final judgment. Thus, as they read the portfolio they did not consciously and explicitly compare Dr. Smith’s goals with his methods, preparation or results.

None of the seven readers explicitly formulated a set of criteria from the beginning. They did appear to implicitly invoke a variety of criteria as they read, in the course of making evaluative statements.

On the surface, their review processes did not seem to match the expected, systematic pattern. The next sections take a closer look at the reading and review processes that these academics used.

4.1.1. Comparing/contrasting against a norm

Readers frequently compared or contrasted Dr. Smith’s stated practices and approaches against a stereotype or what is “normally” done. Sometimes these comparisons were as simple as how much weight was given to laboratory assessments, whether students were given a full syllabus summary instead of merely a lecture timetable, and whether the teacher frequently reviewed the text in light of newer textbooks. For example:

[John] does look fairly frequently at new books and discusses the relative merits of various books with other staff members on a regular basis, which is good. There is a tendency, (emphasis added) once you’ve got a book which you like, to just stick with it.

In some cases, though, a norm was expressed in more complex terms. For example, one reader, who specializes in science communication, used the review as an opportunity to offer extensive critiques of traditional science teaching. At several
points that reader acknowledged his bias, as in the example below, or indicated that he was critiquing not just this teacher or this particular course, but an entire tradition of science teaching.

The fact that it totally fails to realize the potential that science has for educating people in the best sense—for bringing out their critical faculty …it’s all expressed in terms of “big jug” “little mug”—predetermined substance, predetermined content. But content is all the hard crust of received knowledge at this instance in time and even where there are magnificent opportunities to… [for example] take the textbook and refer to a paper that had blown a bit of the textbook apart—these are not realized.

While in some cases the “norm” to which the readers compare and contrast is a negative stereotype of a science teacher, sometimes this “norm” is more benignly seen as merely “traditional.” “It’s a very traditional way of doing it, but he does that traditional way very, very well.” Labeling Dr. Smith’s teaching as “traditional” while recognizing that he does the traditional quite well, seemed to cause a dilemma for some of the readers who wanted to define “excellence” for an award as requiring innovation and unique approaches. As one person said frankly, “I would recommend it for an award, but I probably wouldn’t expect it to get one because it lacks a little in innovation.”

This dilemma may be rooted in the desire to judge the course on its own terms, while simultaneously disagreeing with the way the course and its aims were framed. This idea will be discussed further in a later section.

4.1.2. Comparing/contrasting with own practice

The most common strategy that the academics used was testing the reviewee’s practices against their own practice or experiences. The readers seemed to be asking themselves, “Would I do it this way?” They frequently used phrases such as, “I agree with this” or “That’s exactly the sort of approach that I think I would take.” Also common were explicit statements of “I would/ wouldn’t do X or Y,” as well as shifts to describing what they themselves do in their own classes. These kinds of comments were given in response to every aspect of the teacher’s materials in the portfolio, including the level of detail in the lecture notes, the use of an essay, the construction of the final exam, and the logic of the course content and design.

Generally, if the practice accorded with what the reviewer did or “would do” then it was evaluated positively. If the reviewers would not do it that way, then typically the practice or approach was evaluated negatively or with some reluctance and indecision as they weighed pros and cons. It may be difficult for reviewers to acknowledge a practice or goal as very different from their own, and still see that approach as effectively contributing to student learning.

The most striking example of this approach came from a reader who had taught a similar course. Rather than just critique methods used, he also took issue with the basic aims of the course as stated by Dr. Smith. He argued that the subject-specific aims of the course should be more clearly enumerated.

It’s very important for students at this basic level of biochemistry in second year to get a feel for the experimental basis of the subject… My aim was to reveal the experimental underpinning of a discipline and the development of arguments, the development of concepts to integrate many different experimental findings and the modifications of those concepts as new experiments revealed greater subtleties…

Furthermore, there were very few instances in which the readers acknowledged a practice described in the portfolio as new to them and something worth trying themselves. While one might have expected that reviewers could learn new techniques or approaches from reading an experienced colleague’s portfolio—techniques which they could try in their own classes—revelations of this kind were rare in this study. There are three explanations that could account for this apparent lack of learning from the exercise. First, many of the readers were very experienced teachers themselves. Second, the
readers were from the same department and already knew something about the course described in the portfolio and about Dr. Smith’s methods. Third, most of his approach was, as they indicated, somewhat traditional. So, there may have been little left that was “new” to them.

4.1.3. Holding up an ideal

There was variability in the degree to which the reviewers owned up to the subjectivity of their responses. Comments in which the readers made explicit “I” statements or couched their critiques in terms of their own preferences were easily categorized above. However, a number of statements seemed to refer less to respondents’ own grounded experiences and more to an idealized vision of how the course could or “should” be. Such statements may be based on personal theories, rather than actual practice. Such comments ranged from simply stated, almost common sense conclusions, “It would be more productive for students to develop some sort of rapport with a given tutor over much of the course so they could become like a mentor”, to more elaborated ideas about how the course should be taught:

I think a lot of the principles of Regulation perhaps should be more up the top and more fitting of it into the syllabus as the course goes on. Admittedly it’s called “Biochemistry and Metabolism and its Regulation”, so I suppose in a sense it deals with a lot of the basic aspects of Metabolism and, I guess tries to pull Regulation together at the end…. [But] you can’t really talk about metabolism without talking about its regulation… So I’d prefer to integrate Regulation much more tightly into the whole thing and therefore deal with regulation at the start of the syllabus rather than at the end.

Or as another critical reader said:

Here is a golden opportunity being lost where they could be given something from a textbook—I mean how about having a textbook from the 1950s and having an exercise in which they are to critique the hell out of one section…. I could think of all sorts of ways in which the essay and all sorts of other parts of the course could be used to show how science was about discovery and not about there being decided conclusions.

4.1.4. Moderating evaluations with contextual factors

While the readers had their own preferred ways of doing things, whether they were approaches that they had actively employed in their own teaching or suggestions for how the course “might be” or “should be” conducted, harsh judgments were often moderated with acknowledgement of contextual constraints. Readers noted constraints such as class schedules, class size, timing, or financial and staffing resources in the department. For some, there was an acceptance of the practical constraints as legitimate reasons that the ideal was not realized.

I think a really good unit is highly integrated and just has a natural flow about it. I think this one’s a bit stilted and broken up. For example, last year certain individuals were going away overseas at certain times so the lectures were reorganized to accommodate them. And I think it’s detracted from the curriculum. So I guess that’s a practical thing that all departments face.

At other times, though, while the practical constraints were acknowledged, the reader seemed to hold fast to a vision, refusing to accept the practical constraints as an excuse:

I suppose I dream of a course on metabolism where the first four lectures might map out a terrain and where the students then might be allowed to express some enthusiasm as to where they’re focused, where the rest of the course would focus. Obviously you could think of a thousand different reasons why this would be impossible. And I would reject them completely.

Another reader’s comments encapsulate this issue as he refers to reductions in the number of support staff available to academics:

Considering that sort of duress that all science departments have been under… they’re still
doing a very good job… . But inevitably one has to criticize something from a more absolute standard and it’s a pity that in all universities we’re losing the ability to do things as well as we used to be able to do, particularly in science departments.

4.1.5. Drawing on insider knowledge

As noted above, all participants were members of the same department as the teacher whose portfolio was being reviewed. This approach made sense because the first level of peer review in many summative contexts requires nomination and approval of the department, or at least the department head or chair (Ory, 2000). Seeing whether and how academics integrate information about the teacher and the course that they have acquired from other interactions beyond the reading of the portfolio may be useful. There were three main ways that this “insider knowledge” was brought to bear.

First, one of the most common dimensions of “insider knowledge” was an understanding and appreciation of the context of the course. The reviewers knew more about how the course was organized and the departmental decisions that led up to those structures than could be gleaned from the portfolio. For example, one of the remarks quoted above could only have been made with inside knowledge of the way the course tutorials were organized:

The way the course is structured is that for any given topic, there are a number of tutorials assigned to each [topic] section of the course. So, rather then being the Humanities model, where a given set of students in a tutorial session have a [single] tutor all the way through, one person will tend to handle all of the classes for their section [of content/topic specialisation]. So any particular student will have half a dozen [different] people leading their tutorials [over the course of the semester]… . I actually don’t think that is a good model. We’ve just discussed this recently for how we can improve our tutorials at second-year level… . And I think we can expect other staff members and students to handle, at this level, the whole

variety of [topics in] the course. And it would be more productive for students to develop some sort of rapport with a given tutor over much of the course so they could become like a mentor.

Informal feedback gathered through conversations with students was the second main type of evidence used frequently by the reviewers. For example:

I am familiar with John’s 5-minute breathers [in the middle of a lecture]. It’s an interesting concept. I understand why he does it. [But] I have actually had some feedback from students on that. Do you want me to talk about that? I think in general [that] they find it rather peculiar. I think they find it slightly contrived…. So, I personally question the need for that breather.

A third source of evidence that the departmental colleagues used was observations of the teacher giving lectures. Several of the reviewers had the opportunity to observe the teacher and mentioned their impressions from such observations. These interactions clearly allowed interpretations that went beyond the information available in the brief portfolio:

The students clearly enjoy the informality and the degree of subjectivity—I’ve been to the lectures. EXCELLENT. But again, I’m worried by this—and in fact I’m very familiar with John’s rather sort of jocular, folksy manner, which is appealing. But John often uses this to push his own agenda, rather than to open up discussion.

It is apparent, then, that the reviewers were not relying solely on the information presented in the portfolio. In some cases, it seemed that their knowledge of the department added layers of interpretation and meaning to the sketchy information in the portfolio. In other cases the reviewers seemed to have already formed a judgment about a particular practice they had observed or heard feedback from students about before even looking at the portfolio. Sometimes,
though, the portfolio gave them a context for the teacher’s intentions that allowed them a richer understanding of the practice than they might have had before. This point is illustrated above by the reviewer who said approvingly of the five minute breather that he “understands why [John] does it.”

4.1.6. Reading the portfolio text critically

The readers did not simply accept what was presented in the portfolio at face value. For example, most of the reviewers questioned one or more of the claims made by the teacher: that he memorized all the names of the students, that he could get “candid views of students” from informal face-to-face feedback, or whether some of his goals sounded contradictory or unrealistic for the time constraints.

Only one or two particularly critical readers read between the lines of what was written, noting hidden assumptions in the metaphors and language that he used. The reader who most heavily analyzed the discourse and language of the portfolio was an academic who specializes in science education and science communication. This type of reading may be more common among those more accustomed to talking about teaching and learning in scholarly circles or may be more common among humanists who are accustomed to textual analysis.

4.1.7. Brevity and adequacy of information

As the reviewers read the portfolio, they made few comments about wanting or needing further information, despite the sketchiness of the short portfolio used in the study. Only one of the seven readers concluded that there was not enough information to make a judgment about the award. The rest seemed comfortable with either their positive or negative decisions. When the readers completed the survey at the end of the interview, more comments about the completeness and level of satisfaction with each piece of evidence emerged. In general, the readers’ comments at the end of the interview, prompted by questioning about the usefulness of each piece of evidence to the overall portfolio, brought forth more critique and commentary about what a portfolio should contain than their initial reading had. Given this opportunity, most readers thought that more information would be valuable. Thus a brief portfolio that contained enough information so that all but one reader could confidently reach a binary judgment, was still considered fraught with shortcomings when the readers turned their attention to the critique of the genre of documentation. Put in another way, even though the sample was brief and left questions about the course in the readers’ minds, six of the seven reviewers still confidently reached a decision based on the information presented and their inside knowledge of the teacher, the course and the department.

4.2. Interpretation of the reading process

From the theoretical framework, I expected that academics would engage in a systematic search for key information in the portfolio. These academics, though, did not flip back and forth through the portfolio to confirm or disconfirm hypotheses or compare one piece of evidence with another. This finding seems to suggest that academics relatively inexperienced in the review of teaching do not naturally use the same type of systematic, scholarly approach as they may use with their other scholarly work.

As the readers were going through the portfolio, they were not explicitly evaluating the evidence against Dr. Smith’s goals as outlined in his reflective statement, as might be expected from Glassick and colleagues’ “Assessing Scholarship” (1997) framework. Instead, they focused more on his methods, comparing his practices to norms, ideals or to their own personal experiences. Relying on prototypes or precedent cases is common in many settings, including classic studies of expert chess players who remember thousands of other chess games and use those precedents to think through new situations as they encounter them. Like the chess players of earlier studies, these readers seemed to have pictures of case examples in their heads to which they compared Dr. Smith’s practice. Many of these examples were drawn from their own first-hand experience. Other images represented a generalization of past experience as observers and students, a sort of vision of what is “normally” or “typically” done. The only
apparent heuristic, of the type Wineburg (1991) noted in studying historians’ thinking, was a tendency to contextualize the materials in their knowledge of the external constraints within which the teacher worked.

However, as noted above, these academics did apply a critical eye to what they were reading. They did not merely accept the teacher’s claims at face value. Furthermore, the responses to the portfolio contained a range of evaluations. Three of the seven readers concluded that they would not recommend the course for an award; one did not have enough information; and three said they would recommend it for an award. One might have expected that a senior member of the department, whose teaching had been nationally recognized, would receive unanimously positive reviews, especially because he was in the position of presenting his own teaching from his own perspective. Yet, some of the academics were critical of his approach, offering eloquent reasons for their decision.

The main part of this study focused on the processes readers used as they read and evaluated the contents of the portfolio. The psychological literature on human judgment, problem solving and decision making advises researchers to focus on spontaneous thinking aloud because retrospective introspection on our thinking process is fraught with biases and miscalculations. In a search for other evidence to shed light on the ways in which readers evaluate portfolios, I looked at two other related questions: how did these academics make and justify their final decisions? What did they pay the most attention to in their reading? While both of these secondary analyses delve into the suspect areas of retrospective introspection, they nonetheless may help us better interpret how academics make sense out of portfolios.

5. Secondary analysis

5.1. Making the final decision

At the end of the interview, the participants returned to the hypothetical task of determining whether they would recommend Dr. Smith’s course for an award. The readers made this final decision after having read through the entire portfolio. At this point, their attention shifted from the reading of parts of the portfolio to making a judgment about the whole. In some ways this might be considered a retrospective introspection on their reasoning throughout the reading of the portfolio, and thus suspect for its accuracy. Nonetheless, this is precisely the reasoning that is most often shared with colleagues. As such, it is useful for practical application to see how academics’ explanations and reasoning at this stage are similar to or different from their spontaneous reactions during their initial reading.

There are striking differences between the process the readers used in their reading of the individual portfolio items, and the process they used in making the final decision. When making the final decision about whether to recommend the course for an award, most of the readers returned to a justification of their decision in terms that would match the Boyer project’s processes for the assessment of scholarship (Boyer, 1990; Glassick et al., 1997). Six of the readers either implicitly or explicitly couched their final decisions in terms of what goals were achieved through the course. The three readers who rated the course positively expressed willingness, on the whole, to take the course on its own terms. For example, one reader consciously set aside her own biases in favor of the internal consistency between Dr. Smith’s philosophy and his approach:

Well, there is a very strong concern for students... coming over in the whole philosophy behind the course. And there are certain elements that I probably wouldn’t utilize in my own teaching but the fact that they’re there indicates a very strong concern for students has been thought about. Whether I agree or not about a particular approach is irrelevant, I guess.

Two of the three readers who would not recommend Dr. Smith for the hypothetical award disagreed explicitly with his goals. They criticized
him for failing to achieve other important instructional goals. One explained:

It probably doesn’t have a sufficient emphasis on the experimental underpinning... I wouldn’t recommend it for the award—I think the ‘touchy-feely’ things—they’re overemphasized—you know, tricks of communication, self-development—all this sort of thing rather than rigorous pursuit intellectually of something that is quite demanding... . See, too many revealed answers here and too little excitement about where the next set of answers are coming from. And how one would get there experimentally.

5.2. Interpretation of analysis of readers’ final decisions

The readers seemed to use a more rational, scholarly approach when they stepped away from the actual documents to make and explain their final decision, than when they were thinking aloud in the process of reading the documents. Interestingly, while the portfolio did not contain samples of student work, only one reader noted the need for such information in order to make a decision. Thus, although the readers couched their final judgment in terms of the goals, their analysis of the achievement of those goals focused on the methods as evidenced in the syllabus, samples of teaching-related documents, and in the students’ evaluation of the course. For all but one reader, student learning outcomes were apparently either inferred from the documents presented or regarded as unnecessary to reach a decision.

5.3. To what did the readers pay most attention?

While the main focus of the study was on the reasoning process used by academics when reading a teaching portfolio, the study also raised questions about what the readers paid most attention to in their reading process. To address this question, I did two additional analyses, using different methods.

First, I analyzed the results of the survey, studying the readers’ retrospective ratings of the importance of each of the items in the portfolio. A mean importance rating was calculated across the seven readers for each portfolio item and is displayed in Table 1. The reviewers rated the student evaluations of instruction as most important in determining whether to grant an award for teaching to the course represented in the portfolio. The teacher’s self reflection was ranked as the second most important piece of evidence, while the syllabus and the guidelines for students on writing the essay assignment were also rated highly.

The second measure treats the percentage of words spoken about each section as the indicator of attentiveness (see Table 2). I divided each transcript into sections which matched the sections of the portfolio, ran a word count for each section, and calculated the percentage of words in the entire interview that were devoted to each of the items in the portfolio. Overall, respondents thought out loud most about the self-reflection

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Retrospective ratings of importance of each item in the portfolio (1–7 scale) by participant (P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>P2</td>
</tr>
<tr>
<td>Self reflection</td>
<td>5</td>
</tr>
<tr>
<td>Syllabus</td>
<td>6</td>
</tr>
<tr>
<td>Lecture notes</td>
<td>4</td>
</tr>
<tr>
<td>Tutorial test</td>
<td>6</td>
</tr>
<tr>
<td>Practical</td>
<td>4</td>
</tr>
<tr>
<td>Essay</td>
<td>4</td>
</tr>
<tr>
<td>Final exam</td>
<td>6</td>
</tr>
<tr>
<td>Student evaluations</td>
<td>7</td>
</tr>
</tbody>
</table>
by the teacher (38% of words spoken), and the syllabus (30% of words spoken). Each of the other six items in the portfolio accounted for only between 2% and 8% of the words spoken.

5.4. Interpretation of analysis of readers’ attentiveness

We might take the findings of the secondary analysis of the readers’ attentiveness as somewhat supportive of the “scholarship of teaching” rationale that underpins arguments for teaching portfolios. One of the major benefits that portfolios offer is the opportunity for teachers to present their aims, philosophy and thinking behind their pedagogical choices in a self-reflective statement. The self-reflective statement was rated as the second most important piece of evidence in the portfolio and the readers spent the most time on it (thought aloud the most about it). On average, nearly 40% of the words spoken were about the self-reflection. Together, those two measures of attentiveness provide some evidence for the importance of the self-reflection to the review of the portfolio. The conclusiveness of those findings is limited by the fact that the self-reflection was one of the longest documents in the portfolio—it contained more text than the other entries. It was also the first piece of evidence presented. It is possible that the readers were feeling time pressured and spent less time on later sections than they did at the beginning of the session. However, this order was probably realistic, as one might expect to start a portfolio with an introductory commentary from the teacher. Nonetheless, future research could vary the order in which the portfolio entries are presented.

6. Conclusion

There has been much rhetoric about the benefit of peer review of teaching and of teaching portfolios in particular. The rationale for the peer review of teaching (particularly teaching and course portfolios) has grown out of theoretical contributions on the scholarly nature of teaching (Boyer, 1990) and teachers’ knowledge (Shulman, 1986). However, there has been little research on peer review to determine whether and how those theorized concepts actually come into play during the review process by untrained academics. This study borrows cognitive psychological methods (Ericsson & Simon, 1980) used for research on academics’ thought processes and problem solving in other scholarly areas (Elstein et al., 1978; Leinhardt et al., 1994; Wineburg, 1991; Ochs et al., 1993) to study reasoning about teaching portfolios. This study is an important step in the empirical investigation of teacher thought processes in that it adds insight about how faculty makes sense of and forms judgments about a colleague’s teaching documentation.

In this study, academics used a normative, case-based approach to reasoning about the teaching practices represented in the portfolio. As they read, they were likely to compare the teacher’s practices to their own experiences, that of their colleagues, and to a vision of what is typical, normal or traditional in science teaching. This

<table>
<thead>
<tr>
<th>Category</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
<th>P7</th>
<th>Mean</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self reflection</td>
<td>33</td>
<td>46</td>
<td>33</td>
<td>46</td>
<td>24</td>
<td>47</td>
<td>37</td>
<td>38</td>
<td>24–47</td>
</tr>
<tr>
<td>Syllabus</td>
<td>28</td>
<td>29</td>
<td>25</td>
<td>35</td>
<td>34</td>
<td>29</td>
<td>29</td>
<td>30</td>
<td>25–35</td>
</tr>
<tr>
<td>Lecture notes</td>
<td>9</td>
<td>7</td>
<td>8</td>
<td>3</td>
<td>6</td>
<td>0</td>
<td>6</td>
<td>6</td>
<td>0–9</td>
</tr>
<tr>
<td>Tutorial test</td>
<td>4</td>
<td>3</td>
<td>9</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>0–9</td>
</tr>
<tr>
<td>Practical</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0–4</td>
</tr>
<tr>
<td>Essay</td>
<td>8</td>
<td>1</td>
<td>2</td>
<td>6</td>
<td>8</td>
<td>9</td>
<td>6</td>
<td>1–9</td>
<td></td>
</tr>
<tr>
<td>Final exam</td>
<td>15</td>
<td>1</td>
<td>13</td>
<td>8</td>
<td>8</td>
<td>4</td>
<td>7</td>
<td>8</td>
<td>1–15</td>
</tr>
<tr>
<td>Student evaluations</td>
<td>0</td>
<td>9</td>
<td>9</td>
<td>0</td>
<td>12</td>
<td>11</td>
<td>8</td>
<td>7</td>
<td>0–12</td>
</tr>
</tbody>
</table>
approach is consistent with problem solving research findings (Shulman, 1992).

The academics in this study seemed to use few of the sorts of systematic heuristics that Wineburg (1991) observed in historians’ problem solving about historical documents, though. The readers’ linear progression through the documents does not suggest a search for particular pieces of information to confirm or disconfirm hypotheses, to find evidence to support particular criteria, or to check for internal consistency. The readers did, however, tend to contextualize the practices, considering the practical constraints and knowledge they already had about the teacher, the department and the subject matter when evaluating the practices presented in the portfolio.

The faculty reviewers paid more attention to teaching methods as they read the portfolio documents, putting less emphasis on the other important dimensions of scholarship described by scholars from the Carnegie Foundation (Boyer, 1990; Glassick et al., 1997). However, a secondary analysis of the readers’ final decision making processes revealed greater attention to the adequacy and achievement of educational goals. This finding occurred across the academics, regardless of years of teaching experience or degree of expertise in the subject matter. It may be that the conscious process of making and justifying a final decision invoked more of the formal, rational, scholarly processes that Glassick and his coauthors (1997) wrote about. Thinking aloud about the individual items of the portfolio, though, may be a quite different task for academics who are inexperienced in the review of portfolios. In the less conscious, more spontaneous thinking aloud, academic conventions may not dominate.

One speculation is that the task of making a decision about someone’s teaching was more familiar to the readers than studying a teaching portfolio because most of the readers did not have experience in reviewing portfolios. Perhaps academics who are formally educated about teaching and learning, or who have had extensive experience reviewing teaching, might use different processes, or approach the reading of the portfolio differently. Further study would need to be conducted with academics with different backgrounds. Teaching experience alone may not be the same as experience or education regarding the review of teaching documentation—or about a scholarly orientation to teaching. If so, programs that help prepare academics for the reviewer role may be in order.

This study sheds light on the natural tendencies of academics when asked to review teaching. Given the salience of cases from their own experience and that of traditional norms in the field, a focus of a peer reviewer preparation program could be sharing, examining and critiquing case examples, particularly those from the participants’ own experience. One aim of such a program might be to expand the repertoire of examples that the reviewers have to draw on, particularly as those examples relate to key aspects of the review criteria or standards.

Other practical implications can be drawn from this study as well. The study confirms the centrality of the teacher’s self reflection to the portfolio. Much of the rationale for teaching portfolios rests on the importance of academics offering their own commentary on their teaching. Through the secondary analysis reported here, it appears that reviewers do pay close attention to that part of the portfolio. An analysis of the readers’ final decisions also shows that understanding the teacher’s aims is important to reviewers.

Most of these reviewers confidently reached a decision based on the portfolio review, together with their existing knowledge of the teacher, the department or the students. Thus, the portfolio may be best seen as supplementing information that internal reviewers already have. For reviewers outside the department, it may be necessary to provide additional information. Given the way the readers took into account contextual factors, teachers preparing portfolios may want to explicitly address the circumstances in which they teach and the situational constraints they perceive as affecting their pedagogical choices. Furthermore, as academics seem to have a vision of the “typical” or the “traditional” in mind as they read about teaching practices, it may be useful for portfolio authors to provide thorough rationales for practices that deviate from the norm.
This study shows that academics—even those inexperienced with portfolio review—are critical readers. There has been concern about the use of portfolios in summative contexts because readers might be so biased by the teacher’s own positive claims—with materials that the teachers selected and couched in their commentaries—that they would not be able to distinguish good from bad. If that were the case, one might have expected uniformly positive reviews of the teacher in this study, given that Dr. Smith was a senior academic, was recognized nationally for teaching, and presented himself in a positive light in the portfolio. However, only three of the seven readers decided they would (hypothetically) recommend his course for an award for excellence, while three did not (one could not arrive at a definitive judgment on the basis of the materials). The readers read for substance, questioned claims made by the teacher, and made well-articulated final decisions. It is also noteworthy that these critical readers did not appear to bring the same values to their readings, reinforcing the importance of negotiating explicit, shared criteria and standards for the review of teaching.

The research presented here could be extended by further study. Think-aloud methods can be employed to expose the less conscious reasoning processes used by faculty as they read documentation about teaching. Future research might compare the readings offered by those with considerably more experience with the review of teaching documentation and familiarity with the notion of the scholarship of teaching to those like the inexperienced readers here. The task could be altered to specify particular criteria that the readers should use and study the extent to which those criteria were used and applied consistently (Centra, 2000). Readers in a variety of disciplines—beyond just science—might be studied, too. Given the existing research on historical thinking, it might be particularly useful to see whether the heuristics historians use in studying historical documents carries over to the review of teaching documents. Finally, individual cases might be investigated in more detail to provide further insight into positive versus negative judgments about colleagues’ teaching.

Acknowledgements

Research toward this paper was conducted with support from a New Starters Grant from the Australian National University while the author was a lecturer in the Centre for Educational Development and Academic Methods at the Australian National University.

References


