Assessment & Evaluation in Higher Education
Vol. 32, No. 2, April 2007, pp. 159–181

Student responses to criteria-referenced self-assessment

Heidi Andrade* and Ying Du
University at Albany, USA

This paper reports on a study of undergraduate students' experiences with criteria-referenced self-assessment. Fourteen students who had taken a course involving self-assessment were interviewed in focus groups segregated by gender. The findings suggest that students had positive attitudes toward self-assessment after extended practice; felt they can effectively self-assess when they know their teacher's expectations; claimed to use self-assessment to check their work and guide revision; and believed the benefits of self-assessment include improvements in grades, quality of work, motivation and learning. There were indications that some students sensed a tension between their own standards for good work and some of their teachers' standards. There was no evidence of differences in the responses of male and female students. The paper concludes with the suggestion that self-assessment involves a complex process of internalization and self-regulation, and with implications for research and practice.

> A master can tell you what he expects of you. A teacher, though, awakens your own expectations. (Patricia Neal)

Misconceptions about student self-assessment are common. 'Self-assessment', we often hear, 'means letting students grade themselves.' This statement is generally delivered with a raised eyebrow ('What could be more ridiculous?') and followed closely by a critique: 'They'll just give themselves As!' True? Yes and no. If a teacher asks her students to grade themselves and counts those grades toward final grades, then yes, savvy students motivated by grades will give themselves As. If, in contrast, the teacher frames self-assessment as an opportunity to reflect on the quality of students' own work in order to learn more, make improvements and perhaps even earn a higher grade, a very different picture of self-assessment emerges. The purpose of this study is to begin to paint such a picture in order to combat misconceptions that limit the development of a coherent theory of self-assessment, as well as to inform the creation of classroom practices that take advantage of a potentially powerful tool for

*Corresponding author. University at Albany, 1400 Washington Ave., ED 233, Albany, NY 12222, USA. Email: handrade@uamail.albany.edu

ISSN 0260-2938 (print)/ISSN 1469-297X (online)/07/020159–23

DOI: 10.1080/02602930600801928

teaching and learning. The study employed focus groups to investigate undergraduate students' reactions to criteria-referenced academic self-assessment. Our analyses suggest that the undergraduates became quite thoughtful about their self-assessments but some experienced a tension between self- and teacher-generated criteria.

Theoretical framework

A definition of self-assessment

The literature on student self-assessment tends to use the terms 'self-assessment', 'self-reflection' and 'self-evaluation' interchangeably. In order to impose order on the literature, we make the following distinctions in this paper: self-*reflection* takes a global view of learning in terms of one's own general qualities, attitudes and dispositions. When engaged in self-reflection, students may reflect on their achievement over a certain period, or on their interest in a particular subject matter, usually without an established set of criteria (e.g., Stellwagen, 1997; Camps, 1998; Garcia & Floyd, 1999; Walstad, 2001). Self-*evaluation* involves students in making summative judgments of their work that result in a final grade or mark, e.g., the notorious, 'I give myself an A.' Studies of self-evaluation tend to focus on the correlations between self- and teacher ratings, which are generally high (Falchikov & Boud, 1989; Gruppen *et al.*, 1997; Hafner & Hafner, 2003).

Self-*assessment* is a process of formative assessment during which students reflect on and evaluate the quality of their work and their learning, judge the degree to which they reflect explicitly stated goals or criteria, identify strengths and weaknesses in their work, and revise accordingly (Goodrich, 1996; Gregory *et al.*, 2000; Hanrahan & Isaacs, 2001; Paris & Paris, 2001; Andrade & Boulay, 2003). With few exceptions, this definition of self-assessment excludes self-grading. Rather, student self-assessment is a process in which students collect information about their own performance and see how it matches their goals and/or the criteria for their work. The study reported in this paper abides by this definition of self-assessment. Put simply, we see self-assessment as feedback for oneself from oneself.

Characteristics of self-assessment

Although there is no standard definition of self-assessment in the literature, there are several characteristics of self-assessment common to the various definitions. For one, student self-assessment is criterion-referenced. Frederiksen and Collins (1989), Wiggins (1998) and Stiggins (2001) argue that the criteria for student work must be so transparent that students can learn to evaluate their own work the same way their teacher does. The majority of the available studies of self-assessment report using assessment criteria pre-determined by a teacher or instructor (Falchikov & Boud, 1989; Longhurst & Norton, 1997; Garcia & Floyd, 1999; Hanrahan & Isaacs, 2001) but some, such as Dochy and McDowell (1997) and Stallings and Tascione (1996) argue that the criteria should be co-defined by the instructor and students.

A second common characteristic of self-assessment is an emphasis on promoting learning by providing feedback that guides students' efforts and strategies (Adams, 1998; Lewbel & Hibbard, 2001; Paris & Paris, 2001; Horner & Shwery, 2002). A third characteristic is that it is ongoing: self-assessment involves regularly monitoring and regulating one's thinking processes and task performances as they happen (Goodrich, 1996; Andrade & Boulay, 2003).

There are also commonalities in the processes of self-assessment described in the literature. With minor variations, academic self-assessment is scaffolded in the following way.

1. The teacher shares the expectations for the desired performance with students, often by providing a rubric and/or models or examples of student work (e.g., Stallings & Tascione, 1996).
2. Students prepare drafts of the assignment and formally and/or informally compare their work to the rubric and/or the examples (e.g., Hart, 1999; Gregory *et al.*, 2000; Paris & Paris, 2000; Hanrahan & Isaacs, 2001; Lewbel & Hibbard, 2001).
3. Students use the feedback generated by their self-assessments to guide them in making corrective adjustments to their work (e.g., Adams, 1998).

Research on self-assessment

Researchers make a variety of claims about self-assessment and the central role it plays in learning and academic achievement. Many of the claims are related to learner autonomy (Stallings & Tascione, 1996; Hart, 1999; Paris & Paris, 2001), including increases in metacognitive engagement (Rivers, 2001). Similarly, our focus on self-assessment reflects our interest in academic self-regulation, or the ways in which goal-setting, planning, self-judgment and self-reaction can promote achievement (Zimmerman & Schunk, 2001). We see criteria-referenced self-assessment as a key component of self-regulation – 'self-judgment' in Zimmerman's framework – with the potential to scaffold other components, including goal-setting, planning and self-reaction.

Although the role of self-assessment in becoming an autonomous, metacognitive, self-regulated learner has logical appeal, it has little empirical support. Studies of self-*evaluation*, or the correlation between self- and teacher ratings, are reviewed in Falchikov and Boud (1989). Studies of the effectiveness of self-*assessment*, in contrast, are scarce. Stallings and Tascione (1996) employed student self-assessment in high school and college mathematics classes and found, among other things, that most of the students checked their work more readily than students in previous classes who were not exposed to self-assessment practices. In a survey of undergraduates who had engaged in self- and peer assessment, Hanrahan and Isaacs (2001) report that students see benefits of, as well as difficulties with, self- and peer assessment. Benefits included gaining a better understanding of grading, developing critical thinking, developing empathy with lecturers and becoming motivated to do better work in order to impress one's peers. Some students reported difficulties with self- and peer-assessment when they were 'not sure of standards' (p. 59).

Taken together, the studies by Stallings and Tascione, and Hanrahan and Isaacs provide support for our hypothesis that self-assessment can promote the kinds of behaviours typical of self-regulated learners. However, not enough is known about what students actually do, think and feel when they are asked to self-assess, to enable researchers to construct a useful theory of self-assessment or to determine the most effective approaches to self-assessment in the classroom. Because, as Brookhart (2003, p. 6) notes, 'student perceptions are inextricably tied to the classroom assessment experience and ultimately the meaning and use of the information it affords', more evidence of how students perceive of and use self-assessment is needed.

This study attempts to lay the groundwork for future research via an exploratory investigation of issues related to students' responses to self-assessment. Although we are particularly interested in self-regulation and academic achievement, we chose to begin at the beginning by posing broad questions related to how students respond to self-assessment.

1. How do students react to criterion-referenced self-assessment? What do they think, feel and do when required to self-assess?
2. What are students' attitudes toward and beliefs about self-assessment after extended experience with it?

This study was also designed to pursue the possibility of gender differences in students' responses to criterion-referenced self-assessment. Some research suggests that such differences may exist (Goodrich, 1996; Andrade & Boulay, 2003), while other studies found no differences (Dweck *et al.*, 1978; Roberts & Nolen-Hoeksema, 1989). Given the mixed results related to gender and self-assessment, we included our third research question:

3. Do male and female students respond to self-assessment in different ways?

Method

Participants

Fourteen undergraduate teacher education students (six female and eight male Caucasian, middle-class US Midwesterners) participated in focus-group interviews. The groups were segregated by gender in an attempt to capture any gender differences in students' responses. Three of the groups included four students. Because two women did not attend the scheduled focus-group interview, one of the female groups had two students. In order to ensure participants had experience with formal self-assessment and enough 'depth of experience with the phenomenon [to] describe their experience articulately' (Hill *et al.*, 1997, p. 530), participants were recruited from the first author's former class lists. Each participant had completed the professor's introductory educational psychology course in 2000 or 2001. The focus group interviews were conducted in spring 2002.

The educational psychology course had involved the participants in regular, formal self-assessment according to rubrics and checklists like those in Appendix A. For each

assignment, students had been required to check the criteria or circle the gradation of quality that best described their work, then attach the rubric or checklist to their written work to hand in. The self-assessments had not counted toward students' grades but had been required in order to have the work graded.

Because a small number of individuals who are acute observers and well informed is 'more valuable many times over than any representative sample' (Blumer, 1969, cited in Fontana & Frey, 1994, p. 365; see also Rubin & Rubin, 1995), the participation of students who had been particularly reflective and forthcoming with their opinions in class was solicited by the professor. The sample was a purposeful one, chosen for its potential to illuminate areas in need of further study, not to represent a larger population. Each participant signed a consent form before the interview, and was offered a $20 stipend.

Procedures

Focus groups were used because they may permit participants to make more critical comments than they would in one-on-one interviews (Kitzinger, 1995), because the format of a focus group tends to create a permissive, non-threatening environment in which participants can share ideas and perceptions (Krueger & Casey, 2000), and because 'young people are often stimulated to talk more expansively when others of their age join them' (Bogdan & Biklen, 1998, p. 100). Each session was both video- and audio-taped, and an advanced doctoral student took field notes.

The interviews were conducted by the first author. The decision to have the participants' former professor interview them about elements of that professor's class was a calculated risk. The fact that the interviewer and interviewees had already developed the kind of personal relationship valued in qualitative research (Rubin & Rubin, 1995) was weighed against the possibility of the data being tainted by social desirability. The risks of social desirability affecting responses were minimized because the students were not in a class with the interviewer at the time of the interview and knew they would never again have that professor for a class.

The interviews were semi-structured. The interview questions that were used to initiate discussion can be found in Appendix B. Transcripts of the interviews were sent to all participants for confirmation and clarification and revised, as appropriate, in light of their comments.

Data analysis

Focus-group data analysis is similar to the analysis of other qualitative self-report data, with an added emphasis on the impact of group dynamics and the interactions between participants (Kitzinger, 1995). An adapted version of the consensual qualitative research methodology (CQR) (Hill *et al.*, 1997) was used. CQR involves a team of researchers in coming to a consensus during five analytic steps: (1) developing domains or topic areas; (2) coding the data; (3) constructing core ideas across cases while examining the data for confirmatory and disconfirmatory evidence; (4) charting

the results; and (5) writing a narrative summary. Codes (see Appendix C) were defined in terms of the content of participants' comments, rather than by length of utterance. Information about focus group (1–4) and placement within the interview were included with each code in order to allow for an examination of representativeness of students' views (Reed & Payton, 1997). An analysis of gender differences was conducted separately, using a similar process.

Results

The analyses revealed eight main findings. We will review briefly the first six findings, which are relatively unsurprising, and discuss the last two findings in detail, as appropriate.

Finding 1: students reported that their attitudes toward self-assessment became more positive as they gained experience with it

Although many participants initially perceived of the requirement to self-assess as 'a big pain', they were unanimous in reporting positive attitudes toward it after having done it. Students' comments suggest that their initial reactions might be explained in terms of their lack of prior experience with self-assessment. None of the participants had experienced formal academic self-assessment before. As a result, they reported a perceived inability to self-assess, and they placed a low value on themselves as a source of feedback on their academic work. One male student addressed the latter issue explicitly when he commented, 'maybe I didn't value my own feedback enough.'

Once students overcame these hurdles, however, they were quite positive about self-assessment. One female student said, '… I don't know the belief in it or, something just made me realize that it's just a great thing to do … it's something that I want to do when I start teaching … Not only did it help me, but also I just think that it would help everyone.'

Finding 2: self-assessment and teacher expectations were inextricable

Students said that if they self-assessed academic work at all they usually did it in terms of what they knew about their teacher's expectations: 'with the help of the guidelines [the rubrics and checklists], I think that made it pretty easy to look at what I did personally and was able to, you know, kind of almost try to objectively look at it like it was someone else's paper and what would I say about this if someone asked me to give them feedback on it.' Students reported feeling frustrated when expectations were not communicated. In fact, frustration about unclear expectations was rampant. A male student complained, 'you're judged based on no principles … if you don't have guidelines, there's nothing to base any assessment on so you just kind of write what you have. You may look at grammar if you're an English person. And then you turn it in and hope you got the bulk of the content. Without any guidelines, you can't

assess yourself based on those guidelines, especially if you don't understand what you're reading to begin with.' Similarly, a female student noted: 'My other class is just so vague and the questions are just so open, and you're just out there, and you don't even really know what you're aiming for. I mean, you know what you think you're supposed to answer but, I mean, it's really difficult in other classes when they don't tell you or at least give you some idea. I mean, let you go with it, but just give me an idea what needs to be in it.' Although students admitted that they do not always read written expectations as carefully as they should, they crave clearly articulated requirements, criteria and standards.

Finding 3: students felt they can self-assess effectively and are more likely to self-assess when they know what the teacher expects

'It is a guessing game without the rubric. I mean you can have in your mind what the teacher wants and what you should put into it then you self-assess that, but with [a rubric] you know, you exactly know, there's no guessing.' One female student spoke of attending carefully to her professor's expectations because she did not want to be left out of her own evaluation: 'My geology professor now, we have eight papers due. But he, … nobody was really sure exactly what he wanted and one day he went over it and I have like an entire page of notes because I wasn't going to be left out of the grading. [laugh] I want to do well in the class so I pay a lot more attention to the professors when they talk about grades and papers and stuff.' In contrast, when expectations are not articulated, students report little or no self-assessment: 'Because always before, I'd type up the paper and hand it in. The teacher never told us what she wanted on it.'

Finding 4: students self-assess by checking, revising and reflecting

Students reported using criterion-referenced self-assessment to check on their works-in-progress, to guide revision and/or to reflect on their understanding of the topic: 'As I was writing it, I'd be looking at the rubric, just the whole time. And then afterwards, look at [the rubric or checklist] again. So pretty much, I used them both before, during and after to make sure I had everything covered.' 'I would ask myself, is this in here? Would I give myself an A on this or a B or a C?' 'I would write, and then go back and look at it to see if I was missing anything for the assignment, to see if I had to improve on something on some part of the paper.' Some students admitted that they did not self-assess as often as they should and that, at least at first, they did the formal self-assessments only because they were required: 'I probably wouldn't have used it as much [if it wasn't required], I don't think. Maybe after I saw how much it actually helped. But I know that first paper; [the rubric] would probably never even have been kept. It would have been in my trashcan. Like I said, you were the first professor who ever required something like that.'

Other students said their self-assessment was relatively mindless until they found that careful self-assessment could help them do better work and get better grades.

They also noted that when they did self-assess, they usually used their judgments to guide revision: '… if I say, hmm, I didn't do that so well, I'll try and correct it. It's like getting another shot, sort of.' However, they would use their self-assessments to revise if and only if they had an opportunity to resubmit their work for a new, presumably higher, grade.

Finding 5: students believed there were multiple benefits of criterion-referenced self-assessment

The most commonly cited results were higher grades and better academic work: 'It made you do better.' 'I knew what to expect more, I felt like I was ready, like I knew what I would probably receive. I knew what I earned.' Other perceived results included improvements in the ability to focus on key elements of an assignment: 'it helps you focus, it helps you get to where you need to be, it helps you learn material', and increased effectiveness in identifying strengths and weaknesses in their work: 'It's like you're, I don't know, you take kind of like take a teacher's role. You're looking at it from a different, granted you're the one who wrote it, but you're looking at it from a different point of view.' Participants also commented on increases in motivation, mindfulness, and learning, and a reduction in anxiety: 'Self-assessment … just eases your mind about doing your papers and stuff, it doesn't make you so anxious and you can actually work ahead a little bit.' 'Confidence-wise, it just made it easier to turn your paper in.' Some participants also talked about having developed a habit of self-assessment.

Finding 6: the transfer of self-assessment processes was spotty

A few students reported transferring both the process of and the criteria for self-assessment from their educational psychology class to other classes. Others, however, admitted they were not consistent in self-assessing: 'I do catch myself slipping, like I'll turn in a paper and I'll be thinking to myself, I'm like, that's just a bunch of BS you turned in there. But then the next time I'll be specific.' Most students admitted that they did not self-assess enough or at all in other classes. They cited a lack of motivation and a lack of support for self-assessment among the reasons that 'we slip': 'There are not self-assessment things in others' classes … I haven't been able to apply that very well. It's hard for me to make a rubric on my own, or something like that, or to make a checklist, unless it is spelled out in a course guide.' 'I'm back into the old routine, I see myself going back into where I spew out a five-page paper in a couple of hours. I don't get the opportunity to know where I need to be, and that, it hurts.'

Finding 7: a tension between teachers' expectations and students' own standards of quality

The study surfaced a tension between the notion of self-assessment and of assessment according to teachers' expectations. It appears that some students see a difference between the two, which made us ask, where is the self in self-assessment, and whose

expectations matter? This question first arose when we noticed that some students were troubled by the fact that their teachers' expectations clashed with their own values or standards. For example, when a group of male participants were asked about their perceived abilities to judge their own work, they said it depends on the teacher: (Participants' names have been changed.)

Jason: Not good … I'm horrible at [self-assessment]. For most of the, part of the time.

Nathan: It depends. I usually think I'm pretty good, but this quarter, last quarter, I've had a couple of professors where it seems like, yeah, I know it. We go in to the test. You know, like, it seems we know what we're doing, but they have a totally different expectation, or something like that. Like I'm in math right now. So you'd think in math there'd be no ambiguity, excuse me, whatever it is, ambiguity about it. But you figure it's like black or white. You know, 1 + 1 = 2. But, like, a lot of the classes I have right now involve proofs, so it can be actually, like, pretty stylistic. So what I think is fine, like, one teacher might say, well you didn't do this or something. I thought, well, I could just assume that.

Interviewer: Jason, you say you're terrible at it. Why?

Jason: Well, now that I think about it, maybe not so much that I think I'm terrible. It depends on if the teacher thinks in the manner that I do. [*Sounds of agreement from the group*.] If I had a similar personality and think about things in the same way as the professor, then when I self-assess myself, it's going to be pretty close to what the professor thinks about it. But then like he said, you'll have other professors where it's like cats and dogs and you just don't get along – period. And what I think is top-quality work, they're going to look at and hate it.

Interviewer: Matt, are you a good, bad, or middling judge of your own work?

Matt: Uh, I'd have to say pretty good. Pretty good. Because I can tell, like I just had anatomy last quarter, which was, I got through it but. The first test I went into I, oh man, I thought I did good and then I got it back. I didn't do too good. But I think once I figure out, like after the first test or so, I can figure out. And then, like, after I get it back, you can, like learn from what you did. You know what I mean?

Two of the female participants commented on a mismatch between their own and a teacher's standards, and expressed their discomfort with suppressing their own standards:

Dana: I think you just form your style to the teacher's, like I'm in a class this quarter that we're doing like lesson plans and she's so picky about the layout, and I just go nuts because at this whole end of doing my lesson plans – and no other teacher will ever ask me to do them like that again – and it's just so frustrating, and I'm just doing it to please her and not for me. All teachers want something different and so you conform yourself for that period of time to please them.

Annie: I guess that, as in like Dana's example, about it, like you always are, like even though it's not the way you would do it or what you would want turn in, but not the way you want, but no matter what you're still going to make sure that it's good enough. Well, you think.

A female student in another group expressed a similar discomfort:

> [My professor] has a different way of requiring, like what he requires in a thesis is just so different than what I'm used to. Because he wants you to, I mean I guess you're supposed to spell out everything but it's just to whole new level with him. You have to say, 'In this paper I will be talking,' and for me … I like to … have a nice thesis you know and not just be so [pause]. So that was kind of a traumatic experience but I don't know if it's self-assessment.

The comment, 'I don't know if it's self-assessment' was echoed by a male student who struggled with the distinction between self-assessment and 'giving them what they want'. He said, 'We're trained to spew out what the teacher wants but, and that's where, and I'm not sure if this says that we're self-assessing or that we're simply just breaking down what the teacher wants in the paper. Basically you're just giving them what they want … it is self-assessing but what is it self-assessing, it's self-assessing what the teacher wants in the paper.'

The difference between self-assessment and giving the teacher what he or she wants was a recurring theme. A few students referred to self-assessment in terms of their own expectations. More often, however, students spoke of a tension between their own and their teacher's expectations. Table 1 has a sampling of their comments. Over and over again, students rejected their own judgments of their work in favour of guessing how their teacher or professor would grade it.

Our analyses suggest that the tension between one's own and another's standards does not exist, or at least is not as salient, in non-academic contexts. Students said they were able to judge their performances according to their own criteria and

Table 1. Student comments on the tension between their own and their teachers' expectations

- I don't think I knew how [to self-assess]. I don't think that I knew that I knew how. I think I knew how to evaluate what a professor would expect and want.
- You're trying to prove yourself to the teacher and you're trying to make your work what they want to see and it doesn't really matter too much what you think.
- Right now, it seems like with a lot of my classes, I'm not so much worried about, like, how well I think the work is. I'm just worried about doing work that the professor's going to like … Because, like, we're in groups for some of these classes and we're like doing these projects together. And we're not thinking, like, is this good enough? We're like literally saying out loud, 'OK, what is he gonna want for this. And I don't think that's good enough. He's gonna want that right there.' You know?
- You could go on about like what you, yourself, want to put in the paper, but if it's not, and it could be a great paper, but if it's not to what the teacher wants, you're not going to get the grade that you want, so you constantly, you go to the teacher.
- I wouldn't say like what do I think of it, because the person grading it's the one that gives you the grade that really matters, I mean your feedback doesn't matter if, you know, you say it's an A it doesn't matter, it depends on what the teacher says.
- You are trained to think that feedback is great and that the job is not to do your best and to do your best to yourself, the job is to get the best grade and that's to impress your teacher and to get the best feedback from your teacher. So, maybe if you're not trained to think about you know how, what you think is your best work, it's what someone else thinks is your best work.

standards on the job and in social situations. For instance, one young woman said, 'You know the way you're supposed to act in a job, you know what I mean, just like, this is what I know is appropriate and this is what I'll do and I know if that if I do something inappropriate it will look bad.' In contrast, most students said they rely on their teachers to provide the criteria for academic self-assessment.

Finding 8: no evidence of gender differences

A second surprise in the study was the lack of clear evidence of gender differences in responses to academic self-assessment. Given the differences between male and female students in some earlier research, we predicted that female and male students would report differences in their reactions to and/or perspectives on self-assessment. We combed the data carefully for such differences but did not find them.

Discussion

The data from this study suggest that the participants uniformly endorsed self-assessment after extended practice, and felt they could effectively assess their own work when they knew their teacher's expectations. These findings are relatively unsurprising, given earlier research. White (1998), and Black and Wiliam (1998) have written about the need for students to know what counts in order to self-assess. Black and Wiliam argue, 'pupils can assess themselves only when they have a sufficiently clear picture of the targets that their learning is meant to attain. Surprisingly, and sadly, many pupils do not have such a picture, and they appear to have become accustomed to receiving classroom teaching as an arbitrary sequence of exercises with no overarching rationale' (p. 143). The students interviewed for this study frequently noted that having a rubric or checklist that articulated their teacher's expectations for an assignment helped make those expectations seem clear and fair and scaffolded self-assessment (Andrade & Du, 2005).

Students also claimed to have experienced multiple benefits from the practice of self-assessment, including improvements in the quality of their learning and their work, course grades and motivation. Given these findings, we believe empirical studies of the effectiveness of self-assessment are likely to produce positive results.

This study does not support the gender differences in academic self-assessment noted in some earlier research. Either they do not exist, at least in relation to the questions asked, or our research design was not powerful enough to detect them. Precedent for the former conclusion can be found in the literature on self-regulated learning, which shows that many gender differences in academic self-beliefs disappear when previous achievement is controlled (Pajares, 2002). Given the limitations of this study, however, we cannot claim to have definitive evidence that there are no gender differences in students' responses to criterion-referenced self-assessment.

Designers of future studies are advised to explicitly address the question, 'Where is the self in self-assessment, and whose expectations matter?' Whether or not students see themselves as active and authentic participants in the assessment of their own

work is likely to be a key determinant of the effectiveness of any approach to self-assessment. We must understand the tension students noted between self-assessment and self-assessment according to teacher expectations. Do some students feel uncomfortable 'conforming themselves' to a teacher's expectations because they do not understand the relationship between those expectations and a broader definition of quality, or because the expectations are truly idiosyncratic? Is it a novice/expert problem, a matter of not yet having experience with the criteria and standards commonly accepted by the domain? Does it reflect a performance orientation toward learning? Is it a matter of students trusting what they know (Raider-Roth, 2005), or a matter of power (Tan, 2004)?

We suspect a combination of issues is at work here, and the tension between self-assessment and teacher expectations reflects a complex process of internalization of criteria and standards and the development of self-regulation. For the purposes of the following discussion, we will assume that the tension between student-held standards for a piece of work and truly idiosyncratic, unreasonable teacher expectations is self-explanatory and appropriate. This discussion assumes reasonable, discipline-based teacher expectations that reflect definitions of quality generally accepted by the domain.

From the cognitive developmental point of view, the internalization of criteria and standards involves equilibration. Disequilibrium occurs when cognitive structures are challenged by new information, and the learner is motivated to regain equilibrium through assimilation and accommodation (Piaget, 1975). According to this view, students with different prior knowledge and experiences will experience different levels of discomfort – the 'pain' of disequilibrium – as they adapt to their teacher's expectations for an assignment.

From the sociocultural perspective, the concept of appropriation (Wertsch, 1991) suggests that internalization involves adopting conceptual and pedagogical tools and internalizing ways of thinking in particular contexts. The extent of adoption depends on the congruence of learners' values, prior experiences and goals with those of more experienced members of the same culture (Newman *et al.*, 1989; Wertsch, 1991; Smagorinsky, 1995; Cole, 1996). In this study, the students who readily adopted their teachers' expectations may have experienced congruence, whereas students who experienced a tension were struggling with incongruence. If a teacher's expectations are especially demanding or novel, students may experience incongruence, struggle longer and suffer more 'pain' in order to effectively appropriate those expectations.

From the perspective of self-regulated learning (Zimmerman & Schunk, 2001), internalization can be seen as a matter of transitioning from other-regulation to self-regulation by interaction with others, including teachers. Zimmerman (2002) has identified four milestones in the development of self-regulated skill: observation; emulation; self-control; and self-regulation. At the observation level, learners watch a model perform a task and become able to discriminate between levels of quality in the model's performance. At the emulation level, learners practise the skill being learned by emulating the strategic features demonstrated by the model and blending them into their own approaches. At the level of self-control, a skill becomes internalized but

remains dependent on the standards of the model. By the self-regulated level, learners are presumed to be able to make adjustments to their performances based on self-monitored outcomes. It is at the self-regulation level that learners develop 'their own distinctive styles of performing' (Zimmerman, 2002, p. 8).

Guided by this view, we propose that students' levels of self-regulatory skill will affect how they respond to discrepancies or incongruence with a teacher's expectations. We speculate that students who are in one of the first three levels – observation, emulation or self-control – are less likely to experience incongruence because they are actively attending to and internalizing their teacher's definition of quality. In contrast, students with enough skill at a task to generate their own standards will be more likely to experience a tension when faced with a perceived mismatch between their own and their teachers' standards.

There is some support for our theory in the interview data collected for this study. Two of the young women who expressed discomfort with some of their teachers' expectations also spoke about their preferences for an open-ended self-assessment (see Appendix D), rather than rubric-referenced self-assessment:

> Annie: I really like self-evaluations like [the open-ended questions] because it actually helps you step out, away from yourself I think, and like look at how you're actually doing. More so than these [rubrics]. Because these [open-ended self-assessments] are just information that you're writing down, and I think you kind of have to know how you're doing as a person, you know what I mean? … Those are open-ended questions where you have to actually examine yourself and how you're fulfilling and meeting the questions. You can't just, I mean it's not all just written down, you don't circle the answer. You have to come up with it.
>
> Dana: Yeah … these [open-ended self-assessments] are more self-self because it's your own concern … it's your own standards on these that it's completely based off of you and not really your teacher as much, so it's more your own expectations of yourself.

These young women's comments, which emphasize self-assessment based on their own standards, can be contrasted with Paul's claim that: 'I still would find it difficult to self-assess if I didn't know what the teacher wanted … You can self-assess all you want, but if you don't know exactly what the teacher's looking for, then how do you know?' If we were able to measure Paul, Annie and Dana's levels of self-regulation of the assignment to which they refer, we predict Paul would be at the level of emulation or self-control, and Annie and Dana would be at the level of self-regulation.

We further predict that students at the self-regulated level of control are more likely to experience a power struggle in terms of whose knowledge or standards are legitimate – their own or their teacher's. Tan (2004) describes competing conceptions of power in self-assessment, and concludes that the focus of self-assessment should be on learning rather than on student autonomy or power. We agree. Our definition of self-assessment as informal feedback, not self-grading, is intended to emphasize learning. Nonetheless, this study suggests that power issues need to be managed even in classrooms that emphasize formative self-assessment over self-grading.

We believe that, in general, issues of incongruence and power can be productive if students are given an opportunity to think and talk about the matches and

mismatches between their own and their teacher's definitions of quality, and if the students and teacher co-define quality for a given assignment. We propose that shared definitions of quality can serve as goals for students and teachers. As a result, 'power … can be used for the benefit of students' (Tan, 2004, p. 660) and new opportunities for self-regulation will be available to them.

Practical implications

This study is based on self-report data from a small sample of students. The limitations of the study necessarily prevent us from making concrete recommendations to teachers interested in using student self-assessment, particularly teachers with populations of students very different from ours. We can, however, note tentative implications of the study. One is that student self-assessment is feasible and likely to be beneficial if it is employed as a process of having students critically review their own work with an eye for improvement. In our teaching, we make self-assessment a prerequisite to having an assignment graded in order to ensure that students actually do it. In order to avoid confusing matters of summative evaluation with formative assessment, we do not count students' self-assessments toward a grade. Counting self-assessments toward final grades, we believe, could turn students' attention away from the quality of their work and how to improve it, and toward getting a high grade, thereby compromising their honesty and their focus on learning.

A second implication is that teachers, professors and instructors of all kinds should share their expectations for an assignment and their definition of quality on that assignment in some way. Students in our study frequently complained about a lack of clarity regarding standards of quality, and reported that self-assessment was less difficult when they knew the expectations for an assignment. The obvious implication is to ensure that students know and understand the criteria and standards for each assignment by discussing them, sharing models and providing a detailed rubric or other scoring device. This implication reflects the views of other assessment theorists, including Black and Wiliam (1998), Shepard (2000), Brookhart (2003) and Wiggins (1998), who espouse student-centred approaches to assessment.

One final implication is that teachers should discuss the qualities of effective and ineffective pieces of work with students and use the results of that discussion to create or inform the scoring guidelines that are shared with students (see Andrade, 2000, for details regarding this approach). By so doing, teachers will not only illuminate the expectations for an assignment but might also reveal and address possible discrepancies between their own and their students' conceptions of quality.

Acknowledgements

We thank Jeremy Liu for his assistance with the review of the literature, Emily Sniezyk, Katerina Passa, and Ning Wen for their assistance with the data analysis, and our anonymous reviewers for their helpful comments.

Notes on contributors

Heidi Andrade, EdD, is Assistant Professor of Educational Psychology and Methodology at the University at Albany, USA.

Ying Du is an advanced doctoral student in Educational Psychology and Methodology at the University at Albany. USA.

References

Adams, T. L. (1998) Alternative assessment in elementary school mathematics, *Childhood Education,* 74(4), 220–224.

Andrade, H. G. (2000) Using rubrics to promote thinking and learning, *Educational Leadership,* 57(5), 13–18.

Andrade, H. & Boulay, B. (2003) Gender and the role of rubric-referenced self-assessment in learning to write, *Journal of Educational Research,* 97(1), 21–34.

Andrade, H. & Du, Y. (2005) Knowing what counts and thinking about quality: students report on how they use rubrics, *Practical Assessment, Research and Evaluation,* 10(4). Available online at: http://PAREonline.net/getvn.asp?v=10&n=3 (accessed 1 May 2006).

Black, P. & Wiliam, D. (1998) Inside the black box: raising standards through classroom assessment, *Phi Delta Kappan,* 80(2), 139–148.

Bogdan, R. & Biklen, S. (1998) *Qualitative research for education: an introduction to theory and methods* (Boston, MA, Allyn and Bacon).

Brookhart, S. (2003) Developing measurement theory for classroom assessment purposes and uses, *Educational Measurement: Issues and Practice,* 22(4), 5–12.

Camps, R. (1998) Portfolio reflection: the basis for dialogue, *The Clearing House,* 72(1), 10–12.

Cole, M. (1996) *Cultural psychology: a once and future discipline* (Cambridge, MA, Harvard University Press).

Dochy, F. & McDowell, L. (1997) Introduction: assessment as a tool for learning, *Studies in Educational Evaluation,* 23(4), 279–298.

Dweck, C., Davidson, W., Nelson, S. & Enna, B. (1978) Sex differences in learned helplessness: 2. Contingencies of evaluative feedback in the classroom and 3. An experimental analysis, *Developmental Psychology,* 14(3), 268–276.

Falchikov, N. & Boud, D. (1989) Student self-assessment in higher education: a meta-analysis, *Review of Educational Research,* 59(4), 395–430.

Fontana, A. & Frey, J. (1994) Interviewing: the art of science, in: N. Denzin & Y. Lincoln (Eds) *Handbook of qualitative research* (Thousand Oaks, CA, Sage), 361–376.

Frederiksen, J. & Collins, A. (1989) A systems approach to educational testing, *Educational Researcher,* 18(9), 27–32.

Garcia, J. A. & Floyd, C. E. (1999) Using single system design for student self-assessment: a method for enhancing practice and integrating curriculum, *Journal of Social Work Education,* 35(3), 451–461.

Goodrich, H. (1996) *Student self-assessment: at the intersection of metacognition and authentic assessment*. Unpublished doctoral dissertation, Harvard University.

Gregory, K., Cameron, C. & Davies, A. (2000) *Self-assessment and goal-setting* (Merville, Connection Publishing).

Gruppen, L., Garcia, J., Grum, C., Fitzgerald, J., White, C., Dicken, L., Sisson, J. & Zweifler, A. (1997) Medical students' self-assessment accuracy in communication skills, *Academic Medicine,* 72(10), S57–59.

Hafner, J. & Hafner, P. (2003) Quantitative analysis of the rubric as an assessment tool: an empirical study of student peer-group rating, *International Journal of Science Education,* 25(12), 1509–1528.

Hanrahan, S. J. & Isaacs, G. (2001) Assessing self- and peer-assessment: the students' views, *Higher Educational Research and Development,* 20(1), 53–70.

Hart, D. (1999) Opening assessment to our students, *Social Education,* 63(6), 343–345.

Hill, C., Thompson, B. & Williams, E. (1997) A guide to conducting consensual qualitative research, *The Counseling Psychologist,* 25(4), 517–572.

Horner, S. & Shwery, C. (2002) Becoming an engaged, self-regulated reader, *Theory into Practice,* 41(2), 102–109.

Kitzinger, J. (1995) Introducing focus groups, *British Medical Journal,* 311(7000), 299–303.

Krueger, R. & Casey, M. (2000) *Focus groups: a practical guide for applied research* (3rd edn) (Thousand Oaks, CA, Sage).

Lewbel, S. R. & Hibbard, K. M. (2001) Are standards and true learning compatible? *Principal Leadership (High School Ed.),* 1(5), 16–20.

Longhurst, N. & Norton, L. S. (1997) Self-assessment in coursework essays, *Studies in Educational Evaluation,* 23(4), 319–330.

Newman, D., Griffin, P. & Cole, M. (1989) *The construction zone: working for cognitive change in school* (New York, Cambridge University Press).

Pajares, F. (2002) Gender and perceived self-efficacy in self-regulated learning, *Theory into Practice,* 41(2), 116–125.

Paris, S. G. & Paris, A. H. (2001) Classroom applications of research on self-regulated learning, *Educational Psychologist,* 36(2), 89–101.

Piaget, J. (1975) *The equilibration of cognitive structures: the central problem of intellectual development* (T. Brown & K. J. Thampy, Trans.) (Chicago, IL, The University of Chicago Press).

Raider-Roth, M. (2005) *Trusting what you know: the high stakes of classroom relationships* (San Francisco, CA, Jossey-Bass).

Reed, J. & Payton, V. R. (1997) Focus groups: issues of analysis and interpretation, *Journal of Advanced Nursing,* 26, 765–771.

Rivers, W. (2001) Autonomy at all cost: an ethnography of metacognitive self-assessment and self-management among experienced language learners, *The Modern Language Journal,* 85(2), 279–290.

Roberts, T. & Nolen-Hoeksema, S. (1989) Sex differences in reactions to evaluative feedback, *Sex Roles,* 21(11–12), 725–746.

Rubin, H. & Rubin, I. (1995) *Qualitative interviewing: the art of hearing data* (Thousand Oaks, CA, Sage).

Shepard, L. (2000) The role of assessment in a learning culture, *Educational Researcher,* 29(7), 4–14.

Smagorinsky, P. (1995) The social construction of data: methodological problems of investigating learning in the zone of proximal development, *Review of Educational Research,* 65, 191–212.

Stallings, V. & Tascione, C. (1996) Student self-assessment and self-evaluation, *Mathematics Teacher,* 89(7), 548–55.

Stellwagen, J. B. (1997) Phase two: using student learning profile to develop cognitive self-assessment skills, *American Secondary Education,* 26, 1–8.

Stiggins, R. J. (2001) *Student-involved classroom assessment* (3rd edn) (Upper Saddle River, NJ, Merrill/Prentice-Hall).

Tan, K. H. K. (2004) Does student self-assessment empower or discipline students? *Assessment & Evaluation in Higher Education,* 29(6), 651–662.

Walstad, W. B. (2001) Improving assessment in university economics, *The Journal of Economic Education,* 32(3), 281–294.

Wertsch, J. V. (1991) *Voices of the mind: a sociocultural approach to mediated action* (Cambridge, MA, Harvard University Press).

White, E. (1998) *Teaching and assessing writing: recent advances in understanding, evaluating, and improving student performance* (2nd edn) (Portland, ME, Calendar Islands Publishers).

Wiggins, G. (1998) *Educative assessment: designing assessments to inform and improve student performance* (San Francisco, CA, Jossey-Bass).

Zimmerman, B. J. (2002) Achieving self-regulation: the trial and triumph of adolescence, in: F. Pajares & T. Urdan (Eds) *Academic motivation of adolescents* (Greenwich, CT, Information Age Publishing).

Zimmerman, B. J. & Schunk, D. H. (2001) *Self-regulated learning and academic achievement: theoretical perspectives* (2nd edn) (Mahwah, NJ, Lawrence Erlbaum).

Appendix A. Examples of rubrics and checklists used by students for self-assessment during the educational psychology course and field placement

Learning vignette rubric

Learning vignettes performance rubric

	A	B	C	D/F
Instructional objectives (What your students will learn, 1 pt)	Communicates objectives to audience verbally and in writing, and shows how they connect to the assessment of the project. Objectives reflect the generativity of the topic and include >1 high-level thinking goal(s) (critique, metacognition, analyse, interpret, solve complex problems, apply, etc.). A handout with background info is provided.	Communicates objectives verbally and in writing but doesn't connect them to assessment. Objectives only tend to reflect the generativity of the topic and include 1 high-level thinking goal for students.	Communicates learning objectives to audience by simply saying them or writing a list. Objectives do not reflect the generativity of the topic and/or do not include high-level thinking objectives.	Does not communicate learning objectives effectively, and/or objectives do not reflect the generativity of topic, and/or does not include high-level thinking objectives.
Instructional theories and techniques (How you teach, 7 pts)	Uses a wide variety of techniques that promote the learning objectives, (e.g., modelling, metacognition/thinking skills, attention to misconceptions and motivation, student interaction, wait time, MI, constructivism, ongoing feedback, transfer, reflection on prior knowledge, positive reinforcement, teacher expectations)	Uses a variety of techniques. Most are appropriate for the learning objectives of the lesson. Some may not be well- matched with objectives but none are blatantly inappropriate.	Uses a few teaching techniques. The appropriateness of one or more may be unclear, seem 'crammed in' or random.	Uses only one or two approaches to instruction. The approaches used may be limited to 'traditional' techniques such as memorization or lecture.
Active engagement (What students are doing, 4 pts.)	All or most of the instruction involves active engagement on the part of students. The teacher(s) acts as a monitor and resource.	Most of the instruction involves active engagement. Lecture and seat work, if used, requires thoughtful participation by students.	Lots of teacher talk. Some active engagement is used, but the bulk of the instruction does not rely on it.	Instruction rarely actively engages students in learning. It relies on lecture, worksheets, etc. The teacher acts as director.

Appendix A. *(continued)*

	A	B	C	D/F
Adaptations for students with special needs (3 pts)	Student's behaviour reflects the case profile. Seamless attention to atypical student. The instruction focuses on the student's needs, uses a variety of appropriate strategies for meeting those needs, and creates a supportive environment that fosters self-worth. Is consistent with laws, policies and procedures.	Student's behaviour tends to reflect the case profile. LV focuses on individual needs, uses some appropriate strategies but overlooks others. Some elements of a supportive learning environment are evident but others are missing.	Student's behaviour does not reflect case profile. The teacher may create a dependency on the part of the student. There is recognition of student's needs but the interventions either don't fill it or single the student out by focusing too much on her/him.	Deals only with typical development, or uses only inappropriate strategies (e.g., punishment is the only strategy used with an AD/HD student).
Developmental appropriateness (3 pts)	At least one attempt is made to explicitly promote development by addressing common milestones in cognitive, linguistic, personal, social and/or moral development. All activities and concepts are age-appropriate.	All activities and concepts are age-appropriate.	Most activities and concepts are age-appropriate, but there is one example of content or a teaching technique that is either too simple or too sophisticated.	Several activities or concepts are not age-appropriate.
Presentation (2 pts)	Organized and interesting. Actors know their lines and are professionally dressed. Costumes, scenery, humour and narration are used effectively. Performance is 15 minutes long.	Professional. May over-rely on telling instead of *showing* how techniques are used. Actors talked too fast and/or too quietly.	Some parts were out of character, unpolished and/or unprofessional. The LV was choppy and/or blah. Went over 15-minute time limit.	Inappropriate dress and/or language. No clear attempt to engage audience. Actors read from notes.

Throughline checklist

First throughline checklist

Throughline reflections should have the following qualities:

You	Me	
_____	_____	Address each throughline question in some depth (two or three paragraphs or more) (2 pts)
_____	_____	Reflect your own developing ideas (that is, your writing should be about your own thoughts: it can reflect the fact that many of your ideas are not yet fully 'cooked') (2 pts)
_____	_____	Raise multiple interesting questions and puzzles (things you genuinely wonder about, not just rhetorical questions) (2 pts)
_____	_____	Legible and well-written (typed or handwritten is fine, as long as I can read your writing) (1 pt)

Field placement rubric

Scoring Rubric for Observation Notes – Week 5

Self-assess by circling the items below that best describe your work, and attach this entire sheet to the back of your notes.

A	B	C	D/F
● Supports all conclusions with detailed descriptions of what was seen and/or heard (x 2). ● Makes **explicit** connections to 200 and/or 201 course content (x 2). ● No (or few) problems with conventions.	● Supports most conclusions with descriptive evidence. ● Makes some connections to 200 and/or 201. ● Several problems with conventions.	● Supports few conclusions with evidence. ● Few or token connections to 200 and/or 201. ● Frequent problems with conventions but not enough to interfere with meaning.	● No support for conclusions. ● No connections to 200 and/or 201. ● Extensive problems with conventions make the paper hard to read.

Appendix B. Interview questions

1. What are the most useful sources of feedback about your performance for you? Can you give an example? Why those?
   a. What were the most useful sources of feedback in my class?
   b. What about self-assessment? Did you do it? Was it useful? Why or why not?

2. Did you do any kind of self-assessment before coming to my class?
   a. If so, please tell me about it. Give me an example.
   b. If not, why not?
3. I'm doing this study to start finding out how students respond to self-assessment. Let's start with a reminder of the kind of self-assessment you did in my class …
4. Tell me about your experiences with formal self-assessment in my class. What did you think and feel when you were asked to self-assess?
   a. Did you do it?
   b. Why did you do it or not do it?
   c. If you did it, how did you do it? Give me an example.
   d. What was it like to assess your own work?
   e. What, if anything, did you get out of doing it?
   f. How was it like or unlike the self-assessment you talked about earlier when I asked about feedback?
5. Do you do any self-assessment now?
   a. Do you remember to do it if you aren't required to?
   b. If you do remember, do you care to do it?
   c. If you remember and care to do it, do you feel like you know how to do it?
6. Self-assessment seems to help some students but not others. Can you explain why?
   a. How or why did it not help you if it didn't? Please give an example.
   b. Does it seem possible that male and female students respond differently to self-assessment? If so, can you try to explain it?
   c. Is it possible to structure self-assessment so all students benefit? If so, how? If not, why not? Give me an example.
7. What is self-assessment for? What does it involve?

Appendix C. Final coding domains

Rubrics

- Rp: purposes of rubrics (what they do/are for)
- Ru: use of rubrics (e.g., how students say they use them)
- Rr: perceived results/effects of having used rubrics (pros and cons)
- Rpe: previous experience with rubrics

Checklist

- Chp: purposes of checklists
- Chu: use of checklists (e.g., how students say they use them)
- Chr: perceived results/effects of having used checklists
- Chpe: previous experience with checklists

Self-assessment

- S-a/ir: initial reactions to self-assessment
- S-a/ca: current attitude toward self-assessment

S-a/p: purposes of self-assessment
S-a/u: use of self-assessment (e.g., how it was done or why it was not done)

Effects/results of self-assessment (including lack of effects and pros and cons)

S-a/pe: prior experience with self-assessment
S-a/pa: perceived ability to self-assess
S-a/rep: reporting self-assessments (what was circled on rubrics or checklists, and why)
S-a/rel: reliability of self-assessments with teacher assessments
S-a/exp: explanations of possible lack of or negative effects of self-assessment
S-a/g: explanations of gender differences (including attributions)
S-a/h: evidence of having developed the habit of self-assessment, including the transfer of criteria and standards to other classes/assignments

Self-assessment as a process of adjusting to the teacher's expectations (not one's own)

Personal criteria (what counts or matters to the student him- or herself, not what's on a rubric or what others say)

Feedback

Fv: the value of or need for feedback from teachers or peers
Fs: sources of feedback, and relative value of each
Fp: purposes of feedback
Fpe: prior experience with feedback
Fr: results of getting or not getting feedback

Quality

Qg: grades vs. learning
Qq: quantity vs. quality

Relationships b/w rubrics, checklists and self-assessment

Teacher expectations (articulated or not)

Limitations of the study

Appendix D. Open-ended self-assessment referred to by participants

FINAL SELF-EVALUATION

Field Students

Name:_____ **Date:**_____

Please respond to the following questions honestly, practicing the self-reflection so necessary to good teaching. This information will also help in our joint planning for your continued professional development.

Please answer all the following questions on a separate sheet:

1. Describe your performance during the second half of the quarter regarding:
   a. Your own professionalism (attendance, punctuality, compliance with teacher requests, comportment in school, etc.).

   b. Your ability to establish rapport with and interact positively with students across age, gender, culture and disability.
   c. Any particular personal strengths you have noticed.
   d. Your comfort level in the classroom.
   e. Your opportunities to observe according to syllabus guidelines and topics.
   f. Any comments, suggestions from your homebase or other teachers.
2. How have you responded to teacher comments from mid-term evaluation?
3. What professional improvements do you plan? How will you address areas in need of improvement? What resources will you seek out?
4. Additional comments: