

Issues in Accounting Education
Vol. 13, No. 1
February 1998

A Content Analysis of Teaching Evaluation Instruments Used in Accounting Departments

Brian Patrick Green, Thomas G. Calderon
and Barbara Powell Reider

ABSTRACT: This study surveys 70 actual student evaluations (SEs) used in accounting departments. The results show that many SEs include items that students cannot assess, and more than 30 percent of SEs do not contain demographic and contextual questions. At least 20 percent of SEs capture no data on two of the dimensions of effective teaching—course materials and curriculum design/course development. The authors recommend that accounting departments should redesign their SEs to capture information on the dimensions of effective teaching that students can validly assess. It is also imperative that accounting departments re-evaluate their SEs with the objective of removing items to which students are incapable of responding, and including relevant demographic items that could enhance the interpretation of SE data. SEs, however, represent only one type of information about an instructor's teaching effectiveness. The education literature generally advocates the use of a broad portfolio of information in evaluating a professor's teaching effectiveness.

STUDENT evaluations (SEs) are used by most administrators of accounting departments as a primary information source in evaluating teaching effectiveness (Calderon et al. 1994; Yunker and Sterner 1988). Many rely exclusively on SEs (Calderon et al. 1994; Calderon et al. 1996). Although the information contained in SEs could affect faculty compensation, tenure and other personnel decisions, little is known about the content of SEs used in accounting departments. This study surveys actual student evaluation instruments used in accounting departments and explores two issues related to the validity of SEs for drawing inferences and making decisions related to a

professor's teaching effectiveness. First, the paper examines the extent to which the content of SEs cover the domain of the concept of effective teaching. Second, the paper provides insight into the extent to which SEs used in accounting departments appear, *on the surface*, to

Brian Patrick Green is an Assistant Professor at the University of Michigan-Dearborn, Thomas G. Calderon is an Associate Professor at the University of Akron and Barbara Powell Reider is an Assistant Professor at the University of Alaska Anchorage.

The authors gratefully acknowledge the helpful comments of the editor and two anonymous reviewers.

be valid for drawing inferences about an instructor's teaching effectiveness.

BACKGROUND

Traditionally, evidence about whether an SE embodies the domain of effective teaching would be referred to as *content validity*, while evidence relating to what SEs appear to measure, on the surface, would be referred to as *face validity* (Messick 1989). Though several writers have recently argued for a unified concept of validity that emphasizes the inferences drawn from an instrument (Messick 1989; Cole and Moss 1989; Sanders et al. 1994), both content and appearance are important considerations in evaluating the validity of an instrument for drawing inferences and making decisions.

The validation process is intended to compile "evidence that supports the interpretations and uses of data and information" collected by using a particular instrument (Sanders et al. 1994, 145). An assessment of content validity produces evidence that helps in determining whether an instrument—e.g., an SE instrument—embodies the domain of the constructs and behaviors about which information should be acquired. This type of evidence is essential because it contributes to the interpretation of data derived from the instrument (Messick 1989; Sanders et al. 1994). Similarly, an assessment of face validity helps in determining whether an instrument appears, on the surface, to measure what it purports to measure. Face validity is important because it contributes to stakeholders' acceptance of, and support for, the instrument and the processes it intends to sustain.

Content Validity of SEs

This study assesses the content validity of SEs by evaluating the extent to which they cover the AECC's five dimen-

sions of effective teaching. Based on a comprehensive review of the education literature, the AECC (1993) identified five dimensions that cover the critical tasks of teaching. The model, detailed in table 1, is comprehensive and consistent with other critical task models of teaching documented in the literature (Calderon et al. 1996). An SE instrument that fails to embody the dimensions of effective teaching does not have the prerequisite content for drawing inferences about the full domain of instructors' teaching performance. Such an SE is limited in scope and, at best, provides only a restrictive view of faculty teaching performance. It is neither expected nor desirable that SEs should provide the full array of information needed to completely assess all five dimensions of effective teaching. As discussed in the next section, there are many aspects of teaching that students cannot evaluate. Nevertheless, students can provide *some* information on every dimension of effective teaching (Calderon et al. 1996).

Face Validity

This section presents background information on the following three issues relating to the face validity of SE instruments:

1. Do SEs appear, on the surface, to include items that students cannot effectively evaluate?
2. How are SEs structured?
3. Do SEs include items that can help administrators and others identify factors that may correlate with student ratings?

Items That Students Cannot Effectively Evaluate

There is broad consensus that students cannot effectively respond to SE items that require them to make judgments, generalizations or inferences that are beyond their knowledge and

TABLE 1
Accounting Education Change Commission's
Dimensions and Characteristics of Effective Teaching

- **Curriculum Design and Course Development**
 To effectively design curricula and develop courses the teacher must: set appropriate objectives; develop a useful framework for the conduct of courses and programs; conceptualize, organize and properly sequence the subject matter; integrate courses with other related courses, disciplines and current research; and be innovative and adaptive to change.
- **Use of Well-Conceived Course Materials**
 Effective course materials enhance presentation skills, fulfill course objectives, are consistent with current developments and new technology in the field, create a base upon which continued learning can be built, challenge students to think, and give them the tools to solve problems.
- **Presentation Skills**
 Effective presentation skills stimulate students' interests and their active participation in the learning process, respond to classroom developments as they occur, convey mastery of the subject matter, achieve clarity of exposition, instill professionalism, and engage students with different learning styles.
- **Well-Chosen Pedagogical Methods and Assessment Devices**
 Effective pedagogical methods (e.g., experiments, cases, small-group activities) vary with circumstances (e.g., size of class, nature of the subject, ability or skill being developed). Assessment devices (e.g., examinations, projects, papers, presentations) should be geared to both course objectives and the progress of the course and should have a pedagogical component (e.g., fixing in the student's mind what is most important, learning by thinking through a problem, identifying weaknesses to be corrected, reinforcing acquired skills).
- **Guidance and Advising**
 An effective teacher guides and advises students as appropriate to the level of study and research (e.g., a freshman's exploration of potential careers, a senior's job placement, or a doctoral student's work on a dissertation).

Source: Accounting Education Change Commission (1993).

experience (Seldin 1993; AECC 1993). Calderon et al. (1996) provide a comprehensive list of items related to teaching efforts and accomplishments that students are unable to effectively assess. Among them are items that require students to make judgments in relation to:

1. the sufficiency of course content;
2. whether course materials are current;
3. instructors' knowledge of the subject matter;
4. appropriateness of course objectives and content; and
5. appropriateness of technology used in the course.

Assessment of these items require professional background and are best left to the professor's colleagues (Seldin 1993). An instrument that includes items that appear, on the surface, to require students to make inferences outside the scope of their knowledge and background lacks face validity, and may not

receive the support of faculty as the basis for assessing their teaching performance. Omission of *face invalid* items from SEs avoids the risks associated with using these items to draw inferences and make decisions related to instructors' teaching performance.

Structure of SEs

The structure of SEs includes the wording of items, the mix of open and closed-ended items and, generally, the total number of items appearing on the instrument. The wording of an SE item may be positive, neutral or negative. A negatively worded item requires students to consider a negative assertion about the instructor and then support or refute it. A positively worded item requires students to consider a positive assertion about the instructor, and then support or refute it. In contrast, a neutrally worded item is objectively phrased and solicits one of several possible responses that reflects the traits of both an effective and an ineffective instructor. Table 2 provides an example of each type of item. Because properly phrased,

neutrally worded items require students to select among unambiguous responses that describe specific attributes of the class or instructor, they produce the most direct evidence of students' perceptions of instructor effectiveness or ineffectiveness.

Open- and closed-ended items generate different types of responses. Open-ended items allow students to express an unconstrained opinion about some aspect of the class and/or instructor, and usually provide vital information for formative evaluation (Centra 1993; Marsh 1987). Failure to include open-ended items in an SE instrument negates the opportunity to obtain valuable information about instructors' classroom performances that could help to improve their teaching.

Factors That May Correlate with Student Ratings

There are many factors that correlate with SE ratings (Centra 1979, 1993; Goldberg and Callahan 1991; Green et al. 1994; Marsh 1984, 1987; Marsh and Overall 1981; Scherr and Scherr 1990).

TABLE 2
Examples of Negatively Worded, Positively Worded, and Neutrally Worded Items

Types of Items	Response Scale
<p>Negatively Worded Item <i>Given the topics covered in class, the instructor gave examinations that were unreasonably detailed.</i></p>	<p>Strongly agree Strongly disagree</p>
<p>Positively Worded Item <i>The instructor gave adequately detailed examinations that reflected the aspects of the course emphasized in class.</i></p>	<p>Strongly agree Strongly disagree</p>
<p>Neutrally Worded Item <i>Given the topics covered in class, the instructor gave examinations that were...</i></p>	<p>(1) too detailed (2) highly detailed (3) adequately detailed (4) insufficiently detailed</p>

Table 3 provides a brief review of several factors reported to be associated with SE ratings. Failure to collect data on student characteristics and other contextual factors precludes both administrators and faculty from assessing the impact, if any, on SE ratings. Thus, properly designed SEs identify and measure such variables, and provide administrators with useful data to control for fac-

tors that are unrelated to instructor effectiveness on SE ratings (Cashin 1990).

METHOD

Sample and Data Tabulation

The sample was selected from Hasselback's (1993) *Accounting Faculty Directory*. Requests for a copy of the actual SE instrument used in their programs were mailed to all 86 universities

TABLE 3
Factors that May Correlate With Student Ratings

Demographic Questions	General Findings	Sample References
Gender	Mixed results have been reported in relation to students' gender, but a significant interaction effect between student gender and instructor gender is often reported. Mixed results have been reported in relation to instructors' gender. A few recent studies suggest that female instructors receive higher ratings.	Feldman (1992, 1993); Goldberg and Callahan (1991); Basow and Silberg (1987); Tatro (1995).
Course division (upper/lower division)	Mixed results are reported on the association between SE results and course division. The literature reports no difference between faculty evaluations given by students in graduate and undergraduate courses.	Gage (1961); Goldberg and Callahan (1991); Scherr and Scherr (1990); Cashin (1990).
Course difficulty	Consistent findings of a significant negative relationship across disciplines have been reported. Within disciplines, however, findings are mixed. Some studies show that, for the same course, instructors who were perceived to give more work received higher ratings. Studies show that faculty members believe that course difficulty is negatively associated with SE ratings.	Marsh (1987, 1984); Centra (1993); Marsh and Overall (1981).
Required course (elective/non-elective)	Students generally give higher ratings to courses in their major field and to courses that they elect to take. Required courses outside the major field receive lower ratings. Students with an intrinsic interest in a course—required or elective—give higher ratings to instructors.	Marsh (1987, 1984); Centra (1993); McKeachie (1979); Scherr and Scherr (1990).

(Continued on next page)

TABLE 3 (Continued)

Demographic Questions	General Findings	Sample References
Year in school (sophomore, junior, senior, etc.)	Year in school has not been found to be significantly related to SE ratings.	Braskamp et al. (1984); Centra (1979); Cashin (1990).
Major	Significant differences across majors have been reported. Accounting and Business Administration are rated lower than the humanities. Also, majors in a particular discipline give higher ratings than non-majors.	Cashin (1990); Feldman (1978); Marsh (1987).
Student effort (outside study time)	Studies show that student out-of-class effort has a significant, positive relationship with SE ratings.	Marsh (1984, 1987); Marsh and Overall (1981); Centra (1993).
Student interest in a course prior to taking it	Student interest has a significant, positive relationship with SE ratings.	Marsh (1984, 1987); Marsh and Overall (1981); Centra (1993); Kulik and McKeachie (1973).
Overall GPA	Students with higher GPAs generally give higher ratings to their instructors. Classes with higher average GPAs generally have higher SE ratings. This is, however, interpreted as grading leniency by some studies and as evidence of learning by others.	Marsh (1987); Centra (1993); Marsh and Overall (1981).
Expected grade	Most studies report that higher expected grades are positively associated with SE ratings. There are, however, varying interpretations of this finding, including evidence of grading leniency, evidence of SE validity, and evidence of learning.	Goldberg and Callahan (1991); Kemp and Kumar (1990); Arnett et al. (1989); Marsh (1984, 1987); Centra (1977); DuCette and Kenney (1982); Feldman (1976).

with doctoral programs in accounting and to 86 randomly selected U.S. universities without doctoral programs. SEs were received from 67 schools (a 39 percent response rate). Three of the respondents submitted two evaluation forms used in their programs, bringing the total number of SE instruments received to 70. Table 4 provides selected demographic data on respondents.

All SEs received were analyzed by the authors and the resulting data were tabulated in an electronic spreadsheet. SEs were analyzed to obtain the total number of items and to classify all items into open-ended and closed-ended. All closed-ended items were classified by the authors as being *positively worded*, *negatively worded*, or *neutrally worded*. SEs were also checked for: demographic

TABLE 4
Demographic Characteristics of Schools From Which SEs Were Received

School Type	Total Number of SEs Received	Response Rates ^a
Total SEs Received	70	39%
Private	22	31
Public	48	69
Doctoral Granting	28	40
Non-Doctoral Granting	42	60
AACSB Accredited	38	54
Non-AACSB Accredited	32	46

^a Except for the first row of the table, all rates are expressed in terms of the total number of SEs received. Three schools submitted two SEs used in their programs. Responses were received from 67 of 172 schools contacted.

and contextual items; items that appear, on the surface, to require students to make inferences outside the scope of their knowledge and background; and consistency with the AECC's dimensions of teaching effectiveness.

RESULTS

This section presents results on (1) the structure of SEs, (2) inclusion of demographic and other contextual factors in SEs, (3) the extent to which *face invalid* items appear on SEs, and (4) the extent to which SEs embody the AECC's dimensions of effective teaching.

Structure of SEs

Table 5 (panels A and B) shows that SE instruments contain an average of 22.2 items made up of 2.6 (12 percent) open-ended items and 19.5 (88 percent) closed-ended items. The total number of items on SE instruments range between six and 49. Closed-ended items are the norm, and 17 percent of the SEs reviewed included no open-ended items. The mix of open-ended and closed-ended items is consistent across different types of schools. Most SEs (88 percent) contain a combination of positively worded and neutrally worded items. However, more than 70 percent of items

on SEs are positively worded (see table 5, panel C). About 35 percent of the SEs surveyed contain only positively worded items. Very few SEs (under two percent) contain negatively worded items. Overall, the structure of SEs used by different accounting departments is consistent. The format of items (positive, negative and neutral) is similar, and the number of open-ended/closed-ended items is roughly the same across different schools.

Demographic Items and Other Contextual Factors

Table 6 shows that 73 percent of SEs used in accounting departments include at least one contextual factor that may correlate with SE ratings. On average, SEs include 2.6 such items. However, 27 percent of SEs contain no contextual factors that may correlate with SE ratings. Non-AACSB and private institutions tend to use SEs with significantly ($p < .01$) fewer contextual factors than those used at AACSB and public institutions. Items relating to workload and course difficulty (46 percent), student's year in school (40 percent), whether a course is required or elective (39 percent), and expected grade (31 percent) are the most popular. A somewhat less

TABLE 5
Structure of SEs

Panel A: Mean Percentages of Open-Ended and Closed-Ended Items on SEs

<u>School Type</u>	<u>Total Number of SEs Examined</u>	<u>Mean Percent of Open-Ended Items^a</u>	<u>Mean Percent of Closed-Ended Items^a</u>
Total SEs examined	70	12	88
Private	22	14	86
Public	48	11	89
Doctoral Granting	28	11	89
Non-Doctoral Granting	42	13	87
AACSB Accredited	38	11	89
Non-AACSB Accredited	32	13	87

Panel B: Number of Items on SEs Analyzed by Type

<u>Summary Statistics</u>	<u>Total Number of Items</u>	<u>Number of Open-Ended Items</u>	<u>Number of Closed-Ended Items</u>
Range	6–49	0–17	5–48
Average number	22.2	2.6	19.5
Standard deviation	9.6	4.123	6.928

Panel C: Type of Wording Used on SEs

<u>School Type</u>	<u>Average Number of Positively Worded Items</u>		<u>Average Number of Negatively Worded Items</u>		<u>Average Number of Neutrally Worded Items</u>	
	<u>Number</u>	<u>Percent^b</u>	<u>Number</u>	<u>Percent^b</u>	<u>Number</u>	<u>Percent^b</u>
Among all SEs examined	11.7	73	0.2	1	4.2	26
Private	12.5	71	0.3	2	4.9	28
Public	11.3	73	0.2	1	3.9	25
Doctoral Granting	12.6	71	0.2	1	5.0	28
Non-Doctoral Granting	11.2	73	0.2	2	3.9	25
AACSB Accredited	11.4	71	0.2	1	4.5	28
Non-AACSB Accredited	12.2	74	0.3	2	4.0	24

^a Except for the first row of the table, all rates are expressed in terms of the total number of SEs received. Three schools submitted two SEs used in their programs. Responses were received from only 67 of the 172 schools contacted.

^b Percentages reflect the proportion of SE items that are positively, negatively and neutrally worded. Row percentages may not sum to 100% because of rounding.

popular item is information on overall GPA, which appears on 26 percent of the SEs examined. Thus, not every department that solicits information about expected grades seeks to obtain corresponding information on students' current GPA. Course division (3 percent)

and student gender (10 percent) are the least represented items. Although the literature consistently documents a significant relationship between SE ratings and students' interest in a course prior to taking it, this item appears on only 14 percent of the SEs examined.

TABLE 6
Demographic and Other Contextual Items Appearing on SEs

Panel A: Number of Instruments that Capture Data on Items that May Correlate with SE Ratings

Description of Item	Total		Non-Doctoral	Doctoral	AACSB	Non-AACSB	Public	Private
	Number	Percent						
Student's gender	7	10	4	3	4	3	6	1*
Course division	2	3	1	1	1	1	2	0
Required course	27	39	13	14	17	10	21	6**
Workload/course difficulty	32	46	16	16	22	10*	23	9**
Year (Freshman, Junior, Senior, etc.)	28	40	12	16	17	11	21	7**
Major	18	26	9	9	13	5	12	6
Student effort (outside study time)	16	23	8	8	11	5	12	4**
Student interest	10	14	7	3	7	3	6	4
Overall GPA	18	26	7	11	12	6	13	5**
Expected grade	22	31	9	13	15	7*	18	4**
p-value				$p > .10$				$p < .01$
(Based on Wilcoxon Signed Rank Test)								$p < .01$

(Continued on next page)

TABLE 6 (Continued)

Panel B: Frequency of demographic items

	Number	Percent
Number of SEs that include demographic items	51	73
Average number of demographic items per SE	2.6	
Standard deviation	2.1	
Number of SE instruments with:		
0 demographic and contextual items	19	27
1 demographic and contextual item	10	14
2 demographic and contextual items	9	13
3 demographic and contextual items	5	7
4 demographic and contextual items	11	16
5 demographic and contextual items	9	13
6 demographic and contextual items	3	4
7 demographic and contextual items	2	3
8 demographic and contextual items	2	3
Total SEs examined	70	100

* p-value (based on Chi-square test of homogeneity) < .10; **p-value (based on Chi-Square test of homogeneity) < .05.

Face Invalid Items

Table 7 shows that 39 percent of SEs contain no *face invalid* items. However, roughly 60 percent of SEs contain at least one item that requires students to make inferences that are beyond their backgrounds and experiences. Panel B, table

7, lists examples of *face invalid* items included in the SEs examined. On average, face invalid items appear on SEs at the rate of one item per SE—i.e., an average of five percent of the total number of items on the instrument. No differences across school type are evident.

TABLE 7
Descriptive Statistics on Items that Students Cannot Respond to Effectively
(*Face Invalid* Items)

Descriptive Statistics	All SEs	Doctoral	Non-Doctoral	AACSB	Non-AACSB	Public	Private
Average number of <i>face invalid</i> items per SE	0.9	1.1	0.9	1.0	0.8	1.0	0.8
Standard Deviation	0.9	1.1	0.9	1.0	0.9	1.0	0.8
Average percent of <i>face invalid</i> items per SE ^a	5%	5%	5%	5%	5%	6%	4%
Standard Deviation	6%	5%	6%	6%	5%	6%	4%
t-test of Means		p-value = .3645		p-value = .4298		p-value = .3163	

Number (percent) of SEs that Ask	Number of SEs	Percent of SEs ^b
0 <i>face invalid</i> items	27	39
1 <i>face invalid</i> item	24	34
2 <i>face invalid</i> items	14	20
3 <i>face invalid</i> items	3	4
4 <i>face invalid</i> items	1	1

Panel B: A Sample of Twelve *Face Invalid* Items Appearing on the SEs Examined^c

1. In *what ways* is this course integrated with other courses *in the curriculum*?
2. *Current and timely developments* related to the topics in this course were covered *where appropriate*.
3. Instructor presented *points of view other than his/her own* when appropriate.
4. How *well* does the instructor *know the subject*?
5. The instructor *has extensive knowledge* of the field.
6. The instructor was *knowledgeable* about his/her *area of specialty*.
7. The instructor *has current knowledge* of the field.
8. *Relevance* of course content was....
9. *Suitability* of assigned work to the course objective....
10. The course presented a *comprehensive body of information*.
11. The instructor incorporated *appropriate* global issues into the course content....
12. This course is a *valid requirement* for my major.

^a Average percent of *face invalid* items is defined as $1/N * \sum[(\text{Number of } \textit{face invalid} \textit{ items on } SE_i / \text{Total number of items on } SE_i) * 100]$ where N is the total number of SEs examined.

^b The "Percent of SEs" column does not add to 100% because of rounding.

^c Italics were added by the authors to highlight the judgment required in responding to each item.

AECC's Dimensions of Effective Teaching

Results on the extent to which items on SEs cover the five dimensions of effective teaching are provided in table 8. More than 90 percent of SEs collect data on presentation skills, pedagogical methods and assessment devices, and

guidance and advising. SEs average between three and six items on each of these dimensions. Presentation skills are the most emphasized dimension, and only one of the SEs reviewed included no items on this dimension. By contrast, more than 20 percent of SEs examined contained no items relating to the other

TABLE 8
Coverage of the AECC's Dimensions of Effective Teaching

Panel A: Mean Number of Items Relating to AECC's Effective Teaching Dimensions

	<u>Total</u>	<u>Non- Doctoral</u>	<u>Doctoral</u>	<u>Non- AACSB</u>	<u>AACSB</u>	<u>Public</u>	<u>Private</u>
Curriculum Design and Course Development	1.8	2.1	1.7	1.8	1.9	1.8	1.8
Course Materials	3.0	3.8	2.5*	3.2	2.8	3.0	3.1
Presentation Skills	6.1	6.4	6.0	6.1	6.3	5.9	6.5
Pedagogical Methods and Assessment Devices	4.1	4.7	3.9	4.1	4.3	4.1	4.3
Guidance and Advising	3.2	3.1	3.3	2.1**	3.7	3.1	3.5

Panel B: Percent of SE Instruments that Do Not Include Items Relating to AECC's Effective Teaching Characteristics

Curriculum Design and Course Development	28%	35%	24%	32%	23%	33%	18%
Course Materials	24	15	29	19	29	26	18
Presentation Skills	2	0	2	0	3	0	5
Pedagogical Methods and Assessment Devices	7	8	7	5	10	9	5
Guidance and Advising	6	8	5	8	3	6	5

* p-value (based on t-test) < .10; ** p-value (based on t-test) < .05.

two dimensions—material and curriculum design/course development. SEs with no items on course material tended to come from private, non-AACSB accredited institutions without doctoral programs in accounting. SEs from accounting departments with doctoral programs included significantly ($p < .10$) more items relating to course material than their non-doctoral counterparts.

DISCUSSION AND CONCLUSION

The results show that the structure of SEs from different accounting departments is consistent. The format of items (positive, negative and neutral) is similar, and the average number of open-ended/closed-ended items is roughly the same across different accounting departments. The typical SE used in accounting departments contains roughly 22 items made up of approximately 19 closed-ended, positively worded items, and three open-ended items. Although well-designed, neutrally worded items can be used to solicit clear, unambiguous responses from students, they are seldom used in accounting departments. Similarly, despite the consistent observation in the literature (e.g., Centra 1993; Sanders et al. 1994) that open-ended items are an invaluable source of information for formative evaluation of teaching, SEs used in several accounting departments do not include open-ended items.

In general, SEs used in accounting departments include a wide array of items to which students can effectively respond. However, more than 60 percent of SEs include at least one item that appears, on the surface, to require students to make inferences that are beyond their knowledge and experience. Such items diminish the *face validity* of SEs, and may adversely affect faculty confidence in, and support for, the teaching evaluation process and the associated deci-

sions. The problem is heightened because over 30 percent of SEs do not include demographic and other contextual items and, therefore, provide no opportunity to identify factors that correlate with student ratings. While the average number of demographic and contextual items included in SEs is generally low, SEs used in accounting departments at private institutions contain substantially fewer demographic and contextual items than those used at public institutions. Thus, it can be inferred that many accounting departments, particularly at private institutions, do not control for possible sources of bias in SE results. Yet, the literature recommends that known sources of bias—such as course difficulty, major, student effort, student interest in a course prior to taking it, and whether the course is required or elective—should be controlled for when using SE results (Cashin 1990).

Considering their exclusive role in evaluating faculty teaching performance (Green et al. 1994; Yunker and Sterner 1988; Seldin 1990, 1993), it is ironic that SEs used in many accounting departments do not provide any information relating to materials used in teaching and curriculum design and course development. A calculated decision to exclude these two dimensions from SEs is logical, and would not necessarily result in invalid inferences if (1) activities associated with the dimensions of effective teaching not covered by SEs are the responsibility of a group of faculty; (2) administrators restrict SE use to only the dimensions they cover with some degree of validity; and (3) other types of information are used, in conjunction with SEs, to evaluate teaching effectiveness. However, the available evidence indicates that SEs are the primary source of information in assessing faculty teaching performance and there is relatively little use of other types of information in

the evaluation process (Calderon et al. 1996; Calderon et al. 1994; Yunker and Sterner 1988).

Accounting departments should re-evaluate their SEs with the objective of removing items to which students are incapable of responding, and including relevant demographic items that could enhance the interpretation of SE data. They should also ensure that SEs are designed to allow students to contribute information to all aspects of the five dimensions of effective teaching that students can validly assess. It should be recognized, however, that while properly designed SEs can provide useful information on instructors' classroom performance (Centra 1993), they do not provide the complete array of information required for assessing the multiple dimensions of effective teaching. Evaluation of a professor's teaching performance should be based on a portfolio of

information drawn from multiple sources, including information from students (Calderon et al. 1996; Calderon et al. 1994; AECC 1993; Edgerton et al. 1991; Seldin 1991; Seldin and Associates 1993; Shore et al. 1991; O'Neil and Wright 1993; Cashin 1990; Sanders et al. 1994). Thus, improvements in the content and appearance of SEs represent only a subset of the ways in which the quality of inferences and decisions about a professor's teaching effectiveness may be enhanced. A more comprehensive, albeit less popular, approach for both summative and formative evaluation requires the use of portfolios that document professors' activities and accomplishments related to the various dimensions and critical tasks of effective teaching (Calderon et al. 1996; Edgerton et al. 1991, Seldin 1991; Seldin and Associates 1993; Shore et al. 1991; O'Neil and Wright 1993).

REFERENCES

- Accounting Education Change Commission (AECC). 1993. *Evaluating and Rewarding Effective Teaching*. Issues Statement No. 5. Torrance, CA: AECC.
- Arnett, K., D. R. Arnold, and D. S. Cochran. 1989. Improving business school student evaluation of faculty performance. *Journal of Education for Business* 64 (March): 268–270.
- Basow, S. A., and N. T. Silberg. 1987. Student evaluations of college professors: Are female and male professors rated differently? *Journal of Educational Psychology* 79: 308–314.
- Braskamp, L. A., D. C. Brandenburg, and J. C. Ory. 1984. *Evaluating Teaching Effectiveness: A Practical Guide*. Newbury Park, CA: Sage.
- Calderon, T. G., A. L. Gabbin, and B. P. Green. 1996. Summary of promoting and evaluating effective teaching. *Journal of Accounting Education* 14: 367–383. In *1994 Ohio Regional Meeting Proceedings*.
- , B. P. Green, and B. P. Reider. 1994. Extent of use of multiple information sources in assessing accounting faculty teaching performance. American Accounting Association/Ohio Regional Meeting.
- Cashin, W. E. 1990. Assessing teaching effectiveness. In *How Administrators Can Improve Teaching*, edited by P. Seldin. San Francisco, CA: Jossey-Bass.
- Centra, J. A. 1977. Student ratings of instruction and their relationship to student learning. *American Educational Research Journal* 14: 17–24.
- . 1979. *Determining Faculty Effectiveness: Assessing Teaching Research and Service for Personnel Decisions*. San Francisco, CA: Jossey-Bass.
- . 1993. *Reflective Faculty Evaluation*. San Francisco, CA: Jossey-Bass.
- Cole, N. S., and P. A. Moss. 1989. Bias in test use. In *Educational Measurement*, edited by Robert L. Linn, 201–221. New York, NY: American Council on Education and MacMillan Publishing Company.

- DuCette, J., and J. Kenney. 1982. Do grading standards affect student evaluations of teaching? Some new evidence on an old question. *Journal of Educational Psychology* 74: 308–314.
- Edgerton, R., P. Hutchings, and K. Quinlan. 1991. *The Teaching Portfolio: Capturing the Scholarship in Teaching*. Washington, D.C.: American Association of Higher Education.
- Feldman, K. A. 1976. Grades and college students' evaluations of their courses and teachers. *Research in Higher Education* 4: 69–111.
- . 1978. Course characteristics and college students' ratings of their teachers and courses. *Research in Higher Education* 9: 199–242.
- . 1992. College students' views of male and female college teachers: Part I—Evidence from the social laboratory and experiments. *Research in Higher Education* 33: 317–375.
- . 1993. College students' views of male and female college teachers: Part II—Evidence from student evaluations of their classroom teachers. *Research in Higher Education* 34: 151–211.
- Gage, N. L. 1961. Appraisal of college teaching: An analysis of ends and means. *Journal of Higher Education* 32: 17–22.
- Goldberg, G., and J. Callahan. 1991. Objectivity of student evaluations of instructors. *Journal of Education for Business* 66 (July/August): 377–378.
- Green, B. P., T. G. Calderon, and B. P. Reider. 1994. Biases in student evaluations of faculty. In *1994 Northeast Regional Meeting: Collected Abstracts and Papers*. American Accounting Association/Northeast Regional Meeting.
- Hasselback, J. R. 1993. *Accounting Faculty Directory*. Englewood Cliffs, NJ: Prentice Hall.
- Kemp, B. W., and G. S. Kumar. 1990. Student evaluations: Are we using them correctly? *Journal of Education for Business* 65 (November/December): 106–111.
- Kulik, J. A., and W. J. McKeachie. 1973. The evaluation of teachers in higher education. *Review of Research in Education* 3: 210–240.
- Marsh, H. W. 1984. Students' evaluations of university teaching: Dimensionality, reliability, and validity, potential biases, and utility. *Journal of Educational Psychology* 76: 707–754.
- . 1987. Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. *International Journal of Educational Research* 11: 255–379.
- , and J. U. Overall. 1981. Validity of students' evaluations of teaching effectiveness. *Journal of Educational Psychology* 72: 468–475.
- McKeachie, W. J. 1979. Student ratings of faculty: A reprise. *Academe* 65: 384–397.
- Messick S. 1989. Validity. In *Educational Measurement*, edited by Robert L. Linn., 13–104. New York, NY: American Council on Education and MacMillan Publishing Company.
- Moore, R. 1990. Student evaluations of teaching. *The American Biology Teacher* 52: 260–262.
- O'Neil, C., and A. Wright. 1993. *Recording Teaching Accomplishment: A Dalhousie Guide to the Teaching Dossier*. Halifax, Nova Scotia: Office of Instructional Development and Technology, Dalhousie University.
- Sanders, J. R. et al. 1994. *The Program Evaluation Standards*. 2nd edition. Thousand Oaks, CA: Sage Publications.
- Scherr, F. C., and S. A. Scherr. 1990. Bias in student evaluations of teacher effectiveness. *Journal of Education for Business* 65 (May): 356–358.
- Seldin, P. 1990. Faculty evaluations: Surveying policy and practices. In *Faculty Performance Appraisal*, edited by A. L. Gabbin, S. C. Cairns, and R. L. Benke, Jr. Center for Research in Accounting Education, James Madison University.

- . 1991. *The Teaching Portfolio*. Boston, MA: Anker Publishing Company, Inc.
- . 1993. The use and abuse of student ratings of professors. *The Chronicle of Higher Education* (July): A40.
- , and Associates. 1993. Successful use of teaching portfolios. Bolton, MA: Anker Publishing Company, Inc..
- Shore, B. M., S. F. Foster, G. K. Knapper, G. G. Nadeau, N. Neil, and V. Sim. 1991. *The Teaching Dossier: A Guide to its Preparation and Use*. Ottawa, Canada: The Canadian Association of University Teachers.
- Tatro, C. N. 1995. Gender effects on student evaluations of faculty. *Journal of Research and Development in Education* 28 (Spring): 169–173.
- Yunker, P., and J. Sterner. 1988. A survey of faculty performance evaluation in accounting. *The Accounting Educators Journal* 1 (Fall): 63–74.