

Student Evaluation of Instruction: A Primer for Developmental Educators

MICHAEL PREUSS
UNIVERSITY OF NORTH CAROLINA
AT GREENSBORO

The use of student evaluation of instruction (SEI) is common place in American higher education. However, there is an open and continuing debate regarding SEI, its reliability and its validity. It is important for developmental educators to be aware of the concerns that exist regarding SEI in order for them to make wise decisions regarding the use of these instruments in their programs and the areas in which they will seek to apply the results of SEI.

“For many years, educators have agreed that the fundamental purposes of teacher evaluation are both quality assurance and professional development” (White, 2002, p. 10). Therefore, student evaluation of instruction (SEI) is presented by proponents as fulfilling these two necessary roles. However, SEI opponents object to this characterization. A key concern for both groups is the use of SEI in the realm of personnel decision making. As a result a “debate revolving around what kind of measures should be used for...making personnel decisions...[regarding] retention, promotions, tenure, or salary increases, and...faculty effectiveness” (Hobson & Talbot, 2001, p. 3) continues. This has “stimulated intense debate, research, and action at various levels,” (Bangura, 1994, p. 1) including discussions of the methodology, reliability and validity of SEI.

Faculty often believe the primary purpose of SEI is use as a “formative evaluation measure” (Szeto, 1994, p. 9), “to help faculty members improve and enhance their teaching skills” (Hobson & Talbot, 2001, p. 2). SEI serves this purpose when four conditions are met: something new is learned, the new information is valued, the new information can lead to improvement, and faculty are either intrinsically or extrinsically “motivated to make the improvements” (Hobson & Talbot). However, these conditions are not often met. The second use of SEI, summative evaluation, in which they are seen as a “rational, equitable basis for making

personnel decisions” (Szeto, 1994, p. 10) or as useful “in evaluating the overall effectiveness of an instructor” (Hobson & Talbot) complicates the picture. These coincidental but distinct purposes for SEI, formative and summative assessment, underline the importance of understanding the characteristics of SEI.

SEI Methodology

The arguments advanced for the use of SEI are often pragmatic. Since feedback is desired and students are a primary source of participant feedback regarding instruction, SEI or some other measure is seen as necessary. It is argued that SEI can be performed at “...relatively low cost, [and provide] reduction in biasing error, greater anonymity, and considered answers” (Bangura, 1994, p. 2). Further, it is argued that student evaluations can serve as a catalyst for faculty and administrative consideration of teaching and learning by gathering student input regarding educational programming and instruction (Szeto, 1994, p. 8). Proponents believe that if they are “employed adequately” (Szeto, p. 7), SEI can improve teaching, increase faculty and student satisfaction with teaching, and lead to personal growth for the faculty member. Each of these arguments might be granted if the assumptions of a valid, reliable, consistently implemented instrument could be affirmed. But, these points are all contested.

Detractors attack the utility of SEI. Layne, DeCristoforo and McGinty argue that student evaluations are “time-consuming and costly to administer” (1999, p. 222), produce questionable results because of “the lack of survey administration standardization procedures” (p. 222), and are “often hurriedly completed” (p. 223) in pressure-packed and uncertain circumstances minimizing the quality of the data collected. In addition, student evaluations are anonymous. The lack of respondent accountability allows vengeance to be sought by students and tomfoolery to be practiced which would render SEI results questionable (Bangura, 1994; Fish, 2005). These circumstances from general research in higher education are applicable to the developmental education (DE) classroom as the circumstances under which SEI is employed in both instances are parallel. Further, Bangura states that standardized measures and quantitative methodology limits

student expression and the ability of the survey to measure the “considerable individual variation in frames of reference, values, and levels of understanding” (Bangura, 1994, p. 3) among the respondents. Bangura believes that SEI utilized with multi-ethnic, multi-national or mixed socioeconomic student groups actually suppress student expression by imposing “a constraint” on information gathered and the manner of expression. In many DE programs every class section is multi-ethnic, multi-national or has mixed socioeconomic student groups. Employing SEI in such settings results in a “pervasive disregard of [the] respondent’s social and personal context of meaning... in the questionnaire... and in the modes of interpretive theorizing about responses” (Bangura). Considered together, these arguments show SEI methodology involves assumptions which may not be valid and which may impact an instructor’s or administrator’s ability to identify or address concerns related to a unique classroom setting.

Methodology is not the only element of SEI that has been a point of contention. Concerns related to the reliability and validity of SEI also exist. It is in respect to these two important topics, and one’s beliefs about them, that the arguments regarding SEI turn.

SEI Reliability

The reliability or “consistency of... results” (Linn & Gronlund, 2000, p. 74) of SEI is debated. Proponents of SEI look at results for a single instrument and have demonstrated that ratings for one instructor are stable over time and across populations, that class average ratings are stable, and that the same instrument yields similar results (Obenchain, Abernathy & Wiest, 2001; Olivares, 2003). However, detractors would argue that these are one of three things: 1) examples of the reliability of the instrument in measuring the perceptions of students about instructors’ practices and personality (Hobson & Talbot, 2001); 2) examples of the reliability of the instrument as a measure of some characteristic that is yet to be determined since SEI has low validity (Hobson & Talbot, 2001); or, 3) a representation of the law of averages (Olivares, 2003). These differences in opinion exist as SEI detractors are concerned about “the reliability of the student as an evaluator” (Obenchain, Abernathy & Wiest, 2001, p.3) whereas proponents have focused on

the ability to replicate results with a given instrument. Studies have found that fewer than “one-third of... students... are... consistent in their evaluations” (Obenchian, Abernathy & Wiest, p. 4) of the same instructor when using different instruments. Given this result, the “aggregated reliability measures” reported by proponents “are giving faculty a false sense of security” (Obenchian, Abernathy & Wiest) in general and in DE settings in particular. While instruments can be developed which yield consistent results, “What remains unclear is the reliability of individual students in evaluating faculty teaching effectiveness” (Obenchian, Abernathy & Wiest, p. 6). As “reliability is a necessary... condition for validity” (Linn & Gronlund, p. 75), the absence of evidence for consistency in individual student’s evaluations of instructors gives one pause. What is being sought and is missing is evidence of consistent measures of an instructor’s effectiveness provided by the same student when rating the same course on different but related instruments.

SEI Validity

In addition to questioned reliability, the validity of SEI as a measure of teacher effectiveness is not supported. Validity “addresses... [the] level of confidence that student evaluations are reflections of an instructor’s effectiveness rather than” (Hooper & Page, 1986, p. 4) some other construct. This is the case as “Teacher performance is a dynamic criterion predicated on an ill-defined notion of teacher effectiveness” (Olivares, 2003, p. 237). “Supporters and critics of [SEI]’s concur that ‘teacher effectiveness’ has not been adequately defined and operationalized” (Olivares) by educators and scholars. Put simply, to gather and interpret information one must be certain that the construct being addressed exists, has been clearly defined “that it differs for other constructs, and that the results provide a measure of the construct that is little influenced by extraneous factors” (Linn & Gronlund, p. 83). Each of these concerns will be addressed briefly below in respect to SEI.

Defining Teaching Effectiveness

There is a question regarding “the adequacy of the definition of teacher effectiveness” (Olivares, 2003, p. 234) employed for SEI.

Such a definition “should reflect a set of teacher behaviours that are universally acceptable ‘across the whole range of subjects, levels, students, and circumstances’ and reflect an equally acceptable definition of teacher effectiveness” (Olivares) for without this “it logically follows that any inferences drawn regarding the validity of data or processes to assess teacher effectiveness are seriously compromised” (Olivares, p. 236).

While there is agreement that teaching effectively is a multidimensional construct, there is not general agreement in respect to the components of that construct or the number of characteristics of the enterprise which should be considered when seeking to measure teaching effectiveness. The suggestion that “a set of teacher behaviors that are universally accepted across a wide range of students, contexts and pedagogical methods and reflect an equally acceptable definition of teacher effectiveness” (Olivares, 2003, p. 238) should be employed in SEI to measure teacher effectiveness is sound. However, it becomes problematic as Medley traced the definition of teacher effectiveness through four stages prior to 1986 (Hooper & Page, 1986). These stages focused on different sets of characteristics to define effective instruction.

Even if one is limited to the present emphasis on discovering the characteristics of an effective teacher, there is a wide variety of views. No general agreement exists regarding “the nature and number of dimensions” (Shevlin, Banyard, Davies & Griffiths, 2000, p.2) to include. Two component definitions of teaching effectiveness were created by Swartz and Lowman and Mathie, yet these have no common elements (Shevlin, Banyard, Davies & Griffiths). Some definitions of teaching effectiveness include three items. One such definition was developed by Brown and Atkins and another by Patrick and Smart (Shevlin, Banyard, Davies & Griffiths). Orpen identified “seven teaching dimensions” (Orpen, 1981, p. 6). “Other researchers have suggested... seven factors... or nine factors of effective teaching” (Shevlin, Banyard, Davies & Griffiths). Yet, even when expanded to lists of nine characteristics, there is little overlap as portrayed in the lists created by Marsh (Bosshardt, 2001, p. 2) and Centra (Hooper & Page, 1986, p. 57-58). At the extreme end of the spectrum, “Feldman (1988) identified

twenty-two ‘instructional dimensions’ of effective teaching in his research” (Hobson & Talbot, p. 2). Still others argue “that the specific attributes of good teaching vary across courses and instructors” (Bosshardt, 2001, p.2) or that the multidimensional nature of teaching should be expanded from a primarily cognitive emphasis to include “social, civic and personal outcomes” (Shavelson & Huang, 2003, p. 12).

Other scholars question whether the use of SEI to measure teacher effectiveness is not a circular and self-perpetuating system. “One of the issues to consider is whether we are measuring the most important variables of teaching effectiveness or whether some variables are becoming more important just because they are measurable” (Shevlin, Banyard, Davies & Griffiths, p. 1). Ultimately, it is important to understand that there is agreement that teaching effectively is a multidimensional construct and, second, that we lack a general agreement regarding the nature of the construct and the number of characteristics of the enterprise which should be considered when seeking to measure teaching effectiveness. All of this is related to one point. “Supporters and critics of [SEI] concur that ‘teacher effectiveness’ has not been adequately defined and operationalized” (Olivares, p. 237) by researchers and educators leaving its measurement in SEI “seriously compromised” (Olivares, p. 236).

Yet, students are asked to employ a definition of teaching effectiveness when completing an SEI. “There is considerable evidence that suggests that students do not hold a common view of teacher effectiveness (Chandler, 1978; McKeachie, 1979) and students are prone to judgment biases (e.g. Scullen et al., 2000; Stanfel 1995)” (Olivares, 2003, p. 237). “Students’ holistic rankings represented their own perceptions of quality teaching with no parameters set by a standardized evaluation instrument” (Obenchain, Abernathy & Wiest, 2001, p. 4). Whether these evaluations are based upon a “personality theory of a good instructor” (Obenchain, Abernathy & Wiest), student “self-interests...within an organizational context” (Olivares), satisfaction of “academic goals” (Olivares) or some other factor or combination remains undetermined, the result is the same. The “‘objectivity’ of students’ evaluative judgments” (Olivares)

is suspect and the “subjectivity in student ratings of teachers is illimitable” (Olivares). “To think that students, who have no training in evaluation, are not content experts, and possess myriad idiosyncratic tendencies, would not be susceptible to errors in judgment [when completing SEI] is specious” (Olivares).

SEI Content

The content of the SEI is another important consideration. “The goal in the consideration of content validation is to determine the extent to which a set of assessment tasks provides a relevant and representative sample of the domain of tasks about which interpretations of assessment results are made” (Linn & Gronlund, 2000, p. 78). This concern is related to SEI as “questions might be relevant in some teaching situations, but not in others” (Fish, 2005, p. 4) and investigations of SEI have found “ambiguous items, positively or negatively skewed items, and items that had no correlation to classroom teaching performance” (Obenchain, Abernathy & Wiest, 2001, p. 1). These factors indicate that content validity may be absent in some SEI instruments.

Influence of Irrelevant Factors

When SEI is used as a measure of teacher effectiveness, one must consider if “it is...unaffected by potential biasing variables” (Olivares, 2003, p. 236). If influenced by “factors that are ancillary or irrelevant to the construct” (Linn & Gronlund, 2000, p. 83), called “construct-irrelevant variance” (Linn & Gronlund), the validity of SEI results is diminished. The question is whether “teacher effectiveness is being measured as opposed to, for example, course difficulty or differences in disciplines, student characteristics, grading leniency, teacher expressiveness, teacher popularity or any number of other variables” (Olivares, 2003, p. 236).

Researchers have found that SEI outcomes are influenced by factors which are not components of teaching effectiveness. These factors are, however, all components of general and DE classrooms. Among these are the level of ease in grading (Wilson, 1998), the student’s “reason for taking the course” (Shevlin,

Banyard, Davies & Griffiths, 2000, p.3), “student emotional states” (Olivares, 2003, p. 238), student “grade expectancies” (Hobson & Talbot, 2001, p. 4), difficulty of the subject, prior preparation of the students (Olivares), student perception of the instructor’s charisma (Shevlin, Banyard, Davies & Griffiths, p.1), “class size” (Shevlin, Banyard, Davies & Griffiths, p.3), the “academic discipline” (Olivares), and even the perceived sexiness of the instructor (Felton, Mitchell & Stinson, 2004, p. 1). As Shevlin, Banyard, Davies and Griffiths concluded, “overall, research on the effects of extraneous variables on the validity of [SEI] suggests the need for caution in the interpretation of... data” (Shevlin, Banyard, Davies & Griffiths, 2000, p.3).

Summation of Concerns

The concerns with the use of SEI to measure teacher effectiveness extend far beyond identifying potential influences on the results. They include questions regarding the methodology of SEI, the content of various instruments, and a valid definition of teaching effectiveness. As Hobson and Talbot wrote, citing multiple researchers, “Validity [for SEI]...is especially difficult to establish because researchers concede that there is no universally accepted criteria for what constitutes effective teaching” (2001, p. 4). As a result “any inferences drawn regarding the validity of data or processes to assess teacher effectiveness are seriously compromised” (Olivares, 2003, p. 236) in all settings including developmental education.

The debate over the validity of SEI will continue. This is due, primarily, to the proponents and detractors generating research on two different but related tracks, the reliability of the instrument and the reliability of the student as an individual evaluator respectively. SEI will also continue to impact American higher education due, in part, to an increasing emphasis on accountability. It is important to note that negative impacts of SEI which have been found by surveying faculty are “reduction in coursework demands on students... lowering grading standards...[that] 50% of respondents [indicated that they] had attempted to improve their ratings in ways they considered inappropriate” and grade inflation (Olivares, 2003, p. 241). Based upon these factors, it is possible to conclude

that “High student opinion survey scores might well be viewed with suspicion rather than reverence, since they might indicate a lack of rigor, little student learning, and grade inflation” (Felton, Mitchell & Stinson, 2004, p. 1). All of these are characteristics to be strongly avoided in general and in developmental education in particular. “Data suggest that the institutionalization of [SEI]’s... as a method to evaluate teacher effectiveness has resulted in students learning less in environments that have become less learning- and more consumer-oriented” (Olivares, p. 243).

An Approach to SEI

A “lack of validity does not mean the [SEI]’s are not useful; rather, it just suggests that [SEI]’s are not measuring what they are intended to measure and therefore inferences regarding teacher effectiveness or student learning should be constrained” (Olivares, 2003, p. 240). This is the case as SEI is not based on an accepted understanding “of teacher effectiveness across instructional settings, academic disciplines, instructors and course levels and types” (Olivares, p. 236), students don’t “hold a common view of teacher effectiveness, and are [not] objective and reliable sources of teacher effectiveness data” (Olivares), the questions asked may not reflect on teaching effectiveness, and there are numerous “potential biasing variables” (Olivares) for SEI. Developmental educators must consider these concerns and means to mitigate them when seeking to utilize SEI, weigh the results, and interpret the significance of these results.

Suggestions for Use of SEI

Given that “In general, student evaluations can [only] be taken to report... student perceptions...[and] Perceptions are not necessarily accurate representations of the objective facts” (Hobson & Talbot, 2001, p. 5; see also Obenchain, Abernathy & Wiest, 2001 and Olivares, 2003), SEI should be approached based upon an informed plan developed in collaboration between faculty and instructional leaders. In this process, faculty members, DE departments, instructional divisions and institutions should clearly define their purposes in using SEI and seek assurance that

their SEI can accomplish these purposes. For example, some institutions conduct SEI early to mid-semester to gauge student perception and to allow for appropriate alteration of the remaining planned instruction and then again at the end of the semester seeking evidence of consistency or change if it was deemed necessary. Divisions and institutions may wish to develop multiple instruments which can be used interchangeably or in conjunction with each other. Certainly, it is imperative that a clear definition of teaching effectiveness in operational terms be developed upon which the instrument(s) are then based and that this definition be communicated to students who are completing the evaluations. For SEI to be employed in a formative manner, a situation which facilitates the four conditions noted in the second paragraph of this article must be established and maintained. Each of these suggestions will involve review of SEI and its content and periods of refining and revision. In all these processes, it is imperative to have the informed input of faculty (DE and general). This will require planning at the institutional and division level and may require professional development to facilitate understanding by all parties of the constructs involved. These suggestions will not change the fact that what is being gathered is information about student perceptions. However, they will create and foster an informed, focused, collaborative and progressive investigation of what students perceive about the instruction they receive, a worthwhile undertaking.

References

- Bangura, A.K. (1994). The focus group approach as an alternative to collecting faculty evaluation data to improve teaching. Center for Educational Development and Assessment Conference on Faculty Evaluation, November, 1994. Retrieved February 4, 2005, from Education Resources Information Center.
- Bosshardt, W. & Watts, M. (2001). Comparing student and instructor evaluations of teaching. *Journal of Economic Education*, 32(1), 3-18. Retrieved February 4, 2005, from the Professional Development Collection on the Ebscohost database.
- Felton, J., Mitchell, J. & Stinson, M. (2004). Web-based student evaluations of professors: The relations between perceived quality, easiness and sexiness. *Assessment & Evaluation in Higher Education*, 29(1), 91-109. Abstract retrieved February 4, 2005, from the Professional Development Collection on the Ebscohost database.
- Fish, S. (2005). Who's in charge here? *ACADEME TODAY: The Chronicle of Higher Education's Daily Report for subscribers*, February 4, 2005.
- Hobson, S.M. & Talbot D.M. (2001). Understanding student evaluations. *College Teaching*, 49(1), 26-32. Retrieved February 4, 2005, from the Professional Development Collection on the Ebscohost database.
- Hooper, P. & Page, J. (1986). Measuring teacher effectiveness by student evaluation. *Issues in Accounting Education*, 1(1), 56-65. Retrieved February 4, 2005, from the Professional Development Collection on the Ebscohost database.
- Layne, B.H., DeCristoforo, J.R. & McGinty, D. (1999). Electronic versus traditional student ratings of instruction. *Research in Higher Education*, 40(2), 221-235. Retrieved February 4, 2005, from the Professional Development Collection on the Ebscohost database.
- Linn, R.L. & Gronlund, N.E. (2000). *Measurement and assessment in teaching*. Upper Saddle River, NJ: Prentice Hall.
- Obenchain, K.M., Abernathy, T.V. & Wiest, L.R. (2001). The reliability of students' rating of faculty teaching effectiveness. *College Teaching*, 49(3), 100-105. Retrieved February 4, 2005, from the Professional Development Collection on the Ebscohost database.
- Olivares, O.J. (2003). A conceptual and analytical critique of student ratings of teachers in the USA with implications for teacher effectiveness and student learning. *Teaching in Higher Education*, 8(2), 233-245. Retrieved February 4, 2005, from the Professional Development Collection on the Ebscohost database.
- Orpen, C. (1981). Student evaluation of lecturers as an indicator of instructional quality: A validity study. *Journal of Educational research*, 74(1), 5-7. Retrieved February 4, 2005, from the Professional Development Collection on the Ebscohost database.
- Shavelson, R.J. & Juang, L. (2003). Responding responsibly to the frenzy to assess learning in higher education. *Change*, January/February, 2003. Retrieved February 4, 2005, from the Professional Development Collection on the Ebscohost database.
- Shevlin, M., Banyard, P., Davies, M., & Griffiths M. (2000). The validity of student evaluation of teaching in higher education: Love me, love my lectures? *Assessment and Evaluation in Higher Education*, 25(4), 397-408. Retrieved February 4, 2005, from the Professional Development Collection on the Ebscohost database.
- Szeto, W.F. (1994). Perceptions of faculty performance evaluation among faculty across academic disciplines at a selected university. *Annual Meeting of the Mid-South Educational Research Association*,

November 9-11, 1994. Retrieved February 4, 2005, from Education Resources Information Center.

White, M.E. (2002). Evaluating the bilingual teacher: The monolingual administrator's challenge. *An Imperfect World: Resonance From the Nation's Violence* (Proceedings of the annual meetings of the NAAAS, NAHLS, NANAS, and IAAS in Houston, Texas, February 11-16,2002). Retrieved February 4, 2005, from the Professional Development Collection on the Ebscohost database.

Wilson, R. (1998). New research casts doubt on the value of student evaluations of professors. *Chronicle of Higher Education*, 44(19), A12-A15. Abstract retrieved February 4, 2005, from the Professional Development Collection on the Ebscohost database.

Michael Preuss is a Grant Specialist in the Office of Sponsored Programs at the University of North Carolina at Greensboro.