

Assessing the College Classroom with Unobtrusive Measures

Ruth V. Russell

Agnes Kovacs

Department of Recreation and Park Administration
Indiana University

Abstract

This paper suggests ways to expand opportunities for evaluating the college course. While exams and papers are useful, an over-dependence on them is problematic because their overt nature allows for factors other than learning (cheating, test taking strategies, attitudes) to impact assessment results. As such, we recommend the addition of unobtrusive measures to assessing what happens in the college classroom because they provide balance to the results derived from more traditional measures. The discussion includes examples of physical trace, archival, and observation sources of unobtrusive measures of student learning outcomes.

Keywords: unobtrusive measures, nonreactive measures, student learning, impact of teaching

Learning is often viewed as the unifying goal of teaching, research, and service for higher education. Yet, while learning is of paramount importance to college faculty, it is difficult to precisely define, and thus difficult to measure. Learning has been described as knowing and interpreting the known, discovering the new, and bringing about desired change in behavior (Watson & Stage, 1999). In terms of teaching, this usually means we want our students to think, perceive, feel, react, or change in a new way.

It follows, then, that if learning is the primary measure of college and university teaching productivity, what and how students learn must be the criteria by which the value of higher education is judged (American College Personnel Association, 1994). This is no small challenge, and assumes we have the wherewithal to demonstrate, in systematic ways, how what we do affects students' learning success.

When the assessment movement first gained momentum in the 1970s there were many who thought it was another "fad" that would quickly fade (Upcraft, 1999). They were wrong. In the 1980s several national reports within higher education called for an even greater emphasis on assessment and accountability for student

learning in colleges and universities. Through the 1990s to today, tough questions have been asked: What should one get from a college education? And, how do we know? (Terenzini, 1989). The effort to answer such questions is now a requirement of the accreditation of higher education institutions and curricular programs.

In spite of this, customary ways of measuring learning have remained rather constant throughout the higher education history. Most of us give examinations, assign papers and presentations, and request course and teaching evaluations. More recently we have added case studies, technology driven competency acquisition, focus groups, and other approaches to assess how students are affected by our courses. Based on the results of these measures we approximate how much and how well students have learned in our courses, or at least how popular our delivery methods have been.

The point of this paper is not to critique these customary ways of measuring what happens in the classroom. We applaud their considered and broad use and encourage their continual refinements. Indeed, as previously discussed in this journal (Dyment & O'Connell, 2003; Anderson, Schroeder & Anderson, 2001; Ross, 1998; DeGraaf & Jordan, 1996; Gladwell, Dowd & Benzaquin, 1995; McLean & Hill, 1993) important improvements in the use of methods such as portfolios, importance-performance analyses, mentoring, and focus groups have contributed in important ways to our ability to assess student learning outcomes. As well, suggestions on assessment in on-line course delivery (Hopp, Morden, & Ostiguy, 2004), in intergenerational education programs (Dupuis, 2002), and for travel education (Brault & Reidy, 2004) have added considerably to our commitment and accuracy in understanding what and how college students learn.

Instead, our purpose is to expand opportunities for determining the outcomes of teaching even more. Our point is that while exams, papers, and other customary methods are useful, an over-dependence on them is problematic because of their uniform source of measurement invalidity. In this paper we recommend the triangulated addition of unobtrusive measures because their biases provide balance to the results derived from more traditional methods.

A Rationale for Unobtrusive Measures

Typical ways of measuring what students have learned are reactive. That is, the measurement strategies, and the motives behind them, are obvious to students and thus student attributes unrelated to learning may influence results. Reactive measures run the risk of creating as well as measuring attributes (Russell & Kovacs, 2003). For example, a well written paper may reflect a student's ability to "cut and paste" from Web resources rather than actual composition skills. In this case the assessment of concept synthesis and writing is clouded by cheating ability. Likewise, a teaching evaluation scale may produce more positive (or negative) responses only because stu-

dents know that these are tools for faculty promotion and tenure decisions. Measures such as these often produce invalidity because of the reactivity of the measurement itself.

Because they are reactive, data collected through examinations, paper assignments, course evaluation scales, and other common assessment measures can contain a considerable amount of error. For example, awareness of being tested (Webb et al., 2000) and pretest sensitization (Campbell & Stanley, 1963) are sources of invalidity that come from the students themselves. As well, students who are aware of the assessment process can select roles (such as “good student” or “bad student”) that are not necessarily “true” selves or behaviors (Orne, 1962).

Another source of error coming from students is response sets (Webb, et al., 2000). This is a tendency to respond in a particular way regardless of the question asked. For example, many college students have learned how teachers devise tests and answer questions in a strategic manner. A common response set, acquiescence, suggests a student will more frequently agree with a course evaluation statement than disagree with it. Other response sets include preferences for strongly worded statements or a tendency to use a particular response pattern (such as all third choices).

Another source of potential invalidity in reactive measurement tools is the teacher. For example, students respond differentially to characteristics of the teacher. These cues might include voice tone, eye contact, age, race, ethnicity, gender, and physical size. Simply being the teacher sets up possible reactions from students to our efforts to assess their learning. Imagine, for example, how college students might react differently to a focus group facilitator who is their course instructor as opposed to a classmate. Instructor behaviors and choices can also create reactivity to assessment. When faculty use the same or similar examinations over several years, for example, fraternity, sorority, and individual student files often become rather complete with old tests. Thus, students who take tests based on the study of previous versions are measured for being in the right social circles rather than their comprehension of course concepts.

Adding nonreactive measures to the mix of classroom assessment tools helps to temper and complete what we know from reactive measures. Based on the work of Webb and associates (Webb, Campbell, Schwartz, & Sechrest, 1966; Webb, Campbell, Schwartz, Sechrest, & Grove, 1981; Webb, Campbell, Schwartz, & Sechrest, 2000), there are alternatives to assessing what is happening in the classroom – ones that are not obvious to students.

Webb and his colleagues recommend the incorporation of unobtrusive measures. Whereas reactive measures either call upon students to respond to a stimulus presented by the teacher (for instance, a question), or to cooperate with the teacher (by meeting an assignment deadline), unobtrusive measures call upon the teacher

either to find naturally occurring assessment data, to observe without being noticed, or to create situations in which the students are unaware of being assessed. Unobtrusive methods are non-reactive because they are used to collect information without asking for it (Scholl & Lago, 1994).

For example, comparing the number of students who withdraw the first week and again at mid-semester may help distinguish perceived difficulty of the course vs. actual difficulty. Library usage rates could add to what is known about students' intellectual curiosity and technology use could contribute to an understanding of student aptitudes. Student-determined changes in class seating patterns over time help us understand increasing or decreasing commitment to the course. An analysis of the content of lecture notes prepared by students who are having difficulty with the course may show those concepts that are inaccurately portrayed and/or missing. Campus bookstore records on the "over-purchase" of course textbooks compared to the enrollment count for the course can help reveal interest in the subject matter elsewhere on campus. Used simultaneously with more typical measures, unobtrusive measurement strategies offer an additional, validating source of evaluation.

Types of Unobtrusive Measures

How do we measure unobtrusively? There are many ways to measure nonreactively, but for ease of presentation we group these measures into the categories of physical traces, archives, and observation (Webb et al., 2000). Most of these strategies are easy to administer, requiring minimal instructor and institution resources.

Physical traces. Physical tracing is the study of physical evidence – those pieces of data not specifically produced for the purpose of a course assessment. For example, the cleanliness of the classroom after the class is dismissed may offer clues to how well the instructional module of "leave no trace" in outdoor recreation transferred to student everyday situations. The dust on reserved books in the library could be a hint of student participation in assigned readings.

An illustration of the usefulness of physical traces comes from a study of academic misconduct (Pullen, Ortloff, Casey, & Payne, 2000). In this study discarded "cheat sheets" were analyzed. The investigators felt that a cheat sheet demonstrates not only an intent to possess disallowed information in testing areas, but also what course concepts students don't understand. Over the course of several years discarded cheat sheets were retrieved from academic building classroom floors, hallways, the trash, and outside bushes and walkways. The information extracted from these physical traces included which disciplines are most frequently represented (business), what kind of information is recorded on cheat sheets (facts, definitions, diagrams, and formulas), how cheat sheets are constructed (grouped lists – mnemonic devices were rare), and how cheat sheets are concealed and used (hidden in the palm, used more near the end of the term).

Webb et al. (2000) distinguish between two broad classes of physical traces. First, there are erosion physical traces, which can be either natural or controlled. Second, there are accretion physical traces, which can also be either natural or controlled.

Natural erosion physical traces focus on the degree of selective wear. Libraries often measure the utility of different titles by noting the wear on the books, particularly on the corners where the page is turned. Which brochure holder racks are most frequently empty could provide an indication of the current popularity of departmental majors, minors or specialty options. *Controlled erosion physical traces* suggest the material has been adapted to be better able to pick up the trace. For example, a small glue spot on the pages, placed close to the binding, of reserve readings in the library illustrates a controlled erosion trace measure. After the reading assignment deadline has passed the readings are checked to see whether or not the seal is still intact for each pair of pages.

Third, *natural accretion physical traces* involve the deposit or accumulation of material. For example, the discipline of tutors most often advertised for on campus bulletin boards can reveal students' unmet needs for instructional support. In addition, the number of newspapers left after class might hint at lack of satisfaction with class material or instruction. As with erosion measures, it can be useful to tamper with accretion materials. An example of *controlled accretion physical traces* might include purposefully littering the classroom before students arrive and noting student reaction as an additional measure of the acquisition of a conservation ethic taught in class.

Although including physical trace measures in classroom assessment might enthrall the Sherlock Holmes in all of us, there are limitations to their usefulness. First, accretion measures vary in their survivability and tendency to be deposited. For example, chalk graffiti on sidewalks and paper trash have fleeting durability. Also, many variables influence the nature of traces left. For instance, the reserve readings may be located in a very public place, thus inhibiting students' ability to concentrate or focus long enough to completely read the material. Another limitation is the scant knowledge available about the students producing the traces; their anonymity, the very nonreactivity of the measure, prohibits us from knowing anything else about them. For example, were men or women more likely to use cheat sheets? What were the feelings or reactions of students to the most frequently used books in the library? These are the sort of questions that can only be answered by more reactive measures, such as paper assignments. This is why we recommend partnerships between reactive and nonreactive assessment measures.

Archives. Archives are written records produced for reasons other than assessment, but which can be used as an unobtrusive evaluative measure. For example, one way to determine student perceptions of the relevancy of the course material for their future careers is the book re-sale percentages available from campus bookstore records.

Another illustration of archival sources of measurement is cited by Koester (1989). Non-reactive archival measurement was used to assess 4-H textile exhibits. Exhibits were counted and compared on specific qualities to previous years' entries to yield new ideas for 4-H curriculum development, as well as marketing. For example, the difficulty and workmanship of the exhibits were noted and observations made to discover what topics were of greatest interest to the public.

Besides the low cost of acquiring even a massive amount of pertinent information, a major advantage of archival material is its nonreactivity. To demonstrate, we cite a study by Briggs (1997) on how students use information technology. The study took advantage of the advent of object-oriented technology and the integration of this technology into such tools as spreadsheets and databases. As a result the work patterns of 627 students on four class assignments in an introductory college course were analyzed. Students had complete control over the length and persistence of their computer supported work sessions on the assignments. Even-driven code made it easy to measure the "when" and "how long" of student work, as well as the correctness of that work. From this, conclusions could be drawn about procrastination, persistence, efficiency, and effectiveness in completing the assignments.

Webb et al. (2000) considered two forms of archives: the running record, and the episodic and private record.

Running records are the ongoing, continuing, and routine records that are particularly useful for longitudinal assessment. Running records tend to be institutional and public. College campuses are a wealth of running record data, including such administratively published information as student age, gender, race, country of origin, admission test scores, high school rank, campus housing type and employment rate, as well as judiciary, advising, police, affirmative action, and health center records. An illustration of using such institutional data is the consideration of Registrar archives to evaluate teacher performance by Green, Prather, and Sturgeon (1983). The indicator of performance was the degree to which instructors developed a following among students measured by how often they returned to the instructor for additional elective courses.

From a curriculum evaluation perspective, the distribution of courses taken by size and type of instruction (e.g., lecture, seminar, lab, independent study, internship) might reveal the nature of the formal educational experience of students. The guiding question could be how many opportunities were there for graduating seniors in our program to study in small numbers with a senior faculty member? As another example of running records, newspaper accounts of student activism could be evaluated according to the majors of the featured students as a measure of adoption of civic responsibility taught in the curriculum.

Episodic and private records tend to be discontinuous and seldom in the public domain. These typically include personal documents (e.g., diaries, letters, and draw-

ings). Applying this to records often kept by instructors, student patterns of social conservatism could be produced by charting across years the topics about which students choose to write in a leisure and diversity course. Another example could be the percentages of late assignments and late arrivals to class as a way of understanding student ability to be self-sufficient and self-managed. As well, the percentage of students attending professional organization meetings could provide an additional measure of student career preparation sincerity. Further, a comparison of the number of students participating in a course web-based chat room when extra credit is offered versus the number of participants when no extra credit is offered could suggest student intrinsic interest in the course subject matter.

Archival sources of unobtrusive measurement are easily obtained in colleges and universities, and when used in conjunction with other methods, add to an assessment's validity (Russell & Kovacs, 2003). Limitations of archival records, however, are authenticity, representativeness, and accuracy (Goodwin & Goodwin, 1989). Authenticity concerns whether the records are real. What is their history? How were they obtained? Are they complete or abridged? Representativeness refers to the record's ability to yield a true likeness of the issue. Such factors as selective recording and selective survival affect their representativeness. For example, e-mail messages praising an instructor are more faithfully filed than are critical ones. Also, how certain data are kept in campus offices can change across time. For example, categories of student ethnicity labels could be different today compared to twenty years ago. Finally, the accuracy of archival records is of particular importance. Who prepared the records, using what information sources, and for what purpose?

Observation. A third type of unobtrusive measure, observation, simply refers to watching people without their knowledge. Two approaches suggested by Webb et al. (2000) are simple observation and contrived observation.

Simple observation is when the observer is passive, having no control over the behavior being observed. For example, a comparison of the percentage of students who "pre-pack" notebooks into back packs before class is officially dismissed both early and late in the semester could be an indication of changes in respect for the instructor. The manner of dress for formal class presentations when a dress code is not mandated could suggest how well students identify themselves as professionals in the subject matter. Observations of the ethnicity of study groups or class seating patterns might help us measure inclusion. Observation of student non-verbal communication during class possibly demonstrates interest level in discussions. Are they leaning forward? Are they making eye contact with each other?

On the other hand, *contrived observation* involves a more active observer who deliberately varies the setting or uses mechanical devices (e.g., video cameras, audio tape recorders, eye or body movement recording equipment). Still, these interventions can be non-reactive: when handled well, students do not detect them and the apparent naturalness of the situation is preserved.

To illustrate, a video camera can be set up in the classroom with the admitted purpose of studying the teacher's behavior, but instead can focus on recording student reactions to specific teaching interventions. That is, the nature of student responses and answers to instructor initiated questions or probes can be recorded to determine willingness to participate in class discussion as well as such intellectual abilities as critical thinking skills.

Of course, there are limitations to observation measures (Russell & Kovacs, 2003). For simple observation these include the chance that much of what is observed is irrelevant. Further, only limited random sampling is possible (Goodwin & Goodwin, 1989). For contrived observations the cost and obviousness of some recording equipment can be a limitation. Analyzing and reporting results can be expensive too, if photographs or tapes are involved. The most important limitation for contrived observations, however, concerns a host of ethical issues, discussed next.

Summary of the Cautions in Using Unobtrusive Measures

Unobtrusive measures collect naturally occurring information. In comparison to more overt measuring, the non-reactivity of these methods entails less risk that those studied will react abnormally. In classroom assessment this means students do not know they are being evaluated and their permission is not acquired. Because of this non-reactivity advantage, unobtrusive measurement raises both validity and ethical concerns.

First, an important apprehension is the data collected might not actually measure the target educational outcome. Does a newspaper strewn classroom really indicate the degree of interest and attention students paid to the lecture? Because the measure is unobtrusive, students cannot be asked about the meaning of the information they provide (Adler & Sedlacek, 1988) – we do not have access to explanations or control over alternative explanations. Nonreactive measures of the classroom can thus require more extrapolation and interpretation by the instructor. The solution to these validity concerns, of course, is to use unobtrusive measures not as a replacement of other assessment methods but rather to supplement and cross-validate them.

No approach to understanding student experiences in the classroom is without bias. Thus, the reason for adding unobtrusive measures is their ability to avoid the same sources of invalidity inherent in other types of measures. In other words, even though unobtrusive measures have their own innate biases, they counterbalance the systematic error characteristics of standard procedures because their bias is different (Russell & Kovacs, 2003). Adding unobtrusive measures to classroom assessment, then, increases confidence in those conclusions that may be similar even though they emanate from different measurement methods. Measuring changes in student beliefs about altruism, for example, is more cogent if assessed not only by a paper assignment that asks about this belief, but also by counting memberships of students in civic

and humanitarian campus organizations, or measuring student involvement in community volunteer projects after the semester when such was assigned.

Ethical considerations are also of concern in the use of unobtrusive measures (Russell & Kovacs, 2003). Are privacy and the right to informed consent violated when students are not aware of the data collected about them? This is a tough call for college faculty because on the one hand we are properly reluctant to make surreptitious observations, and on the other hand worried about the fidelity of information collected otherwise. There are several defenses to such ethical issues.

First, while it might be preferable to collect data about students in other ways, some information is simply not available any way other than unobtrusively (Page, 2000). For example, the excitement levels of students for specific course materials may only be available to instructors through observation of their non-verbal behaviors.

A second defense to ethical concerns is that perhaps it is time to recognize that the doctrine of informed consent has been asked to play too great a role in assessment situations (Page, 2000) because the information being consented to is often of minor importance relative to privacy. In many instances, protection of students' rights is applied to such an extent that much inquiry cannot be undertaken at all. Further, is the implied permission we have from students when we grade their papers and exams fully within the informed consent philosophy in the first place?

A final defense is that unobtrusive methods vary in their degree of ethical concern, and may at times be criticized more because of their reputation than because of any actual ethical transgression (Page, 2000). The broader need for valid information about what and how college students are learning must be weighed against the ethical inelegance of, say, observing the formation of student friendship patterns. We maintain that as long as the observation does not harm, embarrass, or otherwise disadvantage students or faculty, the question of ethics is a minor one.

In fact, Webb et al. (1981) have concluded that the right to privacy falls on a continuum, from observing public behavior of public figures to spying on private behavior in private places. To them, one extreme in this continuum is not an invasion of privacy while the other extreme is. It is the area in the middle of the continuum, of course, that provokes the most controversy. A key question for Webb and his associates was whether the people being observed clearly expect that their behavior will be un-regarded. For example, use of information gained through an overheard conversation in a campus restroom admitting to plagiarism on a paper assignment might be considered an invasion of privacy because people enter restrooms with the expectation "that their behavior will be studiously ignored" even though they are in a "public" place (p. 147).

Conclusion

Greater use of unobtrusive measures in classroom assessment would enable college and university faculty and administrators to move beyond sole dependence on reactive information and in a cost-effective way significantly enhance their understanding about how college affects students.

Our suggestion is to incorporate unobtrusive measurement in student learning assessment as both a preliminary cue to follow-up warranted by more reactive measurement such as surveys and rating scales, as well as a post-hoc confirmation of tentative conclusions made from the results of more reactive measurement. We must use all available options for taking seriously how well and about what students are learning in our courses, even if this means using a full "package" of measurement tools that counterbalance the measurement error inherent in each.

References

- Adler, R.M. & Sedlacek, W.E. (1988). Nonreactive measures in student affairs research. *Journal of College Student Development, 29*, 158-162.
- American College Personnel Association (ACPA). (1994). *The student learning imperative: Implications for student affairs*. Washington, DC: Author.
- Anderson, L.S., Schroeder, T., & Anderson, D.A. (2001). The use of portfolio advising with recreation and leisure services majors. *Scholar: A Journal of Leisure Studies and Recreation Education, 16*, 107-123.
- Brault, R. & Reidy, J. (2004). Travel education: Demonstrating effective learning through the seven principles of good practice in undergraduate education. *Scholar: A Journal of Leisure Studies and Recreation Education, 19*, 11-26.
- Briggs, C.M. (1997). Understanding the way students work: Unobtrusive measures and the effect of effort on performance. <http://hsb.baylor.edu/ramsover/ais.ac.97/papers/briggs2.htm>; retrieved 9/10/01.
- Campbell, D.T. & Stanley, J.S. (1963). *Experimental and quasi-experimental designs for research*. Skokie, IL: Rand McNally.
- DeGraaf, D.G. & Jordan, D.J. (1996). The use of a portfolio system as a pedagogical tool. *Scholar: A Journal of Leisure Studies and Recreation Education, 11*, 37-45.
- Dupuis, S.L. (2002). Intergenerational education programs in leisure and aging courses: Older adult and student experiences. *Scholar: A Journal of Leisure Studies and Recreation Education, 17*, 73-86.

Dyment, J.E. & O'Connell, T.S. (2003). "Journal writing is something we have to learn on our own" – The results of a focus group discussion with recreation students. *Scholar: A Journal of Leisure Studies and Recreation Education*, 18, 23-37.

Gladwell, N.J. Dowd, D.A. & Benzaquin, K.O. (1995). The use of mentoring to enhance the academic experience. *Scholar: A Journal of Leisure Studies and Recreation Education*, 10, 57-65.

Goodwin, W.L. & Goodwin, L.D. (1989). The use of nonreactive measures with preschoolers. *Early Child Development and Care*, 41, 173-194.

Green, J.E., Prather, J.E. & Sturgeon, J. (1983). Using administrative data as unobtrusive indicators of teaching performance. Paper presented at the Annual Forum of the Association of Institutional Research, Toronto, Ontario.

Hopp, R., Morden, P. & Ostiguy, L. (2004). Creating an environment for online courses in recreation and leisure studies programs: A case study. *Scholar: A Journal of Leisure Studies and Recreation Education*, 19, 73-82.

Koester, A.W. (1989). Learning from state fair exhibits. *Journal of Extension*, Spring, 23, 32.

McLean, D.D. & Hill, J.M. (1993). Supporting internship preparation: A Case Study in computer-based support. *Scholar: A Journal of Leisure Studies and Recreation Education*, 8, 37-49.

Orne, M.T. (1962). On the social psychology of the experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, 17, 776-783.

Page, S. (2000). Community research: The lost art of unobtrusive methods. *Journal of Applied Social Psychology*, 30:10, 2126-2136.

Pullen, R., Ortloff, V., Casey, S., & Payne, J.B. (2000). Analysis of academic misconduct using unobtrusive research: A study of discarded cheat sheets. *College Student Journal*, 34:4, 616.

Ross, C. (1998). Importance-performance analysis: A model for faculty and course evaluations. *Scholar: A Journal of Leisure Studies and Recreation Education*, 13, 1-19.

Russell, R.V. & Kovacs, A. (2003). Unobtrusive measures for college campus inquiry. In F.K. Stage & K. Manning (Ed.), *Fact-Finding, Knowledge Generation, and Action Research: Inquiry on the College Campus*. Sage.

Scholl, J. & Lago, P. (October 1994). Using non-reactive methods to study and improve 4-H programs. *Journal of Extension*, 32:3, <http://www.joe.org/joe/1994october/tt3.html>, retrieved 5/9/05.

Terenzini, P.T. (1989). Assessment with open eyes: Pitfalls in studying student outcomes. *Journal of Higher Education*, 60:6, 644-664.

Upcraft, M.L. (1999). Assessing student learning. In: F.K. Stage, L.W. Watson, & M. Terrell (Eds.), *Enhancing Student Learning: Setting the Campus Context*. Lanham, MD: The American College Personnel Associations.

Watson, L.W. & Stage, F.K. (1999). A framework to enhance student learning. In F.K. Stage, L.W. Watson, & M. Terrell (Eds.), *Enhancing Student Learning: Setting the Campus Context*. Lanham, MD: The American College Personnel Association.

Webb, E.K., Campbell, D.T., Schwartz, R.D., & Sechrest, L. (1966). *Unobtrusive measures: Nonreactive research in the social sciences*. Chicago: Rand McNally.

Webb, E.K., Campbell, D.T., Schwartz, R.D., Sechrest, L., & Grove, J.B. (1981). *Nonreactive measures in the social sciences*. Boston: Houghton Mifflin.

Webb, E.K., Campbell, D.T., Schwartz, R.D. & Sechrest, L. (2000). *Unobtrusive measures*. Thousand Oaks, CA: Sage.