

College of Saint Benedict and Saint John's University

DigitalCommons@CSB/SJU

OARCA Staff Publications

Office of Academic Review and Curricular
Advancement

10-2006

Assessment and the fear of punishment: how the protection of anonymity positively influenced the design and outcomes of postsecondary assessment

Philip I. Kramer

College of Saint Benedict/Saint John's University, pkramer@csbsju.edu

Follow this and additional works at: https://digitalcommons.csbsju.edu/oarca_pubs



Part of the [Educational Assessment, Evaluation, and Research Commons](#)

Recommended Citation

Kramer, P. I. (2006). Assessment and the Fear of Punishment: How the Protection of Anonymity Positively Influenced the Design and Outcomes of Postsecondary Assessment. *Assessment & Evaluation In Higher Education*, 31(5), 597-609.

This Article is brought to you for free and open access by DigitalCommons@CSB/SJU. It has been accepted for inclusion in OARCA Staff Publications by an authorized administrator of DigitalCommons@CSB/SJU. For more information, please contact digitalcommons@csbsju.edu.

Assessment and the fear of punishment: how the protection of anonymity positively influenced the design and outcomes of postsecondary assessment

Philip I. Kramer*

College of Saint Benedict and Saint John's University, Minnesota, USA

This study examines how a state system of higher education, located in the Rocky Mountains area of the United States, allowed faculty members from the state's public institutions of higher education to control the design and implementation of a pilot assessment of statewide general education courses. One of the primary reasons faculty were allowed to control the assessment process was to assuage the commonly held fear by faculty, academic departments, and institutions that they could be punished for the poor assessment results of their students who took the pilot assessment. An important conclusion of this study is that the promise of anonymity, given by the state board of regents (and tacitly by the legislature) to individual faculty members, academic departments, and institutions, was essential to the success of the assessment process.

Introduction

This study examines how faculty, academic departments, and institutions in one US state system of education, fearful of being punished by the state's board of regents and the legislature for the potentially poor performance of their respective students, planned, designed, and conducted a large-scale pilot assessment of general education¹ curricula. Furthermore, this study explores how the promise of anonymity made to faculty, departments, and institutions was key to the success of the pilot assessment.

A literature of different and often competing goals

Higher education assessment stakeholders often have different and even competing goals. Legislators, professors, students, parents, employers and others may be in

*College of Saint Benedict and Saint John's University, St. Joseph, MN, 56374, USA.
Email: philipikramer@gmail.com

agreement about some assessment goals (e.g., increasing student knowledge; demonstrating student acquisition of knowledge) but in disagreement about other assessment goals (e.g., use of standardized, norm-referenced assessment instruments; determining responsibility for teaching and learning). During the past few decades in America, some authors have discussed the diverse purposes of postsecondary assessment and the tensions inherent in planning, designing, and conducting it.

Banta *et al.* (1993), for example, have described the ongoing tension between those supporting assessment to increase accountability and those supporting assessment to improve teaching and learning. According to Banta *et al.*,

There is tension between the two principle purposes of assessment: demonstrating accountability and improving curricula, instruction, and student services. External constituencies impel us to undertake assessment for the former purpose; but to the members of the academy, only the latter purpose can provide sufficient intellectual justification for the investment of their time and energy in new methods of measuring student learning and satisfaction. (p. 359)

The tension surrounding the duality of assessment purpose has led some US faculty members, at times, to become resistant to the assessment process. Faculty members may be skeptical and resistant to the use of assessment when assessment is used to increase the public accountability of higher education institutions and systems (Nettles *et al.*, 1997; Nettles & Cole, 1999) rather than as explicit vehicle to demonstrate that teaching and learning have occurred. Nettles and Cole (1999) found ‘resistance [to assessment] at the institutional and/or faculty levels hindered the effectiveness of states’ assessment policies’ (p. 11). This finding, according to Nettles and Cole, is ‘a rather provocative indictment of the faculty that merits further consideration and a chance for the faculty to respond to the criticism’ (p. 12).

As economic pressures (e.g., reduced state budget expenditures) and questions about the worth (e.g., the notion that a university degree is a personal and not a public good) of a postsecondary education in America continue, external higher education stakeholders will ‘redefine relationships by pressuring institutions to become more accountable, more efficient, and more productive in the use of publicly generated resources’ (Alexander, 2000. p. 411). Clearly, this is evident today as increasing numbers of legislators, governors, members of boards of regents, and other postsecondary stakeholders in the United States question whether the public had received the greatest return for its fiscal investment in public higher education (Ewell, 1994).

It seems useful, in light of the tensions (both purposeful and economic) affecting US assessment stakeholders, to briefly remind ourselves of the advice some scholars have offered to those planning and conducting assessment:

- Include not only those who have assessment expertise and academic leaders but also those internal and external stakeholders who affect or are affected by assessment (Sell, 1989; Palomba & Banta, 1999; Delandshere, 2001). Others (e.g., Warren, 1988; Ratcliff, 1994; Appelbaum, 1992; Palomba & Banta, 1999) have specifically mentioned the importance of including students in the design and implementation of assessment.

- Establish and agree on the goals of assessment (Thomas, 1991). This is vital to ensure clarity about what assessment is expected to measure, how it will be measured, and what the assessment results are expected to measure. It is also important to remember that different audiences have different values and standards (Thomas, 1991). Similarly, giving thought to these differences when reporting the results and using caution when reporting the results (McMillan, 1988; Stark, 1990) are advised.
- Consider the social context of the assessment (Delandshere, 2001) and embed the assessment in the institutional culture (Banta *et al.*, 1996) being assessed.
- Reflect upon who will receive the results and how those results will be communicated (Erwin, 1991).
- Assessment can be a political process (Erwin, 1991) where both the political purpose and the political effects of assessment must be considered (Fong, 1988).
- During implementation, it is important not only to monitor the assessment (Banta, 1991) but also to listen carefully to feedback from the stakeholders (Mentkowski, 1991).

Study purpose and rationale

The purpose of this study was to examine how the protection of anonymity, given to faculty, academic departments, and institutions—fearful of being punished for the poor assessment results of their students—positively influenced the design, organization, and implementation of a pilot assessment of four common general education courses in a state system of higher education in a western US state. The findings of the study may inform the discussion about the importance of the political processes inherent in designing, organizing, and implementing postsecondary assessment. Specifically, the study may help assessment practitioners understand the influence inter-stakeholder relationships have on the assessment process, address the perception that faculty are resistant to assessment in higher education, and provide assessment stakeholders additional tools to plan and implement less threatening and more successful postsecondary assessment.

Methods

Research question

How did the fear of punishment for poor assessment results and the protection of anonymity influence the design, organization, and implementation of the pilot assessment of postsecondary general education courses?

Procedures

This study was conducted in 2002 in a western Rocky Mountain state in the United States. Although this study used a qualitative research design in which in-depth

interviews were the primary source of data, the extant² quantitative results from the general education pilot assessment, collected and analyzed by this researcher in 2001, are also presented.

Participant observation

The researcher was a participant observer during this study. The researcher was a member of the faculty task force that helped plan, design, and implement the pilot assessment. As such, the researcher was the primary 'measuring instrument' (Gall *et al.*, 1996, p. 554).

Participants

The researcher used purposeful sampling to select participants. In purposeful samples, participants are selected who have intimate knowledge of the phenomenon studied. In this case, participants were selected because of their knowledge of or expertise in the pilot assessment project.

Seventeen individuals were invited to participate in the study. Three individuals declined to participate. Those who declined said they were too busy to participate. The 14 participants chosen for this study were 'elite participants' (Kvale, 1996; Marshall & Rossman, 1999). Elite participants, according to Marshall and Rossman, are 'those considered to be influential, prominent, and/or well-informed people in an organization or community; they are selected for interviews on the basis of their expertise in areas relevant to the research' (1999, p. 113). Five participants were tenured task faculty force members representing a community college, a research university, a metropolitan state university, and a metropolitan state college; two were members of the office of the commissioners of higher education; three were members of the state board of regents; two participants were legislators (one Democrat and one Republican) and two were chief academic officers of institutions (representing a community college and a research university).

Data collection

In-depth interviews were chosen for this study to provide a highly descriptive and rich understanding of the process of planning, designing, and implementing the pilot assessment test. An interview guide (Gall *et al.*, 1996, p. 313), specifying the interview questions, the order in which they were asked, and the opening and the closing remarks, was developed. Care was taken to ensure that the interview questions were structured so that they were open-ended, neutral, singular, and clear (Patton, 1990).

The interview format included (a) an opening statement detailing the purposes of this study and a researcher introduction statement, (b) human participant protection statements, (c) the interview questions themselves, and (d) closing remarks statements. Interview data were collected by note taking and tape-recording. Interviews were transcribed by a paid transcriptionist and the researcher.

Interviews were conducted from late July 2002 through early September 2002. Twelve interviews were conducted in person. The other two interviews were conducted via telephone. The schedules of the two telephone participants were such that only telephone interviews were possible. Individual interviews ranged in time from 25 minutes to more than 2 hours in length.

Data analysis

The interview transcripts were coded into descriptive codes, interpretative codes, and pattern codes (Miles & Huberman, 1994). Descriptive codes ascribe ‘a class of phenomena to a segment of text’ (Miles & Huberman, 1994, p. 57). Interpretative codes added interpretation and insight to a segment of text. Pattern codes are ‘even more inferential and explanatory’ and add ‘an emergent leitmotiv or pattern that you have discerned in local events and relationships’ (Miles & Huberman, 1994, p. 57).

Coding was an iterative process. A list of codes was developed before fieldwork was begun as recommended by Miles and Huberman (1994); the goal was to conceptually link the research questions to the data. Codes were revised numerous times as the collection of data progressed from the first interview to the last interview and throughout the analysis process.

To increase the confirmability of the study and to prevent the possibility of making incorrect generalizations, the researcher purposely selected subgroups of participants organized by their respective roles relative to the assessment process, rather than only selecting participants with the same role (e.g., faculty members).

Bias reduction

To prevent researcher bias, the researcher (a) made sure his ‘intentions were unequivocal’ (Miles & Huberman, 1994, p. 266) for participants, (b) conducted the interviews in a congenial environment, (c) kept thinking conceptually, (d) avoided being co-opted by spreading out the site visits over more than a two-month period, (e) found a confidential informant—intimately familiar with the pilot assessment—who provided background information and confirmed my data analysis, and (f) frequently contacted an anonymous panel of experts to confirm my results and conclusions.

Findings

The assessment climate: demanding accountability and fearing punishment

In the late 1990s, influenced by the national trend demanding greater accountability in higher education, conservative leaders in the legislature of this state began questioning the academic efficacy and the economic merit of the state system’s rule requiring all undergraduates to complete general education courses at the state’s colleges and universities. The state board of regents, in response to the pressure from the state legislature to demonstrate greater accountability in the state’s public higher education

system in general and the worth of general education in particular, ordered the administration of a nationally normed, standardized commercial general education assessment.

The nationally normed commercial assessment instrument was administered to more than 3100 students completing their sophomore year at one of state's public colleges or universities. The results of the assessment were disaggregated. The assessment results, disaggregated by the company who administered the assessment, allowed the state's assessment stakeholders to compare the student results by institution, academic department, and individual faculty member. Although the assessment results indicated that students generally scored above the national average, the assessment was considered a disappointment by faculty statewide for two reasons.

First, the faculty said the commercial assessment was (a) not aligned with course content or objectives, (b) did not correlate with students' grade point averages, (c) was too general, (d) failed to offer any guidance to faculty, and (e) was not tied to any specific courses or degree requirements. Faculty felt the administration of the assessment had been imposed upon them essentially by the legislature with very little faculty input. Faculty members believed they 'owned' the curriculum and were disappointed they had not been more involved in the selection of the commercial assessment instrument. Faculty believed they—better than any other stakeholder group—knew what curricula should be taught, what was actually being taught in the classroom, and how best to assess student learning.

Second, faculty were fearful the results of any assessment administration—without the direct involvement of faculty representatives—could be used to compare individual faculty members, academic departments, and even institutions. Faculty and institutional leaders perceived the disaggregated results would be used by legislators as a whip to control the general education curricula that faculty had worked so hard to articulate.

Faculty throughout the state, already participating in collaborative discussion regarding the goals of general education courses, asked the regents for permission to create their own assessment instrument to measure the general education knowledge of their students. In many ways, the faculty felt the need to demonstrate the worth of their collective general education courses by specifically aligning an assessment with what was taught in general education classrooms throughout the state. The regents and state legislators realized using a nationally normed assessment (or any commercial assessment instrument, for that matter) was not possible; politically, it was not worth the effort. The regents agreed and authorized statewide faculty to create and administer course relevant, content-embedded assessment instruments of the state's general education curricula.

Designing the assessment

An ad hoc general education faculty task force, originally appointed by the chief academic officers of the nine state institutions to examine statewide articulation of courses, asked the board of regents to allow faculty members to create their own

assessment instruments to measure the general education knowledge of their students. The regents agreed with the task force and authorized it to create and administer course relevant, content-embedded, value-added assessment instruments. Four common courses in the statewide general education curriculum were selected for the pilot assessment: college algebra, the economic history of the US, the civilization of the US, and the government of the US.

Participants from the higher education commissioner's office and the faculty said the decision was made at both the commissioner's office and the task force level to design a systemwide assessment test so that only systemwide results would be reported. These participants said the reasoning behind reporting systemwide results rather than departmental or institutional results was to maintain anonymity of the institutional and individual faculty participants. Faculty, the chief academic officers, and members of the commissioner's office made it very clear that they did not want institutions, academic departments, or individual faculty members compared or to compete with each other. Faculty were especially fearful that those faculty members teaching the courses would be assessed would be punished for poor student results.

The faculty, led by the ad hoc general education task force, succeeded in thwarting the possibility that institutions, academic departments, or individual faculty members could be punished for the assessment results of their students. Thus, any invidious comparisons or concerns about sanctions for low scores were intentionally 'designed out' of the pilot assessment; the anonymity (and arguably, the responsibility) of the faculty and the institutions was guaranteed.

The creation of a systemwide assessment with individual, departmental, and institutional anonymity was realized not only to prevent the possibility of punishment but also to engender greater participation among institutions and faculty from throughout the state. The logic was that by ensuring anonymity among all participants, institutions and individuals would be more inclined to participate in the assessment process. Faculty argued that the way to determine whether student learning was occurring was to create an assessment environment where faculty participation and cooperation in the assessment process was high. The only way to do this, according to the faculty, was hold the faculty, academic departments, and institutions harmless for the results. 'The motive for anonymity', said one faculty participant, 'was neither cowardice nor modesty, but rather a desire that the assessment experiment work'. There was great concern by academic leaders and the faculty that without the promise of anonymity, faculty would not volunteer to participate, faculty and institutions would argue over which questions to use in the assessment, and quite possibly, the monumental and valuable effort at created course relevant, content-embedded assessment would fail.

Not all of the participants, however, agreed with the decision to allow the faculty to design, implement, and conduct the pilot assessment or to report only systemwide results. One legislator participant, for example, had serious concerns about not reporting the scores of the students of individual institutions and professors. This state legislator was skeptical that faculty could design and implement their own assessments. The legislator preferred to have the state's assessment of teaching and learning designed and controlled by outside forces, like a consultant.

One member of the board of regents, acknowledging the existence of academic resistance to assessment and accountability, said there was a variety of impacted constituencies who did not always share a common perspective on assessment. This regent made the point that it was important to bring all assessment constituents to the table when planning and conducting assessment. Failing to do that, she said, meant ‘the mission of assessment is going to encounter roadblocks and resistance’.

Results of the pilot assessment

Data were collected and reported from all nine public institutions of higher education in the four content areas. The task force reported the aggregated percentage of improvement results for the four content areas. Individual student and institutional results were calculated but not reported in accordance with the agreement to maintain anonymity. Improvement ratios were calculated using the following formula, Percentage of Improvement = ((post-test - pre-test) ÷ (pre-test)). The results follow:

- College algebra: results were calculated from 699 general education student pretest and posttest assessments. The data analysis indicated that each (unreported institution) showed statistically significant student improvement from pretest to posttest with all t tests highly significant ($t = 8.15$ to 17.8 , $p < .001$). College algebra students averaged individual improvement ratios of 169% across all students in all institutions; collectively then, students more than doubled (and nearly tripled) their scores.
- Economic history of the US: results were calculated from 164 general education student pretest and posttest assessments. The data analysis showed that only two of the institutions reporting results showed statistically significant student improvement from pretest to posttest ($t = 2.76$, $p < .05$ and $t = 7.21$, $p < .0001$). Individual student improvement ratios averaged 68% in these two institutions.
- Civilization of the US: results were calculated from 1,207 general education student pretest and posttest assessments. The results were remarkably consistent across institutions. The data analysis indicated statistically significant improvement from pretest to posttest ($t = 7.10$ to 15.65 , $p < .0001$ for all). Across the institutions, the mean pretest score was 12.91 ($n = 514$); the mean posttest score was 17.69 ($n = 500$). Individual student improvement ratios, calculated as indicated above, averaged 36%.
- Government of the US: only one institution provided complete data for the government course, involving only 71 students. The data analysis indicated statistically significant improvement from pretest to posttest ($t = 3.59$, $p < .01$). Individual student improvement ratios averaged 62%.

Discussion

Limitations of the study

The scope of the study was limited by choosing (a) to only examine general education assessment in this Rocky Mountain state, (b) to study the process of creating the pilot

assessment after the standardized national assessment instrument and results were rejected by statewide faculty and the board of regents and a decision was made to create a 'homegrown,' content-embedded general education assessment instrument, (c) to only interview some of the participants who participated in the general education assessment process, and (d) to only interview elite participants.

A faculty driven pilot assessment: designed to avoid punishment

The assessment process was driven and developed by faculty in the specific disciplines. The faculty discussed issues of test items and whether tests would be identical across institutions. Data reports provided statewide results without any specific faculty or institutional information. The faculty in the four disciplinary groups were in general agreement that (a) test items were best developed within a respective discipline, (b) identical tests across institutions were preferred by some but not all groups, (c) identical tests were to be administered on a pretest at the beginning of the term and a posttest at the end of the term, and (d) anonymity of faculty and institution was essential for the success of the pilot assessment.

The primary reason for designing, organizing, implementing a systemwide assessment and only reporting aggregated systemwide results was to avoid the possibility of punishing individual faculty members, academic departments, or individual institutions for poor student results. Faculty members and institutional leaders demanded and received anonymity. With the promise of anonymity and the concern alleviated about institutional, departmental, and individual competition and comparison over, the assessment planners were able to receive strong faculty support and participation.

It was clear faculty greatly influenced the assessment process. Because of the strong influence of the faculty over the planning, designing, administering, and reporting of the pilot assessment, faculty were in a position to shape not only the process but also the outcomes of the assessment and how those outcomes were reported. Of course, neither the task force nor the faculty in general had total control in planning, designing, or implementing the pilot assessment. The commissioner's office and the board of regents had considerable influence in the pilot assessment process: they also had considerable trust in the efforts of the task force and individual faculty members to plan, design, and implement a meaningful measure of student learning.

Concerns about responsibility

For the most part, the faculty, administrators, and members of the commissioner's office thought the design of the assessment was a positive experience. They believed that it allowed for a risk free exploration of the teaching and learning process. These participants talked about the importance of linking the goals of the curriculum (and general education, in particular) to assessment. Faculty were keenly aware of the importance of accountability and emphasized the importance of assessment in articulation, accreditation, and curricular coherence. Yet, in spite of the agreement by faculty that accountability was a major reason to conduct assessment, faculty

expressed concern about the harmful effects assessment efforts could have on the faculty if the faculty were not placed in charge of future assessment efforts.

Other stakeholders, however, were concerned about the high degree of faculty control over the process and the fact that faculty, the institutions, and students were not as accountable as they could have been had a test been administered that provided for disaggregated assessment results. Some stakeholders were concerned that not all assessment stakeholders (including students) participated in the planning, designing, and implementation of the pilot assessment.

Because this was a pilot, the decision was made by those planning the assessment (including faculty, members of the board of regents, and members of the office of the commissioner of higher education) to not hold students accountable for their respective assessment results. In effect, students were 'let off the hook' during this assessment process just as the faculty had been off the hook. Arguably, because it was a pilot, both the internal and external stakeholders were more concerned with process than outcomes—the 'system' tested whether it could conduct curricula relevant, course-embedded assessment. In this sense, the pilot was a success. The merit of this success, however, must be questioned in light of the fact that all student, faculty, departmental, and institutional responsibility was engineered out of the process.

Many (e.g., Terenzini, 1989; Gaff, 1997; Angelo, 1999) have written about the distinctions and the conflicts between accountability and improving teaching and learning. Others (e.g., Palomba & Banta, 1999; Schneider & Shoenberg, 1999), although acknowledging the tensions and distinctions between accountability and improving teaching and learning (Banta *et al.*, 1993), have suggested that accountability must be linked to the improvement of teaching and learning.

This study has illustrated that there was misunderstanding and distrust between some higher education assessment stakeholders. Some members of the legislature did not (and presumably, still do not) trust faculty to monitor and evaluate themselves. Although the demonstration of system accountability was very important to most stakeholders, including faculty, improving teaching and learning was arguably the most important task for the majority of assessment stakeholders.

Faculty teach so that students may learn. Faculty do not necessarily teach to prove that they are teaching effectively (this, however, is a required consequence of teaching). Of course, the demand for proof that faculty are teaching effectively and students are learning effectively will not subside anytime soon. Accountability, like accreditation, is, for better or worse, here to stay. Accountability systems can be designed to improve all teaching and learning, including teaching and learning that ranges from exemplary to poor.

Recommendations

The politics of assessment leadership is an area that merits further research. It would be useful for practitioners, researchers, and policy-makers to learn more about the influence various stakeholder groups have in planning, designing, implementing, and reporting assessment, how that influence is exercised and shared, and to what extent,

if any, having influence over the assessment process influences inter-stakeholder relationships and assessment outcomes.

In this study, the influence of the assessment process had major implications for what was assessed and how it was assessed. Because of their influence, the faculty were able to effectively design and implement a pilot assessment and the positive, albeit, aggregated report of the results. Research examining the influence different stakeholders have (or do not have) in the assessment process may be helpful in understanding how political processes and group dynamics influence assessment practices and results. It would be interesting to know, for example, whether a relationship can be established between influence, process, and outcomes.

Being *punished* for poor assessment results is not the same as being *responsible* for poor assessment results. As least in this study, faculty and institutional concern about being punished for poor student assessment results was the main reason for not reporting student assessment results assigned to institutions, academic departments, or individual faculty members. It would be interesting to know if other research could establish a relationship between inter-stakeholder distrust and the fear of punishment and what that relationship looked like. Specifically, would faculty, departments, and institutions be less concerned about the possibility of punishment if inter-stakeholder trust increased?

While the indictment of faculty (Nettles & Cole, 1999) may be justifiable, one must not forget that the resistance by faculty and academic leaders toward assessment was essentially eliminated when the promise of anonymity was guaranteed. One wonders, as external demands for assessment and accountability in higher education presumably increases, whether the concern about punishment by faculty and others will also increase or whether further inter-stakeholder conversations about assessment can lead to mutual trust and common purpose.

Note on contributor

Philip I. Kramer is the Director of Academic Assessment, and an Assistant Professor in the Education Department, at the College of Saint Benedict and Saint John's University in Minnesota.

Notes

1. General education courses are courses at the states' institutions that satisfy the freshman and sophomore requirements for fulfillment of the associate degrees, bachelor degrees, and other degree, program, or certificate requirements.
2. Pretest assessments were administered early in the 2001 spring semester. Posttest assessments were administered at the end of the same semester.

References

- Appelbaum, M. (1992) Assessment through the major, in: C. Adelman (Ed.) *Performance and judgment: essays on principles and practice in the assessment of college student learning* (Washington,

- DC, US Department of Education: Office of Educational Research and Improvement, Office of Research), 117–138.
- Alexander, F. K. (2000) The changing face of accountability: monitoring and assessing institutional performance in higher education, *The Journal of Higher Education*, 71(4), 411–431.
- Angelo, T. A. (1999) *Doing assessment as if learning matters most*. Available online at: <http://www.aahe.org/bulletin/angelomay99.htm> (accessed 10 January 2002).
- Banta, T. W. (1991) Faculty-developed approaches to assessing general education outcomes, *Assessment Update*, 3, 1–4.
- Banta, T. W. & Associates (1993) *Making a difference: outcomes of a decade of assessment in higher education* (San Francisco, CA, Jossey-Bass).
- Banta, T. W., Lund, J. P., Black, K. E. & Oblander, F. W. (Eds) (1996) *Assessment in practice: putting principles to work on college campuses* (San Francisco, CA, Jossey-Bass).
- Delandshere, G. (2001) Implicit theories, unexamined assumptions and the status quo of educational assessment, *Assessment in Education*, 8(2), 113–133.
- Erwin, T. D. (1991) *Assessing student learning and development: a guide to the principles, goals, and methods of determining college outcomes* (San Francisco, CA, Jossey-Bass).
- Ewell, P. T. (1994) Statewide testing in higher education, *Change*, 33(2), 20–27.
- Fong, B. (1988) Assessing the department major, in: J. H. McMillan (Ed.) *Assessing students' learning—new directions for teaching and learning* (San Francisco, CA, Jossey-Bass).
- Gaff, J. G. (1997) Evaluating learning in individual courses, in: J. G. Gaff & J. L. Ratcliff (Eds) *Handbook of the undergraduate curriculum: a comprehensive guide to purposes, structures, practices, and change* (1st edn) (San Francisco, CA, Jossey-Bass), 591–607.
- Gall, M. D., Borg, W. R. & Gall, J. P. (1996) *Educational research: an introduction* (6th edn) (White Plains, NY, Longman Publishers).
- Kvale, S. (1996) *Interviews: an introduction to qualitative research interviewing* (Thousand Oaks, CA, Sage Publishers).
- Marshall, C. & Rossman, G. B. (1999) *Designing qualitative research* (3rd edn) (Thousand Oaks, CA, Sage Publications).
- McMillan, J. H. (1988) Beyond value-added education: improvement alone is not enough, *Journal of Higher Education*, 59(5), 564–579.
- Mentkowski, M. (1991) Creating a context where institutional assessment yields educational improvement, *The Journal of General Education*, 40, 35.
- Miles, M. B. & Huberman, A. M. (1994) *Qualitative data analysis: an expanded sourcebook* (2nd edn) (Thousand Oaks, CA, Sage Publishers).
- Nettles, M. T. & Cole, J. J. K. (1999) *State higher education assessment policy: research findings from second and third years* (Palo Alto, CA, National Center for Postsecondary Improvement).
- Nettles, M. T., Cole, J. J. K. & Sharp, S. (1997) *Assessment of teaching and learning in higher education and public accountability* (Palo Alto, CA, National Center for Postsecondary Improvement).
- Palomba, C. A. & Banta, T. W. (1999) *Assessment essentials: planning, implementing, and improving assessment in higher education* (San Francisco, CA, Jossey-Bass).
- Patton, M. Q. (1990) *Qualitative evaluation and research methods* (2nd edn) (Newbury Park, CA, Sage Publishers).
- Ratcliff, J. L. (1994) Assessment's role in strengthening the core curriculum, in: D. F. Halpern & Associates (Eds) *Changing college classrooms: new teaching and learning strategies for an increasingly complex world* (San Francisco, CA, Jossey-Bass).
- Sell, G. R. (1989) An organizational perspective for the effective practice of assessment, in: P. J. Gray (Ed.) *Achieving assessment goals using evaluation techniques* (vol. 67) (San Francisco, CA, Jossey-Bass), 21–41.
- Schneider, C. G. & Shoenberg, R. (1999) Habits hard to break: how persistent features of campus life frustrate curricular reform, *Change*, 31(2), 30–36.
- Stark, J. (1990) Approaches to assessing educational outcomes, *The Journal of Health Administration Education*, 34, 96.

- Terenzini, P. T. (1989) Assessment with open eyes: pitfalls in studying student outcomes, *Journal of Higher Education*, 60(6), 644–664.
- Thomas, A. (1991) Consideration of the resources needed in an assessment program, *North Central Association of Colleges and Schools Quarterly*, 66(2), 23–46.
- Warren, J. (1988) Cognitive measures in assessing learning, in: T. W. Banta (Ed.) *Implementing outcomes assessment: promise and perils—new directions for institutional research* (San Francisco, CA, Jossey-Bass).

