

ASSESSMENT OF THE FIRST-YEAR EXPERIENCE: SIX SIGNIFICANT QUESTIONS

Joseph B. Cuseo

Introduction

Karl Schilling (2000) suggested that effective assessment involves asking the right questions. His suggestion reminded me of an old poetic adage penned by Rudyard Kipling: “I keep six honest serving men. They taught me all I know. Their names are what & why and when & where and how & who.” These half-dozen queries can serve as a useful framework for (a) organizing assessment practices in the first year, and (b) stimulating focused, campus-wide discussion of first-year assessment.

While not all of Kipling’s questions have profound theoretical significance, they all have *practical* significance, and they may be adapted to guide the quest for an effective, comprehensive first-year assessment plan. With this purpose in mind, I raise and attempt to provide answers to the following six questions in this manuscript:

- 1. Why?** (What is the *purpose* of assessment?)
- 2. What?** (What are the intended *outcomes* of assessment? What is the intended *target* [*unit or level*] of assessment?)
- 3. When?** (At what *times* in the college experience should assessment be conducted?)
- 4. Who?** (Who will be the *assessor[s]*? *From whom* will assessment data be collected?)
- 5. Where?** (Where will assessment data be *found*? In what locales will data be *collected*?)
- 6. How?** (How will assessment data be *analyzed and summarized*? What *research designs* and *methodologies* will be employed?)

ASSESSMENT QUESTION #1. **WHY?**

This question refers to the *purpose* of assessment; it is the first question that needs to be addressed because its answer should guide and drive decisions relating to all other assessment questions. As Upcraft and Schuh (1996) note, “Perhaps the most important of assessment principles is contained in the question, ‘Why are we doing this study?’” (p. 316).

When answering this question, the classic distinction between “formative” and “summative” assessment (Scriven, 1967) is an important one to keep in mind. If the purpose of assessment is *formative*, its goal is to obtain information that can be used as feedback to improve or fine-tune an existing program. In contrast, *summative* assessment is designed to “sum up” a program’s overall value or impact – for the purpose of making bottom-line decisions about its adoption, retention, or expansion. FYE assessment efforts may have to be more intentionally designed for summative purposes in order to generate the type of “value-added” evidence that will support their program adoption and survival. Supporting this argument is a conclusion reached by the Educational Commission of the States in 1995:

The first years of undergraduate study – particularly the freshman year – are critical to student success. Yet the pattern of resource allocation at most colleges and universities strongly favors upper-division work. Comprehensive efforts to integrate first-year students into the mainstream of collegiate experience often are treated as auxiliary experiences, just the reverse of what a growing body of research indicates as “best practice.” (p. 6)

More recently, Betsy Barefoot (2000) reached a similar conclusion in her review of trends relating to first-year programs.

A pervasive and central problem is that many of the programs and activities that constitute the ‘first-year experience’ are in a continuous battle for status within the academy... never becoming a central, sustainable part of the institution’s fabric. First-year programs often have a single champion rather than broad-based institutional support and frequently operate with a minimal budget or no budget. (p. 17)

This suggests that Trudy Banta’s FYA-list (5-16-00) request for “strong arguments to convince our skeptical colleagues” applies to many first-year initiatives, and engaging in assessment to generate these strong arguments should be a major goal of FYE assessment. Efforts to incorporate this goal should be made from the early stages of assessment design, and done with the following questions in mind: What assessment outcomes will command the most attention from key decision-makers who control resource allocations? What assessment methodologies would generate the most compelling evidence for the value-added impact of first-year programs? How should the results of assessment be packaged, presented, and delivered so as to maximize their persuasive power?

ASSESSMENT QUESTION #2. **WHAT?**

There are two components to this question: (a) What are the intended *outcomes* of assessment? and (b) What is the intended *target* (unit or level) of assessment? The first question relates directly to an important question that Randy Swing posed on the FYA-List, namely: “What are the most important outcomes which should be a core part of our assessment efforts?” (5-22-00). There is a broad range of important outcomes that assessment efforts can address. For simplicity’s sake, they may be divided into two general categories: *Institution-Centered Outcomes*, and *Student-Centered Outcomes*. Institution-Centered Outcomes include those outcomes that have more direct benefit to the college than to students, such as program-generated revenue, program efficiency or cost-effectiveness, and program impact on promoting professional partnerships and campus community among faculty, staff, and administrators. *Student-centered outcomes* encompass outcomes that pertain more directly to student learning and development. Examples include (a) *academic* outcomes, such as academic skill development, academic performance, student persistence to academic program or degree completion, time taken by students to complete program or degree requirements, and student advancement; and (b) *personal (holistic)* development outcomes, which include “non-academic” domains or arenas of development (e.g., social, emotional, ethical, physical, vocational).

Student-centered outcomes may be assessed in terms of three key types or dimensions of developmental change, sometimes referred to as the “ABCs” of outcome-assessment measures: (1) *attitudinal* (e.g., change in student opinions or values with respect to diversity), (2) *behavioral* (e.g., incidence or frequency with which students use support services), and (3) *cognitive* (e.g., gains in knowledge or critical thinking skills).

Unlike traditional academic program and administrative functions that are discipline-centered and office-driven, most FYE programs have emerged as part of a student-centered movement and has been deliberately designed with the intention of promoting student development and student success. This distinctive feature of FYE programs is reflected in the

following description of FYE initiatives provided by John Gardner during the movement's formative stages of development, "The freshman year experience efforts are manifested by their deliberateness, their effort to make things happen by design, not by accident or spontaneity, i.e., those things that must happen if students are more likely to be successful." (1986, p. 267). Since student-centered FYE programs have been deliberately designed and delivered with the explicit intention of promoting student development and success, their intended outcomes should be directly learner-centered and readily amenable to clear identification and subsequent validation.

With respect to the second "what" question for assessment (What is the targeted *unit* or *level* of assessment?), potential assessment targets could be said to fall on a continuum ranging from "micro" to "macro" levels/units, such as the following:

- (a) individual
- (b) course
- (c) department
- (d) program
- (e) institution

John Gardner has suggested that we look at the FYE as a "forest, the entire freshman year, as opposed to the individual trees." (FYA-List, 5-23-00). To assess the impact of the whole "forest," we may need to reach some consensus on what are the "trees" (or, at least, the most influential trees) that comprise the forest. As Astin (1991) suggests, there are many "*subenvironments*" within a college or university that can affect FYE outcomes. Attempting to identify the important subenvironments or component experiences that comprise the total FYE, would be the critical first step toward tackling the challenging task of assessing how these components combine to affect end-of-first-year outcomes.

I have identified seven general subenvironments, and there are undoubtedly additional sub-experiences nested within each one. These seven subenvironments are: New-Student Orientation; Academic Advising; Classroom Instruction; The Curriculum; The Cocurriculum; Academic Support Programs; and Psychosocial Support Programs. Each of these subenvironments may be assessed with respect to their *content*—what the program "covers," *structure*—how the program is organized, and *process*—the manner of program delivery.

My point here is not to complicate the process of FYE assessment by posing all these potential permutations. I acknowledge that the whole (FYE environment) may be greater than the sum of its parts (subenvironments), but we still need to identify all the important parts of the FYE, and how their individual effects add up (or multiply) to produce the total effect that is the FYE.

ASSESSMENT QUESTION #3. **WHEN?**

This question refers to the *timing* of assessment, or more accurately, the timing of assessments because it is clear that a one-time, outcomes-only assessment will not provide useful information for assessing program impact or making program improvements. As Terenzini and Upcraft (1996) advise, "It is not enough to know simply whether change occurs, we must know when it occurs and why it occurs. These requirements suggest the need for several collections throughout the college career of an entering cohort of students." (pp. 221-22).

As an initial stab at implementing this recommendation, I offer the following taxonomy of data collection points for consideration.

- Data-Collecting Point #1. *College Entry (Beginning of First Year)*

Data collection at this point can provide a baseline or pre-test needed for making meaningful comparisons with, and interpretations of, outcome data. The importance of entry data collection is underscored by Randy Swing in an earlier contribution to the FYE-List: “Baseline data on students contains key information which undergirds all other components of assessment...and the way we organize those data should not be understated.” (5-3-00).

It is probably safe to say that more data are collected on students at college entry than at any other time in the college experience (e.g., student admissions data, placement-test data, CIRP data). This college-entry data can provide a “baseline” that may be used in conjunction with outcome assessment to detect pre-to-post longitudinal change in student development that may take place between college entry and exit, otherwise known as “value-added” assessment. As Peter Ewell suggests, “so much depends on establishing a good set of data on baseline conditions from which to conduct an ongoing longitudinal study.” (FYE-List, 5-10-00).

Furthermore, since assessment is commonly defined as the systematic collection of information for the improvement of learning, college-entry assessment becomes valuable not only as a baseline for longitudinal outcomes assessment, but also as a vehicle for helping us know who our learners are and what they are like when they enter our doors. With this knowledge in hand, we can more intentionally and proactively design first-year programs to support student success.

- Data-Collecting Point #2. *End of the First Year*

John Gardner has suggested that we look at assessing the first year as a “unit.” This recommendation implies that, in addition to and in conjunction with college-entry assessment, another key data-collection point is the end of the first year – to assess outcomes of the total first-year experience. The relevance of Gardner’s recommendation is certainly supported by national research, which indicates that student attrition is highest during the first year of college, making it a “critical period” for retention-intervention efforts. Since much improvement still needs to be made in the area of first-year retention (Barefoot, 2000), this practice would be consistent with Ted Marchese’s call for us to prioritize our assessment efforts by focusing them on areas which “beg for improvement.” (1993, p.4)

Also, there is accumulating evidence that the first year of college may be a critical period for student learning and cognitive development (MacGregor, 1991). This suggests that the first-year experience may represent a “window of opportunity” for promoting learning that would be missed if we do not proactively “front-load” our best learning resources and educational interventions during this pivotal period of college development. Given this finding, comprehensive assessment of the FYE will likely be a high-yield endeavor, as it has the potential to shed light on two of our most important educational outcomes: student retention and student learning.

- Data-Collection Point #3. *End of Sophomore Year/Beginning of Junior Year*

Admittedly, this assessment point takes us beyond the first-year experience. However, I mention it because it may provide an important *intermediate* assessment point in the undergraduate experience. In particular, assessment at this midpoint in the college experience may serve the following useful functions:

- (a) Assessing the impact of general education (lower-division) course work by serving as point for re-administration of instruments that were initially administered at college

entry (e.g., ACT).

- (b) Assessing the transfer readiness of students who are transitioning from 2- to 4-year institutions.
- (c) Diagnosing the academic readiness of students who are about to begin upper-division course work in their major field of study.
- (d) Providing an entry-assessment point for upper-division coursework, which may be used as baseline data to compare with outcome data gathered at graduation, thereby providing information on the impact of specialized work in the student's specific academic major or field of pre-professional study.

- **Data-Collection Point #4. *End of Senior Year***

This brings us well beyond the first-year experience, but the senior year is an important transitional experience that represents the culmination of a college education. As such, it provides an opportunity for assessing the potential long-term impact of first-year programs by measuring their durability throughout the remaining years of college.

Predicting and assessing the long-term effects of first-year programming is consistent with the “utilization-focused evaluation” model posited by Patton (1978). According to this assessment model, student development is cumulative or hierarchical, so any educational program that increases the accomplishment of immediate and intermediate outcome goals (such as a first-year experience program) also has the potential for achieving ultimate goals.

This model is congruent with Alexander and Stark's contention that college education is an “iterative process with current outcomes for influencing future achievement.” (1986, p. 24). Thus, outcomes attained during the first year of college may, in turn, exert a positive “ripple effect” or “mediating effect” on the realization of additional positive outcomes that are manifested at college completion.

Historically, first-year program assessment has focused heavily on immediate or intermediate goals (e.g., retention through the first semester or first year). A useful direction for future assessments of first-year programs might be to explore their potential for building on these shorter-term accomplishments and to document their impact on broader, long-term student outcomes assessed at the end of the senior year. This would represent a powerful source of evidence for the durability of FYE-program impact.

ASSESSMENT QUESTION #4. WHO?

This question refers to the persons who may be involved in the assessment process; it embraces two component questions: (a) Who will be the assessor(s)? and (b) From whom will assessment data be collected?

Regarding the first question, an almost axiomatic answer is to have someone conduct the assessment who is not associated with the first-year program being evaluated, and who has no vested interest in its outcome—i.e., an “external” or “third-party” assessor who will neither be biased for or against the program. This practice guards against charges of “evaluator bias”—the tendency of the individual who designs or conducts the study to unwittingly skew the findings in the direction of its intended (hoped for) outcome. At the University of South Carolina, evaluation data that provided pivotal support for continuation of its University 101 (first-year seminar) program was conducted by a “highly respected researcher who has no official relationship to or responsibility for the program being studied” (Gardner, 1986, p. 271). This practice is recommended for the assessment of any first-year program, because it will serve to

enhance the credibility of its findings, as well as their potential for commanding institutional attention and promoting institutional change.

Turning to the more specific issue of what particular member(s) of the college community should conduct first-year program assessment, John Gardner has mentioned that he would like to see more institutional research professionals involved as FYE assessment agents. Additionally, there are several other individuals who are well positioned to assist in the collection, analysis, or interpretation of assessment data. First, faculty in the departments of education or social and behavioral sciences (as well as those in computer science, mathematics, and statistics) may be good candidates, because they have graduate training in research methodology and statistics. Second, student development professionals may have both the interest and ability to function as assessors, because of their graduate work in educational assessment. Finally, graduate students from relevant academic departments or from student development programs, as well as upper-division undergraduates, could assist in FYE assessment, perhaps serving as research assistants on faculty-student or graduate-undergraduate student research teams.

Turning now to the second “who” question (From whom will assessment data be collected?), all of the following individuals are potential sources: (a) students, (b) alumni, (c) faculty, (d) student development professionals, (e) administrators, (f) administrative staff/assistants, and (g) trustee members.

Naturally, since the FYE is a student-centered movement, students have been and are likely to continue to be the primary focus population for first-year assessment. However, the undergraduate student population is not a homogeneous entity, and the impact of any educational intervention may vary for different students subpopulations, resulting in “conditional effects” or “interactions” which may be masked if assessment data are simply aggregated and collapsed into a singular “average” measure. Thus, it is recommended that FYE assessment efforts examine the differential effects of its programs on various student subpopulations.

Assessment of non-student populations (e.g., faculty, administrations, staff) can allow for comparison between their responses and those of students (e.g., on surveys of institutional or program satisfaction). Significant discrepancies or “gaps” that happen to emerge between the responses of students and non-students often have the potential to create a sense of cognitive “disequilibrium” or “dissonance” among college decision-makers, could provide the impetus for initiating administrative action on behalf of first-year students.

A second reason for assessing faculty, staff, and administrators is that student-directed programs may indirectly promote positive outcomes for other members of the college community. For instance, instructional development of faculty may be promoted by a first-year seminar that is accompanied by a comprehensive instructor training-and-development program, which can increase faculty awareness and use of effective teaching strategies for promoting student involvement, learning, and retention. Offering such a faculty development program under the aegis of instructor training for the first-year seminar, and making it available to all faculty members who would like to experience it, may serve to stimulate campus-wide faculty dialogue and instructional improvement.

In addition, first-year programs may have the potential for indirectly promoting such systematic outcomes as the building of campus community and the establishment of professional partnerships across different divisions or units of the college. The community-building potential of a first-year seminar’s instructor training program is reinforced by John Gardner’s early report on the University 101 program at South Carolina. “The program integrates faculty and professional staff at the university in a joint undertaking [which] tends to reduce the barriers

between the faculty and staff on campus, reduces stereotyping...and has promoted better relationships between faculty and especially student affairs staff.” (1980, pp. 6&7). More recently, informal reports from institutions where both faculty and staff are involved in planning and teaching the course – such as Champlain College (VT) and Marymount College (CA) – indicate that such partnerships have enhanced communication and collaboration between the offices of academic and student affairs (Barefoot, 1993).

Perhaps future FYE-assessment initiatives can be more intentionally designed so that these campus-specific “anecdotes” are systematically documented and converted into either qualitative or quantitative “data.” If this can be done, it could provide the foundation for an eventual database of empirical evidence supporting the positive impact of FYE programs on systemic, college-wide outcomes. Such an accomplishment might enable FYE initiatives to meet the criterion of a truly “transformative” learning experience, as suggested in the following question posed by the American Association for Higher Education: “What separates an interesting but isolated educational experiment from a strategy with the potential to transform student learning throughout the institution?” (1998, p. 6).

ASSESSMENT QUESTION #5. **WHERE?**

This question refers to the *location* of data collection: Where will data be collected for use in the assessment process? To begin answering this question, it might be useful to classify the potential location points for data collection into the following categories: (a) office-use data, (b) student program/service-use data, and (c) classroom-based/course-embedded data.

With respect to the first category, Peter Ewell has already recommended that a “data audit” would be a valuable first step in the assessment process (FYA-List, 5-10-00). Campus offices housing data that may be particularly relevant to first-year student assessment include those that are responsible for college admissions, new-student orientation, student placement testing, and CIRP administration.

In regard to the second category, student program/service-use data, Lee Upcraft and John Schuh recommend “keeping track of who participate in first-year student programs.” (FYE-List, 5-26-00). These student “footprints” (as Ewell calls them) may be tracked via:

- *Logs* kept by student-service providers to track such behaviors as: (a) the incidence and frequency with which students utilize support programs, (b) the number of contacts between students and their advisors, or (c) number of “critical incidents” on campus (e.g., number of disciplinary reports filed).
- *Trace audits*, a.k.a., “credit-card measures” of student involvement (e.g., using student identification cards to assess frequency of library use).
- *Transcript analysis* of students’ course-enrollment patterns to assess the extent and nature of student progress toward completion of academic programs and degrees (e.g., incidence and frequency of students changing majors; time taken by students to complete programs or degree requirements).
- “*Student development*” or “*co-curricular*” transcripts of individual students’ participation in campus clubs, student organizations, or college-sponsored community service.

Admittedly, colleges and universities often have not kept careful records of many of the above-mentioned student behaviors, but if such record-keeping is intentionally built into future FYE-assessment plans, it may serve as a spark plug for igniting institutional awareness of the need to engage in the systematic collection and organization of student-involvement data.

With respect to *classroom-based (course-embedded)* assessment, students' "course products or curricular artifacts" (as Karl Schilling calls them) may provide useful assessment data. Such products or artifacts may include students' written products and videotapes of student presentations or performances.

Student work in the *first-year seminar* may provide particularly pertinent data for FYE assessment. Since the seminar has a student-centered focus, which often involves a good deal of self-assessment, these student "products" may prove to be a valuable source of college-entry assessment data. Furthermore, the course can provide a comfortable and relevant venue within which to conduct institution-centered assessment, which is likely to be perceived by students with the course's goals and practices.

Also, student motivation for engaging in assessment, which has been a challenging issue for the assessment movement (Banta, 1988), is likely to be high in the first-year seminar because the assessments are embedded in a credit-bearing, grade-earning course. Furthermore, the first-year seminar can provide a substantial *sample size* of entry students for assessment, as well as ample *time* needed to conduct comprehensive college-entry assessment. Lastly, if the seminar is a required course, then the issue of assessment "volunteers" is circumvented along with the potentially confounding effects of sampling bias associated with the "volunteer effect."

The upshot of these arguments is that the first-year seminar, in addition to its service to students, has the potential to serve the institution by providing a convenient curricular conduit through which the college can access a substantial sample of captive (and motivated) new students, which is needed to collect college-entry assessment in a comprehensive and representative fashion. Analogously, senior seminars may provide a viable and relevant context within which the outcomes component of longitudinal, value-added assessment can take place. As Astin (1991) points out, "Voluntary participation in follow-up testing may lead to a large amount of attrition because of noncooperation. Required participation, however, raises both logistical and ethical issues. For students, participation in cognitive post-testing might be incorporate into the requirements of a course." (pp. 168-169).

ASSESSMENT QUESTION #6. **HOW?**

This question refers to how the assessment data will be collected, analyzed, and summarized, i.e., the research *design or methodology* to be used. A wide range of both quantitative and qualitative methodologies is available to answer this question, each with its own advantages and disadvantages. *Quantitative* methods can generate data amenable to tests of statistical significance and involve research designs that are more amenable to reaching the conclusion that there may be a casual connection between the program experience and program outcomes. *Qualitative* methods are capable of generating data that capture the participants' phenomenological experience and how that experience is conveyed through the eloquence and poignancy of human language. Such data may provide more dramatic evidence than numerical charts and tables, especially for members of the college community who are neither conversant with, nor partial to "number crunching," regardless of its degree of statistical precision or methodological rigor.

An increasing emphasis is being placed on qualitative research in higher education (Fidler,

1992). Some of its more radical proponents argue that it should displace or replace the dominant quantitative paradigm, which they argue, has exerted an almost hegemonic hold on the research methodology used in education and the social sciences (Duffy & Jonassen, 1992). On the other hand, those in the quantitative camp are likely to charge that qualitative research lacks reliability or objectivity, yielding data that are dangerously subject to biased interpretation (Reigeluth, 1992). As is usually the case with such thesis-antithesis dichotomies, an effective synthesis lies somewhere between the two polar positions. While acknowledging that quantitative and qualitative research emerge from contrasting philosophical traditions and rest on very different epistemological assumptions (Smith & Heshusius, 1986), the position taken here is that the data generated by these two styles of inquiry can provide *complementary* sources of evidence, with the disadvantages of one method being offset or counterbalanced by the advantages of the other.

Indeed, two of the most prolific and highly regarded quantitatively oriented researchers in higher education, Ernest Pascarella and Patrick Terenzini, have argued for complementing quantitative methods with qualitative research procedures: “[An] important direction of future research on college impact should be a greater dependence on naturalistic and qualitative methodologies. When employed judiciously, such approaches are capable of providing greater sensitivity to many of the subtle and fine-grained complexities of college impact than more traditional quantitative approaches” (1991, p. 634).

Among program-evaluation scholars, it is almost axiomatic that use of “multiple measures” in the assessment process represents a more reliable and valid practice than exclusive reliance on a single research method or data source (Wergin, 1988). Including multiple measures in the assessment plan for first-year programs increases the likelihood that subtle differences in the effects of the program will be detected. Use of multiple methods also can be used to demonstrate a consistent pattern of results across different methods—a cross-validation procedure known in the assessment literature as “triangulation” (Fetterman, 1991) or “convergent validity” (Campbell & Fiske, 1959). Such cross-validation serves to minimize the likelihood that the results obtained are merely an artifact of any one single method used to obtain them, and it may serve to magnify the persuasive power of the results obtained.

As Dorothy Fidler notes in her *Primer for Research on The Freshman Year Experience*, “All research designs have strengths and weaknesses; therefore no single piece of research can fully answer a research question. Researchers can select between qualitative or quantitative designs, and ideally a body of literature contains both types of designs in order to balance the strengths and weaknesses of each” (1992, p. 11). Consequently, a comprehensive and well-balanced assessment of the first-year experience should include a complementary combination of different quantitative and qualitative methods, such as those described below in the following taxonomy.

I. Quantitative Research Methods

Pre-Test/Post-Test Design

This quantitative method is designed to assess the amount of change in students’ attitudes, reported behavior, or cognitive performance between the onset and completion of an intervention program. The research design involves administering an assessment instrument to students before program participation, which is used as a baseline (pre-test) against which their post-program (post-test) responses can be compared.

To ensure that pre- to post-program changes can be attributed to a first-year program experience in particular, rather than to general student maturation over time, students’ pre- and

post-program responses could be compared with the responses provided by other first-year students gathered at the beginning and end of the same time period.

Experimental Design

This research method involves comparing student outcomes for first-year students who are *randomly assigned* to either one of the following two groups: an “experimental” group of students who participate in the program, or a “control” group of students who do not participate in the program.

Historically, this method has been considered to be the scientifically ideal or “true” experimental design for evaluating educational programs, because it ensures *randomized assignment* of students to both the experimental and control groups, i.e., each student selected has an equal and independent chance of being placed into either one of these groups. This design is ideal for guarding against the “volunteer effect” or “self-selection bias”—the possibility that students who volunteer to participate in an educational program, or select themselves to be part of that program, may be more intrinsically motivated and committed to college than students who elect not to become involved in the program (Fidler, 1992). In the specific case of the first-year experience, this poses a threat to the validity of positive assessment outcomes obtained for any first-year program because of the tendency for more highly motivated first-year students to enroll in the program. Thus, the positive outcomes are confounded by the possibility that they were caused by the participating students’ characteristics rather than the actual influence or impact of the program itself.

One procedure that can be used to circumvent this methodological problem is to seek a larger number of students who are interested in participating in a particular first-year program than the actual number that can be accommodated. Half of those students who express interest in taking the program are then randomly selected to participate in it (to serve as the experimental group), while the remaining half of the students who had expressed interest in participating are randomly placed in the control group and do not experience the program.

The major *disadvantage* of the experimental design is an *ethical* one: Its random selection of students to become program participants or nonparticipants (members of the control group) results in the arbitrary denial of program access to one-half of the students who want to become involved in the program and may benefit from it (Pascarella, 1986). This ethical disadvantage of the experimental design may be tempered somewhat by the argument that it is a justifiable procedure when used to conduct a “pilot study” on only one cohort of students, with the intention of gathering just enough data to serve as supporting evidence for subsequent expansion of a program, thereby ensuring the program’s availability to all future cohorts of entering students.

Another strategy for reducing the ethical disadvantages of random assignment is to assign students to a “control” group who experiences a different support program, rather than no program at all. For example, students could be randomly assigned to a faculty mentoring program (experimental group) or a peer mentoring program (control group). While this strategy might reduce the likelihood of finding statistically significant differences between the experimental and control groups (since the latter experiences an intervention program as well), this strategy has the counterbalancing advantage of creating an assessment design that may more accurately reflect the reality of educational decision-making, whereby choices have to be made between two (or more) potentially effective programs or practices.

Quasi-Experimental Design

This research method involves comparing outcomes for first-year students who *volunteer* to participate in a program (experimental group) relative to a “*matched*” control group, i.e., selected first-year students who have elected not to participate in the program but whose personal characteristics are similar to (“match”) the experimental group on important student variables that may influence the outcomes being measured. For example, in previously conducted first-year assessments, students in experimental and control groups have been matched with respect to such characteristics as (a) high school grade-point average, (b) standardized college-admission test scores, (c) basic-skills placement test scores, (d) predicted GPA derived from weighted scores on college-preparation courses, (e) high school grades and SAT scores, (f) residential or commuter status, and (g) demographic characteristics such as age, gender, race, or ethnicity. Matching program participants with non-participants in this fashion serves to control for, or rule out the possibility that differences in student outcomes associated with program participation could be due to the fact that program participants had personal characteristics that differed significantly from non-participants.

A major ethical advantage of this research design is that it allows all students who express interest in experiencing the program to have access to it, thus circumventing the ethical problem associated with an experimental design in which some students are arbitrarily denied program access so they can serve as a control group.

However, one methodological disadvantage of the quasi-experimental design is that students are *not* randomly assigned to experimental and control groups as they are in a true experimental design (hence its name, “quasi-experimental”). Consequently, this design fails to control for the volunteer effect of self-selection bias, leaving open the possibility that any positive outcomes resulting from program participation may be due to the highly motivated nature of students who elect to participate in the program, rather than to the effects of the program itself.

One possible strategy for addressing this limitation of the quasi-experimental design is to survey students in both the experimental group and matched-control group to assess whether they report differences in their level of college motivation, or in their willingness to participate in programs like the one under investigation.

Time-Series Design

In this research design, outcomes assessed *after* implementation of the first-year program are compared with the same outcomes achieved *prior* to the program’s implementation. For example, freshman-to-sophomore retention rates at the college after adoption of a first-year seminar are compared with freshman-to-sophomore retention rates for the years preceding course adoption.

The advantage of this design is that it provides a type of “historical” control group against which the effects of program participation may be compared, without having to withhold the program from a portion of entering freshmen so they can serve as a “contemporary” control group. Hence, this design circumvents the ethical drawback of an experimental design in which some students are deliberately recruited for program participation and then arbitrarily deprived program access so they can be used as a control group.

However, two caveats must be issued with respect to the time-series research design: (a) The personal characteristics of entering freshmen during years before and after implementation of the first-year program should be similar or matched, so that any changes in student outcomes subsequent to program implementation cannot simply be due to historical changes in the entry characteristics of the entering class (e.g., more academically qualified students entering the

institution during and after implementation of the first-year program). (b) Two or more years of outcome data should be gathered before and after implementation of the program for comparison purposes. Data should not just be gathered in the year immediately before and after program implementation, because any year-to-year fluctuations in student outcomes (e.g., retention) may simply be due to random variations that take place across a period of time (Pascarella, 1986).

Multiple Regression Analysis (a.k.a., Multivariate Analysis)

This statistical procedure, or some variant thereof, has been the favored research design of contemporary scholars interested in assessing how college experiences affect student outcomes (e.g., Astin, 1993; Pascarella & Terenzini, 1991). In short, multiple regression analysis involves computing correlations between *student outcome* variables (e.g., student retention or academic performance) and two other types of variables: (a) students' *input* variables (e.g., entering students' SAT scores) and (b) *college experience* variables (e.g., student participation in a particular first-year program).

For a detailed explanation of multiple regression analysis, consult the appendices in Astin (1991) or Pascarella and Terenzini (1991). The following synopsis of multiple regression analysis has been adapted from Astin (1993) and applied to assessment of the first-year programs.

The first step in multiple regression analysis is to calculate correlations between all influential student-input characteristics and a single student outcome in order to obtain a “prediction score” or “estimated outcome” score for that particular outcome (e.g., the predicted or estimated first-year GPA for college students based on their entering high-school GPA, SAT, and placement-test scores). This estimated outcome score, based on characteristics that students bring with them to the institution, serves as a type of “statistical” control group or baseline against which to compare the effects of later college experiences on student outcomes.

The amount or degree of impact of a first-year program, such as a first-year seminar, can be assessed by computing the correlation between the residual score it produces and the student outcome in question. This partial correlation—called the “beta” coefficient—represents the degree to which the educational experience (e.g., participating in a first-year seminar) and the student outcome (e.g., first-year retention) are statistically related—*after* all other potentially biasing student characteristics have been controlled for. In other words, it represents what the freshman seminar experience adds to the predicted student outcome—above and beyond what can be predicted by student input characteristics.

Thus, it might be said that multiple regression analysis attempts to control for confounding student variables *statistically* by computing and comparing correlations between student variables and outcomes, whereas the aforementioned experimental and quasi-experimental research designs attempt to gain this control *procedurally* by the procedures used to select and assign students to experimental and control groups.

Multiple regression analysis can also be adapted to assess whether the effect of a first-year program (or any other college-experience variable) on a student outcome is either “direct” or “indirect.” A college-experience variable can be considered to have a *direct* effect on a student outcome if its beta coefficient remains statistically significant even after correlations between all other college-experience variables and that student outcome have been included in the regression analysis. This suggests that a particular college-experience variable is making a unique or independent contribution to the student outcome, which cannot be accounted for by other college-experience variables.

A college-experience variable may be deemed to have an *indirect* effect on a student outcome if its beta coefficient, which was significant after input (entry) characteristics were controlled for, is later reduced to nonsignificance when other college-experience variables are added to the regression equation. This suggests that the effect of the particular college-experience variable is accounted for, or mediated by, other college-experience variables.

There are three key advantages associated with the use of multiple regression analysis as a research method for assessing outcomes of the first-year experience:

- (a) It circumvents the disadvantage of a “true” experimental design in which first-year students are denied access to a program so they can be used as a control group.
- (b) It allows investigators to assess whether the addition of individual college-experience variables (such as their participation in a summer orientation program) results in any incremental improvement in the predicted score for a student outcome (Banta, 1988).
- (c) It allows investigators to compute the percentage of outcome variance that may be attributable to a particular student-input or college-experience variable (by squaring the beta coefficient), thus providing an estimate of the variable’s relative influence on the outcome under investigation (Pascarella & Terenzini, 1991).

Two limitations of multiple regression analysis have been cited in the assessment literature:

- (a) The procedure does not allow assessment of how joint or common variance between college-experience variables and student-input variables may interact to influence outcomes (Hanson, 1988).
 - (b) It assumes that any variance in outcome that may be attributed to the joint influence of student-input and college-experience variables is attributed solely to the student-input variable, thus the influence of student-input variables on outcomes may be overestimated while the influence of college-experience variables may be underestimated (Pascarella & Terenzini, 1991).
- However, proponents of multiple regression analysis consider these to be relatively minor limitations that have little adverse effect on the overall validity and interpretability of the results generated by this statistical procedure (Astin, 1991; A.W. Astin, personal communication, October 21, 1992).

If the FYE is to be assessed as an entire unit, as John Gardner has encouraged, then it appears as if multiple regression analysis is our best available methodological tool because the FYE is comprised of multiple environmental variables (orientation, advising, first-year seminar, etc.). How all of these variables combine to affect end-of-first-year outcomes is a complex phenomenon, which requires a comprehensive methodology that is capable of shedding light on how these multiple variables affect multiple student outcomes, and whether these effects occur independent of, or interact with, student entry characteristics. As Terenzini and Upcraft (1996) state categorically, “In order to evaluate the relative importance played by multiple variables, one must use multivariate statistical procedures.” (p. 105).

Concluding this discussion of quantitative research methodologies, there is one final consideration that should be kept in mind when using any statistical measure of the significance or impact of first-year programs. If measures of *student* outcomes associated with a first-year program are not found to reach a level of *statistical* significance, the program’s effect on certain *institutional* outcomes may still have *practical* significance. For instance, a first-year program which results in a very modest 5-10% increase in student retention may generate a gain in the college’s total enrollment number that does not reach a level of statistical significance; however, the revenue gained from this modest increase in additional tuition-paying students may contribute significantly to the institutional budget, particularly at colleges whose operational budgets are heavily tuition dependent. For example, it has been estimated that small private

colleges typically lost \$5500 per year for each missing student (Update, 1983). Campus-specific research reports at Seton Hall University, University of South Carolina and Baptist College (now Charleston Southern University), strongly suggest that comprehensive assessment of a first-year program's overall impact should focus not only its statistical effect on student outcomes, but also its fiscal effect, which is an important institutional outcome.

II. Qualitative Research Methods

Quantitative data, such as survey ratings and behavioral measures, provide evaluative information that can be readily summarized and manipulated numerically. Such data can be scored efficiently by machine or computer and are amenable to statistical analysis (e.g., correlation coefficients or chi-square analysis). In contrast, qualitative data take the form of human *actions* and *words*—e.g., students' written or verbal comments and they are analyzed by means of "human instruments"—e.g., listening, reading, and observing (Kuh, Schuh, & Whitt, 1991). Also, in contrast to the hypothesis testing and scientific methodology that characterizes quantitative research, qualitative research is "exploratory [and] inductive, . . . one does not manipulate variables or administer a treatment. What one does is observe, intuit, [and] sense what is occurring in natural settings" (Merriam, 1988, p. 17).

Analysis of Students' Written Comments

Written comments made on student surveys can provide a good source of qualitative data. These comments may be difficult to summarize and manipulate statistically, but they have the potential for providing poignant, in-depth information on program strengths or weaknesses, and can provide an index of students' subjective feelings about the program. As Davis and Murrell note, "These descriptions provide much texture and offer rich, often powerful images of the college experience" (1993, p. 50).

Historically, surveys and questionnaires have not been considered to be qualitative research methods because they generate quantitative data (numerical ratings). However, written comments made by respondents to clarify their ratings do represent legitimate qualitative data, the content of which can be analyzed and classified systematically. Another potential source of written comments for assessment are personal documents, which can be described broadly as "any first-person narrative that describes an individual's actions, experiences, and beliefs" (Bogdan & Biklen, 1992, p. 132). For example, student journals used in the freshman seminar qualify as personal documents that could be reviewed to gain insight into student feelings about the course and their first-semester experience.

One particular qualitative-research method that can be used to provide a summation and classification of students' written comments is "category analysis," a procedure in which the reader engages in *inductive* data analysis, identifying common themes that emerge among comments as they are being read and then organizes these comments into categories (Lincoln & Guba, 1985). A tally of the number of written comments per category may also be kept and reported along with the identified categories. These category-specific frequency counts can then be used as quantitative data to help summarize and interpret how representative are the written comments (qualitative data).

The validity of category analysis can be improved by having two or more independent readers categorize the comments, which allows for assessment of inter-reader (inter-rater) agreement, as well as serving as a measure of the reliability (consistency) of the identified categories to be demonstrated. Additionally, the independent readers should have no vested

interest in the outcome of the program, thus minimizing the likelihood of evaluator bias.

Open-Ended Interviews

Tierney (1991) provides a succinct description of open-ended interviews, “Prior to beginning the set of interviews, the researcher develops a protocol of general questions that needs to be covered; however the researcher is free to move in any direction that appears interesting and rich in data.” (p. 9). Some initial set of standardized questions are posed to all interviewees, following which the interviewer is free to create her own follow-up questions to whatever responses emerge from the person being interviewed.

Focus Groups

Succinctly defined, a focus group is a small (6-12 person) group that meets with a trained moderator in a relaxed environment to discuss a selected topic or issue, with the goal of eliciting participants’ perceptions, attitudes, and ideas (Bers, 1989). In contrast to surveys or questionnaires, which solicit individual students’ written responses, focus-group interviews solicit students’ oral responses in a discussion-group setting.

Focus-group interviews may be used in conjunction with surveys. For example, the interviews may be used as a follow-up to surveys for purposes of gaining greater insight into the meaning of the survey’s quantitative results. Or, the order may be reversed, with the focus groups conducted first to collect ideas that may later be used to develop specific items for inclusion on surveys or questionnaires. Another, more tacit advantage of focus groups is that they serve to validate students’ personal experiences by sending them the message that someone at the institution is genuinely interested in their feelings, opinions, and concerns. Indeed, qualitative researchers argue that engaging in dialogue with research participants, particularly those who have been marginalized in some way, serves to empower them and encourages them to gain control of their experience (Roman & Apple, 1990).

References

- Alexander, J. M., & Stark, J. S. (1986). *Focusing on student outcomes*. Ann Arbor, Michigan: National Center for Research to Improve Postsecondary Teaching and Learning, University of Michigan.
- Altman, H.B. (1988). Assessment of learning outcomes: Building Faculty cooperation. *Journal of Staff, Program, & Organization Development*, 6 (3), 125-127.
- American Association for higher Education (1998). Organizing for learning. *AAHE Bulletin*, (September), p. 6.
- Astin, A. W. (1991). *Assessment for excellence: The philosophy and practice of assessment and evaluation in higher education*. New York: Macmillan.
- Astin, A. W. (1993). *What matters in college: Four critical years revisited*. San Francisco: Jossey-Bass.
- Banta, T. W. (1988). Promise and Perils. In T. W. Banta (Ed.), *Implementing outcomes assessment: Promise and perils* (pp. 95-98). New Directions for Institutional Research, No. 59. San Francisco: Jossey-Bass.
- Banta, T. W., Lund, J. P., Black, K. E., & Oblander, F. W. (1996). *Assessment in practice: Putting principles to work on college campuses*. San Francisco: Jossey-Bass.
- Barefoot, B. O. (Ed.) (1993). *Exploring the evidence: Reporting outcomes of freshman seminars*. (Monograph No. 11). Columbia, SC: National Resource Center for The First Year Experience and Students in Transition, University of South Carolina.
- Barefoot, B. O. (2000). The first-year experience—Are we making it any better? *About Campus*, 4(6), 12-18.
- Barefoot, B. O., Warnock, C. L., Dickinson, M. P., Richardson, S. E., & Roberts, M. R. (Eds.)(1998). *Exploring the evidence, Volume II: Reporting outcomes of first-year seminars*. (Monograph No. 29). Columbia, SC: National Resource Center for The First Year Experience and Students in Transition, University of South Carolina.
- Bers, T.H. (1989). The popularity and problems of focus-group research. *College and University*, 64(3), 260-268.
- Bogdan, R. C., & Bilken, S. K. (1992). *Qualitative research for education* (2nd ed.). Boston: Allyn & Bacon.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multivariate-multimethod matrix. *Psychological Bulletin*, 56, 81-105.
- Davis, T. M., & Murrell, P.H. (1993). *Turning teaching into learning: The role of student responsibility in the collegiate experience*. ASHE-ERIC Higher Education Report No. 8. Washington, D.C.: The George Washington University, School of Education and Human Development.

- Duffy, T. M., & Jonassen, D. H. (1992). Constructivism: New implications for instructional technology. In T. M. Duffy and D. H. Johassen (Eds.), *Constructivism and the technology of instruction: A conversation* (pp. 1-16). Hillsdale, NJ: Laurence Erlbaum & Associates.
- Education Commission of the States (1995). *Making quality count in undergraduate education*. Denver, CO: ECS Distribution Center.
- Ewell, P. T. (1988). Implementing assessment: Some organizational issues. In T. W. Banta (Ed.), *Implementing outcomes assessment: Promise and perils* (pp. 15-28). New Directions for Institutional Research, No. 50. San Francisco: Jossey-Bass.
- Fetterman, D. M. (1991). Auditing as institutional research: A qualitative focus. In D. M. Fetterman (Ed.), *Using qualitative methods in institutional research* (pp. 23-24). New Directions for Institutional Research, No. 72. San Francisco: Jossey-Bass.
- Fidler, D. S. (1992). *Primer for research on the freshman year experience*. National Resource Center for The Freshman Year Experience and Students in Transition, University of South Carolina.
- Fidler, P. P., & Shanley, M. G. (1993, February). *Evaluation results of University 101*. Presentation made at the annual conference for The Freshman Year Experience, Columbia, South Carolina.
- Gardner, J. N. (1980). *University 101: A concept for improving teaching and learning*. Columbia, SC: University of South Carolina. (ERIC Reproduction Service No. 192 706).
- Gardner, J. N. (1981). Developing faculty as facilitators and mentors. In V. A. Harren, M. N. Daniels, & J. N. Buck (Eds.), *Facilitating students' career development* (pp. 67-80). New Directions for Student Services, No. 14. San Francisco: Jossey-Bass.
- Gardner, J. N. (1986). The freshman year experience. *College and University*, 61 (4), 261-274.
- Gardner, J. N. (1991). Introduction. *Perspectives on the freshman year: Selected major addresses from freshman year experience conferences*. (Monograph No. 2). Columbia, SC: National Resource Center for The Freshman Year Experience, University of South Carolina.
- Gardner, J. N. (1994, July). Comment made at the Seventh International Conference on the First-Year Experience, Dublin, Ireland.
- Hanson, G. R. (1988). Critical Issues in the assessment of value added in education. In T. W. Banta (Ed.), *Implementing outcomes assessment: Promise and perils* (pp. 53-68). New Directions for Institutional Research, No. 50. San Francisco: Jossey-Bass
- Hoffman, E. P. (1986). A review of two studies of elasticities in academe. *Economics of Education Review*, 5(2), 219-224.
- Holsti, O. R. (1969). *Content analysis for the social sciences and humanities*. Reading, MA: Addison-Wesley.

- Kuh, G., Schuh, J., Whitt, E., & Associates (1991). *Involving colleges*. San Francisco: Jossey-Bass.
- Kuh, G., Shedd, J., & Whitt, E. (1987). Student affairs and liberal education: Unrecognized (and unappreciated) common law partners. *Journal of College Student Personnel*, 28(3), 252-260.
- Larimore, L. K. (1974). Break-even analysis for higher education. *Management Accounting*, 56(3), 25-28.
- Lenning, O. T. (1988). Use of noncognitive measures in assessment. In T. W. Banta (Ed.), *Implementing outcomes assessment: Promise and perils* (pp. 41-52). New Directions for Institutional Research, No. 50. San Francisco: Jossey-Bass.
- Lincoln, Y. S., & Guba, E. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.
- MacGregor, J. (1991). What differences do learning communities make? *Washington Center News*, 6(1), pp. 4-9.
- Merriam, S. B. (1988). *Case study research in education: A qualitative approach*. San Francisco: Jossey-Bass.
- Marchese, T. (1993). Understanding benchmarking. *AAHE Bulletin*, 47(8), pp. 3-5.
- Mullendore, R. & Abraham, J. (1992). *Orientation director's manual*. Statesboro, GA: National Orientation Director's Association.
- National Center for Postsecondary Improvement (1999). Revolution or evolution? Gauging the impact of institutional student-assessment strategies. *Change*, 31 (5), pp. 53-57.
- Nowl, L., & Levitz, R., & Associates (1985). *Increasing student retention: New challenges and potential*. San Francisco: Jossey-Bass.
- Pascarella, E. T. (1986). A program for research and policy on student persistence at the institutional level. *Journal of College Student Personnel*, 27 (2), 100-107.
- Pascarella, E. T., & Terenzini, P. T. (1991). *How college affects students: Findings and insights from twenty years of research*. San Francisco: Jossey-Bass.
- Perry, W. (1970). *Forms of intellectual and ethical development in the college years: A scheme*. New York: Holt, Rinehart and Winston.
- Reigeluth, C. M. (1992). Reflections on the implications of constructivism for educational technology in T. M. Duffy & D. H. Jonassen (Eds.), *Constructivism and the technology of instruction: A conversation* (pp. 147-156). Hillsdale, NJ: Laurence Erlbaum & Associates.
- Reinharz, S. (1993). *On becoming a social scientist*. New Brunswick, NJ: Transaction Publishers.
- Rosenthal, R. (1974). *On the social psychology of the self-fulfilling prophecy: Further evidence*

of Pygmalion effects and their mediating mechanisms. New York: M.S.S. Information Corporation, Modular Publications.

- Schilling, K. (2000, May 12). *Assessment questions and first-year job description*. First-Year Assessment Listserv (FYA) Series. [On-line serial].
<http://www.Brevard/fyc/listserv/remarks/schilling600.htm>
- Scriven, M. (1967). The methodology of evaluation. *Perspectives of Curriculum Evaluation*, AERA Monograph Series on Curriculum Evaluation, No. 1. Chicago: Rand McNally & Co.
- Smith, J. K., & Heshusius, L. (1986). Closing down the conversation: The end of the quantitative-qualitative debate among educational inquirers. *Educational Researcher*, 15(1), 4-12.
- Taylor, S. J. & Bogdan, R. C. (1984). *Introduction to qualitative research and methods: The search for meaning*. New York: Wiley.
- Terenzini, P. T., & Upcraft, M. L. (1996). Assessing program and service outcomes. In M. L. Upcraft & J. H. Schuh (Eds.), *Assessment in student affairs: A guide for practitioners* (pp. 217-239). San Francisco: Jossey-Bass.
- Tierney, W. G. (1991). Utilizing ethnographic interviews to enhance academic decision making. In D. M. Fetterman (Ed.), *Using qualitative methods in institutional research* (pp. 7-22). New Directions for Institutional Research, No. 72. San Francisco: Jossey-Bass.
- Turner, J. C., Garrison, C. Z., Korpita, E., Waller, J., Addy, C., Hill, W. R., & Mohn, L. (1994). Promoting responsible sexual behavior through a college freshman seminar. *Aids Education and Prevention*, 6(3), 266-277.
- Upcraft, M. L., & Schuh, J. H. (1996). *Assessment in student affairs: A guide for practitioner*. San Francisco: Jossey Bass.
- Update (1993) Small private colleges begin to feel the effect of enrollment decline. *Change*, 15, 50-51.
- Wergin, J. F. (1988). Basic issues and principles in classroom assessment. In J. H. McMillan (Ed.), *Assessing students' learning* (pp. 5-17). New Directions for Teaching and Learning, No. 34. San Francisco: Jossey-Bass.
- Wrenn, R. (1988). Student-faculty intervention programs. In J. Rehm (Ed.), *Making changes: 27 strategies for recruitment and retention* (pp. 75-77). Madison, WI: Magna.
- Xavier University (1994). E pluribus unum at Xavier University. *The Freshman Year Newsletter*, 6(3), p. 7.