

Next Steps in the Assessment of Collegiate Education*

K. Patricia Cross

From an Educational Testing Service point of view TIME Magazine's Education page has been quite timely recently--having talked about ETS twice in the past two months. But I'd like to start this morning's discussion by quoting a lively exchange in TIME that didn't mention ETS but that I think has a great deal to do with us especially in relation to the comprehensive assessment of collegiate education.

On October 2, TIME wrote as follows:

"Brainy college freshmen under the impression that their next four years would be spent as assembly-line workers in ivy-covered fact factories got a sharp jolt last week. 'You demand facts, facts, facts,' Guest Speaker A. L. Sachar, president of Brandeis, told the entering class at the University of Illinois. What they need just as much, he insisted, is values to serve 'in a world where the harsh voice of unreason cries down the generous passions,' and 'the elasticity of your minds will be a shield.' On campuses across the U.S., college presidents were playing up the value of values.

"Yale President Kingman Brewster, Jr. said that although freshmen were picked for their intellectual ability, their 'moral capacity' had also been taken into account. Success at Kenyon, said Dean Bruce Haywood, ultimately depends on a student's 'individual taste and moral judgment.' 'The collection of knowledge is only the starting point,' echoed Curtis Tarr, president of Lawrence University of Appleton, Wis. At Pomona College, one of the six associated Claremont Colleges of California, President E. Wilson Lyons also greeted freshmen with a call to use knowledge for moral ends.

"The most fervent plea for a 'moral quest' came from Princeton's Robert F. Goheen. He told bright Tiger cubs that if they expected 'only to accumulate knowledge, I would advise you to begin negotiations with another institution where you can attach yourself to a pipeline of inanimate learning and become full, like a storage tank, sealed by a diploma and otherwise useless.'"

* Presented to the ETS Standing Committee on Tests and Measurements, October 22, 1964

And on October 8, a gentlemen from New York replied:

"It's pure hypocrisy for college presidents and deans to talk about morality to the new freshman classes (Oct. 2). Those students were admitted on the basis of grades and test scores, and they will pass or fail on the same basis. The student does not 'demand facts, facts, facts.' His instructor demands them, neatly played back from a term's subject matter. Let the deans weed their own gardens before looking across the fence."

There is probably more truth to the letter-writer's charge than most educators would like to admit. I suppose that were we to ask these college presidents what they are doing in their programs to teach values and how they know when they have done it successfully, we might have some red faces--or perhaps some very sophisticated answers.

Informally, we evaluate the work of the colleges constantly. In the eyes of the general public the "best" colleges are those that are most difficult to get into and/or the most expensive. The accrediting agencies have been criticized for evaluating colleges by focusing on certain quantitative measures which are easily obtained--number of Ph.D.'s on the faculty, number of books in the library, faculty-student ratio, financial resources, etc. College faculty and administration have tended to measure prestige of the institution in terms of scholarly production--number of books and professional journal articles. National Merit research has concentrated on Ph.D. output as a criterion of the quality of undergraduate education. For the most part these are not really goals of education. They are quantifiable variables which may or may not be related to what we're trying to do in colleges. It is regrettable that in some instances these measures have become goals. It is assumed, for example, that a high number of Ph.D.'s on the faculty is a good thing to strive for--a goal in and of itself--even though it might be quite difficult to demonstrate that this has any real connection with the quality of education or even with scholarly production.

At this time improvement in education depends to a very large extent upon what we can do about devising more appropriate yardsticks for evaluation--and

especially for assessing the impact of the college experience on the student.

From a measurement standpoint we have to look first at what it is that educators say they are trying to do and how ETS and similar organizations can devise appropriate measures. What kinds of effects should be studied and how do we ^{from} select the enormous range of material, in a complex experience like college education, those measurements which are most significant?

Superficially at least, the Presidents quoted in TIME have not given us much to go on--except that it is quite clear that increased knowledge is expected by all of them. Sachar said students need values as much as facts; Brewster said freshmen were picked for moral capacity as well as intellectual ability; Tarr maintained that the collection of knowledge is the starting point; and Goheen warned that students should not expect only to accumulate knowledge. Obviously, increased knowledge is a stated goal and it is, we believe, a measurable goal. Later, I'll talk in some detail about ETS's new venture in this area at the college level--the Comprehensive College Tests.

It seems apparent now that all the refinements in measures of intellectual achievement that we can and will continue to make will not completely answer the questions raised in the topic which I have been assigned, "Next Steps in the Assessment of Collegiate Education."

I'm not sure, for example, that we know where to go with the other goals mentioned by the Presidents quoted in TIME, such as the development of "moral capacity" and "individual taste and moral judgment." But since these Presidents did not have our problems in mind when they addressed their freshmen, let's look for a moment at some of the more formally stated purposes of the colleges.

I selected some college catalogues at random and poured over them for stated purposes. My first observation is that, in all likelihood, the catalogues reflect everything from very carefully prepared faculty statements to the prose of the registrar or whoever happens to write that section of the catalogue. Contrast

these purposes for the College of Liberal Arts and Sciences for example:

From the University of Pennsylvania catalogue, we find that:

"The College of Arts and Sciences is concerned with introducing its young men to the fields of knowledge which may lead to the understanding of the most fundamentally important problems of man's life in his universal environments. Our educational program is therefore aimed at giving the individual the best possible opportunity, under guidance, to satisfy basic intellectual curiosities and at the same time to round out an understanding and appreciation of the physical, social, and intellectual world in which he lives."

At the University of Kansas, on the other hand, we find no purpose stated as such, but we can get a good idea of the role of the College of Arts and Sciences from the purposes of the students enrolled in that college. The Bulletin of the University of Kansas describes them as follows:

"Enrolled in the college are those students who have not definitely made up their minds about the field upon which they wish to concentrate, those who are gaining preliminary liberal arts foundations for later specialization in one of the professional schools of the University, and those who are majoring or preparing to major in one or more of the liberal arts and sciences."

Certainly faculty members accepting these definitions as the purposes of instruction are going to teach quite differently, and the evaluation of how well they have done their job is going to be based on different criteria. Kansas gives us a relatively simple measurement problem; we can surely determine whether the student reaches a decision about his major and whether he has the appropriate background for later specialization in a given discipline. The University of Pennsylvania statement of purpose presents a much greater challenge to measurement and perhaps this challenge is worth the effort.

Although the catalogue descriptions are prone to sound idealistic and not easily converted to measurement terms, there is a common theme and educated people know and understand goals such as these:

From St. Olaf:

"The pursuit of learning demands language skills, critical thinking and a social concern which embodies democratic ideals and an appreciation of diverse and foreign cultures."

From Stanford University:

The willingness "to weigh varying viewpoints...is the mark of a man or woman with a true liberal education."

From Swarthmore:

The College does not seek to impose upon its students this Quaker view of life, or any other specific set of convictions about the nature of things and the duty of man. It does, however, have the twofold aim of encouraging conscious concern about such questions and unceasing reexamination of any view which may be held regarding them."

From University of Pennsylvania:

"The primary goal (of the University) is the development of the intellect, insuring the growth of the basic mental processes, and the training of the mind to reach valid conclusions and sound judgments in a spirit of intellectual honesty."

or from Princeton:

"The educational effort in the undergraduate college is directed towards leading the student to be an independent seeker in his own right, not a passive recipient of information, and to assume responsibility for gaining both knowledge and judgment which will enhance his later contribution to society."

In statements such as these, colleges have accepted a responsibility for attempting to develop in their students the capacity for and the commitment to independent and critical thought. Scholars of higher education make the same kinds of sounds when discussing goals. The Presidents' Commission on Higher Education lists the following as one of their objectives of general education. "To acquire and use the skills and habits involved in critical and constructive thinking." Clarence Faust says it in a single sentence. "General education must undertake, therefore, to prepare people to think for themselves."

Here is a goal; it is clearly stated; it's apparently primary; it's universally accepted and as yet no one has really measured this quality in students.

Nevertheless this is one way we might approach the problem of evaluating the outcomes of college education--in terms of the stated goals of individual colleges, of educational scholars, and of study groups such as the Presidents' Commission

on Higher Education. Although this is perhaps the oldest approach and to some extent has not proved as fruitful as we would hope, it seems to me that we have not done as much as we might in developing original instruments which would measure qualities such as critical thinking, open-mindedness, creativity, and other intellectual interests and values which are generally accepted as a mark of the educated person.

For example, in struggling with the design of appropriate instruments for the measurement of the intellectual quality of creativity, ETS is experimenting with a new way of looking at this concept. Instead of conceiving of creativity as a psychological characteristic of the person, Rod Skager has approached creativity as an achievement. This concept, of course, includes environmental and choice factors as well as the presumably enduring personal characteristics related to creative achievement. The measurement of creativity, however, now becomes a relatively simple questionnaire--the Independent Activities Questionnaire--which asks the high school student what he has done, or made, or created, if you will--in art, music, literature, business, science, mechanics, sports or public affairs. One can then derive quantitative scores, which are a simple count of the number of creative accomplishments of the respondent or qualitative scores which reflect the judgment of raters on the respondent's most outstanding accomplishment.

The evidence to date indicates that: Quality and Quantity scores are significantly correlated, but that Quality is related to measures of academic aptitude whereas Quantity is more related to a measure of intellectual stimulation in the home. This instrument provides a realistic way to start a systematic investigation of the variables--whether these be personal traits or environmental circumstances--that are related to the Quality and Quantity of the individual's creative achievements. This work has focused to date upon high school students and home environments but it would seem entirely possible to adapt this concept to the productions in colleges and to college environmental factors. Even a casual look

at campuses will reveal vast differences in opportunities for creative expression. Some campuses have student theaters, literary magazines, open laboratories, poetry readings, and student-run businesses; others have none of these. It seems reasonable to hypothesize and possible to test that colleges will differ not only in the present creative activities of their students but also in the future creativity which might be predicted for students with this kind of college experience.

Thus, where we have agreed-upon goals which are reasonably clearly defined, we can continue to work on imaginative ways of casting them into observable, measurable terms. The Independent Activities Questionnaire is an example of this approach.

It is all too apparent, however, that something else is needed for those more elusive and subtle characteristics which we would like to identify and then perhaps to develop in our students.

In 1958, ETS started a search for the characteristics aside from the traditional criterion of grades which distinguished desirable students. High test scores and good grades do not necessarily add up to an educated man--as the college presidents made quite clear to their entering classes.

To identify other criteria, the ETS College Student Characteristics study went directly to the college faculty who not only have the best opportunity to observe the impact of education on the student but who also presumably have an appreciation for, and an understanding of the nature of higher education. The first step in the CSCS, under the direction of Jay Davis, was to ask faculty members to select desirable students and undesirable students with high grade averages and with low grade averages and to describe these students in whatever terms seemed significant. This resulted in a rating scale of eighty traits which revealed that the faculty employ sixteen relatively independent criteria in their description of students. These can be categorized under five general headings: intellectual qualities, characteristic approach to work, character traits and

values, peer relationships and social style, and personal adjustment.

In this approach where we are literally starting with what we can observe and working back to educational relationships, three sets of characteristics can be identified: those related to academic performance which faculty see as desirable, those not related to grades but seen as desirable characteristics in students, and those characteristics which are simply distinguishable but unrelated to either academic performance or desirability. It is perhaps the second category--desirable but unrelated to academic performance--which is of greatest interest because it has the potential of adding to our traditional criteria for evaluating desirable educational outcomes. Characteristics in this group are: probity (high ideals), likeableness, open-mindedness, altruism, and to some extent, social maturity and self-insight. It is rather interesting to observe that some of these qualities are frequently mentioned as aims of education. If faculty members can indeed observe such qualities in students reliably--a question which remains for study--we may be nearer to measuring the correlates of some of these more elusive goals than we had thought. On the other hand, much remains to be done before this approach furnishes any "answers." For example, traits identified as desirable but unrelated to academic performance furnish some interesting ideas, but they have not as yet been validated nor has any formal attempt been made to justify them as worthy goals of higher education. For example, is helping a student to become more "likeable" a real goal of higher education--or does it simply indicate that faculty are human and like likeable people? The point remains, however, that the criterion development study furnishes a wealth of information about what faculty say they like in students, how variance in traits may be associated with that in grades, SAT, desirability, or desirability apart from grades. And this should lead to new research questions and perhaps to more fruitful approaches to identifying, describing and measuring qualities suitable for evaluating the outcomes of college education.

The third general area I'd like to give some attention to this morning is in many ways the most sweeping and perhaps the most likely to offer immediate rewards in evaluating the outcomes of the college experience. Most colleges, after all, while they may talk in idealistic and noble terms, would be overjoyed if they just knew how to maximize subject matter knowledge among their students. And this is an outcome that we think we're pretty good at measuring. But given our relative expertise in the design of instruments to assess factual knowledge and academic achievement, we have not even scratched the surface of measuring changes in students as a result of college instruction. Longitudinal studies are woefully lacking, as are comparative studies. Everyone knows that Harvard graduates score high on achievement measures but so, after all, do their entering classes. What is important is how much growth have they shown and what is there about Harvard that maximizes this growth or fails to. National Merit is talking about the "cannon ball" and has some evidence to indicate that it doesn't make much difference where the very bright student goes to college; his rate of academic growth is about as good in the less academically oriented college as it is in the colleges known for their academic programs. On the other hand, there is reason to believe that the students in some colleges show considerably more growth than those in others. Surely, the identification of such colleges would give us rich sources of information about other variables which might contribute to the greater growth of students. Although my views do not necessarily reflect those of my sponsors, I am excited to think of the information which could be gained by a growth study at a number of colleges which would use achievement batteries such as CCT or GRE to measure changes in academic growth, college environment measures--such as CUES--to identify college press, and student characteristics scales--such as the College Student Questionnaire--to identify groups of students profiting or not profiting from a given set of circumstances.

Since all three of these scales would seem to have promise in tackling some of the measurement problems of higher education and since they are all ETS-managed instruments, I'd like to spend the rest of the time this morning describing the work on these instruments and getting some reactions from you regarding the next steps in assessing collegiate achievement.

Summary of Mr. Richard Peterson's Description of the
College Student Questionnaire

The College Student Questionnaire is designed to enable colleges to obtain comprehensive descriptions of groups of students--their backgrounds, attitudes, values, activities, and aspirations. The instrument has stemmed for the most part from the perspective of sociology. For example, items about family background serve to locate students' origins in the social system--to which a host of other observations may be related. Other questions have to do with educational and career aspirations, patterns of political and other attitudes and sentiments, types of student subcultures or orientations toward their higher education, and other related matters.

The section pertaining to the identification of student subcultures was suggested by a typology proposed by Burton Clark and Martin Trow. In this section, the student is asked to describe his orientation to higher education, classified in four subcultures identified positively or negatively on the two axes: involvement with ideas and identification with the college. Students in the academic subculture, for example, are positively involved with ideas and positively identified with their college whereas the vocationally oriented student views his college experiences as a pathway to a better job and is not particularly interested in either ideas or in his college as such. The nonconformist identifies positively with ideas but negatively with the college; and the collegiate subculture positively with the college but negatively with ideas.

Preliminary findings show that colleges differ in the proportion of subcultures present on their campus, but that the collegiate subculture certainly predominates in most colleges.

In developing the CSQ a conscientious attempt was made to develop an instrument that could be used throughout the great range of American colleges. It may be

used in a variety of studies directed toward more adequate descriptions of groups of students, e.g., freshmen class, senior class, "under achievers," "dropouts," student leaders, fraternity men, engineers, honor students, etc. Since the CSQ consists of two questionnaires--a pre-college form and a during-college form--changes in student characteristics and orientations can also be studied.

In summary, the CSQ is an institutional research tool, consisting of two questionnaires, comprehensive norms, and an economical system for tabulating, scoring and reporting.

Summary of the Comprehensive College Testing Program

The Comprehensive College Tests consist of two types of examinations: the General Examinations and the Subject Examinations. The General Examinations provide a comprehensive measure of undergraduate achievement in the five basic areas of liberal arts education--English Composition, Mathematics, Natural Sciences, Humanities, and Social Sciences--History. After three years of developmental work, the General Examinations became available to colleges for the first time last spring. They have found wide acceptance for a variety of purposes. Examples of present usage include all of the following:

Credit for independent study:

1. American University granted from 3-36 hours of college credit to entering freshmen on the basis of their performance on the General Examinations.
2. The Commission on Accreditation of Service Experiences has recommended the General Examinations as a replacement for the Tests of General Educational Development--college-level. CASE will determine credit recommendations to colleges in November, 1964.

College equivalency:

1. The Port of New York Authority administers the program to employees wishing to qualify for higher positions.
2. Some graduate schools have recognized the General Examinations as Bachelor's degree equivalents for students with non-traditional undergraduate backgrounds.

Transition to upper-division studies:

State Universities of Florida have conducted experimental testing studies involving their own sophomores as well as those seeking to transfer from other colleges.

Curriculum evaluation:

Both Junior and Senior colleges have participated in two types of evaluation studies; longitudinal growth studies and comparative studies utilizing the norms established for representative national samples of freshmen and sophomore college students.

Six Subject Examinations were developed over the past year for the following courses:

Analysis and Interpretation of Literature
General Chemistry
Introductory Calculus
Introductory Economics
Tests and Measurements
Western Civilization

The norming was completed last spring by administering the tests to a random sample of college students completing one year of college study in the area. The norms are presented in both the traditional percentile rank form and also by grade-in-course distributions. The latter form is designed to give colleges a basis for determining the levels of proficiency which might be expected of students achieving various grades in college courses. Although it is too early to report usage of the Subject Examinations, they were designed as proficiency examinations for students studying independently, by TV or correspondence, or any other non-traditional form of college-level study.