When it comes to identifying excellence in undergraduate education, in common parlance, "we have good news and we have bad news." The good news is that an increasing number of intelligent, dedicated, and articulate individuals are involved in major efforts to identify quality in undergraduate education. Given public calls for accountability, and the fact that students and their parents bear a major burden in financing the postsecondary enterprise through tuition and/or tax dollars, this undertaking is most welcome. The bad news is that some of these efforts ignore the vast evidence on college impact and are based on a naive understanding of just how difficult it is to accomplish in a valid manner what they claim. In this article, I want to critically examine three of the most common ways of identifying excellence in undergraduate education: 1) institutional excellence as reputation and resources; 2) institutional excellence as student or alumni outcomes; and 3)
A more serious problem with the national magazine rankings is that from a research point of view, they are largely invalid.

Institutional Excellence as Resources and Reputation

Undoubtedly the most visible approach assumes that excellence in undergraduate education is a direct function of institutional resources and reputation and finds its popular expression in national magazine rankings of the nation's so-called "Best Colleges." These rankings typically take various measures of subjectively-judged reputation and institution-reported educational and human resources (such as endowment, expenditures per student, faculty salaries, student test scores, selectivity, and the like), and form them into a weighted composite score. The institution with the highest composite score wins, and all other institutions line up behind it in a ranked-ordered queue based on their own composite scores. The logic of such rankings appears to be that easy-to-measure factors like institutional resources and how hard it is for a student to get admitted are valid proxies for the quality of the actual education students receive. Lots of resources, plus selective admissions, equals "excellence" in undergraduate education.

The many shortcomings of the resources and reputation approach have been frequently and trenchantly pointed out by others (see Resources box). From a social scientist's perspective, though, I would argue that two major problems largely invalidate the resources and reputation approach as a method for identifying institutional excellence in undergraduate education. The first, and lesser of the two, has to do with the assumption that the national magazines can actually rank institutions in a way that implies there are discernible differences in the "quality" of institutions that are relatively close to one another in the queue. What's the real difference between the 5th- and the 20th-ranked institution, or between the 20th- and the 35th-ranked school? Is there any real difference at all?

My office in the Lindquist Center at the University of Iowa is about 150 feet from Iowa Testing Programs, which produces the Iowa Tests of Basic Skills, widely regarded as a gold-standard among national achievement tests. Unlike the magazine rankings, Iowa Testing Programs directly measures things like reading comprehension and mathematics achievement through actual samples of behavior, not with distal proxy variables. Yet, despite over 70 years of doing so, they emphatically reject the practice of ranking individuals or schools based on test scores. Their reasons center on issues of the reliability and validity of any score. Not only does every measure possess a considerable margin of error (unreliability) in any score, there is also the problem of whether it taps all the important dimensions of the trait being measured. Bob Brennan, the director of Iowa Testing Programs, even has a coffee cup that candidly sums up their view on the issue. It reads: "We Do Precision Guesswork."

The contrast here is telling. Iowa Testing Programs is in the business of measuring learning, and they measure it directly. Nevertheless, they eschew ranking individuals and schools because of the danger of portraying differences among them that may not be real. On the other hand, the national magazines measure undergraduate educational excellence with what are at best proxy indicators of actual student learning. Yet they have sufficient hubris to assume they can rank order institutions as though individual differences in rank represent discernible differences in educational quality. What is most insidious is that the public, and a substantial part of academia, frequently behave as though they believe these rankings reflect reality.

A more serious problem with the national magazine rankings is that from a research point of view, they are largely invalid. That is, they are based on institutional resource and reputational dimensions, which, at best, have only minimal relevance to what we know about the impact of college on students. Over the last 15 years, my colleague, Pat Terenzini, and I have spent a considerable amount of time reading and synthesizing literally thousands of studies on the factors that are important in shaping a college's impact on its students. If I were forced to summarize that research in a single statement, it would be that within-college experiences tend to count substantially more than between-college characteristics. The quality of teaching, the extent and nature of interaction with faculty and peers, the effectiveness of student affairs programming, the focus and intensity of academic experiences, and the overall level of student engagement, to name several important dimensions, are much more important in defining excellence in undergraduate education than the reputation, selectivity, or resources of the institution attended. This is not to say that such factors as student body selectivity or resources have no role in shaping institutional impact. In some situations they do, at least indirectly. However, the weight of evidence indicates that their impact is substantially less than what a college does with the students and resources that it has.

The fact that the explanatory variables on which national magazine rankings are based don't actually "explain" very much would be sufficient to drive a stake through their heart, if we were functioning within the accepted rules of evidence of scientific disciplines. But, as has often been said, the academy has the unfortunate tendency to apply scientific standards of evidence to every field of study except itself. Indeed, when it comes to the national rankings, a virtual avalanche of evidence of their invalidity will probably have little impact on how institutions behave. As Don Hossler cogently points out, institutional leaders actually reinforce the credibility of the national magazine rankings, and do their own schools a disservice, when they promulgate policies directed at improving their institution's yearly rank in the queue, instead of focusing on
improving the educational experiences that make a genuine difference for their students.

At a recent national conference on higher education, I was part of an informal discussion on what would have to happen to bring an end to the obsessive focus of many colleges and universities on the national rankings. One suggestion, borrowed from the Knight Commission on intercollegiate athletics, was that a consortium of prestigious, national institutions, such as the 60 plus members of the American Association of Universities (AAU), would have to agree as a group not to participate. At first blush, this sounds like a reasonable strategy, but it has an obvious flaw. It is members of the AAU, including my own institution, that benefit most from the national magazine rankings because of their ample supplies of resources and reputation. Asking such institutions not to participate in the national rankings contest when they have most of the advantages would be roughly the equivalent of asking Congress to vote itself out of power. Iowans, appropriately, have a characteristic estimate of both possibilities: "When pigs fly."

It is thus unlikely that the national magazine rankings of colleges and universities will disappear any time soon. Yet, intellectual honesty, not to mention truth in advertising, should dictate at least two changes in how they conduct their business. First, they should stop pretending that individual rankings represent meaningful differences in the quality of undergraduate education. Everything we know about the difficulties inherent in accurately measuring any social phenomenon—and particularly one as comprehensive and ephemeral as excellence in undergraduate education—indicates that the magazines simply cannot make such a claim. To continue to portray the top-ranked institution as discernably better than the 20th-ranked institution which, in turn, is clearly better than the institution ranked 40th, is egregious misleading. A more honest approach might be to take the top 40 or 50 institutions and, within that group, list schools alphabetically. Of course, even this, or any similar, approach is still plagued by the problem of a potentially unreliable "cut score."

How do we know, for example, that the 50th-ranked school, which would be included in the top group, is really any different from the 51st-ranked school, which would be excluded?

Second, the national magazine rankings should stop pretending that they are actually identifying the "best colleges" with respect to undergraduate education. Since their measures of what constitutes the "best" in undergraduate education are based primarily on resources and reputation, and not on the within-college experiences that we know really make a difference, a more accurate, if less marketable, title for their enterprise might be "America's Most Advantaged Colleges."

**INSTITUTIONAL EXCELLENCE AS STUDENT/ALUMNI OUTCOMES**

Acknowledging that institutional resources and reputation don't tell us much about excellence in undergraduate education, an alternative approach is to infer institutional quality or excellence on the basis of what alumni know or what they have accomplished in their lives. On first reading, this outcomes-oriented approach seems to make a lot more sense as a measure of quality than an institution's resources or reputation. If an institution provides a truly "excellent" undergraduate education, it should follow that the results of that education should be visible in things like the cognitive capabilities, activities, and accomplishments of its alumni. But unfortunately, the alumni outcomes approach is also fatally flawed from a methodological standpoint. In fact, it may reveal little more about the actual educational impact of an institution than the resources and reputation approach. Furthermore, the alumni outcomes approach is particularly insidious in its potential to mislead, precisely because it sounds like a very rational way of identifying educationally influential institutions.

The major problem with making inferences about the quality of undergraduate education from differences in the capabilities or accomplishments of alumni from different institutions is that one has to assume that all institutions start with the same kinds of students. As an overwhelming body of evidence clearly indicates, however, there are huge differences in the characteristics of students who select, or who are selected by, different kinds of institutions. And, if we know anything from the field of educational psychology, it is that input is the best predictor of output. An individual's characteristics when he or she begins an educational intervention are far and away the best predictors of outcomes on completion. For example, by far the best predictor of students' Graduate Record Examination (GRE) scores when they graduate from college is their ACT/SAT scores when they enter college.

So, if we were to predict the average GRE score of seniors at 100 institutions with the average ACT/SAT score of the same students entering each institution four years prior to that, it is likely that the correlation would be in the .85 to .95 range, conceivably even 1.00. An analogous argument can be made for almost any kind of outcome (such as test score performance, principled moral reasoning, educational attainment, career success, community involvement), although the resulting correlations will likely vary in size.

Finding discernible differences in alumni outcomes among a given set of institutions thus has a very high probability of being simply a result of differences in entering-students' characteristics. It will probably not be a valid indicator of differences in the actual educational impact of these institutions. Unless alumni outcome differences among institutions are adjusted for student input differences among institutions, valid inferences about institutional effectiveness in undergraduate education are not likely to emerge. When it comes to identifying institutional excellence in undergraduate education, therefore, simply comparing alumni outcomes is not much of an improvement over the resources and reputation approach.

Even if we were to assume the alumni outcomes approach has validity (and only for the sake of argument), there is another nontrivial problem. How would we decide what outcomes to measure? What particular set of competencies, activities, and accomplishments can actually be attributed to the undergraduate experience? Again, the expertise of my colleagues in the Iowa Testing Programs is pertinent. Deciding what to measure is by no means a trivial enterprise. Determining the construct or content validity of any measure or set of measures can be a Herculean task, particularly if one is using those measures to infer differences among individual educational settings like colleges and universities.

Because of their recent prominence, it is also worth mentioning that the problems inherent in the alumni outcomes approach are also a liability in state-level proposals to monitor the educational effectiveness of public postsecondary institu-
If the effective educational practices approach leads to comparisons among institutions, representative samples of student responses from each institution are a *sine qua non*.

Institutional Excellence as Effective Educational Practices or Processes

A third approach to identifying institutional excellence in undergraduate education is to try to assess the practices and processes within a college or university that we know are linked to important cognitive and noncognitive outcomes. This approach differs fundamentally from looking at student/alumni outcomes in that it does not simply use the tested outcomes or accomplishments of graduates to indicate educational excellence. Instead, it focuses on those institutional practices that have been shown to influence real outcomes or accomplishments in carefully controlled studies. While certainly not a perfect methodology, this approach does have some conceptual strengths that render it decisively preferable to either the resources/reputation or the alumni outcomes approaches. Perhaps its major strength—essentially absent in the other two approaches—is that it attempts to focus directly on student experiences that define an impactful undergraduate education (for example, the focus and quality of teaching received, interactions with peers and faculty, writing experiences, involvement in coursework, level of academic and social engagement, and the like).

The assumption here, and it is not an unreasonable one, is that an excellent undergraduate education is most likely to occur at those colleges and universities that maximize good practices and enhance students' academic and social engagement or effort. If such an approach were eventually to lead to institutional comparisons—and we have not been very successful in resisting this temptation so far—at least the comparisons would be made on the basis of indicators that really matter, instead of on proxy measures of highly questionable validity. Given institutional leaders' fascination with ranking, in fact, one potentially beneficial byproduct of the effective educational practices approach is that it might refocus their attention on the educational processes that make a difference in the growth and development of students and prompt them to improve.

Although the effective educational practices approach is, I believe, a major step in the right direction, several criteria need to be met in order to maximize its validity as a method for identifying institutional excellence. First, if you are going to measure "effective practices," it is imperative to measure practices that are empirically, and not just rationally, linked to student growth and development. This means that the effective educational practices approach needs to be guided by the body of evidence on college impact, and not just by factors that *seem* as if they should be important. Fortunately, this body of empirical evidence exists and can be used as a template to help identify which practices to assess.

Second, because the effective practices approach relies on the information provided by random samples of students within each institution, a high response rate is critical to ground valid inferences. Adjusting for the nature of the information nonrespondents might have provided, but didn't, is one of the most difficult problems in the social sciences. And while sample response rates in the 40 percent to 50 percent range might be permissible when one is searching for general associations among the variables that constitute a large research database, one would question the appropriateness of response rates in this range to make credible comparisons among institutions. Moreover, vastly different response rates among participating institutions can introduce yet another source of serious and unwanted bias to any institutional comparison. In short, if the effective educational practices approach leads to comparisons among institutions, representative samples of student responses from each institution are a *sine qua non*. Thus, every effort needs to be made to ensure consistently high response rates among all participating colleges and universities. Response rates in the neighborhood of 70 percent to 80 percent, or higher, would be highly desirable.

Third, and finally, a lot of caution must be used when making institutional comparisons based on student self-reports—especially if they require a substantial level of inference on the part of the respondent. Most evidence suggests that students are quite accurate in reporting low-inference, factual data (such as Baird, 1976; Trusheim, 1994). Thus, there is probably not a great problem in asking students about the things they do and the amounts of time that they spend doing them (for example, number of papers written, amount and type of extracurricular involvement, interactions with faculty, time spent in preparing academic assignments, and the like). But, making comparisons among institutions on less "factual," higher-inference items, like the quality of teaching received, the nature of the institution's intellectual/academic environment, or self-reported growth in college, can be potentially misleading. It isn't that self-reports about such matters fail as reasonably reliable and valid indicators of students' perceptions. In fact, they have modest but consistently positive correlations with objective measures, and have been employed as worthwhile indicators to help answer a range of important research questions. Rather, the problem with such high-inference judgments on the part of students is that they substantially increase the danger of confounding estimates of real institutional impact with the characteristics of the students who are attending the institution.

As an example, consider a hypothetical situation in which undergraduate students at Liberal Arts College A report a more challenging intellectual environment and greater intellectual growth than do students at Liberal Arts College B. One conclu-
tion might be that College A provides a more influential undergraduate intellectual experience than College B. But, an equally plausible explanation, however, is that College A simply enrolls students who enter the institution more open to the impact of a liberal arts education to begin with. Therefore, the actual quality of the education provided, and the value added, by the two colleges might be indistinguishable. In order to determine whether the observed difference in student self-reports is an institutional effect or simply a recruitment effect, we would need to control for the influence of salient precocleage differences among students. However, this is a much more difficult task when the relevant data are self-reports rather than more objective measures, though. In the latter case, a pretest usually correlates highly with a post-test and can provide a strong statistical control for differences in individual student characteristics.

The bottom line is that if the effective educational practices approach is used to make comparisons among institutions, the validity of those comparisons will be enhanced if student responses are based as fully as possible on low-inference measures. The more student responses require high-inference judgments, the greater the danger of confusing institutional differences in educational impact with differences in the characteristics of the students enrolled.

**Conclusion**

My original question implied in the title of this article is whether we are “even close” to identifying institutional excellence in undergraduate education. My answer is that we are moving closer. I argue that the resources/reputation approach, which uses proxy variables of questionable validity, and the alumni outcomes approach, which may simply reflect institutional recruitment practices, are fundamentally flawed as methodologies for identifying institutional excellence. I further maintain that the effective educational practices approach—if carried out with a sensitivity to empirical, methodological, and measurement issues—is a major step forward. Hopefully, however, we will not stop here. It may be possible to bring about something of a synthesis of these three approaches that would provide a level of validity in identifying institutional excellence far beyond that achieved by any single approach.

I'm not really saying anything new here; Alexander Astin has been arguing this for decades with his Input-Environment-Output model. For example, with appropriate adjustment for student inputs (precoclege characteristics), the alumni outcomes approach becomes a much more viable method for identifying particularly influential undergraduate institutions. Similarly, if carried out periodically, and on a broad scale, something akin to the Input-Environment-Output approach could be invaluable in updating knowledge about those educational practices that predict important outcomes, even when controls are made for student input characteristics. Certainly the logistics (not to mention the required funding) of conducting this type of longitudinal assessment on a national scale are nontrivial. But so, too, are the potential rewards. The road ahead is daunting, but it is also rich with possibilities for gaining a better understanding of what really constitutes institutional excellence in undergraduate education.