

Discussion Draft 7-06-06

***Elements of Accountability for Public
Universities and Colleges***

Peter McPherson, President

and

David Shulenburger, Vice President for Academic Affairs

National Association of State Universities and Land-Grant Colleges

INTRODUCTION

Our earlier paper, *IMPROVING STUDENT LEARNING IN HIGHER EDUCATION THROUGH BETTER ACCOUNTABILITY AND ASSESSMENTS*, explored the nuances of public university accountability. It has enjoyed exposure and stimulated conversation, particularly among NASULGC and AASCU Presidents, Chancellors and Provosts. This paper comes out of these conversations and further develops the elements that appear to be important in a true national system of public university accountability.

What problem are we attempting to solve with this paper? Every public university is engaged in a serious and ongoing accountability appraisal with significant time and resources dedicated to the task. The academy's commitment to accountability is real and we should probably say that more often. Unfortunately, universities are regularly asked to report on additional accountability measures because influential public officials from time to time are not persuaded that the ones currently in use are sufficiently useful to serve their needs. Many worry that this never-ending cycle will some day result in the imposition of federal accountability standards that will inhibit university educational innovation, not be changeable as lessons are learned and in the end not really satisfy the public's desire to know what we are accomplishing.

If the academy can agree on a meaningful but manageable small set of accountability measures, we should be able to generally satisfy the public by providing more useful information. Also we should be able retain for the academy the ability to choose measures that we can use to improve our ability to educate our undergraduate students. There would be economy and value in that result.

Accordingly, our goal is to suggest here a reasonable and helpful set of accountability measures for universities. It contains three elements of accountability data: consumer data, campus climate or student engagement data, and educational outcome measurements. The set certainly does not contain all of the accountability information one might imagine but it does contain key measurements that should help stakeholders assess undergraduate student learning and provide other useful information. Transparency in collecting and presentating this information is important if we hope to have an accountability system acceptable to the public. We urge that accountability data be used only to compare specific universities with their own past performance and with the performance of comparable universities. We of course know that public data is just that and will be used by others. Finally, we think that regional accreditors will want to consider using the set of accountability measures to help satisfy their standards.

Public universities have significant graduate programs not discussed here. Universities currently report on those graduate programs and should continue to do so.

We suggest a voluntary standard of accountability as opposed to a government or other outside system. No doubt there would need to be some community process to monitor and make appropriate modifications as time goes on.

Because of the complexity of the process of measuring outcomes, individual universities would report accountability data no sooner than two years after the university had begun the process.

We begin with the set of first principles that our correspondents felt to be important.

FIRST PRINCIPLES ABOUT ACCOUNTABILITY MEASUREMENT DERIVED FROM THE DISCUSSIONS

- 1- Because we are educational institutions our first obligation is to develop and use those accountability measures that provide the maximum assistance in improving the educational outcomes of our students.
- 2- The interests of higher education are served best if our accountability measures focus on the increase in educational value for which we are responsible and not on the initial intellectual and educational endowments our students have when they arrive at our universities.
- 3- All of our accountability measures should be transparent. By that we mean that both the methods by which they are derived and the results of those measures should be publicly disclosed.
- 4- The immense diversity of American higher education institutions is a valuable attribute of the U.S. system of higher education. For that reason universities should be compared against their own past performance and with other universities with similar missions, academic programs and admissions practices. Even comparable university comparisons should be limited to individual accountability measures, not indices composed of multiple accountability measures.
- 5- Because our resources are used best when directed toward our educational, research and service missions, only those resources absolutely required should be devoted to this enhanced accountability effort. Consistent with this:
 - The rigorous set of accountability measures proposed here should be evaluated by accreditors and governing boards and where possible, in order to help contain cost, substituted for the accountability measures they now use.
 - Where appropriate, sampling should be employed in measuring outcomes and campus learning climate rather than subjecting every student to measurement at significant costs.

CONCLUSIONS REACHED ABOUT THE GROUPS TO WHOM WE ARE EXPECTED TO BE ACCOUNTABLE

There was widespread agreement that we should be and/or are expected to be accountable to three groups:

- 1- Prospective students, current students and their parents,
- 2- Faculty and campus support staff, and
- 3- Public policy-makers and funders of higher education.

1-Prospective students, current students and their parents. The impact of higher education lasts a lifetime. In choosing a university for undergraduate study, students and their parents have made a decision that will have fundamental effects. We owe to them the kind of information that would be helpful in making this choice. Our *IMPROVING STUDENT LEARNING...* paper called this kind of information “consumer information.” When one reflects on the questions asked by students and parents what is meant by consumer information becomes clear. *How much will it cost to go to your school? Are there costs beyond the required base tuition? What are they? What kind of jobs do your graduates get and how much do they make? How much does it cost to live in the residence hall? May I live off campus during my freshman year? How long will it really take me to earn a degree? How many students fail to receive a degree? Are student jobs available on campus?* These questions and many others are regularly directed to admissions personnel, presidents, provosts and other faculty.

In a world in which the burden of paying for the cost of undergraduate education increasingly falls on students or their parents, we should provide clear and concise answers to their questions such that they can compare institutions by the measures most relevant to them. Much, but not all, of what they want is the consumer information. The ultimate result of providing such information should be to improve the fit between student and university, resulting in greater satisfaction and improved educational success for the student. Improved initial fit should reduce transfers among four-year colleges and ultimately reduce the cost of education to the student.

2- Faculty and campus support staff. Because improving student learning is our primary goal we must be accountable to those who teach and interact most with our students. Faculty and staff need feedback about the campus climate and their success in educating students. The kinds of data from the National Survey of Student Engagement (NSSE) or the Cooperative Institutional Research Program (CIRP), which we label “Campus Learning Climate Data,” help faculty and staff to know whether their students are being expected to read, write and participate in class discussions as much at their school as at comparable universities. The data helps tell whether their university’s environment creates students reasonably engaged with their institutions. This information helps benchmark campus and faculty practices against those in comparable universities.

In this more controlled environment, change can be introduced and the impact of that change on student engagement with the campus and learning evaluated.

Information on student learning, as measured by the Collegiate Learning Assessment (CLA) and like instruments, provides helpful data to faculty and staff. Also, providing key measures of educational effectiveness at other institutions is valuable in determining if curricular innovations, changes in methodology and revised expectations are likely to have the desired results. Clearly, providing this kind of information to faculty and staff is important to improving the university.

3- Public policy-makers and funders of higher education. Public policy-makers and funders can choose to bestow their attention and funding on higher education or on the institutions and myriad entities that compete for funding with universities. Clearly included in this grouping are legislators and members of boards of trustees but also foundation board members and, importantly, alumni. Each wants to know the conditions under which the public is provided access to higher education, who is receiving the benefits and something about how much education is occurring.

COMPONENTS OF ACCOUNTABILITY AND THEIR ATTRIBUTES

Our conversations generated interest and support for the conclusion that each of these three groups can and do expect accountability. We suggest that a bundle of information could provide that accountability to each group if it included 1) consumer information, 2) campus learning climate data of the sort found in NSSE or CIRP and 3) educational outcomes information such as that generated by the CLA. Our respondents observe that we currently generate most, and in some universities, all, of this data. Indeed, every NASULGC and AASCU member generates at considerable expense enormous amounts of accountability information for their own needs, for the needs of specialized and regional accreditation agencies, for governing boards, for state and federal legislatures and agencies, for granting agencies, etc. Why then are the three groups to whom we agree that we owe the duty of accountability not satisfied?

We conclude that the accountability data our institutions generate fail to satisfy many stakeholders for two important reasons: 1) The data and/or the presentation of the data vary from institution to institution and therefore do not permit comparison of like institutions and 2) the data for some or all of the accountability measures are sometimes not made publicly available.

Why is the data not comparable?

Universities for many reasons gather accountability data by various means using many different definitions. Some of the reasons for the variety of data used are simply mechanical in origin, e.g., no commonly accepted definition of a data element has been agreed upon. We can deal with the mechanical problems by agreeing upon definitions and measurement practices. Similarly, the lack of agreement on process and outcomes measures are starting to be dealt with through the emergence of standardized measuring instruments like NSSE, CIRP and CLA.

Other reasons have to do with the inherent differences across our institutions that make common measurement difficult, e.g., liberal arts universities simply do not have graduates taking the pharmacy licensing exam and cannot generate data on pass rates; dance conservatories may not wish to measure the math ability of their students because math skill development is not part of their curricular intent. We need to deal with these issues by gathering data that is of common value and interest, fully realizing that there is more specialized data that individual schools will want to collect for their own use.

Why are accountability data not made public?

Sometimes there is a fear that incomplete or poorly conceived data will be misleading. A good example of misleading data is the graduation rate data required to be reported by the federal government. Even though the definition used to report graduation rates is common because it is specified by the U.S. Department of Education, concern remains because graduation rates as officially defined reflect only the experience of those who enter institutions as true freshmen (transfer students are excluded). The result is that those students who enter as freshmen and later voluntarily choose to transfer to other universities reduce graduation rates just as much as do those who simply drop out of college or fail. Additionally students who transfer into the university and earn degrees do not improve its overall reported graduation rate.

This concern affects universities differentially. Urban institutions, for example, often admit only a minority of their students as true freshmen while traditional state universities generally admit the majority of their students as freshmen. The reported graduation rate reflects the performance of a small minority of students at the former set of institutions and the majority at the latter. Universities for which the graduation rate does not reflect the actual experience of their students generally would rather not have the data reported as the data do not appropriately reflect the school's performance of its mission. This government data requirement has not served anyone very well and that experience hopefully would be in contrast to a university community system with a built-in process to make adjustments as lessons are learned.

Many universities view the results of NSSE/CIRP and CLA as diagnostic in nature and choose to use them to help guide the internal agenda for change rather than have them as a public indicator of their performance. Many universities that use such measurement instruments do report their scores. Many universities, especially those that have previously not had the benefit of information from such instruments, will want to have an opportunity to adjust their practices and operations once they administer the instruments. Accordingly, we recommend that data like this not be made available during the first two years of the operation of any nationally agreed upon accountability system.

In other cases universities simply feel that a single measure fails to reflect the full performance of the school and they choose not to release individual data elements that might in isolation be used inappropriately to measure performance. It is indeed true that any measure of student learning will fail to reflect the complex package of learning, leadership, independence, other life skills, etc., that the university experience produces.

What we need to discuss is the components of a reasonable, but not too complex, system of data.

We certainly do understand the issues about making such information public. If we could work out a set of information that is fair and helpful for both the universities and the stakeholders, probably most would agree that information should be made available to those stakeholders. As we step back and consider all of this we suppose that the real point is that we can not sustain the trust of others unless we are prepared to provide helpful information to the public.

What is suggested is essentially a bundle of data whose individual elements deal with legitimate requests. The hope is that those that use the elements will be able to gain a fair and useful view of important facets of each participating university. We probably can agree on a set of measurements and data but it will take some careful and hard work. It is critical is that we have a system that the university community can modify as we learn lessons, which we certainly will as time goes by.

Only appropriate comparisons should be made. Universities will want to consider using the same format and terminology for the core of accountability measures and public reporting. The best comparison is of the university with itself over time. Such comparison can be helpful to the individual university.

We strongly argue that there is no intelligent way to enable the comparison of institutions that are essentially different. Only comparison among *comparable universities* is appropriate. In order to enable appropriate comparison it is necessary to appropriately group universities. We recommend that the data set include figures on both entering average SAT and ACT scores and entering average GPAs. One would then be able to use this data to group universities by admissions practices for the purposes of making these comparisons. We also encourage individual universities to inform the public of the set of schools that they believe to be their peers and encourage comparison with this group.

We vigorously oppose creating any overall ranking scheme based on the bundle of accountability measures we recommend here. Of course, we understand that public data is just that. We advocate that common accountability data be presented by institution with the user of the data encouraged to place whatever weight on the individual data elements she/he prefers. Given that missions, student selectivity, urban/rural location, student body composition, etc., differ by university, combining the accountability data into any form of index for the purpose of ranking these universities would represent a disservice to them. Such indices may also constitute a disservice to students, parents, faculty and staff, and policy-makers and funders as the individual data elements are tailored to answer specific questions for specific groups. Combining data into any sort of index will serve to obfuscate what the structure of the data set proposes to clarify.

A Brief Focus on the Set of Measures

Selection of Specific Accountability Measures

We envision developing a process by which universities desiring to articulate and support a set of public university accountability standards would work with the national associations representing them (or by some other means) to decide upon the specific elements of consumer information, the specific campus learning climate measure to be used and the specific educational outcomes measure to be utilized. This paper proposes no specific mechanism for accomplishing this end but does recognize that there will need to be such a process. Once that bundle of measures is agreed upon, individual universities would then choose voluntarily whether to utilize them.

This process would take some time to develop and the measurement process would take time to implement and universities need time to adjust to results of initial measurements. We do not envision the first complete reports from individual universities subscribing to the national standards of accountability becoming available to the public until at least two years after agreement is reached to proceed with this notion.

Three Components of Public University Accountability

As described above there was interest/agreement that three specific components should be included in a nationally agreed upon accountability data set:

1. Consumer Information
2. Campus Learning Climate
3. Educational Outcomes

1. Consumer information. This accountability category primarily focuses on students and their parents. Accordingly, this data should be chosen after first asking the question: *“What information would be most helpful to prospective students and their parents in deciding which university best fits their educational wants and needs?”* Clearly various cost figures, degree offerings, living arrangements, graduate placement statistics, graduation rates, transfer rates, employment facts (see the Appendix for a discussion of earnings data), etc., would be included here. To enable comparing groups of universities with similar degrees of selectivity in admissions, the data sets would also include entering test scores such as average ACT/SAT and entering GPA.

Any valid accountability data set must include student success measures such as graduation rates. Unfortunately, a true picture of graduation rates cannot be generated from data currently available, as information on where students transfer within the higher education system cannot be reliably ascertained from existing data sets. The academy is short-changed by not having robust data available to develop the complete picture. Therefore we support serious conversations among higher education, congressional and

U.S. Department of Education leaders leading to the development of a data set that will overcome the current limitations on measuring graduation and transfer rates.

2. Campus Learning Climate. Both the NSSE and the CIRP are proven programs that enable campuses to compare student perceptions about the total campus environment for learning. This data has proven valuable to faculty and staff by enabling them to make alterations in both campus and classroom environments to provide greater stimulation to students. One of these measures should be part of the accountability package if faculty and staff are to have the tools needed to alter the learning environment. The value of these instruments is described in some detail in our *IMPROVING STUDENT LEARNING...* paper.

3. Educational Outcomes. Each of the groups to whom we need to be accountable has asked for such data, sometimes often. General education is a primary component of undergraduate educational outcomes and it probably can be measured for the accountability purposes under discussion here. This measurement is not a trivial task because the components of general education differ considerably from university to university. However there do appear to be some common outcomes that general education is expected to produce at most NASULGC/AASCU institutions. These include developing critical thinking skills, analytical reasoning, and written communication. Fortunately, tests like the CLA, Measure of Academic Proficiency (MAP) and the Collegiate Assessment of Academic Proficiency (CAAP) have been designed to measure such general education outcomes. Research on the ability of standardized tests to measure these general education outcomes is ongoing and must continue until sufficient numbers of students at institutions of varying sizes and types have taken the exams. Only then can we reach firm conclusions on the adequacy of the various tests to measure these general education outcomes.

What we do know about performance on standardized tests like the CLA is that it reflects both the initial ability of the student entering the institution and the educational value added by the institution. For accountability purposes, only a measure of the latter is desired. One can directly adjust the performance of seniors on general education assessment exams for entering ability (as measured by ACT or SAT scores when the students were accepted for admission to the university) or one can administer the test the general education instrument twice (freshman/senior years) and use the difference in test scores to judge how much educational value is added by the institution. Of course many things happen to a student between the freshman and senior years that are unrelated to the education provided by the university, e.g., travel, development of a wider social network, and summer and academic year jobs, and each of these may have an effect on standardized test scores. Nonetheless, it is clear that selecting one of the normalization techniques is required to refine the measurement such that it comes closer to approximating only the value added by the university.

We are aware that controversy surrounds the measurement of value-added. Both freshmen and senior CLA test results are highly correlated with incoming ACT/SAT

results, leaving little additional value to be attributed to the university educational experience. This does not diminish our resolve that value-added is the appropriate outcome measure upon which to focus. It does mean that the value-added measure must be sensitive to these findings and that a value-added measure initially chosen may have to be reconsidered as additional research results are amassed.

A serious problem that remains for general education assessment testing is the difficulty of ensuring the students tested are motivated to perform at their best level while taking the tests. This problem is generally not a matter of concern when students are taking classroom exams or the ACT/SAT or GRE exams as the test takers are motivated by self-interest to do well. General education standardized exam performance, by contrast, has no impact on the test taker but may have profound consequences for the university. In a sampling scheme like that employed by the CLA, having only a small percentage of the typical sample of 100 students treat the testing experience as a lark or malevolently decide to do harm to their institution, can seriously bias downward the general education outcome reported for the campus. Various methods have been suggested to correct for this problem ranging from embedding the test into a capstone course with performance on the exam counting as part of the course grade, to awards of iPods to top performers on the exam. To avoid this motivation bias problem some campuses may even decide that they will not use a sampling strategy in measuring general education gains but will instead administer CLA-like measures to every student in a required capstone course. There are many measures available to avoid such bias and we do not endorse any one measure here. It is clear that every university will need to tailor a method to its peculiar requirement to ensure that motivation problems do not bias test results.

Similarly, where sampling is utilized, uniform sample selection practices will have to be implemented to ensure that the tests are administered to a random sample of students. Every campus naturally wants to show its best face and thus there will be the temptation to administer the test to non-random samples of students, e.g., to samples of high-ability students. Unless uniform sample selection procedures are agreed to and rigorously observed on every campus, the willingness of campuses to participate in any general education assessment venture and particularly their willingness to make results public will be undermined.

Unfortunately, there are no standardized tests that might be used to measure campus-wide the value added for the entire undergraduate educational experience. Constructing such a test is impossible as campuses have diverse sets of majors and degree programs. There are, however, outcome measures that indicate overall campus educational outcomes for *specific* purposes. They include the LSAT, MCAT, GMAT and the GRE and specialized disciplinary exams. We are not now recommending using those measures as part of this educational outcome suite because they measure the experience of only very small and highly self-selective samples of the student population at some universities. We also do not know how those these measures relate to general educational outcome measures such as the CLA. Further research could help conclude whether including these measures in the bundle would be justified.

Measurement Frequency

Data on each of the measures certainly does not need to be collected annually. Many users of NSSE administer it every two or three years as campus climate seldom changes suddenly and schools wish to allow time for interventions they implement to have an opportunity to work prior to read ministering the test. Similar considerations may lead to the decision to administer other measures on a biennial or triennial schedule. Explicit decisions about the frequency of testing must be made so that the expense of administering the instruments or collecting the data is commensurate with the value of the data obtained.

FINAL THOUGHTS

The Role of Regional Accreditors. Regional accreditation associations serve a quality assurance role of major importance to American higher education and the public. We wish to enhance and not replace their assurance role with a national set of public university accountability standards. Accordingly, the six regional associations should consider substituting the resulting set of accountability measures for the measures they now require. Doing so enhances the goal of achieving transparency on a national level. We understand that most regional accreditation standards already permit the university seeking accreditation or reaccreditation to choose a variety of ways to satisfy those standards and that this request is therefore minimal. In order to further promote adopting these accountability standards, the regional accreditation associations must be convened at the appropriate time to consider this request.

For the sake of Parsimony. If universities agree on a set of national accountability standards they are agreeing to gather and supply a considerable volume of detailed information on a regular basis at significant expense. University budgets are not plush and all would rather expend as many scarce resources as possible directly on the educational mission. If such measures are put in place we believe the universities using them will have a strong basis for asking state and federal government agencies, governing boards, regional accrediting agencies and others to agree to a moratorium on requests for more accountability measurement and to begin a review of existing accountability measurements with a view to determining whether some might be eliminated. If the provision of consistent, clear and comparable accountability data satisfies the various publics, some of the cumbersome, unfocused and expensive measures that are now in place should be withdrawn. It is our hope that the measures chosen would replace most existing measures, not to add to them.

The Need for a Continuing Process. Agreeing to create a national set of accountability standards and implement them cannot end this process. What is an adequate accountability process today may not be adequate tomorrow. Thus a mechanism must

also be created to provide oversight to the agreed upon set of standards and to permit the system to be responsive to changes in the environment.

CONCLUSION

The set of public university accountability standards suggested here amounts to self-evaluation by those universities that choose to subscribe to it. This self-evaluation system arises from the same tradition as does our voluntary accreditation system that has served both the public and the academy well. It preserves the diversity of the U.S. higher education system while permitting the public to judge the various facets of quality of individual universities.

While these suggestions are based on feedback from a substantial number of correspondents who read the predecessor paper, *IMPROVING STUDENT LEARNING IN HIGHER EDUCATION THROUGH BETTER ACCOUNTABILITY AND ASSESSMENT*, we hold no illusion that it cannot be improved. We ask for your comments and suggestions. Please send them to dshulenburg@nasulgc.org or pmcpherson@nasulgc.org.

Appendix:

The Unique Case of Employment and Earnings Measures

Developing reliable data on employment and earnings of university attendees and graduates for inclusion in the accountability databases is particularly problematic. Current attempts to determine placement success of graduates vary because of the difficulty in tracking increasingly mobile graduates and persuading them to respond to surveys. As a result considerable skepticism exists about reported placement and continued employment success of graduates now reported by various universities. No one doubts that the survey results demonstrating very large and increasing returns for earning a university degree are valid. Those findings are robust and vary little from study to study. Unfortunately these findings are not based on samples large enough to permit estimating the earnings gains enjoyed by graduates of each individual university.

Generating a suitable measure to consistently judge job placement and subsequent employment success for graduates of individual universities will require establishing another national database. An accurate database on employment within the United States and its territories could be developed from data already collected for other purposes. For example, the unemployment insurance system in each state collects information on everyone in insured employment in that state. As upwards of 97 percent of employment is covered by the system and quarterly earnings reports must be filed for covered employees, it is possible to track employment for nearly all college attendees, whether they graduated or not, by the SIC code of their industry of employment. Several states now use this database to track earnings of graduates within their states. The combination of state databases into a single national database would provide reliable employment information for the entire population of university attendees and graduates. The unemployment insurance system does not enable tracking of total earnings paid because employer earnings are reported in the system only up to a preset maximum level. Note also that profits from entrepreneurial activities, equity trading, real estate profits, etc., are excluded from these earnings data and these may be important components of earning for some university graduates. Other federal databases, e.g., Social Security and Internal Revenue Service, contain more complete information on earnings and could be used for this evaluative purpose if it is deemed politically palatable to do so. Note that any such domestic data set will exclude the large and growing number of graduates in this flat world who spend some or all of their careers abroad.

While we do not wish to prejudge the conclusion about the specific employment data that ought to be included in the accountability data set, we do wish to highlight the very real difficulty of generating accurate data on this important outcome of university education. Since data sets designed to evaluate the outcomes for higher education could be used for many purposes beyond higher education evaluation, we believe construction of those data sets is something that ought to be considered by Congress and the administration, not simply by NASULGC and AASCU.