QUALITY IN THE UNDERGRADUATE EXPERIENCE

A Discussion Document

December 4, 2015

This framing document is intended to stimulate thinking about key issues related to the quality of undergraduate education. It elaborates on the five themes identified in the workshop invitation: the measurement of student learning; qualitative factors often cited as important outcomes of undergraduate education; the importance and challenges of assessment; federal policy implications of assessing quality; and the importance of context with regard to institution type, learning environments, and student goals. We don't intend to address all facets of the quality challenge, nor do we mean to suggest that this is the only way to unpack it. Indeed, we are interested in identifying and filling gaps and hearing other perspectives. We hope advance thinking about these (and other) issues will help stimulate a rich and generative discussion at the December 14-15 workshop hosted by the National Academies.

1. **Measures of Student Learning.** *Much of the focus on "quality" in undergraduate education has been on input factors or a variety of outcome measures: reputation, entrance examination scores and admissions selectivity, financial resources, graduation rates, graduates' employment and earnings, and other attributes that can easily be measured but that say little about student learning--that is, the acquisition of important and relevant market-valued knowledge, skills and abilities (KSAs) and the ability to apply those KSAs in real-world settings. How can we change that? Are there approaches and metrics that can accurately speak to student learning?*

When reflecting on measures of student learning, one can focus on the specific course experience, formative in nature, or on the overall academic experience, a more summative departmental or institutional perspective. In this section we choose to focus on the latter. Traditional measures such as graduation, retention, campus resources, graduate employment and more, while relatively straightforward to measure and of value, are limited in their ability to provide evidence of student learning. Broader and deeper methods and approaches are needed to help departments and institutions clearly define the value they provide their graduates. Here we highlight three specific approaches, noting that there are many others. We discuss the Association of American Colleges and Universities (AAC&U) VALUE Rubrics, the PULSE Vision and Change Rubrics, and the Degree Qualifications Profile (DQP). These three approaches, outlined below, can help us think about the dimensions of the problem and how a substantial number of institutions are moving forward in an iterative journey of self-exploration to define the value they bring to the student learning dimension.

The AAC&U sponsored VALUE Rubrics (Valid Assessment of Learning in Undergraduate Education) provide tools to help assess students' work produced across the students' varied learning pathways and institutions, "to determine whether and how well they are progressing toward graduation-level achievement in learning outcomes that both employers and faculty consider essential." Dimensions considered by VALUE include intellectual and practical skills (inquiry and analysis, critical thinking, creative thinking, written communication, oral communication, reading, quantitative literacy, information literacy, teamwork, and problem solving); personal and social responsibility (civic engagement, intercultural knowledge and competence, ethical reasoning, foundations and skills for lifelong learning, global learning); and integrative learning. Rubrics were developed by faculty and other professionals from more than 100 institutions. According to the project website, "Each rubric was developed from the most frequently identified characteristics or criteria of learning for each of the 16 learning outcomes" (http://aacu.org/value/rubrics). A recent multi-state collaborative is looking at student work from 69 participating 2- and 4-year campuses. In addition, the Voluntary System of Accountability (VSA; http://www.voluntarysystem.org) includes two VALUE rubrics on the list of assessment tools that participating institutions can use to demonstrate student learning.

The PULSE Vision and Change Rubrics provide a structure for departmental reflection, self-assessment and discussion regarding a host of topics relevant to program transformation. While the focus has been on the life sciences, the process is equally applicable for any STEM field. The current rubrics have criteria immediately applicable to all STEM fields except for the disciplinary core concepts, which are available for life sciences only. The rubrics and suggested activities suggest a process by which faculty and appropriate learning and technology staff work collaboratively to maximize their collective transformative change.

The Degree Qualifications Profile (DQP) helps frame an institution's mission and overall goals when granting degrees. The DQP identifies five essential areas of learning that should be incorporated in any post-secondary degree, with increasing complexity based on the degree obtained. The five areas are specialized knowledge, broad and integrative knowledge, intellectual skills, applied and collaborative learning, and civic and global learning. Specialized knowledge outlines what students in any specialization should demonstrate with respect to the specialization, or major, with proficiencies within each field determined by each field through a process called "Tuning" to describe the particular concepts, knowledge areas, methods, skills and accomplishments necessary. Broad and integrative knowledge asks that students are able to consolidate and utilize knowledge across multiple areas to discover and explore questions that span multiple fields of study. Intellectual skills are defined as evidence-based reasoning across fields of study and include: analytic inquiry and operations, use of information resources,

engaging diverse perspectives, ethical reasoning, quantitative fluency, and communicative fluency. Applied and collaborative learning focuses on how students can utilize what they know to innovate and move beyond classroom level work as individuals and in groups. Civic and global learning refers to student preparation to engage and contribute to political, social, environmental and economic challenges. Overall the DQP asks university stakeholders to engage in a process of ensuring that students are both competent (can demonstrate a certain level of skill in a course/experience) and proficient (summative ability gained through multiple course experiences with commitment to ongoing learning) as relevant to their field of study and level of degree attainment.

These are just three example approaches, by no means meant to be limiting. There is a good deal of other work on 21st century learning skills, including the Next Generation Science Standards and the P21 Framework for 21st Century Learning, among others. Traditional standardized test approaches are also being used to assess more diffuse and generic learning outcomes. For example, in addition to VALUE rubrics, institutions participating in the VSA can demonstrate learning gains using ACT's Collegiate Assessment of Academic Progress, the Council for Aid to Education's Collegiate Learning Assessment, or ETS's Proficiency Profile.

The assessment of learning is complicated by the emergence of new providers and new approaches to provision such as competency-based learning, problem-focused field experiences such as internships, and other programs, Do these change the ways we should approach the definition and assessment of learning? If we say we are assessing "competency" and "proficiency," how do they differ and how does that affect the validity of the assessment tools that are chosen?

Regardless of the specific approach, there is increasing consensus that student learning needs to be at the core of our thinking about educational quality as educators, administrators, taxpayers and global citizens.

2. **Qualitative Factors.** *"Not everything that can be counted counts, and not everything that counts can be counted." – attributed to Albert Einstein. How do we begin to define, identify and measure the qualitative elements of a high-quality undergraduate education?*

A high-quality undergraduate education involves more than the accumulation of factual knowledge and intellectual skills. There is wide agreement that it should also inculcate a range of diffuse skills and habits of mind that prepare students for lives of engaged citizenship, intercultural competence, social responsibility, and continued intellectual growth. While some of these capacities and habits are addressed by the VALUE rubrics

described above, their measurement defies precise and consensually accepted methods. Indeed, the list itself is subject to debate. Excluding these outcomes from the quality discourse risks marginalizing them at a time when there is increasing recognition of the importance of so-called "soft skills." But if we are to tackle these behavioral traits that go beyond traditional declarative and procedural knowledge, we have to become more knowledgeable about how individuals learn these "behaviors" and about how to measure such learning.

How might these important outcomes be incorporated into the quality discourse? One approach is to gather information about the activities and experiences of alumni at designated time points (say, one, five, and ten years after graduation). But graduates don't live their lives in a bubble--they are exposed to other influences after graduation through employment, further education, family formation, and so on. This introduces substantial inferential challenges: Can colleges and universities properly claim credit for their graduates' achievements five, ten, or more years after graduation? Attributing institutional responsibility for alumni outcomes may require unrealistic assumptions--or alternatively, sophisticated analyses that attempt to rule out confounding factors (and that undermine simplicity and transparency from a consumer information perspective).

Another important qualitative dimension involves the student population itself. As the U.S. progresses toward a pluralistic "majority minority" society, can any institution be deemed high quality if (a) it does not serve a student population that is reflective of the broader population, and (b) distinctive educational opportunities and salutary outcomes are not enjoyed across student populations? How should we measure an institution's achievement of these important goals beyond coarse measures of compositional diversity? Should we try to measure the impact of the diversity on student interaction and student learning?

An additional dimension for consideration is whether there is a role for other interested parties, such as those who work with an institution's graduates (employers and graduate program faculty) to provide information about graduates' level of preparation. We often hear of employer complaints regarding student preparation, and more rigorously collecting and analyzing this information could help us understand the specific needs that are unmet. In using such information it is important to be mindful of when in the graduates' work and life the information is acquired. It is perfectly plausible to conclude that learning outcomes that prepare students well for jobs right out of college are not the best preparation over longer time periods, and vice versa.

3. **Assessment.** *In a system strongly guided by norms of professional judgment, peer review, and evidentiary support, quality is closely linked to processes of diagnosis and improvement. For an institution to be judged high quality, should criteria include the*

presence of a rigorous program of outcomes assessments and continuous improvements that are found in other industries? What should such a program look like, and who should judge its adequacy? What should be the role of faculty, employers, governments, and students/parents in establishing desired outcomes?

Assessment is a dish best served formatively – asking what went well, what did not and where can improvements be made. Unfortunately assessments are often used summatively to judge, sort and separate, not to promote growth, especially for students. Individual faculty, curricular committees, and departments are often in a similarly poor situation where assessments are either non-existent, non-actionable or of poor quality. In many of our institutions, adjunct faculty are hired or fired based on teaching evaluations that emphasize student satisfaction with instructor "performance" rather than clear measures of student learning and capability. While appointment, promotion, and tenure decisions for tenure-line faculty involve a wider range of performance criteria, the teaching quality component typically relies on the same evaluations. So what can be done? One potential approach at the course level calls for: 1) clear, measurable learning goals that are agreed upon between instructors of the same course and communicated to students, 2) multiple forms of feedback, including low (quizzes, graded or ungraded assignments) to high (midterm and final exams) stakes assessments, that help students gauge their learning and guide their improvement, 3) instructional approaches appropriate to the student population and the learning outcomes to be achieved and based on evidence of educational effectiveness, 4) agreed-upon approaches for instructors to reflect on their instruction and their students' learning, and to engage in a process of continuous improvement.

Instructional quality can also be assessed at the departmental level, which would involve an expansion of assessment of individual courses. The individual course level elements considered can be summed up for all the courses offered by a department to gauge variation and engagement with evidence-based teaching practices at the level of a course series/sequence and through the entire degree program. Additional elements worthy of evaluation can include TA instructional preparation and assessment, articulation between courses within a department and with related courses in other departments, and other factors that bear on student learning.

At the end of the day, if learning is regularly measured in a systematic way, analyzed, acted upon, and the cycle iterated with an emphasis on increasing student learning and capabilities, chances are the system is working. The "system" can be considered an individual instructor and her/his instructional team, the instructional teams that teach a given course, the teams involved in providing a course series, all the way to the department's educational mission.

One can easily see how this approach can be summed up for a collection of departments, or a college, and these groups can be brought together to reflect an entire institution. What is deemed "fair," "standard," or "exceptional" in terms of overall quality and outcomes is open to debate but as long as there is continuous improvement guided by student outcomes within and outside of the institution, as well as evidence-based instructional practices, chances are high that quality is present. Communicating to stakeholders the evidence that an institution is actively engaged in measuring its aspirations against its actions and is using that information for a process of improvement is likely to be very powerful information for all relevant stakeholders as long as it is sufficiently detailed and transparent. Indeed, one can argue that the quality discourse would benefit from a greater focus on the rigor and consistency of assessment and improvement processes in place than on the specific outcomes (scores) on any given assessment. But this may challenge the system to create quality performance assessments to demonstrate that students are indeed acquiring the knowledge and skills that a department, program or college has promised to deliver.

In addition, numerous approaches exist to help institutions gauge their overall impact on students, including student surveys that assess the learning experience (NSSE, CCSSE, UCUES/SERU) and the use of electronic portfolios to document student progress. These approaches help provide the big picture and can help to identify areas of strength and areas in need of improvement. It may also be helpful to look at other industries, such as manufacturing and health care, which have extensive experience in continuous quality improvement to produce better products and outcomes with less "recall" or "error."

4. **Understanding Quality from a Federal Policy Perspective.** *For understandable reasons, federal policymakers concerned with quality focus on measurable quantities (e.g., completion rates of Pell recipients, employment and starting salaries of graduates). This is important but insufficient and can have perverse consequences.*

The recently updated College Scorecard represents an attempt by the federal government to improve the consumer information available to students planning to attend a 2- or 4-year institution. It displays summary information drawn from existing federal data sources on program offerings; student body composition (achievement test scores, enrollment status, race/ethnicity, and percent receiving Pell grants as an indicator of socioeconomic diversity); net cost of attendance (broken down by income bracket); financial aid and debt (limited to college completers); retention and graduation rates; and earnings ten years after entry (limited to federal aid recipients; including all who ever attended the institution, regardless of duration, whether they graduated, or

where they graduated from). Where relevant, displays compare an institution's result to the national average. Several states have undertaken similar transparency-focused efforts. For example, Indiana produces "College Completion" and "Return on Investment" reports comparing results for public institutions in the state.

The indicators used in these efforts suffer from limitations, but the more important question for the present purpose is: What do they tell us about the quality of education delivered? A college's outcomes are highly influenced by who attends and what they study. Students are not randomly assigned to colleges and universities, so to a considerable extent institution-level variation in outcomes such as retention, graduation, and earnings reflect differences in the characteristics of those who enroll. Employment and earnings are affected by major field of study, so the mix of majors produced by an institution also accounts for differences in these outcomes. But those who turn to resources like these for authoritative consumer information may not be sensitive to these nuances, leading to improper inferences about institutional effectiveness and "quality." Thus, one aspect of any effort to help students choose among institutions should be the implementation of programs to help students and their families to decode and interpret information about institutions and programs.

The federal government is involved in quality assessment in another important way: by recognizing accreditation agencies that employ a peer-review process to ensure that institutions satisfy a designated set of quality standards. Accreditation has been criticized for lack of transparency, failure to motivate optimal performance or adequately penalize poor performance, and inadequate attention to student learning outcomes. Nevertheless, it remains the nation's formal quality assurance process for higher education. It attends to facets of institutional performance that do not lend themselves to easy measurement in a comparative framework, and the use of external peer reviewers is a common approach to performance assessment in professional domains that rely on expert judgment to navigate complexity and specialized knowledge. Can the accreditation process be modified to provide more useful consumer information and to provide federal policy makers with better information about institutional performance, while continuing to serve its core purpose? This is a key question that needs significant attention by organizations and individuals who seek significant federal policy influences on improving the quality of undergraduate education.

5. **The importance of context.** *Does the meaning of quality depend on what students and other payers "hire" colleges and universities to do for them?*

Higher education--whether in a community or technical college, a private liberal arts college, or a public research university--represents a significant investment by students (both direct expenditure and opportunity cost) and by taxpayers (whether direct

institutional subsidy or student financial aid). Students and society deserve to be confident that those investments pay off.

But the diversity of U.S. higher education--encompassing both institution types and students' reasons for attending--means there can be no "one size fits all" solution to the quality question. Across this variety, however, we can and should ask whether colleges and universities provide their students access to quality experiences and outcomes that correspond to their goals. At its most basic, this means asking whether colleges and universities are delivering what students and other payers expect--or as some have put it, what they are hired to do.

This raises the question of the proper level of analysis for the quality question. Where do students experience educational quality? An increasing share of the student population attends more than one institution on the way to a credential. More research is needed to understand the reasons for attending multiple institutions. Research examining the student experience and student outcomes finds far more variability between students within institutions than it does between institutions. These facts call into question whether the institution is the right focus for the quality discussion. Should we look higher, at the level of a state's collection of public and private providers? Or should we be assessing quality at the program level, which is where students engage most directly with educational experiences? Or should we look at all three levels, and be careful to describe what it is that we are discussing and how the levels interact, so that assessment and information can be well used?

Conclusion

Our purpose in this document is to raise a number of the key issues that have surfaced over the past several years among educators, policymakers and others who are seeking to advance both our thinking and our institutional and public policies around defining and measuring quality in undergraduate education. We offer these concluding questions to help guide the discussions during the December 14-15, 2015 National Academies workshop. These represent only a few of the core questions that need attention, and are intended as starting point for action among the various stakeholders, particularly during the next 24-36 months:

1. What actions are required in the next year or two to move us from current models of measuring student learning (e.g. VALUE Rubrics, PULSE, and DQP) that are being implemented on an *ad hoc* basis to a *system* of quality measurement whereby a group of like institutions adopts a standard set of indicators and reports their results, keeping in mind the work of the Voluntary System of Accountability and the related community college effort, the Voluntary Framework

of Accountability? What are the next steps in the process of implementing such a system, even on a pilot basis?

2. Now that the College Scorecard has been released, what further steps should the federal government (and, possibly, state governments) take to improve public information about the quality of undergraduate institutions? Are there improvements to the College Scorecard that are feasible and desirable in the near term? If so, who should be responsible for implementing them? What structures need to be put in place to assure that the College Scorecard is well-curated and that it can improve over time?

3. Can--and should--a group be assembled to create a core set of principles to guide the development of a general framework for measuring quality in undergraduate education--one that can be adopted by nearly any type of institution, e.g. 4-year university, 2-year college, online institution, "boot camp," etc.? If so, who should be involved in that process, who should lead it, and who should fund it? How could such an entity build on many of the existing rubrics and tools that have recently been developed?

4. What might be the most appropriate role, if any, for the National Academies? Could it, for example, serve an *integration* and *synthesis* role, bringing together and leveraging the good work that is underway (including the DQP, VALUE, VSA, and perhaps other emerging programs)? Might it also seek to broaden the emphasis from defining competencies and outcomes to working out the quite thorny assessment and consumer information components?