Peer review as a tool for external validation of programs in higher education

A White Paper prepared for the University Committee on Educational Policy (UCEP), a standing committee of the Assembly of the Academic Senate of the University of California, by Gregg Camfield, Professor of Literature, University of California, Merced

EXECUTIVE SUMMARY: We acknowledge that various constituencies of higher education have been asking for more accountability, and we believe we should judiciously respond to these calls. We believe, however, that our responses to these calls must be congruent with the fundamental nature of higher education or we will lose more value than we gain. Over time, higher education has developed peer review to balance innovation and high standards of professional performance in our scholarship. Peer review, we believe, is equally appropriate as a way to improve and externally validate the quality of student performance, as it can be similarly used simultaneously to reward innovations that improve student learning and to ensure that student performance meets high standards. We make no new claim, here. Indeed, the fundamental principle in accreditation has always been peer review, though the degree to which such review has been effective has been challenged in recent decades, perhaps because some see peer review as subjective and self-serving. It is not, but for peer review to work, the final result must be publication, which under accreditation has been lacking. Hence, calls for transparency might be interpreted not necessarily as a call to abandon peer review so much as to strengthen it. WASC's responses in the previous handbook revision and in the important changes adopted in November 2011 have done all that is necessary to drive institutional improvement and accountability by using peer review appropriately. As we see it, appropriately means that WASC has required faculty to be intentional and reflective as teachers and has required institutions to support constant improvement by (1) requiring programs to develop and post Program Learning Outcomes, which become the basis for ongoing study of student learning, (2) annually assess student learning using direct evidence drawn from authentic student work, (3) require periodic peer review of these studies in Program Review, (4) continue periodic institutional peer review in re-accreditation, which enables WASC to ensure that an institution as a whole is rewarding improvements at the program level and to encourage institutions to aggregate program level data to study how well institutional goals are being met, and (5) publish the results of accreditation review. We believe this last is the crucial step to "close the loop" of accountability and believe that the next ten years will consolidate the improvements generated by the 2001 and 2008 Handbook Revisions and, equally importantly, publicize where we are—and are not—improving. Thus, any move to require institution-wide benchmarking would be at best premature, at worst counterproductive by tipping the balance between innovation and standards irreparably toward the static state of current standards.

Introduction

WASC's proposal to require institutions to benchmark student learning by "externally valid" comparators, such as are ostensibly articulated in the Lumina Degree Qualifications Profile, is a drastic shift from the traditional practice of peer review in accreditation. It essentially externalizes the accreditation function, making WASC an organization charged merely with compliance.[1] Such a radical shift from WASC's constitution must be fully justified before the institutions that constitute WASC — and the faculty of those institutions who are charged with developing and delivering curricula — should accede.

During its meeting of November 3, 2011, the WASC Commission took action on many of the proposed changes in the Handbook, but chose not to act on the plan to seek external benchmarks of degree quality. Instead, recognizing that the proposal was controversial, it chose to seek further input. The Academic Senate of the University of California welcomes this invitation, and we present here our position that WASC has already established appropriate external validation of degree quality, and that moving to "benchmarks" as a way to externally validate degrees would be unnecessary, probably counterproductive, and inappropriate.

It would be (I) unnecessary because the processes already established require institutions to use peer review to externally validate student learning. Because peer review is second nature to academics, we rarely see the need to defend it. But in light of the recent kinds of calls for accountability, we must understand that peer review does pose a public relations problem. Many see peers, at least in accreditation reviews, as a mutual admiration society in which everyone wins the laurel. Exams and "benchmarks" appear by contrast to be objective and thus more trustworthy. Thus, we must explain clearly the value of peer review for validating the quality of baccalaureate education:
   A) Peer review's public relations problem is not a functional problem
   B) Peer review works because it creates an optimal balance between high standards and innovation
   C) When used properly, peer review can provide holistic external validation of student learning
   D) WASC has already created the conditions to use peer review properly for external validation.

As much as we believe WASC's current approach to validate educational effectiveness is working and, with the refinements of 2008 and 2011, will work well, the general outlines of the approach have not been in use long enough for WASC to draw valid conclusions about effectiveness. We encourage WASC to address the public relations problem presented by peer review rather than abandon an approach that promises to achieve the goal of better student

---

[1] WASC sought to shed such an approach to accreditation with the 2001 Handbook revision.

learning. It would violate the goal of simplifying accreditation to add another layer to a process that serves WASC's primary end.

External benchmarks may simplify the public relations problem of showing the quality of higher education, but requiring institutions to display degree quality against external benchmarks is likely actually to be (II) counterproductive for all of the reasons that standardized tests are counterproductive:
   A) By definition, benchmarks provide absolute, rigid performance standards.
      1) The rigidity of benchmarks will stunt innovation
      2) The rigidity of benchmarks will homogenize higher education.
   B) The benchmarks in the Lumina Degree Qualifications Profile (DQP) are predicated on a faulty understanding of what a degree is and can be.
   C) Institutional benchmarks shift the focus of accountability to a level that will impede action to improve student learning because both measurements and outcomes are too far removed from the daily work of faculty to facilitate improvements

Thus, benchmarking degrees is likely to detract from educational quality.

Finally, to impose benchmarks as a form of external validation is (III) inappropriate to WASC's mission and to the educational enterprise as a whole.
      1) If the goals of accreditation, including external validation, are already being met, then it is inappropriate to add another level of assessment.
      2) The common complaints lodged against higher education are almost always put in narrowly economic terms. WASC should be particularly vigilant in resisting such a narrowing of higher education, as WASC's stated "purpose is to promote the welfare, interests, and development of . . . higher education," and that means in its broadest possible construction.

We will conclude this position statement with some observations about higher education's multiple purposes, particularly from the point of view of faculty of the University of California, though we share our basic point of view with faculty from other traditional colleges and universities built around study of the arts and sciences.

I. PEER REVIEW AS THEAPPROPRIATE APPROACH TO EXTERNAL VALIDCATION

We believe that the Commission should consistently differentiate between "external validation" and "benchmarking."[2] "External benchmarking" is a subset of the methods one can use for

---

[2] This has not always been the case in the documents published by WASC, as in the publication of the Commission's November actions: "In its articulation of goals for the redesign of the accrediting process in November 2010, the Commission called for external benchmarking of key graduation proficiencies to provide a more consistent approach to evaluating student learning. This approach was affirmed by the Task Force on Levels of Learning and the Degree Qualifications Profile, which recommended that there be external validation or benchmarking of writing, critical thinking, and quantitative skills at graduation. This recommendation was modified by the Steering Committee to propose that at least two of the five graduation proficiencies in CFR 2.2 (see #3 above) be externally validated or benchmarked. In

"external validation;" peer review is the way academics have long "externally validated" scholarship, and insofar as assessment of student learning is a research problem, peer review is the best way to validate assessment findings.

A) **Peer Review's Public Relations Problem:**

While its public stature is still the envy of other national institutions, higher education has recently been subjected to substantial criticism.[3] Some specific charges are justified: many college students end their attendance carrying too much debt; too few students who begin a bachelor's degree program finish; many of those who do finish take too long. Few would argue that these are not problems; whether or not public higher education can solve them alone is an entirely different question.[4] But many of the proposed solutions, such as graduation exams or the Lumina Degree Qualifications Profile, imply a premise and a conclusion: because colleges and universities are incapable of self-regulation, their students are not learning what they should.

The premise that a profession cannot regulate itself is not restricted to higher education. Many of the learned professions are facing similar challenges. For example, the State Bar Act of the California Business and Professions Code has recently been changed to force the Bar regularly to evaluate the effectiveness of professional codes of conduct to serve the public good.[5] The medical profession, too, has been forced in recent years to do a better job of respecting patients' rights, rather than giving physicians *carte blanche* to prescribe medications and procedures. In these cases, as in the case of higher education, the push toward accountability is salutary; officers of institutions tend to defend the perceived interests of the institution, and unless they are reminded of their responsibilities, it is too easy for them to behave as if the public's interest were not their own. But in the cases of law and medicine, the state continues to

---

discussions throughout the region, the Commission received a range of responses from support for this proposal to expressions of concern. The Commission considered making external validation of graduation proficiencies optional, but there was a range of views within the Commission about how institutions would demonstrate effective learning without external comparisons outside the department or institution. The Commission decided, therefore, to continue to study this issue and to invite additional input from the region within the framework of the other actions taken by the Commission, especially #3 above, regarding the role of external validation, methods for undertaking validation and alternative approaches if external validation or benchmarking were not required, but made optional."

[3] According to the 2010 Harris Confidence Index (Corso, 2010), higher education is the third ranked major institution in the United States. According to a Pew Foundation survey published in May of 2011 (Pew Research Center, 2011), higher education still appears to be well respected and valuable even if it is now considered too costly.

[4] As we mentioned in our memo of December 7, 2011, we appreciate WASC's willingness to address some of these problems by studying whether accreditation of proprietary institutions should require approaches that differ from accreditation of traditional non-for-profit and public colleges and universities.

[5] "Section 6001.2. (a) On or before February 1, 2011, there shall be created within the State Bar a Governance in the Public Interest Task Force comprised of 11 members appointed by the President of the State Bar, seven of whom shall be attorney members of the board and three of whom shall be public members of the board. . . .
  (b) On or before May 15, 2011, and every three years thereafter, the task force shall prepare and submit a report to the Supreme Court, the Governor, and the Assembly and Senate Committees on Judiciary that includes its recommendations for enhancing the protection of the public and ensuring that protection of the public is the highest priority in the licensing, regulation, and discipline of attorneys, to be reviewed by the Assembly and Senate Committees on Judiciary in their regular consideration of the annual State Bar dues measure."

believe that professionals are in the best position to regulate themselves for the simple reason that the expertise involved in the profession is best judged by experts. Despite the vexing conflict of interest, peer review remains central to professional regulation even as the scope of that review is subject to statutory constraints. We believe that higher education is analogous in that it takes expertise to evaluate the quality of education, and that therefore peer review, no matter how much it is subject to external bodies setting the goals and obligations of the profession, must remain the core of accountability.

But education is not like law or medicine in one respect. While most people do not believe that having kidneys qualifies them to treat renal diseases, and while most people who know they will die do not believe that mortality qualifies them to set up a trust for their children, most people who have been to school count themselves experts in education, telling educators not only what education is for, but how to do it. We will address the ends of education at the end of this paper; here we will address the specific advantages to higher education of using peer review as a form of external validation. That said, inasmuch as peer review is poorly understood beyond the professions, we encourage WASC to explain the value to policy makers and to the broader public.

B) Peer review works because it creates an optimal balance between high standards and innovation

Acknowledging that peer review when abused enables professionals to protect themselves from accountability, we agree that it is important for colleges and universities to make the use of peer review in accreditation as objective and transparent as possible. But changes to improve transparency and objectivity should not damage the fundamental strengths of peer review. To that end, peer review in accreditation should mirror as closely as possible peer review in scholarship.

In scholarship, the point of review is to debate the validity and value of our ideas and practices so that we can extend knowledge and improve practices. While peer review is designed to facilitate change, not to defend the status quo, it is also meant to resist novelty for novelty's sake. Scholarly peer review insists on *standards* of evidence and presentation, both of which are developed and continuously change through shared professional practice.

In this regard, peer review is simultaneously formative and summative, and asks scholars to engage in a continuous dialogue in two phases. The first phase usually begins in informal networks, with colleagues bouncing ideas around, shaping and clarifying questions and modes of inquiry. Intermediate steps often are taken in formal conference presentations, where immediate responses from an audience enable scholars to continue to refine ideas. Finally, print publication culminates the first phase, when an editorial board agrees that the ideas are worthy of wide dissemination. The second phase, the scholarly community's reaction to publication, is just as important, with informal and published reviews helping the community to digest new ideas and practices. This digestion may range from outright rejection to complete incorporation, but the continued collective response improves shared professional practice.

Peer review works because:

1. Practicing experts in a field assess new contributions holistically according to their own feel for the profession. As Dreyfus and Dreyfus (1986) put it in their insightful analysis of expertise, experts are not distinguished by how well they apply rational rules to problems, but rather by how well they *incorporate* practices, recognize the complexity of contexts, and blend intuition with rationality in expert performance. An expert can apply this holistic understanding of a field to the work of other performers in the field. Thus, experts can see how a new idea fits into and potentially changes the field. In contrast, assessment by rubric reduces a performance to a set of parts judged against conventional standards; rule-bound judgment can never rise above the current level of practice.
2. It aligns individual rewards with the collective good of the scholarly community. A practicing scholar must be credentialed to enter the profession, but a credential is not enough to ensure continued employment. Constant deepening of expertise, as validated by peers, is necessary.

Correlatively, however, peer review both amplifies expertise and exaggerates the difficulty of judging expertise from beyond the field. It is easy therefore to become skeptical of expertise in this regard—one wants some indicators that performance within a field can be judged by some external standard. Any person who has ever served on a multi-disciplinary committee charged with evaluating the quality of a candidate for promotion or tenure knows how difficult it is to find such indicators. Number of publications, impact factors of journals, letters of evaluation solicited widely from the field or sub-field never do more than enable outsiders to guess at the quality of scholarship. Accreditation review, even though it is looking primarily at education, not scholarship, is in a similarly difficult position; the search for some surrogate measure of quality is the *ignis fatuus* that leads us to trust untrustworthy measures.

C) When used properly, peer review can provide holistic external validation of student learning

To trust peer review, then, is not to seek surrogates so much as it is to ensure that peers are seeking the right ends. When looking at student learning, that means using peer review to assess how well our students are learning and to ensure that faculty continually improve the conditions that enable student learning. A generation of research into assessment has identified two fundamental conditions for this cycle of improvement. First, assessment needs to be valid. Second, it needs to lead to change, or as WASC has traditionally put it, faculty must act on assessment data in order to "close the loop." For these reasons, many influential assessment experts have come to see the importance of assessing student learning in alignment with faculty expertise.

1) Validity: It is very difficult to assess generalized student learning because all learning is context specific and thus even the metacognitive skills that arise in a disciplinary education cannot be readily measured by performance beyond the discipline. As Thompson and Jonson (n.d.) put it in their summary of the research:

> [E]vidence indicates that learning is highly disciplinary specific (Beyer et al., 2007). If skills like writing or critical thinking are disciplinary specific, then measures that treat those skills as generalizable have little meaning. Furthermore, there is evidence that there is greater variance in the performance of students within an institution than differences in measures of performance between

institutions (Borden & Young, 2008). This is not a matter of developing a better test, creating a more authentic context, or assuring adequate student motivation to perform optimally on tests, but rather a matter of interpretative validity relating to the use of a test or measure (Borden & Young, 2008). Such value added measures cannot be meaningfully interpreted to reflect the educational quality of one institution let alone differences across institutions. (p.1)

While much of this summary refers to studies specifically of standardized tests (e.g. Bordon & Young, 2008), the findings are equally applicable to any effort to standardize assessment of institution-wide learning outcomes. Indeed, findings drawn from both the National Survey of Student Engagement (NSSE) and the University of California's University of California Undergraduate Experience Survey (UCUES) suggests that any cross-institutional comparisons are much less useful than intra-disciplinary comparisons (NSSE, 2011; Thompson & Douglass, 2009; Chatman, 2007).

This is why AAC&U, in responding to calls for institutional accountability, developed 15 rubrics to measure a wide range of learning outcomes. More importantly, in publishing these outcomes, AAC&U encourages institutions to adapt the rubrics to local circumstances. That is to say, one cannot apply disciplinary knowledge to assessment without having disciplinary knowledge, and one cannot apply a rubric to disciplinary knowledge without adapting the rubric to specific circumstances.

2) "Closing the Loop:" Assessing in the disciplines not only draws on faculty expertise, it encourages faculty to act on what assessment shows. According to George Kuh and Stanley Ikenberry (2009), "program-level assessment data – especially in large, organizationally complex universities – are more likely to be actionable, to get the attention of faculty, and to point to specific improvement needs and opportunities in teaching and learning."

None of this means that assessment of student learning is being done to the standards of peer review, but it does suggest that such assessment is amenable to peer review. The best assessment, after all, is done in disciplines where faculty have the expertise needed to review not only their colleagues' work and their students' work, but, in looking at student learning, to review their efficacy as teachers. In order to turn such assessment into peer review, all that is needed is to ensure that assessment is judged externally by disciplinary standards.

D) **WASC has already created the conditions to use peer review properly for external validation**.

WASC's current process (that is the 2008 version of the 2001 handbook plus those aspects of the current revision adopted by the Commission in November of 2011) does so by articulating educational ends and then insisting on a common framework for faculty to show that they are seeking those ends. Standard 1 asks to institutions to have clear, widely shared, and appropriate purposes; Standard 2 requires institutions to ensure that students are learning as proposed; and Standard 4 insists that institutions support continuous improvement. These Standards in general, and CFRs 2.3, 2.4, and 2.7 in particular, create a process that shifts peer

review from being a compliance review by WASC of an entire institution, to a process more analogous to that used by faculty in their scholarship.

In particular, faculty are asked to work together to establish program-level learning outcomes and to determine how to assess student learning by the standard of those proposed outcomes. These requirements turn assessment into a research task, encouraging faculty to develop discipline-appropriate measures for student success. Essentially, WASC's requirements push faculty at the program level to initiate assessment by informally swapping ideas, then working those ideas into a research project, and then publishing the outcome to students and other constituencies. But at this point, there is no peer review as we usually understand it. By engaging in regular on-going assessment, however, faculties develop data that are appropriate for peer review, and when compiling on-going assessment into program review, required in CFR 2.7, faculties in essence put their findings up for peer review.

The University of California's Academic Senate Committee on Educational Policy (UCEP), studied program review practices across the system in 2008, and is currently updating that information to reflect changes as campuses have come up for WASC review. We found that campuses treat program review as peer review, with all but one campus using at least one reviewer external to the campus for each review, and all seeking reviewers with significant disciplinary expertise. While results have not been widely published, they are used by both academic senate and administration to serve formative and summative ends. On each campus, reviews are used to help decide how to allocate resources, so that both formative and summative goals are backed up by important consequences. This, indeed, is one of the points of peer review: to align rewards with institutional goals.

As program review is the next step from regular learning outcomes assessment, so periodic reaccreditation is the next step from program review. WASC's current handbook encourages institutions to see the entire reaccreditation process as a higher order version of programmatic review, aggregating program level data into an institutional picture and subjecting the final product to peer review. We are not informing WASC of this so much as we are reframing WASC's current process to show:
1. That the current process is congenial to the academic enterprise. In time, as faculty grow accustomed to the process and come to see the parallels to their own scholarly practice, they will improve student learning dramatically.
2. That, because it brings expert opinion to bear in program review, the current process is in fact externally validating learning outcomes. Reviewers, as experts, bring to bear their own ideas of what students can and should do. But peer review as validation is better than benchmarking because it is flexible and dynamic.[6] Expert reviewers both internalize current standards and let those standards evolve over the course of reviewing. That is, they not only judge, but also learn by doing reviews and will take what they learn back to their home campuses.

---

3. That the current process enables WASC to approach institutions flexibly and holistically, in accordance with the ideal of peer review and in accordance with WASC's stated values in implementing the 2012 Handbook,

> Maintain flexibility, adaptability, and a continuing posture of experimentation that emphasize collaboration between institutions and WASC. This value served WASC well in the period 2001-2011 and means that multiple models for institutional presentations and of organizing evidence to demonstrate meeting the Standards should continue to be encouraged. This posture of experimentation should also position WASC well to effectively assess the many new organizational forms for teaching and learning that the current revolution in instruction is generating. (Ewell, 2010, p. 1)

4. That the 2001/2008 Handbook excluded the most important aspect of peer review—widespread dissemination of published results.

We believe that some of the public's—and higher education's—distrust of accreditation comes from the prior failure to publish the results of accreditation reviews. Until now, all the public could see is whether or not an institution is accredited, not whether it barely cleared the bar or is sailing in the stratosphere. By its recent vote to publish both team reports and Commission action letters, the WASC Commission solved the problem. It does not need to take the extra step of giving institutions grades. Indeed, such grading, like the API scores given to California public schools, may be counterproductive by narrowing the focus of an institution to those criteria that are to be graded. Again, the value of peer review is that it can give a holistic picture of performance, whereas grades provide a reductionistic view. Holistic peer review enables WASC to remain flexible and responsive to the wide variety of types of institutions; grades reduce all institutions to the same rubric.

CFRs 2.3, 2.4, 2.7, and 4.7 establish the process; publication of the results of accreditation review provides the accountability at the core of peer review. The question remains whether this process can address the learning outcomes articulated in CFR 2.2, especially 2.2a. We believe it can and that it is doing so, primarily through program review. The summary requirement of 2.2a, that baccalaureate degree programs require students "to engage in an in-depth, focused, and sustained program of study," is what program review is designed to assess through peer review. But program review does not assess programs merely at the aggregate level; it also addresses a number of the specific requirements of 2.2a.

In short, program review serves as a broad, nearly comprehensive locus for serious peer review of educational effectiveness. As long as WASC keeps pushing campuses to engage in review of all programs that affect student learning and keeps pushing campuses to aggregate these reviews for regular assessment of institution-wide educational effectiveness, program review serves as a very powerful mechanism for both improvement and accountability. **To repeat the most important point, such accountability *is* externally validated by the highest standards.**

WASC has stated clearly three fundamental values that are guiding the current revisions of the Handbook: "First, the institutions comprising WASC strive to achieve mutual and collective accountability for academic quality. As an integral part of this, they seek to assure the public about academic standards and provide the public with appropriate information about institutional performance. Second, their view of quality centers irrevocably on students—student success in attaining degrees, student learning, and the academic and co-curricular experience. Finally, this commitment is not just about demonstrating quality but upon using evidence to improve it on a continuous basis" (Ewell, 2010, p. 1). We have already shown that peer review enables institutions to achieve mutual collective accountability for academic quality and that peer review enables institutions to improve on a continuous basis. We also acknowledge how difficult it is to use peer review to provide the public a readily understandable accountability measure. But the fact that it is difficult is not a reason to avoid the task altogether, especially since the "benchmarking" alternative will seriously undercut the efforts of institutions to improve student learning.

A) By definition, benchmarks provide absolute, rigid performance standards.
In proposing the use of the Lumina Degree Qualifications Profile to benchmark degrees throughout the WASC region, the members of the work group had lofty goals about how the DQP would be used. They assumed that the DQP would be a starting place for discussion, a framework that would enable WASC institutions collectively to decide what a student with a certain degree would be able to do. "The Task Force also reviewed the proposed Lumina Degree Qualification Profile (DQP) and found it to represent a useful and flexible framework that could guide both institutional conversations and accrediting reviews, especially in dealing with new institutions and new delivery modes." (WASC, 2011, p.1) In the high stakes world of accreditation, however, a "benchmark" becomes not a starting point for discussion, but rather an end-point for performance. Careful reflection should make this clear not just by thinking of how a benchmark would in fact be used in accreditation, but also through the meaning of the term itself.

As with many clichéd metaphors, "benchmark" has lost its clarity and power through sloppy use. But when looking at educational benchmarking as it is developing through the international "tuning" practices developed through the Bologna accords and in the U.S. under the aegis of the Lumina Foundation, "benchmarking" is an apt metaphor. A benchmark in carpentry is a shortcut enabling a carpenter to cut a number of identical pieces by measuring once, marking the dimension on the bench, and then using that one mark to guide cuts on every piece of wood. By analogy, then, a benchmark is not merely a reference point, as is so often meant in common parlance; it is a template for the production of identical products. Benchmarks by definition are rigid. How can WASC be sensitive to institutional contexts if it insists on benchmarking? At least as bad, benchmarks measure parts of a product, not the whole. The overall plan is much larger than any piece cut against a benchmark, and no observer will be able to see the plan from the benchmark itself. Indeed, the pieces cut from a benchmark are more likely than the benchmark itself to give hints about the overall structure.

Thus, by pushing benchmarking, WASC undercuts the very idea of peer review by replacing expertise with rigid, external standards, and replacing holism with reductionism. Again, in the high stakes world of accreditation, it is too risky for the faculty and administration of any institution to assume that a benchmark is in fact a starting point for discussion, and thus if WASC were to require any benchmarks to describe degree quality it would create two major unintended consequences that are at odds with WASC's values:

1) The rigidity of benchmarks will stunt innovation

A benchmark becomes another kind of standardized test, albeit the ones in the DQP are vaguer and more difficult to prepare for. Thus all of the innovation-killing aspects of standardized testing will apply to benchmarking, with the added danger that the very vagueness of the benchmarks will encourage institutions to be extremely conservative in how they interpret the vague parts of the benchmark.

Regarding the damage standardized tests do to innovation, there is so much research that it could take pages to cite. We will instead again turn to the useful summary posted on the Carnegie funded Reinvention Center website:

> [T]he focus on value added measures for the purposes of institutional comparisons diverts attention and resources from developing and implementing innovative curricular and pedagogical approaches and assessment process that improve student learning. The diversions that occurred in K-12 education as a result of accountability have already been recognized (Banta, 2007, Gallagher, 2007). These include a narrowing of the curriculum to focus on test content and test-taking skills and state resources supporting test administrations rather than innovations that literature has shown could create more effective schools. Danger also lies in the association faculty will eventually make between assessment and accountability rather than assessment and improvement because quality is judged on the basis of test scores rather than the implementation of effective strategies for improving student learning. This has led to the recognition of the need to reframe accountability and differentiate assessment from accountability. (Thompson & Jonson, n.d., p.1)

Those who defend the DQP will surely say that, while this may be true of tests, it is not true of benchmarks because benchmarks allow an institution to provide its own evidence of quality rather than using a single instrument. But the same problems apply because an institution must aggregate data to present the learning for *all* students regardless of discipline. Notice how in the DQP itself, as various outcomes are described in the body of the report, they are described with some richness and subtlety, but as they are condensed in the summary and then condensed again in the matrix of outcomes, they become reduced to narrow, rigid, and paradoxically fuzzy generalizations. The next step to making these general outcomes measurable is to cram them into rubrics. If the very standards themselves grow reductive as relatively detailed, rich discussions get "aggregated" into something more readily measurable and reportable, then institutions will end up doing the same thing when they are reporting learning outcomes against DQP-based rubrics. As in testing, this kind of aggregation creates

a potential for an "ecological fallacy" (Robinson, 1950), wherein the associations described by institutional-level correlations are incorrectly applied to the individuals who comprise the group. The validity of comparisons among institutional aggregate scores assumes there are no differences in effects across students or subgroups of students within the institution (for example, students in different academic programs and therefore exposed to fundamentally different curricula). The institutional score masks potentially important sources of variation within the institution—the very sources that make such assessments useful for internal improvement purposes. Blanket statements about institutional performance can therefore be minimally informative, if not entirely misleading, especially for large institutions that enroll diverse student populations and have diverse curricula. (Borden & Young, 2008, p.30)

If so, and again given that accreditation is a very high-stakes process, how will institutions react? Probably by trying to bypass the validity problems by using old-fashioned surrogates.

Consider an example: "Communications fluency" as an "intellectual skill," which matches one of the core competencies that WASC wants to measure by the DQP. The grid by which Lumina displays its Profile in tabular form shows a student:

[Associate's level] Presents substantially error free prose in both argumentative and narrative forms to general and specialized audiences.
[Bachelor's level] Constructs sustained, coherent arguments and/or narratives and/or explications of technical issues and processes, in two media, to general and specialized audiences. (Lumina Foundation, 2011, p. 19)

The standards for the associate's level simply erase fifty years of research on how students learn to write. Such standards will push associate's-level institutions to use old-fashioned pedagogies in old-fashioned writing classes because innovations in writing pedagogy are not designed to fill this narrow bill. Associate's level students in turn—especially in the University of California and in the California State Universities—transfer to baccalaureate institutions having learned skills that will not translate to the bachelor's degree expectations. Essentially, rather than having integrated, appropriate, and innovative writing instruction across the bachelor's degree program, students transferring from community colleges to U.C. or Cal State campuses will have a fragmented learning experience. And if community colleges are forced into an archaic posture, all institutions downstream will have to compensate. The net effect is to discourage innovation.

2) The rigidity of benchmarks will homogenize higher education
All rhetoric about flexibility notwithstanding, benchmarks will homogenize higher education. Partly, the disincentives to innovate will freeze higher education in its current state. More importantly, the use of a single standard in any dimension of education punishes non-conformity. If, as Banta (2008) puts it, "No set of tasks or items designed to test generic skills like writing and reasoning is content-free. . . . [s]ome [students] will be advantaged by their choice of major, and others will be disadvantaged" (p.4) , then it is equally true that some kinds of institutions will also be advantaged and others disadvantaged when judged by any benchmarks. Art schools, technical schools, schools biased toward the sciences will find themselves looking weaker than liberal arts colleges in "communication skills." This weakness cannot be corrected if

writing is measured as a generalizable skill. Since accreditation is so high stakes, such institutions cannot afford low scores and will be required to institute a curriculum that shadows that of high-scoring institutions. Emulation of best practices has its advantages, but those advantages disappear if "best" is mismeasured.

This argument assumes that benchmarking will be undertaken regardless of validity. If it is to be undertaken in a way that can validly be used for cross-institutional comparisons, then "learning communities" will need to develop controlled experiments across institutions. To improve validity, institutions will have to minimize variables, which is another way of saying that they will need to develop common curricula and common pedagogies and then isolate common assignments for use in assessment.

B) The DQP benchmarks are predicated on a faulty understanding of what a degree is and can be

The DQP articulates greater complexity in outcomes for different degree levels. While on the surface this seems obviously appropriate, the assumptions behind the particular levels should not be taken as matters of faith. It appears that the DQP relies on what has come to be called Bloom's taxonomy of learning. It is worth noting that what most people cite as Bloom's taxonomy covers just one of three cognitive domains identified by Bloom and his colleagues. Bloom and his colleagues, working in the shadow of Dewey's dictum that we should educate the whole person, identified cognitive, affective, and psychomotor domains of knowledge that needed to be addressed in education. When we reduce our discussion of learning outcomes to the one domain, we do violence to the fundamental point. Education should not be just about cognition.

What is more disturbing is that much current research in cognitive science shows that effective education cannot be just about cognition. Despite a long tradition of denigrating the affective domain as less important than the cognitive, new research shows that it is impossible to separate cognition from affect when affect drives cognition and cognition alters affect. And some influential theories of learning insist that the psychomotor is essential, too. In short, studies in curiosity, creativity, and expertise carried out since Bloom's publication create a much more complex picture of learning as a recursive process, suggesting not that these levels are intrinsically lower or higher so much as that a learner grows more facile in integrating "higher" and "lower" intellectual functions, and that this integration is, to some degree, across the domains. All of these possibilities are speculative at this point; what is not speculative is that American higher education at its best has two aims that are far beyond anything Bloom's cognitive taxonomy explains or articulates. One, higher education aims to teach students how, given their own temperaments, abilities, dispositions, and goals, they can learn how to learn, how to confront novel problems in novel contexts. Two, higher education encourages creativity and innovation. Benchmarks cannot measure these most important outcomes. Accountability in education often leads to the dangerous situation in which we cease to value what we cannot measure, and we certainly cannot benchmark what has not yet been invented. In using benchmarks, then, we risk undercutting some of the most important aims of higher education.

Systematic understanding of effective pedagogy has barely begun to catch up with the current revolution in cognitive science; the DQP locks us into a half-century old model of learning. But the rigidity that the model entails is not nearly as important here as the point that the model may in fact impede learning. It will certainly force us to measure "skills" and "knowledge" by standards that may very well be inappropriate measures both of how students really learn and of the most important things we teach.

C) Benchmarks shift the focus of accountability to a level that will impede action to improve student learning.

To cite Kuh and Ikenberry (2009) again, if faculty attention is going to be directed to assessment, it has to be done in a way that faculty can see results. We, too, need positive feedback if we are to learn, and institutional simulacra of learning alienate us. We encourage WASC to pay close attention to the paper it commissioned from Pat Hutchings: "What Can WASC Do to Increase Faculty Engagement?" She lists a number of reasons faculty have been less engaged in institutional assessment than might be ideal. Among them are the "full plate" faculty are carrying in this complex educational and research environment, the ways in which the language of assessment is managerial rather than education and therefore alienates faculty, and the fact that faculty probably do more assessment than is generally acknowledged because many do not use the language of assessment or publish their work.

> It may be, too, that faculty have not yet seen sufficient evidence that assessment makes a difference. There's a chicken-and- egg dynamic at work here; more faculty involvement would presumably make a bigger difference. But the fact remains that the benefits of assessment are uncertain. Indeed, assessment is seen as "redundant" on many campuses, duplicating already existing processes and not yielding additional benefits (Kuh & Ikenberry, 2009, p. 9). Similarly, in the 2009 Faculty Survey of Student Engagement, 75 percent of respondents indicated that their campuses were involved in assessment "quite a bit" or "very much," but only about a third had positive views of the dissemination and usefulness of assessment findings (National Survey of Student Engagement, 2009, pp. 21–22). Faculty might be readier to engage with assessment if its benefits were easier to see. (Hutchings, 2010, p. 2)

Adding yet another layer onto assessment, a layer so bedeviled by problems with validity and so far from the daily work of faculty, is not going to encourage faculty engagement in assessment. By diverting attention, resources and good will to an at best redundant, at worst invalid form of assessment, benchmarking degrees is likely to detract from educational quality.
b

III THE DQP AND OTHER INSTITUTIONAL BENCHMARKS ARE INAPPROPRIATE TO WASC'S MISSION AND TO THE EDUCATIONAL ENTERPRISE AS A WHOLE.

A) If the goals of accreditation, including external validation, are already being met, then it is inappropriate to add another level of assessment.

WASC's mission is to serve the interests of higher education. If current policies effectively serve that mission, it is inappropriate to add another requirement, no matter how alluring the new tool might be, no matter the external pressure "to do something." WASC has deserved praise for leading the assessment movement for some time. Since assessment and accountability are at best uncomfortable bedfellows, WASC might sacrifice its leadership and its legacy in assessment by prematurely endorsing what we believe to be a poor approach to accountability.

B) The common complaints lodged against higher education are almost always put in terms inappropriate to higher education.

Pat Hutchings (2010) speaks true; terms do matter. Even more, ends matter. Regarding the ends of education, the business community has complained effectively for years about K-12 schools, and, recently, about higher education. When business leaders charge that colleges are not adequately teaching students, they speak out of a widely held assumption that education should primarily serve to support the U.S. economy.[7] Politicians seem to amplify these concerns when they speak publicly about the purposes of higher education.[8] As long as the terms of the debate are couched in this one dimension, the perceived interests of business have too much weight.

In this context, The Lumina Foundation has a weighty burden of proof in arguing that the DQP is appropriate to all of higher education. The Lumina Foundation, after all, justifies its goal of increasing college completion rates almost entirely in economic terms. On its web site, the three paragraphs explaining why it hopes to improve college completion rates include just part of one sentence that lists any other outcomes (such as higher voting rates). In referring to the number of college graduates, Lumina even describes education as an industrial process: "Based on current estimates, to reach the 60 percent level by 2025, the U.S. higher education system must produce 23 million more college graduates than are expected at present rates of production" (Lumina Foundation, 2011b).

We could multiply the number of calls for higher education as an economic good almost infinitely, but it is important to notice that this is not the single, nor even the most important, goal of the University of California, or other colleges and universities based on the study of the traditional arts and sciences. The University of California's mission is to

---

[7] See the newly released ACICS effort to coerce higher education to serve narrowly economic, and no other, ends (Bieda, 2011). Not all efforts to connect higher education to business needs are so tendentious. See Bok (2006), Chapter 11 , and AAC&U survey of business leaders (Peter D. Hart Research Associates, Inc., 2008).

[8] Again, we could cite endless examples, but one telling example best makes the point. In *The Audacity of Hope,* Barack Obama(2006) made a compelling case through almost the entire book that American political life suffers because the average citizen knows too little about American history and the political system. Yet when he turns to discuss education, he speaks of the economic need rather than of the civic need. The glaring omission speaks volumes about what a politician must say about education.

serve society as a center of higher learning, providing long-term societal benefits through transmitting advanced knowledge, discovering new knowledge, and functioning as an active working repository of organized knowledge. That obligation, more specifically, includes undergraduate education, graduate and professional education, research, and other kinds of public service, which are shaped and bounded by the central pervasive mission of discovering and advancing knowledge.

This charge is not disinterested; it centers on "long-term societal benefits." Undoubtedly, many of those benefits are economic, but the mission statement argues that how knowledge is applied is not the primary responsibility of the university. The university is not and must not become an outsourced corporate training center.

Regardless, many of the complaints about higher education assume that higher education is merely job training. For example, when employers are surveyed about what they want to see in college graduates, they generally ask for good communication skills, good mathematics skills, and good critical thinking skills. We are acutely aware of the parallel between these stated business needs and WASC's initial proposal to require benchmarking in these three areas. For the sake of our analysis, we will consider one of business leaders' top priorities and top complaints: writing. As much as business leaders say they want entry-level employees to write well, they often complain that college graduates can't. Their complaints are based not only on the faulty assumption that colleges exist to provide free training, but also that there is a universal standard of good writing. Clearly the Lumina DQP works on that assumption, as do all of the standardized tests that profess to measure student writing. The AAC&U VALUE rubric for written communication, however, provides a salutary alternative perspective because it is built not on this faulty assumption, but rather on research, and its framers caution users of the rubric not to misuse it: "The most clear finding to emerge from decades of research on writing assessment is that the best writing assessments are locally determined and sensitive to local context and mission. Users of this rubric should, in the end, consider making adaptations and additions that clearly link the language of the rubric to individual campus contexts." (Association of American Colleges and Universities, n.d., p. 1)

The reasons for this sound advice are threefold. (1) By definition, good writing effectively blends form and content. One cannot, then, evaluate writing without simultaneously evaluating the content. (2) Content is not the only variable; conventions vary radically between various discourse communities. Writing conventions are analogous to etiquette: what's correct in one culture is egregiously wrong in another. (3) Thus, fluency requires both content knowledge and knowledge of norms, which can be acquired only through time. At best, good writing instruction can teach students how to identify the conventions in a novel writing context so that they can teach themselves how to reach the new standards quickly. In other words, good writing instruction teaches meta-rhetorical writing strategies, not situation specific correctness.

Can we expect, then, for employers to be happy with the writing of newly minted college graduates? Few business managers are inclined or able to assess such a meta-cognitive skill. Moreover, most real-world writing requires quick turnaround, so that a writer new to the context cannot find models to emulate, get feedback on drafts from competent peers, and rewrite repeatedly to refine a crude draft into an acceptable piece of writing within the time allotted.

Hence, an employer is likely to see a bad job on a first example of a new employee's writing, and will take that first example as evidence that "this recent college graduate can't write." This kind of anecdotal thinking is not a sound basis for educational policy.

What is true of writing is also true of mathematical reasoning. For those who think of mathematics as black and white, this might come as a surprise, but how math is used in different contexts is not taught as part of pure math. Thus, many disciplines teach math, especially statistics, within the discipline. Statistics courses in ecology differ from those in psychology, etc. And if it is true that "foundational" skills such as writing and math vary among the disciplines, how much more so is this true of other kinds of critical thinking. What counts as evidence, what counts as analysis, and what counts as effective presentation of ideas all vary by domain. Higher education cannot replicate every context graduates will face in their future employment, so even if it were true that colleges and universities were supposed to perform as corporate training, they could not do it to employers' satisfaction.

We recognize that many WASC institutions exist to provide career training, and we have no complaints if such institutions wish to use the DQP or other tools for institutional-level benchmarking student learning proficiencies. But as long as WASC's mission is to support the diversity of institutions in the region, as long as its mission is to further the ends of higher education broadly construed, it should insulate the accreditation process from untoward pressure from any one constituency. When the agency constituted to promote the interests of higher education surrenders the definition of educational ends, it promotes the end of higher education.

Association of American Colleges and Universities. (n.d.) Written communication value rubric. Retrieved from http://www.aacu.org/value/rubrics/index_p.cfm?CFID=33635459&CFTOKEN=13032794

Banta, T.W. (2007) Can Assessment for Accountability Complement Assessment for Improvement? *Peer Review 9(2),* p. 9-12.

Banta, T. W. (2008). Trying to clothe the emperor. *Assessment Update*, 20(2), 15-16. Retrieved from http://onlinelibrary.wiley.com/journal/10.1002/(ISSN)1536-0725

Bieda, A. S. (2011, December). Closing the gap between career education & employer expectations: Implications for America's unemployment rate. The Accrediting Council for Independent Colleges and Schools. Retrieved from http://www.acics.org/events/content.aspx?id=4718

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of Educational Objectives: The Classification of Educational goals; Handbook I: Cognitive Domain* New York: Longmans, Green.

Bok, Derek. (2006). *Our Underachieving Colleges.* Princeton: Princeton UP.

Borden, V. M. H., & Young, J. W. (2008). Measurement validity and accountability for student learning. *New Directions for Institutional Research*, *Assessment Supplement 2007*, 19-37. DOI: 10.1002/ir

Chatman, S. (2007). Institutional versus academic discipline measures of student experience: A matter of relative validity. UC Berkeley: Center for Studies in Higher Education. Retrieved from http://escholarship.org/uc/item/81f4h1nn

Corso, R.A. (2010, March 9). The Harris poll. The Harris Poll, Harris Interactive. Retrieved from http://www.harrisinteractive.com/vault/Harris-Interactive-Poll-Research-Education-Confidence-2010-03.pdf

Dreyfus, H. L., & Dryfus, S. E. (1986). *Mind over machine: The power of human intuition and expertise in the era of the computer.* Oxford: Basil Blackwell.

Ewell, P. (2010). Commission values for implementing the 2012 handbook. Values and Outcomes Statement. Retrieved from http://www.wascsenior.org/redesign/handbook

Gallagher, C.W. (2007). *Reclaiming Assessment: A Better Alternative to the Accountability Agenda*. Portsmouth, N.H.: Heinemann.

19

Hutchings, P. (2010, April). Opening Doors to Faculty Involvement in Assessment (National Institute for Learning Outcomes Assessment Occasional Paper, No. 4). Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment.

Hutchings, P. (2010, October). What can WASC do to increase faculty engagement? Western Association for Schools and Colleges: Accrediting Commission for Senior Colleges and Universities. Retrieved from http://www.wascsenior.org/redesign/conceptpapers

Kuh, G., & Ikenberry, S. (2009). More than you think, less than we need: Learning outcomes assessment in American Higher Education. Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment (NILOA). Retrieved from http://www.learningoutcomeassessment.org/

Lumina Foundation. (2011a). Degree qualifications profile. Retrieved from http://www.luminafoundation.org/publications/The_Degree_Qualifications_Profile.pdf

Lumina Foundation. (2011b). Goal 2025/Lumina's Strategic Direction/The Impact of Attainment. Retrieved from http://www.luminafoundation.org/goal_2025/goal3.html

National Survey of Student Engagement. (2009). Assessment for Improvement: Tracking Student Engagement Over Time--Annual Results. Bloomington, IN: Indiana University Center for Postsecondary Research.

National Survey of Student Engagement. (2011). Fostering student engagement campus-wide – annual results 2011. Bloomington, IN: Indiana University Center for Post-Secondary Research. Retrieved from http://nsse.iub.edu/NSSE_2011_Results/pdf/NSSE_2011_AnnualResults.pdf

Obama, Barack. (2006). *The Audacity of Hope*. New York: Crown.

Peter D. Hart Research Associates, Inc. (2008). How should colleges assess and improve student learning? Employers view on the accountability challenge. A survey of employers conducted on behalf of: The Association of American Colleges and Universities. Retrieved from http://www.aacu.org/leap/documents/2008_Business_Leader_Poll.pdf

PewResearchCenter. (2011, May 15). Is college worth it? College presidents, public assess value, quality and mission of higher education. Retrieved from

http://pewresearch.org/pubs/1993/survey-is-college-degree-worth-cost-debt-college-presidents-higher-education-system.

Regents of the University of California. (1974). Mission Statement. Retrieved from http://www.universityofcalifornia.edu/aboutuc/missionstatement.html.

Robinson. (1950). Ecological Correlations and the Behavior of Individuals. *American Sociological Review,* 15, 351–357.

Thompson, G., & Douglass, J. A. (2009). Decoding learning gains: Measuring student outcomes and the pivotal role of the major and student backgrounds. UC Berkeley: Center for Studies in Higher Education. Retrieved from http://escholarship.org/uc/item/2173006c

Thompson, R. J. Jr., & Jonson, J. L. (n.d.). Accountability for improving student learning. The Reinvention Center. Retrieved from https://www.sarc.miami.edu/ReinventionCenter/Public/Spotlights/

WASC. (2011, April 3). Redesigning WASC**:** Summary of task force findings and recommendations. Western Association for Schools and Colleges: Accrediting Commission for Senior Colleges and Universities. Retrieved from http://www.wascsenior.org/redesign/handbook