

THE COGNITIVE OUTCOMES OF LIBERAL EDUCATION

Henry Braun

Boston College

THE
ANDREW W.
MELLON
FOUNDATION

This essay was commissioned by The Andrew W. Mellon Foundation under the auspices of the Mellon Research Forum on the Value of a Liberal Arts Education.

© 2019 The Andrew W. Mellon Foundation. This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 License. To view a copy of the license, please see <https://creativecommons.org/licenses/by-nc/4.0/>.

The Mellon Foundation encourages distribution of the report. For questions, please write to communications@mellon.org.

Introduction

The purpose of this paper is to organize, examine and interpret the empirical evidence regarding the cognitive outcomes of liberal arts education. Commissioning this paper is part of a broader effort by the Mellon Foundation that comprises, among other activities, a set of essays that review the state of research on outcomes and benefits of a liberal arts education, for individuals and for society at large. The intention is to inform a forthcoming major initiative by the foundation on liberal arts education. (Note that in this paper the phrase *liberal education* will be used to denote what in the literature is often termed *liberal arts education*.)

Although the charge is straightforward, there are several challenges in setting the parameters of the study. These are clarifying what is generally meant by the term *liberal education* and identifying those institutions with a liberal education ethos, as well as those students who, to some extent, have been exposed to a liberal education. Another is delineating the cognitive outcomes of interest, as well as identifying and critically evaluating the instruments that have been used to measure those outcomes. A third challenge, building on the first two, is obtaining credible evidence on the impact of liberal education on student outcomes in contrast to the impact of other forms of undergraduate education. These challenges will be further elucidated in what follows.

Determining which students should be deemed to have experienced a liberal education is not at all straightforward. In a few studies (e.g., Pascarella et al., 2013) liberal arts colleges are explicitly identified and contrasted with other types of colleges. However, it is likely that some students in other colleges experience some form of liberal education. One prominent example is furnished by students enrolled in honors colleges, or colleges of arts and sciences, within larger undergraduate institutions. The academic models of these units are intended to emulate those of liberal arts colleges. Students may also have co-curricular or extracurricular experiences designed to foster learning communities more typical of liberal arts colleges. More generally, students in these or other institutions may enroll in courses whose characteristics are very similar to those experienced by students in liberal arts colleges. Later we report evidence that such “liberal arts experiences” have a discernible impact on learning outcomes (Siefert et al., 2008; Pascarella et al., 2005, 2013).

Sometimes liberal education is characterized by its intended outcomes instead of by the type of institution or the experiences provided. From this point of view, a liberal education is one that aims to nurture the skills, dispositions, critical faculties, ethical precepts, moral sentiments, life purposes, and the like that students are expected to develop in order to lead lives that are productive and meaningful. Liberal education is often contrasted with a more vocationally-oriented education whose primary purpose is to prepare students for a particular niche in the labor market—accounting, engineering, health care, journalism and the like.

Cognitive outcomes of liberal education can be organized in various ways. One can begin with those outcomes that are generic—that is, not tied to specific disciplines. Common examples are critical thinking, written communication, and quantitative reasoning. There are various taxonomies of such outcomes. Although they are not identical, there is considerable overlap. Although outcomes like critical thinking clearly fall in the cognitive domain, others such as curiosity, need for cognition, openness to new ideas, and moral reasoning, have cognitive

components and are certainly highly valued.¹ Thus, it seems appropriate to cast the cognitive net as widely as the data will allow.

Cognitive outcomes are measured through some sort of assessment. It could involve a class project, a portfolio, responses to a locally produced assessment, or responses to an externally crafted assessment. In the latter case, students may have little motivation to apply themselves to the tasks as they usually have no bearing on the students' grades. To the extent that there is differential motivation across colleges and, especially, across institutional sectors, biased estimates and comparisons will result.²

At the same time, even with a much-cited skill such as critical thinking, there are many definitions and many different instruments, involving a variety of approaches, formats, and properties that all purport to measure that skill. These range from self-reports to relatively short assessments to probes that require extended responses. (Liu et al., 2014). Consequently, comparing results from different studies must be done with due attention to this heterogeneity, as well as the substantial methodological variations that are endemic to the literature (Braun & Mislevy, 2007).

In contrast to generic outcomes, an alternative is a focus on discipline-specific cognitive outcomes. To put it in stark terms, one can ask: In what ways, if any, does the typical economics major at Amherst College differ from the typical economics major at the University of Massachusetts, Amherst (or the University of Massachusetts, Boston) with respect to their mastery of the knowledge and skills of economics, as well as their ability to apply "economic thinking" to a broad range of problems?

Even leaving aside likely differences among institutions in their students' pre-college preparation and experiences, this is a difficult question to answer. The question does not refer to generic skills, such as critical thinking, for which there are extant standardized assessments. Rather, it addresses skills grounded in disciplinary content. In this respect, current standardized assessments (e.g., the major field tests offered by ETS) capture only relatively unsophisticated aspects of the construct. Assessments that tap into higher level facets of the construct are usually locally developed and, thus, do not permit cross-institutional comparisons. As explained below, disciplines such as economics are only beginning to achieve a broad consensus on learning outcomes for majors; the development of aligned, standardized assessments lies in the future (Arum, Roksa, & Cook, 2016).

Based on the writings of liberal education's proponents and defenders, one can surmise that they would argue that the essential "value-added" of a liberal education manifests itself most strongly in a student's later life, where long-term success will depend not so much on specific learnings as on the ability and willingness to continue to learn, to adapt to new settings, to work with diverse colleagues, to see the "big picture" and to act accordingly. This is certainly the position taken by the Association of American Colleges and Universities (AAC&U) in many

¹ Need for cognition is defined as "the tendency for an individual to engage in and enjoy thinking" (Cacioppo & Petty, 1982, p. 116). Individuals with a higher need for cognition are characterized by a greater need to understand the experiential world and more inclined toward a search for high elaboration.

² In statistics, bias denotes an estimator that is systematically off-target. This is different from lack of precision.

of its publications.³ Leaders of elite liberal arts institutions have advanced the same argument (Katz, 2008). Some of these learnings are more clearly cognitive than others.

However the outcomes of liberal education are delineated, a focus on goals alone does not specify what a liberal education consists of. What are the curricular requirements, pedagogical strategies, co-curricular activities, and so on, that characterize a liberal education? To what extent do they differ quantitatively and qualitatively from other forms of undergraduate education?

It is important to clarify these issues as current debates center on the value of a liberal education, with an implicit or explicit contrast with more vocationally-oriented undergraduate experiences. If the comparisons are to be meaningful, therefore, we must be able to distinguish those students who have been exposed to a liberal education from those who have not. The difficulty, as Brighouse (2017) notes, is that in addition to characterizing liberal education by its aims and objectives as suggested above, there are at least two alternative approaches to characterizing liberal education: by the characteristics of the institution (e.g., small classes, undergraduate focus) or by the design of the course of study.

Clearly, it is incorrect to argue that only students enrolled in institutions that are self-described as liberal arts colleges have received a liberal education. Some students at traditional liberal arts colleges may emerge after four years without having undergone measurable changes, while students in other institutional settings (e.g., an honors college at a flagship state university) may receive a fine liberal education with demonstrable changes in key domains. Finally, even vocationally or technically oriented schools often have general education requirements that approximate, to differing degrees, some aspects of a liberal education. Unfortunately, the data available rarely permit us to make such distinctions or to place students on a continuum of exposure to a “liberal arts experience.”

Another challenge is that the phrase “the value of a liberal education” implicitly refers to the extent to which the stated goals have been attained as a result of attendance at the institution.⁴ Thus, it is not sufficient simply to document the student’s level of attainment of a goal like critical thinking); rather, one must also have evidence of the student’s level of attainment at matriculation in order to obtain an estimate of the amount of growth (or “value added”) that has occurred over the four years. Unfortunately, such evidence is rarely available and its absence makes assertions regarding the contribution of a liberal education to cognitive growth rather speculative—even in the presence of substantial outcome data. Moreover, there is credible evidence of substantial heterogeneity in the value-added of exposure to a liberal education, even among students enrolled in small, liberal arts colleges.

To appropriately address the charge for this paper, it is essential not only to find evidence regarding cognitive outcomes, but also to evaluate that evidence with regard to quality and relevance. As noted above, extant evidence is generated in myriad ways and varies considerably with respect to both criteria. Indeed, given the importance of the question, one can ask why

³ For empirical support, see Hart Research Associates (2013).

⁴ As one reviewer points out, it is certainly possible that some student learning takes place from simultaneous experiences **outside** the institution that have nothing to do with the institution or its curriculum. At this juncture it would be impossible to disentangle the different contributions.

there is so little high quality evidence of direct relevance to these issues. Part of the answer is how assessment is viewed in higher education.

For the most part, faculty view assessment as part of their prerogative as teachers of record. Especially in advanced courses, faculty develop the curriculum, adopt a particular pedagogy and decide how to evaluate student performance, utilizing some combination of assignments, quizzes, end-of-course tests and final projects. In many cases, students must complete a capstone project or a senior thesis. In general, the assessment strategy is not designed to generate evidence with respect to a publicly available set of competency standards and thus precludes comparisons among institutions. On the other hand, externally developed, standardized assessments do allow for cross-institutional comparisons. However, because of pragmatic and cost constraints, these assessments typically elicit evidence only with respect to some aspects of liberal education outcomes with others left unexplored. Moreover, as noted above, the quality of the evidence obtained is potentially undermined by variations among students in how seriously they approach the assessment exercise.

There are strong indications that large numbers of administrators and faculty in many institutions, including but certainly not limited to self-described liberal arts colleges, are beginning to see value in systematic assessment and in building an assessment culture on their campuses (Kuh et al., 2014; Paris, 2011). At present such initiatives may well be contributing to the quality of teaching and learning but rarely offer direct evidence regarding the “value of a liberal education.” Nonetheless, the science and practice of assessment of changes in specific cognitive skills is further along than that of changes in other valued college outcomes and, consequently, widespread adoption of systematic assessment could yield more useful information on the contributions liberal education makes to cognitive growth.

In sum, although there are promising developments with regard to the assessment and evaluation of higher education learning outcomes generally, the challenges associating with measuring liberal education outcomes in particular are numerous, varied, and often technical in nature. In this paper we review the literature on what is known about the contributions of liberal education to the improvement of cognitive abilities and performance. Further, and perhaps more importantly, we elaborate on some of the challenges in moving the field forward.

The paper is organized as follows: The next section discusses what we know of the cognitive outcomes of liberal education, followed by a section on assessment for comparison, with a full section devoted to the VALUE initiative of the AAC&U. These are succeeded by an extended summary of key studies of institutional effectiveness, as well as brief reviews of (local) college assessments, discipline-based assessments and long-term outcomes. The concluding section provides an overall summary and some recommendations.

Challenges and Possibilities in Measuring the Outcomes of Liberal Education

There have been many attempts to delineate the learning outcomes/goals associated with liberal education, typically blending more and less clearly cognitive elements. An early contribution was made by Thomas (2002, p. 30) who proposed six goals: (i) The ability to use knowledge wisely, to think creatively, to be able to learn independently, to make balanced choices, to exercise

judgment, and to apply these skills for the common good. Unfortunately, Thomas did not discuss how to evaluate whether, and to what extent, a student has attained one or more of these goals. (Indeed there is scant mention in the higher education literature of how one can tell the degree to which the graduating student has actually attained Thomas' goals or those offered by others.)

Subsequently, the AAC&U (2005) proposed three clusters of higher education outcomes: (i) knowledge of human culture and the natural world (science, social sciences, mathematics, humanities, arts); (ii) intellectual and practical skills (written and oral communication; inquiry, critical thinking, and creative thinking; quantitative literacy; information literacy; teamwork; integration of learning); and (iii) individual and social responsibility (civic responsibility and engagement; ethical reasoning; intercultural knowledge and actions; propensity for lifelong learning). Clearly, the first two clusters have a strong cognitive flavor. A second key AAC&U document, *College Learning for the New Global Century* (2007), presents the five "Essential Learning Outcomes" and the case for their critical role in preparing American students for the 21st century.

Another important contributor to this literature is the staff of the Center of Inquiry in the Liberal Arts at Wabash College. Blaich et al. (2004) describe the product of a liberal arts education as an individual who is (i) intellectually open to inquiry and discovery, and (ii) possessing the ability and the desire to adopt a critical perspective on her own and of others' beliefs. Building on work undertaken under the auspices of the Wabash National Study of Liberal Arts Education (WNSLAE), as well as the extant literature, King et al. (2007, p. 5) formulated an expanded set of outcomes. Their Table I is reproduced below:

Table I.
Liberal Arts Outcomes: Wabash National Study of Liberal Arts Education
(King et al., 2007)

Integration of learning is the demonstrated ability to connect information from disparate contexts and perspectives—for example, the ability to connect the domain of ideas and philosophies with the real world, one field of study or discipline with another, the past with the present, one part with the whole, the abstract with the concrete—and vice versa.

Inclination to inquire and lifelong learning reflects a strong desire to learn, ask questions, and consider new ideas. Such learning involves taking initiative to learn, not being satisfied with a quick answer, and possessing intrinsic motivation for intellectual growth. These dispositions lend themselves to a lifelong pursuit of knowledge and wisdom.

Effective reasoning and problem solving involves the capacity to make reflective judgments; think critically and independently; and analyze, synthesize, and evaluate information in order to make decisions and solve problems.

Moral character involves the capacity to make and act on moral or ethical judgments, treating others with fairness and compassion; this capacity includes several facets of morality: discernment, reasoning, motivation, and behavior.

Intercultural effectiveness includes knowledge of cultures and cultural practices (one's own and others'), complex cognitive skills for decision making in intercultural contexts, social skills

to function effectively in diverse groups, and personal attributes that include flexibility and openness to new ideas.

Leadership entails the seven core values of Astin and his colleagues' Social Change Model for Leadership. Within the model, the core values fall into three categories: personal or individual values (consciousness of self, congruence, commitment), group values (collaboration, common purpose, controversy with civility), and a societal and community value (citizenship).

Well-being encompasses four dimensions: subjective, psychological, social, and physical. Subjective wellbeing is associated with happiness, life satisfaction, and life quality. Psychological well-being is the pursuit of meaningful goals and a sense of purpose in life. Social well-being refers to positive social health based on one's functioning in society. Finally, physical well-being is characterized by positive health-related attributes.

Finally, Detweiler (2016) examined the mission statements of 238 liberal arts colleges and derived a set of five common characteristics of liberally educated students: (i) Life-long learners; (ii) Making thoughtful life choices; (iii) Leaders; (iv) Professionally successful; (v) Committed to understanding cultural life.

Two points are worth noting. First, there is considerable overlap among the lists, though the agreement is somewhat obscured by the use of different terms and, second, there is no explicit mention of cognitive skills. On this last point, though, it is clear that cognitive skills are embedded in many of the goals. A more explicit exposition of the cognitive domain can be found in a report issued by the National Academy of Sciences (Pellegrino & Hilton, 2012). The study offers a three-part decomposition of the cognitive domain: (i) cognitive processes and strategies; (ii) knowledge; (iii) creativity. The first is further elaborated as: critical thinking, problem solving, analysis, reasoning/argumentation, interpretation, decision making, adaptive learning, and executive function. These facets of critical thinking are often implicit to some degree in many of the frameworks found in the literature.

The present review of the literature on higher education outcomes, particularly those produced by test vendors, has yielded three key generic, cognitive skills: critical thinking/analysis, quantitative reasoning (literacy), and written communication. An ongoing debate concerns whether and how these skills can be fairly and validly assessed for students without regard to their major or, rather, whether they should be more properly assessed in the context of particular disciplines, with some expectation that the students possess relevant content knowledge. Clearly, purveyors of critical thinking assessments for higher education presume that meaningful assessment is possible irrespective of major (e.g., Shavelson, 2010), but there are other views (e.g., D. Koretz, personal communication, May 30, 2017). A broad review of assessment issues in higher education, both generic and discipline-based, can be found in Zlatkin-Troitschanskaia et al. (2015).

One complication is that, just as is the case with liberal education outcomes in general, there are many different definitions of critical thinking, leading to different assessment frameworks and assessment instruments. Liu et al. (2014) provide a comprehensive review of existing frameworks and current assessments. Despite such differences one can hope that performances on these assessments would be moderately to highly correlated. Indeed, a study by Klein et al. (2009) found reasonable agreement among three commercially available tests of critical

thinking. Nonetheless, it is important to keep such differences in mind when reviewing various studies. Beyond these three generic skills, aside from the reports by Pascarella et al. (2005), Siefert et al. (2008), and Pascarella et al. (2013), discussed below, there appears to be very little data enabling comparisons of outcomes among different types of institutions.

Assessment for Comparison

There is a trend among undergraduate institutions to adopt an assessment strategy to accomplish one or more goals. One goal may be to meet the demand of an accrediting body to demonstrate that its students are achieving the learning outcomes described in its catalog. If the intent is simple compliance then it is unlikely that such efforts will have much impact on teaching and learning. However, if an institution's administration seizes the opportunity to engage the faculty in a broad-based evaluation of learning outcomes with a view to informing curricular (and co-curricular) changes, then there is a greater chance of impact—though the difficulty many colleges face in “closing the loop” that connects evaluation and action is substantial (Banta & Blaich, 2011). A more positive report describes the efforts of a collaboration among some members of the Council of Independent Colleges, using the Collegiate Learning Assessment (discussed below) to modify instructional practices with the goal of enhancing the development of critical thinking among students over the course of their undergraduate years (Paris, 2011).

Colleges can certainly “go it alone”; that is, have faculty develop explicit standards for learning outcomes and develop new assessments, or adapt existing assessments, to gauge student learning with respect to those standards. Colleges can also participate in collaborative efforts that involve many institutions developing assessment frameworks and/or generic scoring rubrics. Examples are initiatives such as the *Degree Qualifications Profile (DQP)* and the *Valid Assessment of Learning in Undergraduate Education (VALUE)* project (discussed below). On the other hand, colleges can adopt standards that have been built by others and/or administer externally developed assessments. Of course, hybrid strategies are possible, with a college employing a combination of local and external assessments. In fact, Kuh et al. (2014) report that a typical institution uses five different assessment approaches, including the use of surveys that rely on student self-report.

With the use of external assessments comes the possibility of making comparisons with other institutions—though such comparisons are subject to misinterpretation because of differences among institutions in how they participate in the assessment exercise. Such differences include the ways in which students are selected to sit for the assessments, how representative they are of the cohort, and how students are motivated to participate. It is important to bear in mind that even accounting for these differences, raw comparisons of outcomes do not directly address the question of relative effectiveness as there is no adjustment for differences in students' pre-matriculation characteristics.

One route to assessing student learning is simply to ask students via a questionnaire to assess their own learning. (The National Survey of Student Engagement [NSSE] is a widely used example.) However, Pascarella et al. (2010) cite research indicating that student self-reported gains on cognitive dimensions are only weakly correlated with their gains on direct measures of those same dimensions. The clear implication is that there may be no substitute for direct

measurement of the skills of interest. NSSE also asks students to report on the character of the educational practices they have experienced. Pascarella et al. (2010) found a more encouraging result: for example, students who reported frequently experiencing high levels of academic challenge did better than others on a test of their ability to reason effectively.

Colleges that are interested in generating evidence regarding education outcomes to facilitate comparisons with other institutions have a number of choices. Most obviously they can select standardized assessments of generic skills offered by well-known vendors such as ACT, CAE, and ETS. Table 2 provides brief descriptions of the most commonly used assessments, including the ETS Major Field Tests and the Critical Thinking Assessment Test developed by the Tennessee Technological University.

Table 2.

Commercially Available Assessments of Liberal Education Outcomes

ACT: CAAP

The CAAP is intended to measure the outcomes of general education and thus assesses both general and specific skills. It comprises six modules that can be administered in any combination. The modules are: critical thinking, reading, mathematics, science, writing skills, and writing essay. With the exception of the writing essay module, the modules comprise machine-scored, multiple choice items. The writing essay module comprises two essay prompts that are scored by human raters following a detailed scoring rubric. The scoring can be done by the college's own faculty (after appropriate training) or by external raters.

Each CAAP module is accompanied by national norms so that a school can compare its score distribution to the national distribution. The latter is derived from a heterogeneous sample of schools and students, aggregated over years. It is possible to obtain norms by year of administration, as well as for subsamples stratified by school type (2 year vs. 4 year), governance (private vs. public), Carnegie classification, and year of study (freshman, sophomore, etc.) Comparison to these norms is arguably more relevant and informative. However, inferences based on these comparisons must be made cautiously because of remaining differences in student pre-matriculation characteristics, as well as student selection and motivation.

There are also linkages between performance on the ACT (college admissions test battery) and the CAAP modules, so that colleges can compare the percentile distributions of the students prior to matriculation and at CAAP administration. Again, caution is in order since the interpretation of the comparison depends on how the reference population yielding the two sets of norms is chosen.

CAE: CLA+

The CLA+ comprises 2 subtests: A performance task (PT) and a Selected Response Question (SRQ) section. The PT presents a document or problem statement and an assignment based on that document and elicits an open-ended response. The response is graded with respect to three skills: Analysis and Problem Solving, Writing Effectiveness, and Writing Mechanics. The

SRQ includes 25 multiple choice questions that yield subscores on three domains; Scientific and Quantitative Reasoning, Critical Reading and Evaluation, and Critique an Argument. PT and SRQ raw scores are separately analyzed to yield sub-test reported scores and an overall (average) score. Score reports for 2015–16 can be found in Council for Aid to Education (2016). Scores are displayed for freshman and seniors, overall, as well as by PT subscores and SRQ subscores, disaggregated by various institutional characteristics. Measures of student cognitive growth and institutional impact are provided as well.

ETS: HEIghten

HEIghten is a computer-based suite of assessments that is not yet fully developed. It will have five modules: Critical thinking, quantitative literacy, written communication, intercultural competency and diversity, civic competency and engagement. Questions or prompts are presented in different formats but student responses are generally multiple choice or multiple select. There is no data available yet.

ETS: Proficiency Profile

This is a single examination available in long (2 hours) and short (40 minute) forms. It assesses four skills: reading, writing, critical thinking and mathematics. The examination comprises multiple choice items. An optional essay module is available.

National norms are published for different institutional sectors (e.g., research universities, liberal arts colleges) by student level (freshman, sophomore, etc.) and by type of test administration (proctored, not proctored). Thus, for example, for the liberal arts college sector one can compare aggregate results between the freshman and senior years. However, since the same students are not measured at different points in time, the comparisons between freshmen and seniors capture not only how much individual students learn but also other differences between the freshman and senior student populations. Similar considerations apply to comparisons between institutional sectors.

ETS: Major Field Tests

This test battery comprises multiple choice assessments in twelve academic majors, as well as three levels of business (associate, bachelor's, master's). They are administered in the senior year and are intended to measure students' mastery of core concepts and skills in the domain. Individual and departmental reports are provided and results can be compared with aggregate results compiled from all participating institutions in that same academic year.

Tennessee Technological University (TTU)

TTU markets the Critical Thinking Assessment Test (CAT), which is designed to assess four facets of critical thinking: evaluating information, creative thinking, learning and problem-solving, and communicating (Stein & Haynes, 2011). The CAT comprises a number of situations, with each eliciting short answer responses from students. These situations are developed by faculty from different institutions and are vetted by TTU staff. Student responses are graded by local faculty who have been trained to score according to detailed rubrics. In this case, comparability of results depends on ensuring consistency in application of the scoring rubrics. Because any given student is only tested once, comparisons of the test scores of freshmen and seniors conflate the amount of learning from individual students with the effect of other ways that the

freshmen and senior classes may differ. The TTU approach addresses some faculty concerns by providing much greater direct faculty involvement in design and scoring. Since scoring is done communally, it is seen as providing opportunities for professional development.

Standardized assessments like those described in Table 2 possess a number of advantages. First, and most obviously, students in different institutions take the same or parallel forms, thereby enabling cross-institutional comparisons. The large investments made in developing these assessments help ensure that scores are reliable and can be consistently interpreted. Nonetheless, a common concern of faculty regarding these assessments is that they are disconnected from students' coursework and may omit assessment of some valuable forms of learning. Moreover, faculty and others may worry that the vendors who develop these assessments do not adequately draw on faculty expertise and perspectives. The educational process may come to focus too much on what is easiest to measure rather than what is most important to know or be able to do.

Standardized assessments of higher education outcomes became more salient (or notorious) with the release of the "Spellings Commission" report (U.S. Department of Education, 2006), which argued for the need for outcomes assessment to provide evidence that undergraduate education was achieving its mission. The report was greeted with strong opposition from much of the higher education community, although not solely because of this recommendation. As a response, in 2007 two consortia of public institutions, the American Association of State Colleges and Universities (AASCU) and the Association of Public and Land-grant Universities (APLU) initiated the Voluntary System of Accountability (VSA). One component of the VSA concerned learning outcomes, specifically generic skills such as critical thinking, analytic reasoning and written communication. Initially, the approved assessments were those offered by ACT, CAE, and ETS. Subsequently the VALUE rubrics developed by the AAC&U were added to the list. Given its prominence, we provide an extended discussion of this initiative.

The AAC&U VALUE Initiative

In many respects, the AAC&U has been the leader in advancing the argument for the essential role of liberal education through publications, convenings and research. It has been particularly assiduous in gathering expressions of appreciation from leaders in business and government for those individuals who enter the workforce with the skills and dispositions that are cultivated in an environment that honors the goals of liberal education (AAC&U, 2002; Hart Research Associates, 2013).

With the release of the Spellings Commission report, faculty and administrators across higher education voiced concerns regarding its recommendations, especially the perceived pressure to adopt externally developed, standardized assessments. In response, AAC&U launched the Valid Assessment of Learning in Undergraduate Education (VALUE) initiative. The underlying rationale was that assessment data is most useful if it is *authentic*; that is, it reflects the work that faculty demand of students in their courses or capstone projects and, moreover, it is evaluated by faculty in ways that reflect faculty-developed standards of achievement. At the same time, in order to achieve a degree of comparability in performance across institutions,

procedures need to be standardized in some respects. In the VALUE initiative, faculty submit student work products from a senior level course that aims at a particular learning outcome and is designed to elicit performances at the highest level. The ensemble of work products is then graded by multiple faculty from different institutions. Student motivation should be high because performance contributes to their course grades. Comparability is achieved through the use of detailed scoring rubrics based on extensive faculty input, as well as training sessions for faculty to apply those rubrics. Student work is evaluated for what it reveals about students' proficiencies across a wide range of "skill domains," including critical thinking, problem solving and fourteen others. (See Table 3.) Appendix I describes the elements of the critical thinking rubric in detail.

Table 3.
AAC&U VALUE Domains

<i>Intellectual and Practical Skills</i>
Inquiry and analysis
Critical thinking
Creative thinking
Written communication
Oral communication
Reading
Quantitative literacy
Information literacy
Teamwork
Problem solving
<i>Personal and Social Responsibility</i>
Civic engagement—local and global
Intercultural knowledge and competence
Ethical reasoning
Foundations and skills for lifelong learning
<i>Integrative and Applied Learning</i>
Integrative learning

A recent publication, *On Solid Ground* (AAC&U, 2017), presents preliminary results from more than 90 two-year and four-year institutions for three of these skill domains: critical thinking, quantitative literacy, and written communication. While this is clearly work in progress, it does reveal some of the possibilities and inherent difficulties of finding a suitable compromise between authenticity and comparability.

The representativeness of the work samples both with respect to the students and to the assignments chosen for submission is very much in doubt. The sample of students whose work is evaluated may not be fully representative of the senior class, by academic major, for example. Moreover, there is no central control of the quality and appropriateness of the assignments. That is, assignments given by some faculty may not call for some of the skills that constitute the framework or they may not be sufficiently demanding to support distinguishing between good and excellent performance. The testing agencies employ designs that minimize problems like

these, but at the expense of failing to present students with an assessment environment that closely tracks their “normal” classroom experience.

The VALUE initiative is an imaginative response to legitimate public concerns regarding higher education outcomes and understandable faculty concerns regarding mandated assessments that are seen as intrusive and of little pedagogical value. But any attempt to satisfy all constituents is bound to fall short. The VALUE staff recognize this and are working to remedy some of these issues, including the issue of scalability. Others are beyond their purview. With regard to the focus of this paper, it is unlikely this initiative will yield data that informs comparisons between liberal education and other types of undergraduate education.

Institutional Effectiveness

We have focused attention so far on the challenging problems of measuring students’ levels of cognitive skills. We turn now to the equally challenging question of whether the gains in cognitive outcomes of students receiving a liberal education are demonstrably greater than those of students receiving other types of education, after making appropriate adjustments for pre-matriculation differences. This is a question related to the relative “value-added” by different institutions.⁵ Randomized experiments are, in principle, an especially attractive way to address questions of this kind, but it is obviously infeasible in almost all circumstances to assign students to colleges randomly. A next-best option is to solicit the participation of a reasonably large number of schools with different characteristics, draw large random samples of freshman students from each, and obtain background information on each student, along with baseline outcomes data. Further outcomes data should be elicited at intermediate points and certainly toward the end of the senior year. The background data and baseline outcomes test scores can be used to adjust the student samples in the different sectors or institutions so that they are more nearly statistically equivalent. A somewhat less attractive alternative is to carry out the statistical adjustment using background data but without the baseline outcomes test scores. In the evaluation literature, such approaches to data collection and analysis are known as quasi-experimental designs.

Longitudinal studies, which follow individual students over time, have many attractions for studying student learning or improvement. Tracking students over the full four years yields evidence on their trajectories and also on what they experienced, both in and outside the classroom, during those four years. Such rich data can help to elucidate causal relationships. However, there are many logistical and practical difficulties in carrying out these studies: Students may refuse to participate in some of the planned data collections, or transfer to other schools or drop out. Although statistical adjustments can be made to partially mitigate the

⁵ Value Added Models have been employed almost exclusively in studies of primary and secondary education. In higher education, the Assessment of Higher Education Learning Outcomes (AHELO) project of the OECD explored the feasibility of measuring learning outcomes internationally (Ewell, 2012). As part of AHELO, a working group was convened to explore the possibility of estimating institutional value-added. The consensus was that this would not be a fruitful direction to pursue for a host of technical and logistic reasons (Braun, 2013). On the other hand, a successful example can be found in Colombia where all college entrants must take a battery of high school leaving tests and, before college graduation, they must also take a battery of college exit tests (Shavelson et al., 2016). The conjunction of pre- and post-assessments is what makes value-added analysis feasible.

effects of these difficulties, they cannot fully compensate for incomplete data. Thus despite their conceptual attractiveness, longitudinal studies of the cognitive effects of liberal education are rare. We are able, nonetheless, to report on four such studies below.

An alternative to longitudinal studies is to obtain cross-sectional, contemporaneous data from a freshman student sample and a senior student sample in the same year. (As we will see, this technique is often used when individual institutions study their own practices.) The differences between assessment scores for the freshmen and the seniors is used, with appropriate statistical adjustments, to estimate the learning gain a typical student achieves. This approach is logistically simpler, less costly and delivers results in one year rather than four. However, factors like changes in admissions standards, differential attrition of students, and curricular changes can raise doubts about whether the usual statistical adjustments are sufficient to make the comparisons trustworthy.

With either a longitudinal or cross-sectional study design, it is important to collect a broad range of data on students as they enter college in order to address the problem of self-selection. Self-selection arises because students are not randomly sorted into colleges. There are systematic differences across colleges in the characteristics of their entering classes that are plausibly related to their outcomes as seniors. Such characteristics include prior academic achievement, family background, etc. Without taking these initial differences into account, simple comparisons of (say) average senior test scores tell us very little about the relative benefits of a liberal education. A related issue is students' self-selection into different academic tracks and majors within schools that very likely also impact their cognitive outcomes. Thus, investigators must take into account these differences in order to produce meaningful comparisons.

Another concern with comparative studies is the nature of the samples of students who sit for the assessments. It is rare that the sample consists of either a census or a true random sample of the students in the cohort. More commonly, the assessment is administered to a voluntary, self-selected sample enticed by the promise of monetary or other rewards. The use of non-representative samples limits inferences from the sample to the full population. That the departure from randomness likely differs across institutions further undermines the credibility of estimates of differences in institutional contributions to student learning.

Another alternative to either a full longitudinal study or a comparison of freshmen and seniors in the same year is to collect comprehensive, pre-matriculation information on students (seniors). Then statistical regression can be used to adjust for differences upon entry in order to obtain a plausible estimate of the average difference in outcomes between liberal education students and others. It is particularly helpful to have one or more measures of academic preparation. Although, in principle, this approach is inferior to a full longitudinal study it does have some advantages. It can be done at a point in time and so is much less costly. Also some adjustments for attrition from the freshman to the senior year can be made that are analogous to the adjustments that are made in a longitudinal study.⁶ Despite the challenges, several high

⁶ This approach has been used by the Council for Aid to Education (CAE) in the analysis of data from administration of the Collegiate Learning Assessment (CLA), a test of critical reasoning (Council for Aid to Education, n.d.). There has been considerable debate regarding the credibility of estimates obtained in this manner

quality longitudinal studies have taken on the liberal education question. We report on four of those studies here.

ASHE Study (Pascarella et al., 2005)

This ambitious study had five overarching research questions, of which four are relevant here. First, does a comparison among institutional sectors yield evidence of an advantage to liberal arts colleges (LACs) with respect to exposing their students to educational practices that are empirically associated with better student outcomes? Second, do such sectoral comparisons yield evidence that LACs actually produce better student outcomes and, if so, to what extent can these results be explained by their advantage in fostering better educational practices? The third question has a more general focus: irrespective of sector, is there a positive association between fostering better educational practices and achieving better student outcomes? Finally, to what extent do the findings for the first three questions hold for all subpopulations of students?

Pascarella et al. (2005) utilized data from the National Study of Student Learning (NSSL). This study involved a heterogeneous set of 16 institutions. The institutional sample comprised five private LACs, four research universities, and seven doctoral-granting regional universities. Students were followed for three years (1992–95). This is a very rich data set with extensive information about students' pre-matriculation characteristics, as well as their curricular and co-curricular experiences in college. Data was collected at the beginning of the freshman year, and at the end of the first, second, and third years. The outcomes collected comprised the five components of the ACT CAAP battery described above, a measure of the students' plans to obtain a graduate degree, and five measures of students' orientations towards learning and educational pursuits. All eleven measures were administered at the beginning of freshman year but not at all subsequent administrations.

This study is noteworthy not only for its authors' recognition of the methodological challenges in properly answering these questions, but also by the effort invested in organizing and analyzing the data to obtain credible estimates of institutional value-added. Thus, for example, the results reported for gains in critical thinking at the end of the first and third years have been adjusted for pre-admission characteristics and critical thinking scores at the beginning of freshman year.

An important finding is that, in comparison to other institutions, LACs do “uniquely foster a broad range of empirically substantiated good practices in undergraduate education” (p. 87) and “that the estimated positive effects of liberal arts colleges on good practices were the most numerous and pronounced in magnitude during the first year of postsecondary education” (p. 88). The authors suggest some mechanisms to explain these findings. Some practices were clearly related to the residential nature of the LAC experience, as well as the smaller class size. But others appear to be influenced by the general ethos of the LAC.

(Banta & Pike, 2007; Klein et al., 2007). A recent contribution to the debate (Zahner & Steedle, 2015) finds substantial differences between true longitudinal and cross-sectional estimates of institutional value-added.

With regard to the eleven outcomes related to intellectual and personal development, the results were decidedly mixed, with LACs having an advantage for some, a disadvantage for others, and no statistical difference in yet others. As the authors note, “although isolated exceptions existed, it was clear that most of the positive impact of liberal arts colleges on dimensions of students’ intellectual and personal development occurred during the first year of postsecondary education” (p. 90). These findings may be due to the fact that the outcome measures employed do not fully capture the impact of a liberal education. Another possibility is that “institutional type (i.e., liberal arts colleges, research universities, regional institutions) may simply be a structural characteristic that is too general and distal to adequately capture the full impact of liberal arts education on students’ intellectual and personal growth during college” (p. 91).

To explore this hypothesis, the authors developed two theoretically and empirically grounded scales that captured an institution’s “liberal arts emphasis” and a third scale that captured an individual’s “liberal arts experiences.” All three scales were composites constructed from student self-reports with respect to a variety of institutional practices and conditions including: intellectual emphasis of the campus, number of essay exams, ratio of liberal arts courses to vocational courses, quality of non-classroom interactions, faculty interest in teaching and student development, academic effort/involvement, integration of ideas, course challenge/effort, and instructional organization and preparation. These responses were aggregated to the individual level to create the “liberal arts experiences” scale and aggregated to the institutional level to create the “liberal arts emphasis” scales.

After introducing a full set of statistical controls, all three measures were strongly associated with student outcomes. The authors’ conclude that “although a liberal arts education emphasis is most likely at liberal arts colleges, it is not exclusive to those institutions. Where it is implemented and nurtured at research universities and regional institutions, it has important impacts on students’ intellectual growth” (p. 92). Similarly, when considering students’ liberal arts experiences, the authors conclude that “... consistent with an institution’s average liberal arts emphasis, a student’s individual liberal arts experiences enhanced intellectual and personal growth irrespective of whether or not one attended a liberal arts college” (p. 92).

With respect to “making the case” for LACs, the authors characterized their findings as a “mixed bag.” One explanation is that even within a relatively small liberal arts institution there is sufficient heterogeneity in students’ experiences and engagement to generate substantial variation in outcomes, net of pre-admission characteristics and achievements.⁷ On the other hand, the findings that students do benefit from liberal arts experiences, irrespective of institutional type, suggests strategies that all institutions can implement to advance students’ intellectual and personal development.

The Effects of Liberal Arts Experiences (Siefert et al., 2008)

⁷ It is commonly found that about one-fourth of test score variance in a pooled sample of students is due to differences among institutions and about three-fourths to the variation among students within institutions.

As noted above, Pascarella et al. (2005) found that irrespective of the type of institution, there was a positive association between students' location on the liberal arts experiences scale and their CAAP scores. This line of research was continued by Seifert et al. (2008), who developed a somewhat different liberal arts experiences scale. The report details a comprehensive study that was conducted to validate this new composite scale.⁸

Siefert et al. (2008) recruited four institutions that were participating in the pilot phase of the Wabash National Study of Liberal Arts Education (WNSLAE) and agreed to administer an additional assessment battery. Approximately 700 students, drawn from all four freshman cohorts on campus, took part in the study. Students filled out a background questionnaire, and answered a series of questions used to locate them on the liberal arts experience scale, and responded to a number of assessments, which are listed in Table 5. These assessments were chosen to represent important outcomes of liberal education. (Note that each student took only a subset of the full battery of assessments.)

Table 5.
Assessment Battery (Siefert et al., 2008)

Assessment	Reference
Reasoning and Current Issues	Wood et al. (2002)
Intercultural Developmental Inventory	Hammer & Bennet (2001)
Scales of Psychological Well-being	Ryff & Keyes (1995)
Defining Issues Test-2	Rest et al. (1999)
Need for Cognition	Cacioppo et al. (1984)
Socially Responsible Leadership Scale	Tyree (1998)

After controlling for background factors and institution, the authors found a statistically significant relationship between scores on the liberal arts experience scale and scores on four of the six outcome measures. Specifically, there was no relationship with measures of moral reasoning and reflective judgment. There were relationships with measures of intercultural effectiveness, inclination to inquire and lifelong learning, psychological well-being, and socially responsible leadership. Of these four, only the second has a strong cognitive component. The authors speculate on possible reasons for the weak relationship with the other two cognitive measures. With regard to moral reasoning they argue that the difficulty may lie with the particular instantiation of the liberal arts experience scale, while with regard to reflective judgment scores, the scale was not very reliable, making it difficult to find strong evidence of

⁸ A comparison between the Siefert et al. scale and the Pascarella et al. scale can be found in Table I of Siefert et al. (2008, p. 113).

association. Another important methodological limitation is that there were no pre-test scores available.

*Academically Adrift: Limited Learning on College Campuses (Arum & Roksa, 2011)*⁹

This study was based on data collected by the Determinants of College Learning project, undertaken under the joint auspices of the Council for Aid to Education (CAE) and the Social Science Research Council (SSRC). The final analytic database contained information from 2,322 students enrolled in 24 four-year institutions. The institutional sample was not randomly selected but did represent a broad range of college types. Each institution was responsible for the random selection of students from the freshman class and retention through the sophomore year. Overall, retention was less than 50%, though the percentages varied considerably across institutions and by student characteristics.

The assessment chosen was the CLA Performance Task, requiring the student to write a 90 minute essay in response to a problem arising from a realistic scenario that incorporates a variety of source documents. On average, student gains on this measure of critical thinking over the first two years of college were a modest seven percentile points.¹⁰ Again, there was substantial variation in results across institutions and by student characteristics.

In a subsequent report (Arum, Roksa, & Cho, 2011) the authors estimate that the typical gain over four years of college was about 18 percentile points, implying that a student who was at the median as a freshman would by the end of college be at the level of a freshman who was at the threshold of the top third of the distribution. This is comparable to the estimated four-year gain recorded in the Wabash study that used ACT's CAAP Critical Thinking test, which comprises only multiple choice items. Arum & Roksa judge this amount of gain to imply weak performance, but it is not clear how much gain it is reasonable to expect.

Liberal Arts Colleges and Cognitive Development (Pascarella et al., 2013)

In the introduction, the authors note (p. 570) that in the earlier study (Pascarella et al., 2005) the investigators were able

“... to introduce statistical controls for an extensive battery of precollege characteristics and experiences, including a precollege measure of critical thinking. With these controls in place, attendance at an American liberal arts college had no significant impact on a standardized measure of critical thinking [CAAP] over the first year, or over the first three years, of postsecondary education.”

This study drew on longitudinal data from the Wabash study (WNSLAE). The school sample comprised eleven liberal arts colleges, three comprehensive, non-doctoral granting, regional

⁹ Note that this study was not undertaken to compare different institutional sectors.

¹⁰ This means, for example, that a freshman scoring about the 50th percentile (median) of the freshman score distribution would, as a sophomore, score about the 57th percentile of that same freshman distribution.

universities and three research intensive, doctoral granting universities. Invited students were either random samples (from the larger schools) or all students (from the smaller schools). The 4,193 rising freshmen filled out an extensive background questionnaire and responded to a measure of Need for Cognition. About half the sample also responded to a measure of critical thinking (CAAP).

In the spring of their senior year, 2,212 students participated in the follow-up data collection, corresponding to a 47% attrition rate. (Adjustments were made to mitigate the impact of such a high level of attrition.) Students retook the outcome assessments they had taken as freshmen. They also responded to instruments that measured their exposures to effective classroom instruction and to deep learning experiences.¹¹ In addition, the authors collected data on circumstances during college that might influence the assessments: place of residence during college, work responsibility, academic major field of study, and co-curricular involvement (Pascarella et al., 2013, p. 577).

In the first phase of analysis, the authors found that "... net of student precollege characteristics and academic major during college, attendance at a liberal arts college significantly enhanced exposure to clear and organized instruction and the use of deep approaches to learning" (p. 578). In the second phase, the authors examined the impact on assessment results depending on whether a student attended a liberal arts college vs. other types of institution and found modest advantages in critical thinking skills and need for cognition for students who attended LACs compared to alternatives. (p. 578)

Pascarella et al. (2013) then note that "... when exposure to clear and organized classroom instruction and deep learning experiences were added to the equations ... the estimated effects of attending a liberal arts college were reduced to nonsignificance. This suggests that a significant portion of the positive effect of liberal arts college attendance on our measures of cognitive growth is mediated through classroom instruction and student use of deep approaches to learning" (p. 578). Although these findings are drawn from a relatively small and non-representative sample of institutions, they are certainly suggestive and in agreement with other findings, such as those in Seifert et al. (2008) and in reports from individual courses. If they are replicated in subsequent studies, they have important implications for policy.¹²

College Assessments (Local)

Over the last 15 years many undergraduate institutions, including liberal arts colleges, have made progress in developing comprehensive strategies to assess student learning outcomes. In

¹¹ For an instructive exemplar of deep learning in the context of a course in philosophy, see Neumann (2014). Neumann emphasizes the importance of learning grounded in rich content.

¹² Laird et al. (2008) studied the relationships between student and faculty reliance on "deep approaches to learning" and type of discipline based on the Biglan (1973) typology. Using senior class data from the 2005 administration of the NSSE, they found systematic differences by discipline cluster. Moreover, irrespective of discipline cluster, there were moderately strong relationships between students' frequency of engaging in deep learning behaviors and their (self-reported) learning gains. Since deep approaches to learning constitute a component of the liberal education experience scale, this study provides support for the further investigation of students' academic experiences and their relationship to both cognitive and non-cognitive outcomes.

part this has been a response to pressure from accreditors and, in part, a recognition that in a competitive marketplace it is advantageous to be able to cite evidence supporting the rhetoric describing the institution's education goals. For a number of liberal arts colleges, sustained support from the Teagle Foundation provided the resources to mount such an initiative and press forward.

Progress on assessment is certainly welcome. From the perspective of this paper, however, there is relatively little to be gleaned as the outcomes are generally neither publically accessible nor designed for cross-institution comparison. Nonetheless, a brief review is in order as over time it may be possible to leverage these efforts to generate evidence that is relevant to the question at hand.

We conducted a (non-systematic) review of assessment efforts at a number of liberal arts colleges and other undergraduate institutions. To construct the sample, we selected institutions that were featured in case studies from the National Institute for Learning Outcomes Assessment (NILOA) and the AAC&U, as well as leading liberal arts colleges according to US News & World Report rankings. We compiled information on each school's assessment practices from its website. In some cases, we found that colleges either did not provide information on assessment activities or there were no systematic assessment efforts. A general summary of the findings, along with some abbreviated case studies, is contained in Appendix 3.

We learned that a number of undergraduate institutions have articulated student learning outcomes (SLOs) that are consistent with their institutional missions. In our sample the number of SLOs varies from three (Augustana College) to twelve (Pace University). Six to eight SLOs is typical. In most cases, each SLO is accompanied by a detailed rubric that indicates the evidence needed to document that a student has achieved the objective.

The principal rationale advanced for the assessment strategy is that it is essential to a system of continuous improvement focused on enhancing teaching and learning. This is consistent with the recommendations of Dwyer et al. (2006) and Liu (2017). Consequently, the emphasis is on providing useful feedback to both faculty and students. In most cases, SLOs at the course, departmental or program level are linked to one or more institutional level SLOs. The goal is to create an institutional culture in which assessment is seen by the faculty as an integral part of their pedagogy and an indispensable tool for improving the quality of student work.

There is substantial variation in how prescriptive colleges are with respect to how individual faculty and departments are to conduct their local assessments. In some colleges, there is a great deal of flexibility as to which courses, and on what cycle, are nominated to provide data for a college-level report. Flexibility also governs what assessments are required and how student work is to be evaluated. In others, certain courses are required to contribute data on a specified cycle. Further, each SLO is accompanied by a detailed rubric that is intended to guide faculty scoring.

Colleges also differ in the kinds of assessments they employ. Almost all colleges we reviewed use locally developed assessments and rubrics. The rubrics are sometimes informed by consultation with materials made available through national organizations such as the Mathematical Association of America and the National Council of Teachers of English (e.g.,

Augustana College), as well as in consultation with/borrowing from the VALUE rubrics. However, this was not always specified in publicly available documents.

Assessment activities are distributed across the four years of the undergraduate program with no particular emphasis on the senior year. At the same time, large numbers of undergraduate institutions do employ standardized assessments offered by various test vendors. In this setting, seniors are often the focal group. As noted in a previous section, unlike the results of local assessments, the results of the standardized assessments facilitate comparisons through norms at the national level or with peer groups. Such comparisons, however, must be made with due caution as the characteristics of the sample of students sitting for these assessments varies considerably across institutions. In addition, many institutions participate in the National Survey of Student Engagement (NSSE) and other national surveys that rely on student self-reports.

Colleges extoll the benefits of an intensive investment in assessment but are not particularly forthcoming about quantifying improvements in student outcomes—whether on local assessments or national, standardized assessments. To the extent they participate in the latter, the published results are usually aggregated by institutional characteristics such as college type, location, etc. Thus, there is little publicly available information on cognitive outcomes at the institutional level.

One interesting exception is a study conducted at Tennessee Technological University (TTU) that evaluated gains in critical thinking as measured by the CAT (Harris et al., 2014). Data were obtained from students in seven courses across five departments. The courses were designed to spur critical thinking through a number of pedagogical strategies. The results were mixed with respect to which courses had impact on which component skills, but two courses, one in civil engineering and one in psychology, did yield quite impressive gains on the total CAT score.¹³ Similar findings were reported at Keene State College (K. Gagne, personal communication, May 25, 2017). These findings are consistent with the results cited above on the relationship between exposure to deep learning and cognitive gains.

Our informal survey of assessment practices suggests that many institutions, not only liberal arts colleges, are engaging seriously in infusing assessment into academic life. Examples are American University, Carnegie Mellon University and Pace University. It would be of interest to compare the strategies across different school types.

Discipline-based Assessments

Although much of the literature on cognitive outcomes focuses on so-called generic skills, undergraduates spend a good portion of their time taking courses in their major or concentration. In principle, then, it would be of interest to be able to quantify the contribution of liberal education to proficiency in the major. Unfortunately, there is very little relevant evidence. The Major Field Test battery offered by ETS includes 15 tests of the knowledge and skills in 13 domains. As is the case with other such standardized assessment programs, it is not possible to draw inferences from published data both because of the level of aggregation and

¹³ It is not evident from the description of the study what distinguished these two courses from the others: course content, instructor quality or student sample. Clearly, further investigation with larger samples would be helpful.

because there is neither control for differential selectivity into various majors nor for variations in student sampling strategies.

One of the many concerns regarding such assessments is that they rely on machine-scorable items that do not capture the higher level proficiencies that are expected of students majoring in the discipline. To that end, specific departments or entire colleges may require (or welcome) senior theses, capstone projects, senior portfolios, and the like. As these are local efforts they do not lend themselves to cross-institutional comparisons. It is possible that some departments informally monitor the year-to-year quality of such productions but it is unlikely that any systematic analysis is carried out—and even less likely that the results would be made public!

As pressure from accreditors intensifies, colleges may require more evidence from departments regarding their majors' learning outcomes. It is (remotely) possible that some colleges will form a consortium to focus on domain-based assessments in parallel to the consortia that now focus on the assessment of generic skills. The K–12 sector is one source of ideas for innovative approaches to assessment that may be amenable to adoption or adaptation by the higher education sector. Game-based assessments are one example (Glasslab, n.d.; Shute & Wang, 2016). For a comprehensive review of assessment strategies in a number of different disciplines see Braun (2016). For a review of international initiatives in higher education assessment see Zlatkin-Troitschanskaia et al. (2016).

Fortunately, there has been some potentially useful groundwork accomplished in this country. For example, in the mid-2000s the Teagle Foundation commissioned six white papers to address the relationship between disciplinary study and the goals of liberal education as articulated in the AAC&U's Essential Learning Outcomes. The six disciplines selected were: English studies and foreign language studies, history, religious studies, biochemistry and molecular biology, classics, and economics. Each white paper was produced by a team of scholars. Short versions of the six white papers were published in the journal *Liberal Education* (Spring, 2009). Each paper dealt with issues of curriculum and pedagogy, as well as with the relationship between the study of the major and the development of the skills and habits of mind typically associated with a liberal education.

Only three of the papers addressed the issue of assessment in any depth. The Religious Studies report offered some examples of departmental initiatives, the Biochemistry/Molecular Biology report asserted the need for more thoughtful assessment, and the History report addressed the tension between formative and summative assessment. In addition, the Economics report argued that the discipline needed to agree on a set of core learning objectives for the major, while strengthening its ties to the broader goals of liberal education. Another Teagle-related contribution is the edited volume that addresses the issues of assessment in literary studies with a variety of perspectives on the promises and pitfalls of being more systematic about assessing learning outcomes (Heiland & Rosenthal, 2011).

More recently, a volume edited by Arum, Roksa, & Cook (2016) describes the results of an ambitious initiative titled *Measuring College Learning* (MCL) that was conducted during the period 2013 to 2015. The aim of MCL was “to provide a platform for faculty to engage in national conversations about student learning and assessment” (p. 6). The focus was specifically on developing learning outcomes and assessments within specific disciplines—as opposed to assessments of generic skills (e.g., critical thinking) neither grounded nor dependent upon

disciplinary knowledge. To this end, the MCL project established national panels in six academic disciplines: biology, business, communication, economics, history, and sociology.¹⁴

Each panel contributed a report to the volume, describing their deliberations, conclusions, and recommendations. Each panel developed a set of *essential concepts* and *essential competencies* that constituted the desired learning outcomes for the discipline. The panel reports also described the current state of assessment in the discipline and offered rationales for the development of standardized assessments that would yield evidence of the accomplishments of majors with respect to those learning outcomes. One rationale is that they found current assessments wanting in a number of respects. A second is that such assessments would provide departments with useful information regarding the success of their program with respect to the full set of learning outcomes valued by faculty in the discipline. Of course, the use of the term *standardized* implies that these assessments could be administered to students in the relevant department at any institution and thereby enable comparisons among departments willing to share their results.

Although the MCL project revealed a general interest by faculty in improved assessment, it is not clear whether and how further progress will be achieved. To this point, the volume ends with commentaries by a number of higher education luminaries. The commentaries recognize the accomplishments of the panels but elaborate on a number of challenges that will confront any effort to build such assessments (Ewell, 2016). The challenges range from the substantive to the political. Left unsaid are the financial, logistical and pragmatic obstacles to mounting a national assessment program.

Long-term Outcomes

As noted earlier, many of the claims made regarding the benefits of a liberal education employ time frames extending well into adulthood and are concerned with outcomes that are not explicitly cognitive in nature; rather, they involve such outcomes as success in the workplace, civic participation, and various aspects of personal well-being. Thus, it is not surprising that the research in this area, meagre as it is, does not focus on cognitive outcomes per se. On the other hand, cognitive functioning likely contributes, directly or indirectly, to a host of adult outcomes and these interactions over the lifespan merit further study (Santrock, 2011).

An early contribution to understanding the long-term effects of higher education, generally speaking, was the work of Hyman, Wright, & Reed (1975). The authors analyzed data from multiple, then extant, social science surveys to compare high school graduates and college graduates over the lifespan on a broad range of measures of knowledge (academic and non-academic), as well as measures of receptivity to knowledge. The latter was inferred from knowledge acquired after leaving school. Not surprisingly, they found substantial advantages accrued to those who earned a college degree. To their credit, the authors acknowledged the many possible confounding factors that might have contributed, at least in part, to these findings and carried out those validity checks their data permitted.

¹⁴ Note there is some overlap with the six disciplines in the Teagle-funded initiative.

More recently, in 2007, in response to the Spellings Commission report and the launch of the VSA, both the presidents of Harvard and Princeton in separate addresses argued that the outcomes of a liberal education could only be determined many years after graduation and did not lend themselves to standardized measurement (Katz, 2008). Katz (2008), however, urged leading institutions of liberal education “to advocate assessment-based evaluation and, where necessary, the reform of liberal undergraduate education”.

To be sure, one of the arguments advanced in support of liberal education is that it better prepares individuals for success, broadly conceived, in later life (AAC&U, 2007). Success includes professional advancement, civic participation, and well-being. The argument is buttressed by anecdotal evidence, as well as testimonials from business leaders extolling the value that liberal arts graduates bring to the workplace. It is also affirmed by other academics and academic leaders. For example, Richard Miller, president of the Olin College of Engineering, argues that supplementing technical engineering courses and project work with a rich mix of liberal arts courses better prepares Olin graduates to function effectively in 21st century work environments (R. Miller, personal communication, March 17, 2017). A similar argument was articulated by Kenneth Osgood, a professor at the Colorado School of Mines (Osgood, 2017). Although these arguments are both plausible and persuasive, systematic, relevant empirical evidence is in short supply. Without substantial relevant information on initial student characteristics, however, there is little that can be said even with good outcome data.¹⁵

An investigation that did generate credible evidence was conducted as part of a larger study discussed above (Pascarella et al., 2005). For this component the authors employed longitudinal data collected by the Appalachian College Association. Individuals five, fifteen and twenty-five years from graduation were contacted and filled out an extensive questionnaire that covered a range of topics.

Graduates of LACs were compared to graduates of two other types of institutions (public regional universities and private masters-granting institutions) on each of the many long-term outcomes surveyed, using statistical controls to adjust for confounding variables. Overall, the results were mixed: LAC graduates exhibited advantages in some domains, disadvantages in others. For example, the authors note that “alumni of baccalaureate liberal arts colleges reported that their undergraduate experience had a significantly stronger positive impact on their learning and intellectual development, the development of leadership and self-efficacy skills, personal and spiritual development, and the development of responsible citizenship than did similar graduates of public universities” (p. 73). On the other hand, “Compared with similar graduates of public universities, alumni of baccalaureate liberal arts colleges were significantly less likely to be employed full time and to be employed in a for-profit business or organization than public regional university graduates. Similarly, liberal arts college graduates had a small but statistically significant disadvantage in annual salary relative to public university alumni” (p. 74).

¹⁵ For example, Detweiler (2016) identified graduates of liberal arts colleges and graduates of other institutions who had graduated from ten to forty years earlier. He compared these two groups with respect to a number of outcomes and found that, on average, the former exceeded the latter. However, with no controls for prior achievement and family background, inferences regarding the advantages conferred by a liberal education are purely speculative.

The measured impacts (both positive and negative) were typically modest. For some outcomes related to intellectual and personal development, some apparent impacts may actually be the result of unmeasured pre-college influences. Interesting and unique as they are, these results are drawn from a specific region of the country and it would be purely speculative (if not foolhardy) to infer from them general propositions about the higher education system as a whole.

Conclusions

“The elaborate rhetoric and anecdotal support, long used to advance liberal arts education as the premier type of education with value for all, is no longer sufficient. The institutional practices and conditions that lead to outcomes associated with a liberally educated student remain an empirical black box.” (Seifert et al., 2008, p. 2)

In the decade since that charge appeared, we have made only modest progress in lifting the lid of the black box. Despite substantially greater amounts of assessment data, we have made even less progress on documenting the (hoped for) advantages of attending a liberal arts college or, more broadly, experiencing some form of liberal education, either at the point of graduation or in the world beyond. This is as true for cognitive gains as it is for other valued outcomes. Ironically, the rationale for a liberal education is now arguably stronger now than ever before but, at the same time, is under attack by those with a more instrumental, and more short-term, view of the goals of college.

As explained above, when appropriate statistical controls are applied to longitudinal data, the record is decidedly mixed regarding the advantage, with respect to a wide range of outcomes, of enrolling in a liberal arts college in comparison to other types of institutions. Admittedly, the samples of colleges in these studies are small and not representative in the usual statistical sense. Thus, these results should be considered suggestive rather than definitive. A common, and plausible, argument posits that the benefits of an intensive liberal education become manifest only after graduation as individuals make their way into and through the world of work. Although there are many supporting anecdotes, this conjecture is even harder to investigate systematically and only a single well-designed study has come to light (Pascarella et al., 2005).

On a brighter note, through the efforts of many different organizations, as well as the pressures exerted on college administrators by accreditors and by legislatures, there has been progress in helping faculty to see the value in being more systematic and intentional with respect to both assessment design and the rating of student work. Although the external pressures are more keenly felt by public institutions, private universities and liberal arts colleges have not been immune to market pressures and many have made progress as well.

For some institutions this entails purely local efforts, while for others it comprises a hybrid strategy of both local assessments and externally developed assessments. Hundreds of colleges have participated in the VALUE initiative that involves the evaluation of student work (responding to locally developed assignments) through communally developed scoring rubrics.

Many others have subscribed to the Degree Qualification Profile that attempts to create a consensus framework on the outcomes of higher education and strategies for assessing those outcomes (Ewell, 2013; Marshall, 2015). Further, large numbers of colleges participate in various standardized assessment programs, often in conjunction with other assessment activities. Notably, although there appears to be increasing investment in assessment by undergraduate institutions in all sectors, the published literature and colleges' public reports contains little evidence of their impact on student learning (Banta & Blach, 2011).

A similar difficulty pertains to the focal topic of this paper. Whether they employ locally developed assessments, externally developed standardized assessments, or some combination, institutional-level results are kept private and so cannot be used to generate evidence to inform a comparative analysis of the cognitive outcomes of liberal education. In effect, each college is a "hermetically sealed system" when it comes to assessment results. Aggregate data that are published have limited utility, even if they are reported by sector, because of the likely differences across participating institutions in the representativeness of the student samples and the lack of information on pre-college characteristics and achievements. More data sharing and a more collaborative approach across institutions and institution types would be desirable.

In sum, the findings we have been able to glean regarding the cognitive outcomes of liberal education suggest that further attempts to compare liberal arts colleges directly with institutions in other sectors are not likely to bear much fruit. There is no reason to expect that future studies will yield findings very different than those already cited. One likely factor is the heterogeneity among institutions within the different sectors, magnified by the variation among students within institutions. On the other hand, there are consistent findings across sectors of the relationships between students' exposure to so-called liberal arts experiences, including deep approaches to learning, and their gains on liberal arts outcomes. This suggests that it may be more productive to investigate these relationships both more thoroughly and more broadly. For example, the scales quantifying liberal arts experiences could be refined through both qualitative research and additional psychometric analyses. Further, the range of outcomes could be expanded to include more discipline-based skills. This approach has the benefit of not only being relevant to most institutions of higher education, but also more likely to generate information that is useful to improvement initiatives in those institutions.

Even were this general strategy adopted, there would still be many difficult choices to make among specific targets. One would be quantifying the impact of a strong liberal arts curriculum on academic performance in STEM majors and, perhaps more importantly, in success beyond college. Another would be adding a focus on more difficult to measure, but highly valued, skills that do interact with cognitive ability, such as collaboration. A rather different one would be enhancing the utility of assessment outcomes for program improvement.

As with the work we discussed earlier, longitudinal designs would be ideal. But studies incorporating longitudinal designs are not only difficult and expensive to mount, but also are beset by logistical problems like sample attrition that reduce their evidential value.¹⁶ Alternative designs, including modified longitudinal designs, may be more feasible and possess evidential

¹⁶ A comprehensive review of the problems with randomized controlled trials, including some longitudinal designs, can be found in Ginsberg & Smith (2016).

value not substantially lower than that of full longitudinal designs. Whatever the choice, it will be very important to consider both institutional incentives and more innovative approaches to student recruitment, engagement and retention in order to strengthen the inferential value of the studies (Schneider, 2016).

Equally important is careful consideration of the assessments employed to measure the selected outcomes, making sure that they more fully represent the target constructs. Here it may be feasible to modify or adapt assessments that have already been fielded (e.g., as part of the VALUE initiative). Similarly, the scoring rubrics developed through the VALUE initiative could presumably be modified to support assessment in the disciplines. This may also be possible with the TTU scoring rubrics, as well as those employed for the CLA.

A theme that recurs in this essay is the tension between what psychometricians refer to as the “validity” and the “reliability” of an assessment. The validity of an assessment is roughly the degree to which the results of the assessment align with what the assessment is intended to measure. Reliability refers to the replicability of the results of an assessment on repeated application. (For example, will two trained graders of an essay arrive at similar judgments?) Proponents of tests comprising multiple choice and/or short answer questions correctly note that, for a fixed amount of testing time, they are much more reliable than tests requiring long, open-ended responses (e.g., essays). They also cite studies where both item types were administered. Typically, total scores on the two item types demonstrate high correlations. Consequently, they argue, there is no need to include the less reliable items that take up precious testing time and add considerably to the cost of the assessment. What this argument misses, however, is that the more reliable test may suffer from low validity. That problem may only be remedied by including well-designed, extended assessments that tap into facets of the construct that are not accessed by multiple choice items. These considerations should be kept in mind when considering the different assessments used to measure higher education outcomes.

In view of the range of possibilities and the complexities in designing such studies, organizations and researchers seeking to make further progress in understanding the impact of liberal education practices on cognitive growth should proceed deliberately: The literature on meta-analysis is replete with accounts of how, out of hundreds of published studies on a particular topic, only a handful meet the methodological requirements for inclusion (Hedges & Olkin, 1985). Consequently, analysts should first formulate a clear set of study objectives and generate alternative study designs. These designs ought then be evaluated with respect to expected findings, feasibility and cost. Such careful preparatory work is essential if the results of the study will prove to be as useful as hoped—and as needed.

BIBLIOGRAPHY

- Arum, R., & Roksa, J. (2011). *Academically adrift: Limited learning on college campuses*. Chicago: University of Chicago Press.
- Arum, R., Roksa, J., & Cho, E. (2011). *Improving undergraduate learning: Findings and policy recommendations from the SSRC-CLA longitudinal project*. New York: Social Science Research Council.
- Arum, R., & Roksa, J. (2014). *Aspiring adults adrift: Tentative transitions of college graduates*. Chicago: University of Chicago Press.
- Arum, R., Roksa, J., & Cook, A. (Eds.). (2016). *Improving quality in American higher education: Learning outcomes and assessments for the 21st century*. San Francisco, CA: Jossey-Bass.
- Association of American Colleges and Universities. (2002). *Greater expectations: A new vision for learning as a nation goes to college*. Washington, DC. Retrieved from <https://www.aacu.org/sites/default/files/files/publications/GreaterExpectations.pdf>
- Association of American Colleges and Universities. (2005). *Liberal education outcomes: A preliminary report on student achievement in college*. Washington, DC. Retrieved from https://www.aacu.org/sites/default/files/files/LEAP/LEAP_Report_2005.pdf
- Association of American Colleges and Universities. (2007). *College learning for the new global century*. Washington, DC. Retrieved from https://www.aacu.org/sites/default/files/files/LEAP/GlobalCentury_final.pdf
- Association of American Colleges and Universities. (2017). *On solid ground*. Washington, DC. Retrieved from <https://www.aacu.org/publications-research/publications/solid-ground>
- Astin, A. W. (2011, February 14). In 'Academically Adrift,' data don't back up sweeping claim. *Chronicle of Higher Education*.
- Banta, T. W., & Blaich, C. (2011). Closing the assessment loop. *Change: The Magazine of Higher Learning*, 43(1), 22–27.
- Banta, T. W., & Pike, G. R. (2007). Revisiting the blind alley of value added. *Assessment Update*, 19(1).
- Belkin, D. (2017, June 7). Many colleges fail in teaching how to think: A test finds students often gain little ability to assess evidence, make a cohesive argument. *Wall Street Journal*.
- Biglan, A. (1973). The characteristics of subject matter in different scientific areas. *Journal of Applied Psychology*, 57, 195–203.
- Blaich, C., Bost, A., Chan, E., & Lynch, R. (2004). *Defining liberal arts education*. Unpublished manuscript. Crawfordsville, IN: Center of Inquiry in the Liberal Arts at Wabash College.
- Blaich, C., & Wise, K. (2011). *The Wabash National Study: The impact of teaching practices and institutional conditions on student growth*.
- Braun, H. I. (2013). Value-added measurement: Report from the Expert Group Meeting. *AHELO Feasibility Study Report* (Vol. 3, Chap. 10, pp. 9–26). Paris: OECD.
- Braun, H. I. (2016). *Meeting the challenges to measurement in an era of accountability*. New York: Routledge.
- Braun, H. I., & Mislavy, R. J. (2005). Intuitive test theory. *Phi Delta Kappan*, 87(5).
- Brighthouse, H. (2017). *Memo on Liberal Arts Education*. Occasional paper. New York: The Andrew W. Mellon Foundation.
- Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment*, 48(3), 306–307.

- Council for Aid to Education (n.d.) *CLA+: Operation details*. Retrieved June 19, 2017, from <https://cae.org/flagship-assessments-cla-cwra/flagship-assessments-cla-cwra/operational-details>
- Delbanco, A. (2012). *College: What it was, is, and should be*. Princeton, NJ: Princeton University Press.
- Detweiler (2016, January). *Presentation at the annual meeting of the American Association of Colleges & Universities*.
- Dwyer, C. A., Millett, C. M., & Payne, D. G. (2006). *A culture of evidence*. Princeton, NJ: Educational Testing Service.
- Ewell, P. T. (2012). A world of assessment: OECD's AHELO initiative. *Change: The Magazine of Higher Learning*, 44(5), 35–42.
- Ewell, P. T. (2013, January). *The Lumina Degree Qualifications Profile (DQP): Implications for assessment*. Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment.
- Ewell, P. T. (2016). A promising start and some way to go: Some reflections on the Measuring College Learning project. In Arum, R., Roksa, J., & Cook, A. (Eds.). *Improving quality in American higher education: Learning outcomes and assessments for the 21st century*. San Francisco, CA: Jossey-Bass.
- Freedman, J. O. (2003). *Liberal education and the public interest*. Iowa City, IA: University of Iowa Press.
- Ginsberg, A., & Smith, M. S. (2016). *Do randomized controlled trials meet the “gold standard”? A study of the usefulness of RCTs in the What Works Clearinghouse*. Washington, DC: American Enterprise Institute.
- Glasslab (n.d.). *Psychometric considerations in game-based assessment*. New York: Institute of Play.
- Hammer, M. R., & Bennett, M. J. (2001). *The intercultural development inventory (IDI) manual*. Portland, OR: Intercultural Communication Institute.
- Harris, K., Stein, B., Haynes, A., Lisic, E., & Leming, K. (2014). *Identifying courses that improve students' critical thinking skills using the CAT instrument: A case study*. Proceedings of the 10th Annual International Joint Conferences on Computer, Information, System Sciences, & Engineering. doi: 10.13140/RG.2.1.1751.3442
- Hart Research Associates (2013). *It takes more than a major: Employer priorities for college learning and student success*. Washington DC: Association of American Colleges & Universities.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- Heiland, D., & Rosenthal, L. J. (Eds.). (2011). *Literary study, measurement, and the sublime: Disciplinary assessment*. New York: The Teagle Foundation.
- Hersh, R. H., & Keeling, R. P. (2013, February). Changing institutional culture to promote assessment of higher learning. *National Institute for Learning Outcomes Assessment*.
- Hyman, H. H., Wright, C. R., & Reed, J. S. (1975). *The enduring effects of education*. Chicago: University of Chicago Press.
- Katz, S. N. (2005, April 1). Liberal education on the ropes. *Chronicle of Higher Education*.
- Katz, S. N. (2008, May 23). Taking the true measure of a liberal education. *Chronicle of Higher Education*.
- King, P. M., Kendall Brown, M., Lindsay, N. K., & VanHecke, J. R. (2007). Liberal arts student learning outcomes: An integrated approach. *About Campus*, 12(4), 2–9.

<https://doi.org/10.1002/abc.222>

- Klein, S., Benjamin, R., Shavelson, R., & Bolus, R. (2007). The collegiate learning assessment: Facts and fantasies. *Evaluation Review*, 31(5), 415–439.
- Klein, S., Liu, O. L., & Sconing, J. (2009). *Test validity study (TVS) report*. Retrieved from http://cae.org/images/uploads/pdf/13_Test_Validity_Study_Report.pdf
- Kuh, G. D., Jankowski, N., Ikenberry, S. O., & Kinzie, J. (2014, January). Knowing what students know and can do: The current state of student learning outcomes assessment in U.S. colleges and universities. *National Institute for Learning Outcomes Assessment*.
- Laird, T. F. N., Shoup, R., Kuh, G. D., & Schwarz, M. J. (2008). The effects of discipline on deep approaches to student learning and college outcomes. *Research in Higher Education*, 49(6), 469–494.
- Liu, O. L. (2017). Ten years after the Spellings Commission: From accountability to internal improvement. *Educational Measurement: Issues and Practices*.
- Liu, O. L., Frankel, L., & Roohr, K. C. (2014). *Assessing critical thinking in higher education: Current state and directions for next-generation assessments*. ETS Research Report Series, 1, 1–23.
- Martin, E. D. (1926). *The meaning of a liberal education*. New York: W. W. Norton.
- Neumann, A. (2014). Staking a claim on learning: What we should know about learning in higher education and why. *Review of Higher Education*, 37(2), 249–267.
- Osgood, K. (2017, May 26). Engineers need the liberal arts, too. *Chronicle of Higher Education*.
- Paris, D. C. (2011). *Catalyst for change: The CIC/CLA consortium*. Washington DC: Council of Independent Colleges.
- Pascarella, E. T., Seifert, T. A., & Blaich, C. (2010). How effective are the NSSE benchmarks in predicting important educational outcomes? *Change: The Magazine of Higher Learning*, 42(1), 16–22. <https://doi.org/10.1080/00091380903449060>
- Pascarella, E. T., & Blaich, C. (2013). Lessons from the Wabash National Study of Liberal Arts Education. *Change: The Magazine of Higher Learning*, 45(2), 6–15. <https://doi.org/10.1080/00091383.2013.764257>
- Pascarella, E. T., Wang, J. S., Trolan, T. L., & Blaich, C. (2013). How the instructional and learning environments of liberal arts colleges enhance cognitive development. *Higher Education*, 66(5), 569–583. <https://doi.org/10.1007/s10734-013-9622-z>
- Pascarella, E. T., Wolniak, G., Seifert, T. A., Cruce, T., & Blaich, C. (2005). Liberal arts colleges and liberal arts education: New evidence on impacts. *ASHE Higher Education Report*, 31(3).
- Pellegrino, J. W., & Hilton, M. L. (Eds.). (2012). *Education for life and work: Developing transferable knowledge and skills in the 21st century*. Washington, DC: National Academies Press.
- Rest, J., Narvaez, D., Thoma, S. J., & Bebeau, M. J. (1999). DIT2: Devising and testing a new instrument of moral judgment. *Journal of Educational Psychology*, 91(4), 644–659.
- Roohr, K. C., Liu, H., & Liu, O. L. (2016). Investigating student learning gains in college: A longitudinal study. *Studies in Higher Education*. <https://doi.org/10.1080/03075079.2016.1143925>
- Roth, M. S. (2014). *Beyond the university: Why liberal education matters*. New Haven, CT: Yale University Press.
- Ryff, C. D., & Keyes, C. L. M. (1995). The structure of psychological well-being revisited. *Journal of Personality and Social Psychology*, 69, 719–727.
- Santrock, J. W. (2011). *Life-span development* (13th ed.). New York: McGraw-Hill.

- Schneider, C. G. (2016). How MCL can make a lasting difference. In Arum, R., Roksa, J., & Cook, A. (Eds.), *Improving quality in American higher education: Learning outcomes and assessments for the 21st century*. San Francisco, CA: Jossey-Bass.
- Seifert, T. A., Goodman, K. M., Lindsay, N., Jorgensen, J. D., Wolniak, G. C., Pascarella, E. T., & Blaich, C. (2008). The effects of liberal arts experiences on liberal arts outcomes. *Research in Higher Education*, 49(2), 107–125. <https://doi.org/10.1007/s11162-007-9070-7>
- Shavelson, R. J. (2010). *Measuring college learning responsibly*. Stanford, CA: Stanford University Press.
- Shavelson, R. J., Domingue, B. W., Mariño, J. P., Molina Mantilla, A., Morales Forero, A., & Wiley, E. E. (2016). On the practices and challenges of measuring higher education value added: The case of Colombia. *Assessment & Evaluation in Higher Education*, 41(5), 695–720. <https://doi.org/10.1080/02602938.2016.1168772>
- Shute, V. J. & Wang, L. (2016). Assessing and supporting hard-to-measure constructs. In A. A. Rupp, & J. P. Leighton (Eds.), *The handbook of cognition and assessment: Frameworks, methodologies, and application*, (pp. 535–562). Hoboken, NJ: John Wiley & Sons, Inc.
- Stein, B., & Haynes, A. (2011). Engaging faculty in the assessment and improvement of students' critical thinking, using the critical thinking assessment test. *Change: The Magazine of Higher Learning*, 43(2), 44–49.
- Thomas, N. (2002). Liberal education for a changing world. *Liberal Education*, 88(4), 28–33.
- Tyree, T. M. (1998). Designing an instrument to measure the socially responsible leadership using the social change model of leadership development. *Dissertation Abstracts International*, 59(6), 1945.
- U.S. Department of Education. (2006). *A test of leadership: Charting the future of U.S. higher education*.
- Wood, P. K., Kitchener, K. S., & Jensen, L. (2002). Considerations in the design, evaluation of a paper-and-pencil measure of reflective thinking. In B. Hofer & P. Pintrich (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Zahner, D., & Steedle, J. T. (2015). *Comparing longitudinal and cross-sectional school effect estimates in postsecondary education*.
- Zlatkin-Troitschanskaia, O., Pant, H. A., & Coates, H. (2016). Assessing student learning outcomes in higher education: Challenges and international perspectives. *Assessment & Evaluation in Higher Education*, 41(5), 655–661. <https://doi.org/10.1080/02602938.2016.1169501>
- Zlatkin-Troitschanskaia, O., Shavelson, R. J., & Kuhn, C. (2015). The international state of research on measurement of competency in higher education. *Studies in Higher Education* 40(3): 393–411.

Appendix I. Value-added Rubric (Sample)

Critical Thinking VALUE Rubric

Definition

Critical thinking is a habit of mind characterized by the comprehensive exploration of issues, ideas, artifacts, and events before accepting or formulating an opinion or conclusion.

	Capstone	Milestones		Benchmark
	4	3	2	1
Explanation of Issues	Issue/problem to be considered critically is stated clearly and described comprehensively, delivering all relevant information necessary for full understanding.	Issue/problem to be considered critically is stated, described, and clarified so that understanding is not seriously impeded by omissions.	Issue/problem to be considered critically is stated but description leaves some terms undefined, ambiguities unexplored, boundaries undetermined, and/or backgrounds unknown.	Issue/problem to be considered critically is stated without clarification or description.
Evidence <i>Selecting and using information to investigate a point of view or conclusion</i>	Information is taken from source(s) with enough interpretation/evaluation to develop a comprehensive analysis or synthesis. Viewpoints of experts are questioned thoroughly.	Information is taken from source(s) with enough interpretation/evaluation to develop a coherent analysis or synthesis. Viewpoints of experts are subject to questioning.	Information is taken from source(s) with some interpretation/evaluation, but not enough to develop a coherent analysis or synthesis. Viewpoints of experts are taken as mostly fact, with little questioning.	Information is taken from source(s) without any interpretation/evaluation. Viewpoints of experts are taken as fact, without question.
Influence of context and assumptions	Thoroughly (systematically and methodically) analyzes own and others' assumptions and carefully evaluates the relevance of contexts when presenting a position.	Identifies own and others' assumptions and several relevant contexts when presenting a position.	Questions some assumptions. Identifies several relevant contexts when presenting a position. May be more aware of others' assumptions than one's own (or vice versa).	Shows an emerging awareness of present assumptions (sometimes labels assertions as assumptions). Begins to identify some contexts when presenting a position.
Student's position (perspective, thesis/hypothesis)	Specific position (perspective, thesis/hypothesis) is imaginative, taking into account the complexities of an issue. Limits of position (perspective, thesis/hypothesis) are acknowledged. Others' points of view are synthesized within position (perspective, thesis/hypothesis).	Specific position (perspective, thesis/hypothesis) takes into account the complexities of an issue. Others' points of view are acknowledged within position (perspective, thesis/hypothesis).	Specific position (perspective, thesis/hypothesis) acknowledges different sides of an issue.	Specific position (perspective, thesis/hypothesis) is stated, but is simplistic and obvious.
Conclusions and related outcomes (implications and consequences)	Conclusions and related outcomes (consequences and implications) are logical and reflect student's informed evaluation and ability to place evidence and perspectives discussed in priority order.	Conclusion is logically tied to a range of information, including opposing viewpoints, related outcomes (consequences and implications) are identified clearly.	Conclusion is logically tied to information (because information is chosen to fit the desired conclusion); some related outcomes (consequences and implications) are identified clearly.	Conclusion is inconsistently tied to some of the information discussed; related outcomes (consequences and implications) are oversimplified.

Appendix 2. Longitudinal Liberal Arts Outcomes Studies

Study	Assessment	Sample	Results
Arum & Roksa (2011)*	CLA	24 colleges (2,322 students)	Small gains
Arum & Roksa (2014)	CLA	24 colleges (1,600 students)	Larger gains in last two years
Blaich & Wise (2011)*	CAAP	49 colleges (approx. 17,000 students)	Small gains in first two years, larger gains over 4 years
Pascarella et al. (2005)*	CAAP	5 Liberal Arts Colleges	LAC had no significant impact
Pascarella et al. (2013)*	CAAP	17 colleges, of which 11 were Liberal Arts Colleges (approx. 2,212 students)	Liberal Arts attitudes related to Liberal Arts experience, which in turn was related to student learning outcomes
Seifert et al. (2008)*	Reasoning and Current Issues Intercultural Developmental Inventory Scales of Psychological Well-being Defining Issues Test-2 Need for Cognition Socially Responsible Leadership Scale	4 Liberal Arts Colleges (approx. 700 students)	Liberal Arts experience related to outcomes
Klein (2009)	CLA		
Roohr et al. (2016)	ETS	1 college (168 students)	Moderate gains in critical thinking after 4-years of college

Appendix 3. College Case Studies

Liberal Arts Colleges

Truman State University

Truman State University has been fostering assessment institutional assessment since 1973. The goal of its assessment program is to measure student progress toward educational goals, to determine academic progress, to improve teaching and learning, and to evaluate institutional effectiveness. The assessment practices are primarily designed for senior students. Truman assessment practices rely on external instruments such as the CIRP, NSSE, and the CLA, as well as internally-developed instruments. The internal instruments include portfolio projects, capstone experiences, student interviews, and a graduating student questionnaire (a complete list of Truman State University's assessment components can be found at: assessment.truman.edu/components/). Additionally, seniors are required to take a test in their major field; however, the results of such tests are not used to make claims about individual students' achievement or knowledge, but rather about students within a major as a whole. Results from these assessments are disseminated annually to campus via the Assessment Almanac. The goal of the Almanac is to improve access to the assessment results, and further the use of these results.

Carleton College

Assessment at Carleton College is focused on measuring six institutional learning outcomes:

- Carleton College graduates should be able to demonstrate that they have acquired knowledge necessary for the continuing study of the worlds, peoples, arts, environments, literatures, sciences, and institutions
- Carleton College graduates should be able to demonstrate substantial knowledge of a field of study and the modes of inquiry or methodologies pertinent to that field
- Carleton College graduates should be able to analyze evidence
- Carleton College graduates should be able to formulate and solve problems
- Carleton College graduates should be able to communicate and argue effectively
- In their chosen field of study Carleton College graduates should be able to conduct disciplinary and/or interdisciplinary research and/or undertake independent work which may not include artistic creation or production

Each of these six learning outcomes is accompanied by a rubric that identifies dimensions of the outcome and describes various levels of student proficiency. Additionally, there are sources of evidence for each learning outcome to be used to measure performance on each learning outcome. These sources of evidence include both direct and indirect, external and internal measures. Examples of external instruments include the CIRP, NSSE, and the CLA. Internal instruments include portfolio assessments, focus groups, and capstone projects. Assessment activities at Carleton College occur across all departments and class years.

Macalester College

Similar to Carleton College, assessment at Macalester focuses on measuring six institutional learning goals:

- Intellectual depth and breadth
- Think critically and analyze effectively
- Communicate effectively
- Demonstrate intercultural knowledge and competence
- Make informed choices and accept responsibility
- Engage with communities

Assessment practices at Macalester fall into three domains: institutional assessment, general education assessment, and academic departments assessment. Institutional assessment includes measures to determine the impact of the entire Macalester experience. Instruments used for institutional assessment include the CLA, the Research Practices Survey, and the Macalester Assessment Instrument (internally developed instrument to measure internationalism, multiculturalism, and service to society). General Education assessment ensures that students are meeting the learning goals for the four general education areas: internationalism, multiculturalism, quantitative thinking, and writing. Instruments used for General Education assessment at Macalester include student portfolios and surveys. Finally, Academic Departments assessment at Macalester informs the general effectiveness of each departments' curriculum and evaluates the success of the department in meeting the learning goals. Each department is responsible for developing learning goals and implementing a plan to ensure that the goals are being achieved. Tools to measure department learning goals include both national and local instruments, such as ETS Major Field Tests, the Science Literacy Concept Inventory, capstone projects, surveys, portfolios, and exit interviews. While primarily geared towards first year and senior students, assessment activities occur in all years.

Universities

Pace University

According to a 2006 report, students at Pace University will participate in at least 12 institution-wide assessment activities before they graduate, including CIRP, NSSE and the CLA, as well as ETS Major Field Tests. These national instruments are directed primarily toward first year and senior students. Pace University fosters a strong culture of assessment, with a particular focus on local and/or indirect measures. For example, there has been a focus in recent years on using e-portfolios as assessment tools for all disciplines. Additionally, several colleges within the university have taken a diverse approach to assessment. The Seldenberg School of Computer Science and Information Systems approaches assessment using a triangular method, relying on quantitative (e.g., Major Field Tests results), quasi-quantitative (e.g., rubric-based data), and qualitative (e.g., interviews) assessment data. Meanwhile, the Lienhard School of Nursing takes a holistic approach to assessment, which involved introducing the Appreciative Inquiry framework, implementing course surveys framed in the Appreciative Inquiry model, and then implementing changes based on the results of the survey. These changes included further

promotion of professionalism, developing a skill set of providing constructive feedback, and promoting accountability.

Bowling Green State University

Bowling Green State University was named one of five *Excellence in Assessment* designees by NILOA in August of 2017, along with James Madison University, Middlesex Community College, Rio Salado College, and Southern Connecticut State University. Assessment practices at Bowling Green State University are primarily focused on gathering data to support the general education learning outcomes, called the Bowling Green Perspective (BGP) learning outcomes. There are 34 learning outcomes that fall within seven domains: English composition and oral communication, quantitative literacy, humanities and the arts, social and behavioral sciences, natural sciences, cultural diversity in the United States, and international perspective. Assessment data is collected from each BGP course each semester in accordance with the approved assessment plan.

In addition to the university-wide BGP learning outcomes, there are also specific learning outcomes for each major. While each department is responsible for the assessment of their designated learning outcomes, these practices are supported by the AAC&U VALUE rubrics for critical thinking, engagement, information literacy, inquiry, oral communication, and written communication. Faculty are encouraged to use these rubrics as criteria for their various assignments.

Finally, assessment at Bowling Green State University also involves the use of national, external surveys such as the CLA+ and NSSE. Assessment activities at Bowling Green State are primarily geared toward first year and senior students.

Regional Colleges

Gustavus Adolphus

Assessment at Gustavus Adolphus College is a continuous activity that is used for feedback and improvement. At the institution level, there are seven institutional student learning outcomes related to the following domains: cognitive practice, intellectual capacities, integration of learning, ethical reflection, intercultural understanding, leadership, and wellbeing. Assessment practices are guided by these seven learning outcomes. Additionally, every academic department is encouraged to have 4–5 student learning outcomes and a plan for assessing these outcomes on a 4–5 year cycle. Departments and faculty are encouraged to incorporate the AAC&U VALUE rubrics into their assessment practices. Assessment activities at Gustavus Adolphus College occur across all departments, courses, and years.

Southern Connecticut State University

Southern Connecticut State University was also named one of five *Excellence in Assessment* designees by NILOA in August of 2017. The assessment system at Southern Connecticut is

broad, and encompasses performance-based tests, rubrics, and surveys that are both internal and external. External assessments include the CLA+, NSSE, and the Multi-State Collaborative to advance Quality Student Learning (MSC). (A list of performance-based assessments used by Southern Connecticut can be found here: <http://www.southernct.edu/assessment-and-planning/Assessment%20Activities.pdf>.) Assessments are both formative and summative and are used at all ages and years. Assessment activities are designed to measure the eight liberal education program learning outcomes:

- Analyze and solve complex problems
- Cogently and articulately express ideas in speaking and writing
- Demonstrate academic habits of mind
- Think independently and creatively from an informed understanding
- Demonstrate ability to synthesize learning throughout the liberal education program curriculum, through application to a culminating experience or project
- Apply the standards and ethics required to enter into the professional world
- Articulate/evaluate multiple perspectives on an issue, acknowledging the potential for complexity and ambiguity
- Engage in the integration of informational resources and technology

Every year, an assessment plan is developed and implemented to ensure there are multiple measures of each of these eight learning outcomes.