

## CHAPTER 2.

# WHAT IS OLD IS NEW AGAIN: A HISTORY OF WRITING ASSESSMENT, SYSTEMIC MANAGEMENT, AND THE NEOLIBERAL UNIVERSITY

At the 2016 Assessment Institute—a conference targeted more toward assessment professionals than writing program administrators—there is a lot of buzz about rubrics. To my surprise, rubrics are being hailed as the wave of the future. An attendee, half-jokingly, calls them “the next high impact practice.” In essence, they appear to be the new, shiny thing—a surprise to me, who first learned about rubrics as a senior in AP English in 1994.

While this project engages with the “life” of one specific rubric—the American Association of Colleges & Universities (AAC&U)’s Valid Assessment of Learning in Undergraduate Education (VALUE) rubric for Written Communication—I begin by asking: what is this particular moment in the history of higher education assessment? Why are rubrics popular now? Where have we come from and how did we get here? Most of this book is focused on capturing moments at specific universities, as seen in the introduction. In this chapter, I aim to establish the context for those stories by exploring how rubrics exist within the larger *institutional* conversation.

Here, I mean institution in terms of higher education at large. Dorothy Smith (2005) explained that institutions are local settings but also larger entities that influence local practice. Local institutions “participate in relations that standardize their operations and generalize them across particular local instances” (p. 206). This complex set of relations and organizational structures produces the “institution” as a larger concept that organizes behaviors across local contexts. In other words, while each of our universities is its own institution, they all—to at least some extent—participate in the power structure of the *institution* of higher education. Higher education as an institution organizes and rules our local practice, and this is becoming increasingly true as we look at national assessment.

National trends in higher education (such as the need to compare universities, assign transfer credit, and design common outcomes) impact our everyday practice in our core curriculum and general education committees, assessment

groups, and even in the running of our writing programs. In order to engage in critical and meaningful practice, writing program administrators (WPAs) as well as others involved in assessment need to be aware of the trickle-down effect of institutional rhetoric, forwarded by large national higher education groups—what Linda Alder-Kassner (2017) called the Educational Intelligence Complex (EIC). These groups have their own agendas that may or may not match our local practice, and they often operate in the background; traces of their influence are lost in a sort of top to bottom game of telephone. For example, rubrics created for national assessment may find their way to the classroom, adopted by faculty who got them from other faculty members or from administrators who at some point—a moment lost and forgotten—got them from some workshop they went to led by a representative from an organization such as the AAC&U. This is often how the EIC operates: a wizard behind the curtain, pulling strings we don't even know that we are attached to.

This chapter seeks to trace some of what has led to our current moment within institutional history where the scales are tipping in favor of rubrics over testing. This moment is complex. In some ways, it is attractive. Its promise of turning toward student outcomes and away from the technocratic Spellings Commission sounds encouraging. Yet as Chris Gallagher (2016) found, the moment can quickly sour as tools composition scholars have traditionally supported (such as e-portfolios) are being co-opted by a “neoliberal agenda whose endgame...is competency-based education” (p. 22). The focus can quickly turn from learning in individual courses to certifying competencies via rubrics. The rhetoric of austerity and neoliberalism is also a part of our current moment, and it makes this a fraught time for writing scholars and administrators. In this moment, it is particularly important that we pay attention to institutional power relations.

Historically, writing scales and rubrics have functioned within these systems of power as a tool for efficiency and social control. Previous scholars have acknowledged the racist nature of large-scale writing assessments (Elliott, 2005; Inoue, 2015). That is not to say that rubrics are always used in this manner or that they cannot do some good in the world. However, more must be done to link our current assessment system to the racist past of American education and to link our classroom practice to large-scale, institutional initiatives. Tracing the origins of any genre can demonstrate cultural shifts and reveal the ideological underpinnings of a form (Devitt, 2004, p. 92). The precursor to the modern rubric, the writing scale emerged in the early 1900s, a time when a whole new repertoire of managerial genres emerged to standardize and systematize daily work (Devitt, 2004). As Smith (2005) explained, the growth of industry and corporations led to a disconnect between workers and supervisors: “instead of

being ruled directly by individuals whom we've known ... we are ruled by people who are at work in corporations, government, professional settings and organizations, universities, public schools, hospitals and clinics, and so on and so on" (p. 18). Thus, texts became key to enforcing institutional control as individuals became removed from daily interactions with individual supervisors.

So, too, in the age of austerity and accountability endemic to the 2010s, teachers have seen an increase in reports to administrators and forms have emerged to measure and compare students across classrooms. No matter how much we work with our students and our fellow teachers to develop good rubrics for classroom use, we cannot fully separate them from an institutional system that is tied to problematic ideologies. Throughout this book, I focus on the rubric as a text that is such an "instrument of ruling" (Rankin, 2017b, p. 2). This chapter grounds that focus in historical and current political ideologies. I then introduce the AAC&U's two signature movements as relative to this project: Liberal Education and America's Promise (LEAP) and Valid Assessment of Learning in Undergraduate Education (VALUE). By placing these movements within the larger history of education and assessment, we see how they co-construct the current moment that calls for a shift toward rubrics and away from testing while still operating within the confines of neoliberalism.

## UNBALANCED: SCALES OF RACIAL EXCLUSION

The early 1900s represented a shift toward efficiency and standardization across sectors: business, industry, and education. Scholars, such as Joseph Mayer Rice brought back European (particularly German) methods of incorporating science into the study of education (Elliot, 2005). It is no surprise that early talk of evaluating writing on a scale appears in a book by Rice (1914) entitled *Scientific Management in Education*. Rice's article (originally published in 1903) explained his method of scoring composition themes by placing them in one of five piles: "Excellent, Good, Fair, Poor, and Failure" (Hudelson, 1923, p. 164). The idea of using such a system came from a desire to standardize, and thus make more efficient, the evaluation of writing. In fact, Rice (1914) bragged that by using his method he was able to score "60-70 composition themes per hour" (p. 182). These early writing scales were the antecedent to current rubrics.

The drive for efficiency wasn't only about faster grading of classroom themes; however, it was about social control. This early focus on scientific management led to a push for writing ability to be tested scientifically, which brought with it a distrust in the reliability of classroom instructors. Educational scientists questioned the "lack of agreement among teachers as to the merit of their pupils' writing" and proposed scales as a solution (Hudelson, 1923, p. 163). Milo B.

Hillegas is often credited with the first writing scale developed in 1912, the “Scale for the Measurement of Quality in English Composition” (Behizadeh & Engelhard Jr., 2011, p. 194). Hillegas was the student of another key figure in early writing assessment: Edward Thorndike. Thorndike, who worked with Hillegas on their writing scale, was a proponent of eugenics who saw education as a means to weed out the intelligent population from the “dull normals” and “subnormals” and thus make society more efficient (Russell, 2002, p. 139). He sought a method for determining which individuals fell in which category and because an early (1904) study established a link between intelligence and “abilities in English,” evaluating writing was important to Thorndike’s goal (Elliott, 2005, p. 35).

The 10-point Hillegas scale or Thorndike-Hillegas scale (they revised it together) reads less like a modern rubric and more like a set of benchmarks, examples of writing at various levels. In the first stage of developing the scale, Hillegas gathered actual student work, but this work did not represent the full range he wanted on the scale, so he also added artificial samples. On the high end was writing from Jane Austen and the Brontës—surely no found student writing could match this. On the low end of the scale was nearly incomprehensible prose that Hillegas made up. Hillegas had 100 readers (only 73 of whom were reliable enough to make his final cut—perhaps not Austen fans?) arrange the samples from worst to best. He also had teachers and published authors judge the samples. The final set of 27 artifacts were arranged on a scale (Elliott, 2005). Teachers were then instructed to use this scale to compare their own students’ work to the samples and thus grade more reliably.

Teachers themselves valued writing scales (and currently value rubrics) for making their grading more “fair.” However, such measures have also been used to keep teachers in check and compare them to others. The drive for national comparison, fueled by a distrust of teachers, started with an overall drive for efficiency in education during the early 20<sup>th</sup> century. Hillegas heralded the scale for making comparisons across institutions and creating a national standard (Turley & Gallagher, 2008, p. 88). In particular, early writing scales were used at the secondary level to evaluate teachers. Principals evaluated teachers on whether or not their students improved on the Hillegas scale (Turley & Gallagher, 2008, p. 88). Similarly, today’s rubrics are heralded for their ability to standardize the work of teachers and compare students to national benchmarks, but they also are a means of ruling teachers and exercising control over classrooms.

The Hillegas scale came under fire for many of the same reasons rubrics do today. Educational administrator Franklin W. Johnson (1913) complained that the scale was vastly inadequate for evaluating the content of student writing or the originality of thought. He saw it “like using a yardstick to determine the

weight of material in the physical laboratory” (p. 48). Simply put, the tool was not valid for measuring writing. Even Thorndike turned on the scale, saying it was only good for identifying errors in *evaluating* writing, not errors in the writing itself (Elliot, 2005, p. xiv).

Nevertheless, the quest for the perfect writing scale continued. The 1920s saw the development of analytic-point scales to complement general scales and allow for the quality of different elements of writing to be scored independently of one another (Hudelson, 1923, p. 168). Still, Johnson (1913) warned that such scales would not improve student writing ability and anyone who expected them to do so would be disappointed (p. 163). But of course, improvement was never the sole goal. Rather, writing scales were meant to sort, compare, and exclude. The “link” between literacy and intelligence that Hillegas and Thorndike reinforced had real social consequences that continued long after their original writing scale. For example, when it came to the military, literacy tests had real life and death consequences. The Thorndike Reading Scale and accompanying literacy tests were used to test soldiers drafted for WWI and indicate the inferiority of the “negro draft” (Elliot, 2005, p. 70). In WWII, verbal analogies from the SAT were used in The Qualifying Test for Civilians to determine who might be trained as officers rather than placed in front-line combat (p. 118). Finally, in 1951, President Truman approved the Selective Service Qualifying Test that used similar questions on verbal relations as well as reading comprehension to determine who might defer the draft and go to college (p. 325). Thus, the consequence of Thorndike and his colleagues’ efforts in the assessment of literacy extended far beyond the college classroom.

## **AMERICA’S LEGACY: ETS, TESTING, AND THE MODERN RUBRIC**

Trends during these early periods are important because they show that educational measurement theorists *have always* constructed writing assessments separate from the teaching of writing (Behizadeh & Engelhard Jr., 2011). Although Hillegas intended his scale for classroom use, he still saw it as a means for keeping teachers in check across institutions. However, it is testing that Norbert Elliot (2005) called “America’s unique contribution to education” (p. 4). And writing scales became essential to scoring any tests that involved writing essays.

First used in 1926, the SAT became a touchstone for the advancement of writing assessment. The emergence of the SAT and the College Board solidified the already underlying connection between writing scales and testing. It also led to the creation of the modern writing rubric and often used norming pro-

cedures. Carl Campbell Brigham, chairman of the College Board in the 1920s and hailed creator of the SAT, was also well-known eugenicist. Like Thorndike, Campbell Brigham's interest in writing assessment was connected to his racism. He used the "proven" connection between literacy and intelligence as evidence of lower intelligence in immigrants and African Americans (Elliot, 2005).

From this problematic history arose procedures for scoring essays that led to current practice. In order to develop reliability among essay scorers, Campbell Brigham used a process where scorers gathered around a table to read essays, each essay was scored twice, and difficult essays were sent to a special group of readers (Elliot, 2005). As resources became tight, Brigham was careful to only select readers that were consistent in their scoring. He also kept adjusting the number of points on the rating scale, which were at one point as high as 35. Even on the 10-point scale, he found that readers only regularly used four of the point values, and he ended up becoming frustrated with a numerical score at all (Elliot, 2005). The idea of using two-raters for an essay as well as the idea of a limited number of ratings (4-5) stuck.

When the Educational Testing Service (ETS) was created in 1947, reliability in essay testing was one of their top priorities. In fact, ETS became one of the primary sponsors of research on writing assessment throughout the 1940s, 1950s, and 1960s (O'Neill, Moore, & Huot, 2009). For their first 20-some years, ETS struggled to find a satisfactory method of scoring essay exams. However, in the 1960s their research led to what we might recognize as the first modern writing rubric. In 1961, an ETS-funded study by Diederich, French and Carlton narrowed writing assessment to five main categories: "ideas, form, flavor/style, mechanics and wording" (O'Neill, Moore, & Huot, 2009, p. 22). These categories became the basis for an analytic trait "rubric," known as the Diederich scale, in which each category was scored separately since readers had difficulty agreeing on overall quality of the essays (p. 22). While modern rubrics have become more nuanced, these categories are likely not unfamiliar to current scholars or practitioners. Thesis, organization, style, and grammar remain some of the most common rubric categories (Dryer, 2013).

Diederich was major force in shifting to what Yancey (1999) dubbed the "second wave" of writing assessment, which focused on direct assessment through holistically scored essays rather than indirect assessment through objective tests. As Diederich said in 1974: "whenever we want to find out whether people can swim, we have them jump in the pool and swim" (as cited in Behizadeh & Engelhard Jr., 2011, p. 202). Unlike his predecessors, Diederich was also concerned with inequity. In particular, he was concerned that southern students failed the Selective Service College Qualifying Test that allowed them to defer military service at a far higher rate than those from northern colleges (Elliot,

2005). Much of his life's work was spent in trying to assess direct samples of writing fairly. Diederich saw writing scales as means to create a common vocabulary and in doing so, ensure reliable and fair assessment (Haswell, 2014). While current writing assessment scholars recognize that "even features that seem generic...that are often found on rubrics and scoring guidelines should be defined by the specific situation" (Adler-Kassner & O'Neill, 2010, p. 64), for Diederich, disagreement among readers was as a matter of individual taste rather than social circumstance (Broad, 2003). In the interest of saving teachers time and confident that common vocabulary could lead to reliable essay assessment, Diederich advocated for employing external graders trained to score essays with high levels of reliability. It is no surprise that this effort was supported by the Ford Foundation whose name is often thought of as synonymous with scientific management and efficiency (Elliott, 2005). Diederich's example is important to our current context because it shows how direct assessment and a drive for efficiency have historically co-existed.

In 1966, a breakthrough occurred in essay scoring that furthered both the goals of direct assessment and the drive for efficiency. Godshalk, Swineford, and Coffman designed the basis for modern holistic scoring, including the process of norming to train readers and monitor their progress (O'Neill, Moore, & Huot, 2009). Each paper had four readings on a three-point (high, average, low) scale. A key addition to their process was norming the group of readers using sample papers, which were discussed as a group in order to reach consensus (Elliot, 2005). The results were strong, above .7, but the process was cut from four readers to two due to cost. Their 1966 work "The Measurement of Writing Ability—A Significant Breakthrough" they claimed that their process certified the essay as a means to assess writing ability (Elliot, 2005, p. 164).

It would be at least another decade until holistic scoring and direct assessment were dominant, but the framework for it had been laid by educational measurement specialists. Direct assessment became the national trend in the late 1970s and 1980s, and it is not a coincidence that its popularity corresponds to the rise of outcomes-based assessment in the 1980s. Outcomes originally seemed like a promising way to assess courses and curricula—one that took into account actual student work, but the history of exclusion and America's preoccupation with testing was never fully left in the past. In order to assess outcomes, raters needed measurements, and thus the writing scale gained dominance in writing assessment. By the early 1980s, it could be assumed that any process of scoring essays would involve the use of a writing scale (Dryer, 2013). While it replaced multiple choice testing, this process was still designed to be used outside of the classroom context, for trained external raters to score essays in a way that ensured reliability and acted as a check on individual teachers.

## AN EDUCATIONAL INTELLIGENCE COMPLEX “TROJAN HORSE”: OUTCOMES BECOME STANDARDS

Two words dominate assessment in the 1980s & 1990s: outcomes and standards. As a field, composition has distinguished the two, but in some circles, they have always been connected. So, too, have we seen outcomes conflated with competencies. Such inconsistencies in language can be disturbing for those who study writing. Chris Gallagher (2016) described a talk from an official at an accrediting agency where the guest switched almost seamlessly from a position on “authentic assessment” that those in our field would support to “validating competencies,” which raises our alarm. This “Trojan Horse,” as Gallagher (2016) called it, is not necessarily nefarious. But it is a way of using language that is often different from the way we see it used by writing scholars. Diving into the history of outcomes-based assessment provides us background for understanding the current discussion on outcomes, competencies, and national-scale rubrics.

William Spady was instrumental in beginning the outcomes-based educational (OBE) movement. With co-author Kit Marshall, Spady defined an outcome as “a successful demonstration of learning that occurs at the culminating point of a set of learning experiences” (1991, p. 70). Spady and Marshall’s vision for “transformational OBE” was that educators would have a “guiding vision of the graduate” that would guide the development of curriculum (p. 70). In a retrospective interview, Spady lamented the loss of his transformational vision for OBE. He called current OBE: “a curriculum-driven system with what people claim to be ‘outcomes’ sprinkled over the top” (Killen, 2016). This “traditional” approach, Spady explained is not outcomes *based* at all. Rather, it simply adds outcomes to existing curriculum in order to meet accountability mandates (Spady & Marshall, 1991, p. 69). Part of the difference between Spady’s dream and the current reality is that Spady wanted outcomes-based *education* rather the outcomes-based *assessment*.

This distinction can be explained in terms of writing scholars’ own views on the role of outcomes. The Council of Writing Program Administrator (CWPA) Outcomes Statement was initially developed to guide curriculum development, not to assess it. Harrington et. al’s (2005) edited collection *The Outcomes Book* is a retrospective on the WPA Outcomes Statement that draws important distinctions between outcomes, which guide curriculum, and standards, which are used to assess it. The authors of this collection agree that outcomes form the basis for designing curriculum while standards provide a check on whether or not that curriculum has been successful (Yancey, 2005). White (2005) explained that “outcomes do not require agreement on a single best way to achieve those outcomes” or agreement on the level to which they should be achieved (p. 5). The WPA Outcomes were meant to give guidance to teachers and programs,



not to standardize writing curriculum (Wiley, 2005, p. 27). Nor were the WPA Outcomes developed with a rubric or even any particular type of assessment in mind. In fact, when the Council of Writing Program Administrators (CWPA) developed the WPA Outcomes in 1999, they did so to strategically *avoid* composition from being targeted by the growing movement for standards in higher education. Patricia Ericsson (2005) remembered: “Developing the Outcomes Statement was an offensive, proactive move” against first-year composition being defined by those outside the field (p. 115).

Despite this initial focus on outcomes as a guide for curriculum rather than as a means of assessment, I would wager that many of those in writing studies have at least seen, if not created, a rubric based on the WPA Outcomes. Outcomes are now about results to be reported rather than goals to define curriculum. Outcomes assessment serves the needs of reporting by “providing nice, clean numbers for university administrators’ spreadsheets” (Gallagher, 2012, p. 46). Shifts in accreditation are certainly one reason for this change. While Linda Adler-Kassner and Peggy O’Neill explained in 2010 that accrediting agencies looked to institutions to set their own standards (p. 28), by 2012, Adler-Kassner noted that accreditors were under attack for this. She explained that government and public agencies were repeatedly criticizing accreditors for “allowing institutions to set their own learning standards and develop their own assessments; the lack of consistent outcomes across institutions; and the lack of comparable data” (Adler-Kassner, 2012, p. 123).

Similarly, we see a shift in a second retrospective collection about the WPA Outcomes (2013) where Paul Anderson et al. noted that many institutions were using writing outcomes for assessment to an external body, such as a state mandate. Outcomes are now inherently tied to assessment, and that assessment often comes in the form of rubrics that are designed for large-scale assessments. Theoretically, the assessment loop would then lead back to curricular development. What we learn from assessment would be used to shape curriculum and thus both assessment and curriculum would be guided by outcomes—something known in assessment circles as “closing the loop.” Too often, however, the focus is on reporting and accountability rather than on teaching and learning. This focus comes from a larger focus on management and quality assurance that parallels the development of outcomes-based assessment and accountability.

## **ACADEMIC MANAGEMENT: THE RISE OF QUALITY ASSURANCE, NEOLIBERALISM, AND RUBRICS**

As we’ve seen, writing scales are not new, nor is comparison between teachers or among schools. At the broader institutional level, the focus on efficiency dates

back to the early 20<sup>th</sup> century when Frederick Taylor applied scientific management to make industry less wasteful, and others applied this efficiency approach to education (Wiley, 2005, p. 26). A drive for quality and efficiency also permeates writing assessment from its origins. The elusive factor of “quality” is in the very title of the Hillegas scale for “the Measurement of *Quality* in English Composition.” Similarly, the question of “quality programs” and their assessment permeates WPA discourse from the beginnings of the CWPA in 1977 (Strickland, 2011). Early meetings of the Conference on College Composition and Communication (CCCCs) took this approach by focusing on systematizing first-year writing through workshops that allowed for more efficient, more accurate reading of student themes (Strickland, 2011).

What changed in the later portion of the 20<sup>th</sup> century is *who* is responsible for quality and efficiency. While early scales had a policing effect on society, early educational scientists also created writing scales so that *teachers* could more efficiently score student work. Yet in our current neoliberal landscape, *society* has become the evaluators. It is now the burden of the public to hold higher education accountable—to evaluate teachers and universities on how quickly they graduate students and how well they compare to their peers. With options like the College Scorecard and the Voluntary System of Accountability (VSA)’s College Portraits,<sup>1</sup> which allow students to compare universities based on factors such as cost, graduation and employment rates, potential students are tasked with finding the “best fit” for them based on the ratio of cost to value. We have made it the responsibility of the “savvy student” to choose well in order to make the most of their student loans and ensure the highest rate of return when graduating (Seal, 2018).

This push toward privatization and individual choice is a key tenant of neoliberalism and part of an overall narrative that individualized instruction is the most cost-effective and efficient means to graduation (Seal, 2018). In the introduction to their edited collection *Composition in the Age of Austerity*, Tony Scott and Nancy Welch (2016) defined neoliberalism as a change toward making public services private, or when they remain public, applying “market logics” to them. While terms like “corporate university” convey a similar market logic, applying the term neoliberalism to the academy links changes to education with overall economic changes in democratic society (Seal, 2018). Neoliberalism has

---

1 Since this research was conducted, VSA has become VSA Analytics. Their service now seems more geared toward providing these statistics to college and universities themselves. However, it appears that many universities place these facts on their websites in the hopes of using them to attract students. As of this writing in February 2022, I found many universities with links to their College Portraits stats that now come up with an error. This shift is itself testament to how such ideas circulate. At this point, College Scorecard is still available.

led to what Scott and Welch (2016) called “audit culture” where “everything must be assessed against institutional benchmarks and comparator/competitor schools are measured for its value added” (p. 12). Agents of neoliberalism shift the focus from solving economic challenges with wide-scale economic solutions to individual accountability and determining the best value for the money. As Scott and Welch (2016) concluded, it’s not about cost-saving, but cost-shifting: from publicly funded universities to student debt. In such an economy, assessment serves not only to certify student success and choices but to hold universities accountable—to compare them so that students can make the best decisions with their money. Andrew Seal (2018) noted that this neoliberal society has taught students to look at their course schedule “like a bond trader looking over a portfolio” to maximize their investments.

This neoliberal shift to apply market logic to education happened gradually, starting in the 1980s. In the 1980s, “‘quality’ became a buzzword in management” (Strickland, 2011, p. 113). Beginning with *A Nation at Risk* (1983), a report by the U.S. National Commission on Excellence in Education, “alarmist reports” about the decline in educational quality captured the attention of both government officials and the general public (McClellan, 2016). So, too, these reports have worked to tie individual success to the success of the United States as a nation. At lower levels of education, the 2002 *No Child Left Behind* Act tied “education’s role as a public good” to the progress of an individual (Adler-Kassner & O’Neill, 2010, p. 24). *A Test of Leadership: Charting the Future of U.S. Higher Education* (2006), commonly referred to as the Spellings Commission because then Secretary of Education Margaret Spellings headed the commission, is known for devaluing of teacher expertise and taking a technocratic approach that insists on external management of education (Adler-Kassner & O’Neill, 2010). Obama’s 2009 *Race to the Top* initiative linked education and individual economic mobility and called for means by which parents and students could compare the product—or school—that they were purchasing with other options (Adler-Kassner, 2012). The underlying thread that connects this national discourse on education is that all students, including those from unprepared and diverse backgrounds, can succeed if they make wise choices. Wanting every individual to succeed is a noble goal, but in these reports, “success” looks the same for every student, and it is never defined *by* the student. Students should be offered different paths to that success, even individualized paths, but what is considered a “quality education” is seen as universal.

The ability to compare the “quality” of schools rests on the idea that all schools should help students reach the same education outcomes and goals. Outcomes are defined by external partners—government, corporate partners, and philanthropists—and the biggest concern is if these outcomes can be accomplished

efficiently. The “endgame,” according to Gallagher (2016), is competency-based education, where everything is based on meeting common competencies, not on taking particular courses. Competency-based education takes outcomes-based education to one particular extreme. As Yancey (2005) explained, outcomes have to do with what students know, but don’t necessarily define a particular level of proficiency. In addition, outcomes allow for individual teaching style and are used for programmatic assessment and curricular development. In contrast, competency-based education develops standards that “act as a check on the students as well as the courses” (Yancey, 2005, p. 20). Unfortunately, the words “outcomes” and “competency” (and “proficiency”) have become somewhat interchangeable at this point (Mette Morcke, Dornan, Eika, 2013).

This shift from outcomes to competencies fits with a neoliberal agenda by linking education and economic order (Seal, 2018). Accountability serves as a “sleight of hand” to distract us from systemic economic and racial inequity to higher “quality” education as a solution to economic distress (Scott & Welch, 2016). As Scott and Welch (2016) aptly put it:

The solution to the *economic* gap is not *economic* restructuring (i.e. restored funding) but instead *educational* restructuring through accountability and efficiency mandates that push foundational changes in curriculum, pedagogy, and—by tying the ‘value’ of a college degree to speed of its completion and the earning of its recipient—what a college degree signified. (p. 10)

We have become focused on whether or not students reach goals, not on which goals are appropriate (Adler-Kassner & O’Neill, 2010). Thus, outcomes—a seemingly solid concept for *planning* curriculum—have become competencies that are “enshrined in the bureaucratic machinery” (Gallagher, 2012, p. 45). They must be included on forms when we propose new courses; they must appear on our syllabi, and the two should always match. And of course, outcomes must *always, always* be assessed. They have become “fetishized” (Gallagher, 2016) and measuring and *reporting* how students meet outcomes has come to serve the needs of academic management, not students (Gallagher, 2012).

Like in other industries with systemic management, the need to report to distant supervisors becomes a dominant force for how assessment work is completed. Reports abstract actual experience and define what is normal or abnormal (Nichols, Griffith, & McLarnon, 2017). The genre of the assessment report “commoditizes, reifies, and obscures the dynamic, messy, material, socially useful, inescapably values-driven labor of teaching and learning” (Mutnick, 2016, p. 39). Competencies serve academic management by allowing administrators

to show how their university compares to others, and when students fail to gain competencies, it is teachers who need disciplined in the form of better training and increased accountability rather than better working conditions. Teachers are the “objects of regulation” in a competency-based system (Mette Morcke, Dornan & Eika, 2012, p. 855). Put in the context of management and reporting, it makes sense that the product of the academic enterprise is a competent student and that the teacher is the employee responsible for the quality of this product.

Rubrics and their claims of reliability became one method of ensuring that success looked the same for different types of students at different colleges and universities. The word “rubric” was first used in 1981 to describe the writing scales used to holistically score essays written for Advanced Placement (AP) English exam (Griffin, 2010). By 1984, national data on student writing performance was also being collected in the United States and was being scored based on the Godshalk research group method (Elliot, 2005). Applebee, Langer, and Mullins wrote a 1984 report *Writing Trends across the Decade, 1974-1984* that further solidified rubrics, particularly those focused on primary trait scoring (using a scale for one particular trait in writing) as the gold standard for reliability (Elliot, 2005, p. 197). Elliot (2005), however, explained that minoritized students as well as Title I, lower-class schools did poorly on these assessments. Thus, we come full circle. Literacy is tied, as it was from the beginning, to the success and intelligence of an individual, and minoritized groups “test” below others. And since the individual’s success is tied to the nation’s success, national standards are created and incentivized to raise the “quality” of education at large.

In the absence of indirect testing, rubrics serve as a neat and clean way to report on whether or not those competencies are being met—to report on the quality of education. Bob Broad (2003) noted that scoring guides and rubrics serve to *document* the evaluation of student writing. They have become a means of communication about writing—a public record—within a larger system of academic management, yet they only capture a fraction of the values at work when evaluating student writing (Broad, 2003). Outcomes-based assessment using rubrics is now seen as “common sense” within academic management, something we accept, however begrudgingly, as a part of our work as academics (Gallagher, 2012, p. 48). While many scholars have focused on either the value or the detriment of the rubric to students, teachers, and writing programs (Anson et al., 2012; Balester, 2012; Broad, 2003; Crusan, 2015; Turley & Gallagher, 2008; Wilson, 2006), what remains relatively unexplored is the role of rubrics within this larger *institutional system*. That isn’t to say outcomes assessment or writing rubrics are all bad, simply that they play an *institutional* role in governing our work.

## ENTER THE AAC&U

This book explores how one national rubric—the AAC&U’s VALUE rubric for Written Communication governs the work of writing assessment nationally and particularly at two small colleges. In this section, I discuss the role of AAC&U as a key organization that has historically defined what it means to be an institution of higher education, specifically what it means to receive a “liberal education.” Their current overall mission is “to advance the vitality and public standing of liberal education by making quality and equity the foundations for excellence in undergraduate education in service to democracy” (AAC&U, n.d., “About”). Their Liberal Education and America’s Promise (LEAP) movement institutionalizes the AAC&U’s vision for liberal education through common outcomes and rubrics. These texts then go on to be propagated by administrators and faculty in higher education, few of whom connect them back to the AAC&U and their larger mission defining liberal education.

The definition of liberal education has shifted significantly over time, but one consistent thread has been the goal to “train good citizens to lead society” (Crowley, 1998, p. 47). However, what this training looks like, who should be trained, and what benefits society has shifted significantly since the origins of the term. Liberal education is often confused with general education, but the two movements were originally distinct (Crowley, 1998). Liberal education was first associated with training gifted individuals to master traditional subjects and read the canon of “great books,” while general education was concerned with providing skills to a broader base of students to succeed professionally (Crowley, 1998). David Russell (2002) described the warring of factions defining higher education, where one side believed that a good citizen was a cultured citizen, and culture was synonymous with White and Western. The other faction was that of social efficiency, which maintained that all students needed specific skills and qualifications to be strong citizens. Social efficiency won out and this influenced the teaching of writing as a skill that could be tested and quantified (Russell, 2002).

Since the 1940s, the AAC&U has concerned itself with defining liberal education, and they have continued to re-define that term for the 21<sup>st</sup> century. The definition of liberal education that the AAC&U now subscribes to is quite broad. It is simply an approach to education that “empowers individuals with broad knowledge and transferable skills, and a strong sense of values, ethics, and civic engagement” (AAC&U, 2006, p. 2). Rather than a “great books” approach, we now have a “great skills” approach—liberal education means teaching skills that are transferable to careers. This definition builds on the history of social efficiency—the quicker these great skills can be achieved the sooner a student can

move into a career and benefit society. Social efficiency has also now been linked to equity. Rather than making sure the best and the brightest read the great books, we now must make sure that all students acquire these great skills—yet as I show throughout the book, these skills still come from White values. It may be a positive step to move away from historic approaches to writing assessment that linked the knowledge of White canonical texts and SEAE to intelligence, but the AAC&U’s version of equity still represents a neoliberal vision. Rather than acknowledge and fight systemic, structural impediments of education, this neoliberal vision maintains that if all students simply receive similar instruction, they will all succeed. The notion that “productivity=success=equity” is prevalent in discourse about both faculty and students within higher education (Adsit & Doe, 2020, p. 90). For the AAC&U, the solution to inequity is to define success by common outcomes (LEAP), ensure success by assessing it on common rubrics (VALUE), and thus prepare every individual for citizenship and the workforce (McConnell & Rhodes, 2017). Thus, LEAP outcomes and VALUE rubrics are seen as a part of AAC&U’s overall mission as an organization to foster liberal education for all students.

### **LIBERAL (OR NEOLIBERAL) EDUCATION AND AMERICA’S PROMISE(D)**

LEAP stands for “Liberal Education and America’s Promise,” and the product of the movement is a set of national outcomes for liberal education. According to the AAC&U, the “promise” embedded in LEAP is one made to students that higher education will be worth their time and money and lead to “a better future,” that no matter what school they attend, they will acquire these particular skills (AAC&U, 2007, p. 1). But so, too, are students promised to employers as ideal future employees who graduate with the training and skills they need for the workforce. The LEAP outcomes connect these two interests and define what skills society at large can expect all students to acquire in post-secondary education.

The LEAP Essential Learning Outcomes are less a set of *measurable* outcomes than a categorization of the learning that should be valued in an overall liberal education. The 2007 executive summary stressed that different types of institutions and programs would apply the outcomes differently. Furthermore, the initial LEAP Outcomes do not read the way we’ve come to expect outcomes to read—as measurable goals beginning with clear verbs. Rather, they include four broad categories to help students “prepare for twenty-first century challenges:”

- Knowledge of Human Cultures and the Physical and Natural World
- Intellectual and Practical Skills

- Personal and Social Responsibility
- Integrative Learning

Under each of these broad categories there are more specific bullet points. For example, under “Intellectual and Practical Skills” the AAC&U lists six bullets, including critical and creative thinking, written and oral communication, and information literacy (AAC&U, 2007, p. 3). Under this list is a line noting that these should be “practiced extensively, across the curriculum,” rather than accomplished in one specific course (AAC&U, 2007, p. 3). The LEAP outcomes, as with the original outcomes-based movement, were thus meant to *inform* curriculum, not to standardize it. Nevertheless, like the outcomes-movement as a whole, the rhetoric behind LEAP has both shaped and been shaped by neoliberal views about the purpose of higher education within the economic system of capitalism and political system of U.S. democracy.

Although the AAC&U has argued that their LEAP Outcomes are not “just about the economy” but about all areas of life, including “environmental, civic, cultural, imaginative, ethical” spheres, the outcomes were based on a survey of employers, not students or teachers (AAC&U, 2007, p. 17). The LEAP National Leadership Council, which consisted of heavy hitters from corporate settings<sup>2</sup> as well as colleges, formed the outcomes from a 2006 survey of employers about what they felt graduates of higher education needed (AAC&U, 2007). The council’s report aimed to shift the focus away from the conversation about access, affordability, and accountability and toward the consensus of what a college graduate should “know and be able to do” (AAC&U, 2007, p. 1). To do so, LEAP relies on a narrative of consensus among educators and employers.

By including teachers in this consensus, the AAC&U seeks to separate their LEAP Outcomes from other contemporary reform initiatives, particularly the Spelling Commission Report. Although the development of LEAP’s Essential Learning Outcomes began several years before the Spellings Report was released, the AAC&U officially released the LEAP outcomes later that same year, 2007, and offered them as a counter narrative to the distrust of teachers conveyed by the Spellings Commission. Rather, the AAC&U asserted that teachers should be central to educational reform (AAC&U, 2007). Nevertheless, the argument presented by the AAC&U is that teacher expertise is valuable not for what it represents within the classroom but for how it helps graduates meet the needs of

---

2 The council favored representation from Ivy League schools but did include strong community college representation. On the employer side it had a leaning toward legal professions. The group was ethnically and racially diverse as well, including strong Black, Hispanic, Asian, and Middle Eastern leadership. The group had a democratic bent, including several of those who worked for the Clinton administration, as well as those who have fought for racial and gender representation in higher education.



future employers. As is common in neoliberalism, individual success, whether defined economically or otherwise, is tied to the success of the institution and society at large.

When the LEAP council calls for “a new compact, between educators and American society” that puts “the future of democracy at the center” of education (AAC&U, 2007, p. 5, p. 9), make no mistake that it is a part of the larger narrative that Adler-Kassner (2017) called the Education Intelligence Complex (EIC)’s story of “The Problem with American Education and How to Fix It” (p. 320). The rhetoric of the AAC&U and the LEAP movement may resist some of the “technocratic narrative” of Spellings (Adler-Kassner & O’Neill, 2010, p. 85), but ultimately their approach still exists within this larger national frame. As we turn to examine the VALUE rubrics more closely, it is important to remember that alternative assessment does not automatically counter the agenda of neoliberalism, nor does it tackle the systemic issues inherent to having a national standard in the first place. A different method for “how to fix it” (rubrics instead of testing) does not resist the narrative that the structure of higher education needs fixing or that external stakeholders are the ones to develop the solutions.

## DEFINING THE VALUE OF OUTCOMES THROUGH RUBRICS

The LEAP Outcomes were designed to be adapted by individual schools, and originally, there was no means to measure how well each school incorporated the outcomes or how well each student achieved them. However, the Spellings Commission refocused the national conversation on *measuring* the outcomes, and outcomes education as a whole moved toward accountability and assessment (Gallagher, 2016). The AAC&U looked to answer the Spelling Commission’s call for accountability with an alternative to testing by developing the Valid Assessment of Learning in Undergraduate Education (VALUE) initiative in 2009 (McConnell et al., 2019; Rhodes, 2010). Then Vice President for Quality Curriculum and Assessment at the AAC&U, Terrel (Terry) Rhodes wondered: could an alternative means be used to represent the work of higher education, one that captured more of the “rich and varied dimensions” of individual institutions? (Rhodes, 2010, p. 1).

The VALUE rubrics held the promise of a different way to quantify the success of higher education. AAC&U President Geary Schneider (2015) presented the VALUE rubrics as a “more specific” means of accountability that accounts for the complexity of learning in higher education (p. vii). The goal of the VALUE rubrics was to develop an accountability measure that says something “significant about learning”—to “respect the complexity” of higher education and “embrace multiple essential learning outcomes” (Schneider, 2015, p. vii). Sulli-

van (2015) added that metrics about access, completion, and earnings of graduates do not say anything significant about learning with the implication that the AAC&U's new VALUE initiative does. Whether or not the VALUE rubrics have met these goals—whether any rubric can—is debatable, however, a *belief* that they could is the reason that the AAC&U turned its attention to the creation of the VALUE rubrics.

Resisting the “technocratic” narrative of the Spellings Commission also fits with the role AAC&U sees for itself as a steward of liberal education. Adler-Kassner (2008) explained that technocrats and stewards are two historically competing views of liberal education. Stewards focus on “the cultivation of critical intelligence by means of inductive, nurturing education” while technocrats see the need for that intelligence to be managed from above (Adler-Kassner, 2008, p. 44). Although the word steward is not prominent in the current AAC&U literature, it seems telling that it comes up in the AAC&U's description of their role in relationship to the VALUE rubrics. They see themselves as “the intellectual and logistical *steward* [emphasis added] of the VALUE rubrics” (McConnell et al., 2019, p. 2). Since the role of steward is tied to fostering individual intelligence rather than managing education externally, the AAC&U continually stresses that their rubrics were designed to be adapted to local use.

From the beginning, the rubrics were meant to be “meaningful for local purposes” and “local pedagogical needs” (McConnell et al., 2019, p. 2). Rhodes and Finley (2013) rejected the language of standardization:

Precisely because they are not standardized, the VALUE rubrics can be readily adapted to accommodate the language used to frame learning goals on individual campuses and to reflect different institutional missions and program variations.  
(p. 3)

Rhodes (2012) called the VALUE rubrics “meta-rubrics,” rubrics to be adapted and used by multiple institutions. In 2009, the AAC&U released 15 of these meta-rubrics, including one for Written Communication. The notion of meta-rubrics to be adapted by local institutions is in keeping with AAC&U's philosophy as a steward—a guide in higher education, not a technocratic manager.

Yet, over time the role of the AAC&U has shifted to be more managerial, and the VALUE rubrics have become more standardized. As of 2019, the VALUE Institute now offers an external service for evaluation where institutions may send samples of student work to be scored by external raters (AAC&U, 2017). Thus, the AAC&U seems to be walking a well-worn path in the history of assessment. As seen historically, the use of writing scales shifted from classroom consistency to external reliability, as Diedrich and others trained external graders

for higher levels of agreement (Elliot, 2005). So, too, have the VALUE rubrics evolved beyond a local, adaptable tool. They have come to play a significant role in creating *standard* tools that are used across institutions to assess and compare student (and therefore university) performance.

Whether or not the AAC&U's motivation was originally (or is currently) to manage higher education, building *national* outcomes and rubrics has a normalizing effect. Such texts shape practice, "mediating idiosyncrasies and variability in local settings" (LaFrance, 2019, p. 43). Inoue (2015) explained that writing assessment is an ecology where "individual actions by students or a teacher or rubric" do not work in isolation. Rather, they "may be instigators" within a larger ecology "that determines what possible outcomes, effects, changes, or products" (p. 120). This chapter has laid the historical foundations that form the current assessment ecology in which organizations, individuals, and texts interact.

A central question of this book is *how* rubrics normalize our discourse about writing and what power we have over this as administrators and teachers in higher education. In order to answer this question, I follow the stories of Kristen and Dwayne and their colleagues at Oak and St. Rita's as they interact with, and sometimes adapt, the VALUE rubric for Written Communication. In addition, I continue to engage with the discourse and texts of the AAC&U about the VALUE rubrics and their role in higher education. As I explain in Chapter 3, institutional ethnography provides a methodology for connecting the everyday experiences of individuals to larger power structures that is useful for understanding how rubrics function on the institutional level. Institutional ethnography examines "how individuals take up texts and coordinate their actions, so they produce *the particular institution's standard sequences*, its decision, policies, and outcomes" (Turner, 2006, p. 140). When Kristen and Dwayne decide to "take up" the VALUE rubric for Written Communication, they knowingly or unknowingly operate within the larger, historical forces of the writing assessment ecology.