VALUE@10
Valid Assessment of Learning in Undergraduate Education



WE HAVE A RUBRIC FOR THAT

The VALUE Approach to Assessment

Kathryne Drezek McConnell, Erin M. Horan, Bethany Zimmerman, and Terrel L. Rhodes

AAC&U Association of American Colleges and Universities

WE HAVE A RUBRIC FOR THAT

The VALUE Approach to Assessment

Kathryne Drezek McConnell, Erin M. Horan,
Bethany Zimmerman, and Terrel L. Rhodes

VALUE@10
Valid Assessment of Learning in Undergraduate Education

AAC&U
Association of American Colleges and Universities



Association of American Colleges and Universities

Association of American Colleges and Universities,
1818 R Street NW, Washington, DC 20009

© 2019 by the Association of American Colleges and Universities

All rights reserved.

ISBN 978-0-9961404-7-8

To order copies of this publication or to learn about other AAC&U publications, visit www.aacu.org, e-mail pub_desk@aacu.org, or call 202-387-3760.

Contents

Acknowledgments iv

Foreword by Lynn Pasquerella v

Part I. An Introduction to the Evaluation of the VALUE Methodology. 1

Part II. The Interpretive Argument for VALUE: Claims and Assumptions Associated With VALUE Rubric Design and Development. 11

Part III. The Interpretive Argument for VALUE: Claims and Assumptions Associated With VALUE Rubric Application, Score Interpretation, and Use 27

Part IV. The Validity Argument for VALUE: Strengths, Limitations, and Future Work. 43

References 46

Appendix 57

About the Authors 58

Acknowledgments

This publication would not have been possible without the generous support of a host of collaborating funding organizations. The research that serves as the foundation of this report was made possible (in part) by a grant from the Spencer Foundation (#201600020), an organization dedicated to the improvement of education around the world. The views expressed are those of the authors and do not necessarily reflect the views of the Spencer Foundation. Furthermore, the support provided by the Bill and Melinda Gates Foundation, Lumina Foundation for Education, and the Sherman Fairchild Foundation enabled us to interrogate the VALUE approach's *value* in order to improve it.

Our ability to provide the field with a volume addressing VALUE from a methodological perspective depended on the psychometric prowess, critical thinking, and kindness of some of the best scholars in the field. We are forever grateful to Marianne Perie for providing us with the theoretical and organizational framework that serves as the foundation of this work; Gary Pike for his close examination of the dependability of the VALUE approach, good counsel, and sense of humor; and John Hathcoat for his help evaluating the reliability of VALUE while offering critical feedback from his vantage point implementing the VALUE approach through his institution's participation in the Multi-State Collaborative.

We owe an incalculable debt to the myriad higher education, state organizational, and individual collaborators, including but not limited to: the State Higher Education Executive Officers (SHEEO), the National Institute for Learning Outcomes Assessment (NILOA), technology partner Taskstream (now Watermark), Indiana University's Center for Post-Secondary Research (IUCPR), the Multi-State Collaborative's 13 states and nearly 100 institutions, the Minnesota Collaborative's 10 institutions, the Great Lakes Colleges Association Collaborative's nine institutions, approximately 400 VALUE scorers, and all of the individual presidents, provosts, assessment directors, faculty developers, cocurricular professionals, librarians, and faculty of all stripes. We also want to recognize AAC&U's stalwart and talented Communications staff who have assiduously shepherded this manuscript to successful completion.

Finally, we must thank the students whose work and quality learning are the focus of the VALUE Initiative and this publication.

Foreword

By Lynn Pasquerella, *President of the Association of American Colleges and Universities*

For the past decade, the Association of American Colleges and Universities (AAC&U) has shaped the landscape of higher education through championing the Valid Assessment of Learning in Undergraduate Education (VALUE) approach to student learning outcomes assessment. Designed in response to the 2006 Report of the U.S. Department of Education's Commission on the Future of Higher Education (widely known as the "Spellings Report"), VALUE offers an alternative to the report's recommendation that standardized tests, administered to first-year and graduating students, be utilized as the primary method for evaluating student progress and proficiency. Calling into question the efficacy of the statistical methodology embedded within the scope of narrowly focused, commercially available standardized tests, AAC&U instead proposed the expansion of learning outcomes assessment across the curriculum and cocurriculum, using artifacts of student work from a variety of courses and disciplines. Teams of faculty representing the rich diversity of AAC&U's membership, from community colleges and four-year residential liberal arts colleges to large research universities and mid-sized regional colleges, took up the charge to develop rubrics corresponding to each of the 16 Essential Learning Outcomes that emerged from AAC&U's signature initiative, Liberal Education and America's Promise (LEAP).

Reflecting AAC&U's deep and abiding commitment to quality and equity as the foundations for excellence in undergraduate education, the trailblazing nature of the VALUE project resulted in a paradigm shift. Grounded in the assumption that achieving high-quality education for all students, at all types of institutions, requires assessment data derived from authentic student work carried out within courses and cocurricular programs, VALUE positions faculty at the center of the assessment process. Leveraging their expertise, faculty across the country are engaged in creating and scoring student work in course-embedded assignments. As higher education professionals, they understand that student learning develops over time and that demonstrated success requires not only challenging students with increasingly complex problems, but also providing them with opportunities to engage in self-reflection about their work from their first to final semesters. In the end, this type of comprehensive assessment of applied learning outcomes across a broad range of disciplines supplies faculty with the crucial data necessary to improve course assignments and overall pedagogy. It is a process that has the capacity to contribute to faculty professional development and student success alike.

From its inception, there has been increasing demand for participation in and access to the robust assessment protocol, data and resources provided by VALUE and targeted to the needs of specific institutions. Ongoing efforts to enhance scale were catalyzed by the establishment of a VALUE Institute. Launched in fall 2017, AAC&U has been deeply gratified by the level of interest exhibited by institutions of higher education from around the world, and the Institute welcomes all programs, colleges, universities and consortia to engage in the VALUE approach to assessment. To that end, the evidence-based research contained in *We Have a Rubric for That* supports the VALUE methodology, offering a compelling rejoinder to the notion that standardized tests are the only and most reliable method for assessing student learning outcomes. At the same time, it provides a model for developing an asset-based approach to assessment that focuses on positive performance rather than on what students are lacking, alongside a roadmap for undertaking future research and exploration in the realm of learning outcomes assessment.

Amid the growing skepticism regarding the value of liberal education and a lack of confidence among the American public that higher education is having a positive influence on our society, the collective enterprise being carried out through the VALUE Initiative is more critical than ever. We look forward to working together toward achieving our shared objectives of advancing student success and fulfilling the promise of American higher education.

PART I

An Introduction to the Evaluation of the VALUE Methodology

With this publication on assessment for learning and quality, the Association of American Colleges and Universities (AAC&U) provides its most comprehensive methodological consideration of the Valid Assessment of Learning in Undergraduate Education (VALUE) Initiative to date, representing the culmination of a decade's worth of work instantiating the VALUE approach to assessment. First introduced in 2009 as an alternative to the predominant educational assessment paradigm (i.e., standardized tests), the VALUE rubrics have arguably moved from the periphery of student outcomes assessment discussions to the center of conversations about the quality of student learning within and across institutions. As of 2018, the rubrics have been downloaded more than 70,000 times by individuals representing more than 5,895 organizations, including more than 2,188 colleges and universities. The rubrics were used in the first-of-its-kind national scoring initiative—described in the previously published AAC&U report, *On Solid Ground* (McConnell & Rhodes, 2017)—which evolved into the VALUE Institute (www.valueinstituteassessment.org), launched in fall 2017.

From its inception, VALUE has been guided by a core set of fundamental assumptions (Rhodes, 2010c):

- In order to achieve a high-quality education for all students, valid assessment data are needed to guide planning, teaching, and improvement. This means that the work students do in their courses and the cocurriculum is the best authentic representation of their learning.
- Colleges and universities seek to foster and assess learning outcomes beyond the three or four typically addressed by currently available national standardized tests.
- Learning develops over time, is nonlinear, and should become more complex and sophisticated as students move through their curricular and cocurricular educational pathways within and among institutions toward a degree.
- Good practice in assessment requires multiple assessments over time. Well-planned electronic portfolios (ePortfolios) provide opportunities to utilize college data from multiple assessments across a broad range of learning outcomes and modes for expressing learning, while guiding student learning and building reflective self-assessment capabilities.
- Assessment of student work in ePortfolios can inform programs and institutions on their progress in achieving expected goals for external reporting and, at the same time, provide faculty with information necessary to improve courses and pedagogy.

It is the relationship between these assumptions—all pointing toward the goal of improving instruction and, by extension, student learning (Perie, 2013)—and the methodology behind the VALUE approach to assessment that this publication explores. While aspects of the validity and reliability of the VALUE rubrics have been addressed previously (e.g., Finley, 2011; Rhodes & Finley, 2013), this publication attempts to make meaningful connections between and among various studies conducted by AAC&U, as well as the rich extant body of literature that has emerged about rubrics broadly and the VALUE rubrics specifically.

It is critical to note AAC&U's positionality regarding the VALUE rubrics. As the leading national higher education association dedicated to promoting quality and equity in undergraduate education, AAC&U is continuously scanning the environment for evidence of best practices in teaching, learning, and assessment, as well as securing funding to engage campuses in project-based work intended to foster the design and

implementation of best practices through local and collective experimentation. AAC&U does not claim to have single-handedly created the Essential Learning Outcomes or the VALUE rubrics; rather, AAC&U catalyzed their articulation and continues to serve as the intellectual and logistical steward of the VALUE rubrics. While AAC&U has done a great deal of centralized work with the VALUE rubrics, significant work has been done on the VALUE rubrics by others.

This was the plan from the very beginning of VALUE: the diffusion of the rubrics was coupled with an explicit call by AAC&U to the field to take the rubrics and make them meaningful for local purposes. This arguably helped to promote campus-based adoption of the VALUE rubrics. That said, this approach also signals a loosening of control—from modification and implementation to scoring and interpretation of data—that appears to be unique to the VALUE approach and stands in particular contrast to protocols associated with commercially available national standardized tests. As an approach to assessing student learning, VALUE must balance local pedagogical needs with methodological control.

The Argument-Based Approach to Evaluating Validity

This nuance plays out in varying ways throughout this exploration of the VALUE methodology. *We Have a Rubric for That* is not a catalog of evidence. Instead, it employs an *argument-based framework* that synthesizes evidence from various sources first against an interpretive argument and second to form a validity argument (Kane, 2006; Perie, 2013). As defined by Kane (2006), "an interpretive argument specifies the proposed interpretations and uses of test results by laying out the network of inferences and assumptions leading the observed performances to the conclusions and decisions based on the performances" (p. 17); the validity argument is, in turn, the "evaluation of the interpretive argument" (p. 17). In her working paper for AAC&U, *Developing a Validity Argument for the VALUE Rubrics*, Perie (2013) offered a further explanation to distinguish the interpretive argument from the validity argument. She likened the interpretive argument to a "theory of action" that delineates the relationships between the various components of the assessment under consideration. In contrast, the validity argument examines the sum of the interpretive argument's parts by considering the "soundness" of each individual relationship to make a judgment about the interpretive argument as a whole (Perie, 2013).

Establishing the validity of the VALUE approach to assessment using this argument-based framework requires:

- "explicitly specifying the proposed purposes and uses of the [VALUE rubrics],
- developing an interpretive argument consistent with these proposed uses, and then
- critically evaluating the plausibility of these inferences and assumptions" (Perie, 2013, p. 2).

After first providing a brief synopsis of the history of VALUE (2009–2019), we address each of these requirements by describing the interpretive argument of the VALUE validity evaluation and then exploring each claim of the argument in relation to available evidence. Finally, we consider the totality of the interpretive argument in relation to the purpose and goals of VALUE to evaluate its validity, identifying areas of relative strength as well as explicitly delineating concrete steps to address methodological gaps and concerns, including but not limited to discussing plans for the revision of the VALUE rubrics. The argument-based framework outlined here "provides a pragmatic approach to validation that can be followed by test developers and test users" (Perie, 2013, p. 2). This is a critical feature of the work, as AAC&U is not and should not be the sole arbiter of the soundness of the VALUE approach. Each institution must consider the important features of VALUE in relationship to their intended use(s). It is our hope and expectation that the work encapsulated here supports and extends work transpiring in local contexts across the diversity of institutions that are leveraging VALUE to improve instruction and enhance student learning.

Finally, we acknowledge and embrace the dynamism that is inherent to the VALUE approach. VALUE was purposefully designed to be both an articulation of commonly held essentials representing a collective understanding of the ultimate purposes of a quality undergraduate education as well as an intentionally malleable approach to making sense of learning at an individual institution, one that could be changed as necessary to meet local needs. As such, the assertion that validity "results should not be interpreted in a black and white context of the assessment is working or the assessment does not work" (Perie, 2013, p. 4) is particularly salient to the interpretation and use of VALUE. Instead, it is AAC&U's hope and expectation that the broader community of faculty, administrators, cocurricular professionals, methodologists, scholars, employers, policy makers, and students will help AAC&U use the evidence contained herein to continue to refine and improve VALUE.

From Inception to Institute: A Brief History of VALUE (2009–2019)

Following the release of the U.S. Department of Education's Commission on the Future of Higher Education (2006) Spellings Report, the country engaged in widespread debate regarding the use of standardized testing in higher education. The Spellings Report posited that student learning must be assessed in higher education to show student learning from college entrance to graduation. Educators agreed. The disagreement occurred over the "who should and how" to assess student learning. The Spellings Report recommended the use of major, commercially available standardized tests typically administered to first-year college students and/or seniors to demonstrate learning. In response to this recommendation, many education experts and methodologists pointed to the inherent problems with standardized testing to assess student learning as well as the statistical shortcomings of a "value-added" statistical methodology popularized at the time by one of the newer standardized tests (Banta & Pike, 2007, 2012). Regarding standardized testing, researchers point to the inability of standardized tests to effectively capture the learning from a specific course of study connected to the student's curriculum. Moreover, the tests were low-stakes exams, typically taken by a small sample of volunteer students, which affects those students' motivation to do well and therefore calls question to the validity of the scores on the test (Banta & Pike, 2012).

This debate highlighted the need for an effective measure of student learning connected to the actual work done and courses taken. It also prompted the U.S. Department of Education to issue a call for proposals to suggest alternatives through its Fund for Improvement of Postsecondary Education (FIPSE). Thus, AAC&U embarked on the Valid Assessment of Learning in Undergraduate Education (VALUE) project (Rhodes, 2010a). The project's goal was to determine if it was feasible to provide colleges and universities with an alternative to standardized tests for assessing student learning outcomes. In response, AAC&U led the development of a collaborative effort with the American Association of State Colleges and Universities (AASCU) and the Association of Public and Land Grant Universities (APLU; at the time, APLU was still known as the National Association of State Universities and Land-Grant Colleges, or NASULGC) to apply for a grant in 2007 (USDOE Office of Postsecondary Education, 2007, n.d.). The successful award to these three organizations was designed to facilitate the development of new tools to measure student learning outcomes as noted in the Spellings Report (USDOE Commission on the Future of Higher Education, 2006), but with input from colleges and universities and, more specifically, their faculty, who at the time had not been included in conversations about the adequacy of ensuring student learning (Lederman, 2007).

AAC&U used the existing faculty-developed and employer-endorsed Essential Learning Outcomes (Figure 1) to facilitate the creation of 16 rubrics to assess student learning using existing student work. Leveraging the expertise of faculty to both create and score assessments of student learning through course-embedded assignments combatted many of the common issues reported with standardized assessments. The resultant rubrics were released in 2009 and are freely available on the AAC&U website (https://www.aacu.org/value-

Figure 1. The Essential Learning Outcomes.

The Essential Learning Outcomes

Beginning in school, and continuing at successively higher levels across their college studies, students can prepare for both responsible citizenship and a global economy by achieving the Essential Learning Outcomes.

★ ★

★ Knowledge of Human Cultures and the Physical and Natural World

- Through study in the sciences and mathematics, social sciences, humanities, histories, languages, and the arts

Focused by engagement with big questions, both contemporary and enduring

★ Intellectual and Practical Skills, including

- Inquiry and analysis
- Critical and creative thinking
- Written and oral communication
- Quantitative literacy
- Information literacy
- Teamwork and problem solving

Practiced extensively, across the curriculum, in the context of progressively more challenging problems, projects, and standards for performance

★ Personal and Social Responsibility, including

- Civic knowledge and engagement—local and global
- Intercultural knowledge and competence
- Ethical reasoning and action
- Foundations and skills for lifelong learning

Anchored through active involvement with diverse communities and real-world challenges

★ Integrative and Applied Learning, including

- Synthesis and advanced accomplishment across general and specialized studies

Demonstrated through the application of knowledge, skills, and responsibilities to new settings and complex problems

Note: This listing was developed through a multiyear dialogue with employers and with hundreds of colleges and universities about needed goals for student learning; analysis of a long series of recommendations and reports from the business community; and analysis of the accreditation requirements for engineering, business, nursing, and teacher education. The findings are documented in previous publications of the Association of American Colleges and Universities: *College Learning for the New Global Century* (National Leadership Council for Liberal Education and America's Promise 2007) and *The LEAP Vision for Learning* (AAC&U 2011). For more information, see www.aacu.org/leap.

rubrics). Between June 2010 and June 2012, just after the initial release of the rubrics, over 17,000 new individuals visited the VALUE website (Rhodes & Finley, 2013).

Although individual programs and institutions began using and experimenting with the VALUE rubric approach in the five years following the rubrics' release, it was not until 2013 that AAC&U was able to take VALUE to scale. In collaboration with the State Higher Education Executive Officers Association (SHEEO), 13 states and their two- and four-year public institutions agreed to submit samples of student work from their curricula to be scored by a trained set of VALUE scorers using three VALUE rubrics (i.e., Critical Thinking, Quantitative Literacy, and Written Communication) to create a broad, nationwide landscape of student learning achievement. Two other collaboratives, the public/private institution Minnesota Collaborative and the private liberal arts institution Great Lakes Colleges Association (GLCA) Collaborative, also joined the VALUE Initiative Collaboratives. This multi-campus faculty and administrative leadership group developed shared protocols with the following goals in mind:

- create a robust system of important learning outcomes focused on authentic student work as evidence and faculty expertise as the arbiter of quality student learning;
- articulate shared standards for student learning;
- create a transparent system for assessment that provides appropriate data and resources to allow participants (e.g., individual institutions, states, consortia) to benchmark learning; and
- explore the potential for this approach to inform state and institutional policy on issues such as student transfer between two- and four-year institutions.

The resultant aggregated VALUE rubric scores provided normed evidence of the quality landscape of student learning across institutions and states for external stakeholders, while also giving faculty helpful information for improving teaching and learning in courses and programs. The ability to take the approach to scale while not losing its power for local pedagogical efforts to improve learning led to the establishment of the VALUE Institute in fall 2017, opening the opportunity to any program, college, or consortium of institutions across the country (www.valueinstituteassessment.org). Indeed, the first year of the VALUE Institute attracted participation from international universities as well. The VALUE Institute continues to provide local information to individual institutions, which builds local capacity to engage more fully in assessment for learning, as well as provides externally valid evidence of quality of learning for accountability purposes (e.g., regional accreditation).

AAC&U recognizes that as the popularity of the rubrics continues to grow, it is necessary to more fully establish the methodological soundness of the VALUE approach to assessment. Included in this paper are methodological elements related to the validity of the VALUE rubrics: (a) best practices for using the rubrics; (b) documented evidence related to their development; and (c) recently collected data related to rubric application, score interpretation, and use that builds on the earlier AAC&U publication, *On Solid Ground* (McConnell & Rhodes, 2017).

Applying the Argument-Based Framework to the VALUE Approach to Assessment: Establishing the Initial Interpretive Argument

As addressed first in *On Solid Ground* (McConnell & Rhodes, 2017), the VALUE approach is philosophically, pedagogically, and methodologically complex. Data generated by scoring with the VALUE rubrics may be considered categorical, quantitative, or qualitative data, depending on the purpose of the end user. This level of flexibility allows faculty—the ultimate arbiters of student learning at the majority of institutions—to align VALUE with their own disciplines' standards of quality and rigor, while at the same time providing institutions

with a common framework with which evidence of learning may be aggregated and considered across disciplinary and programmatic boundaries.

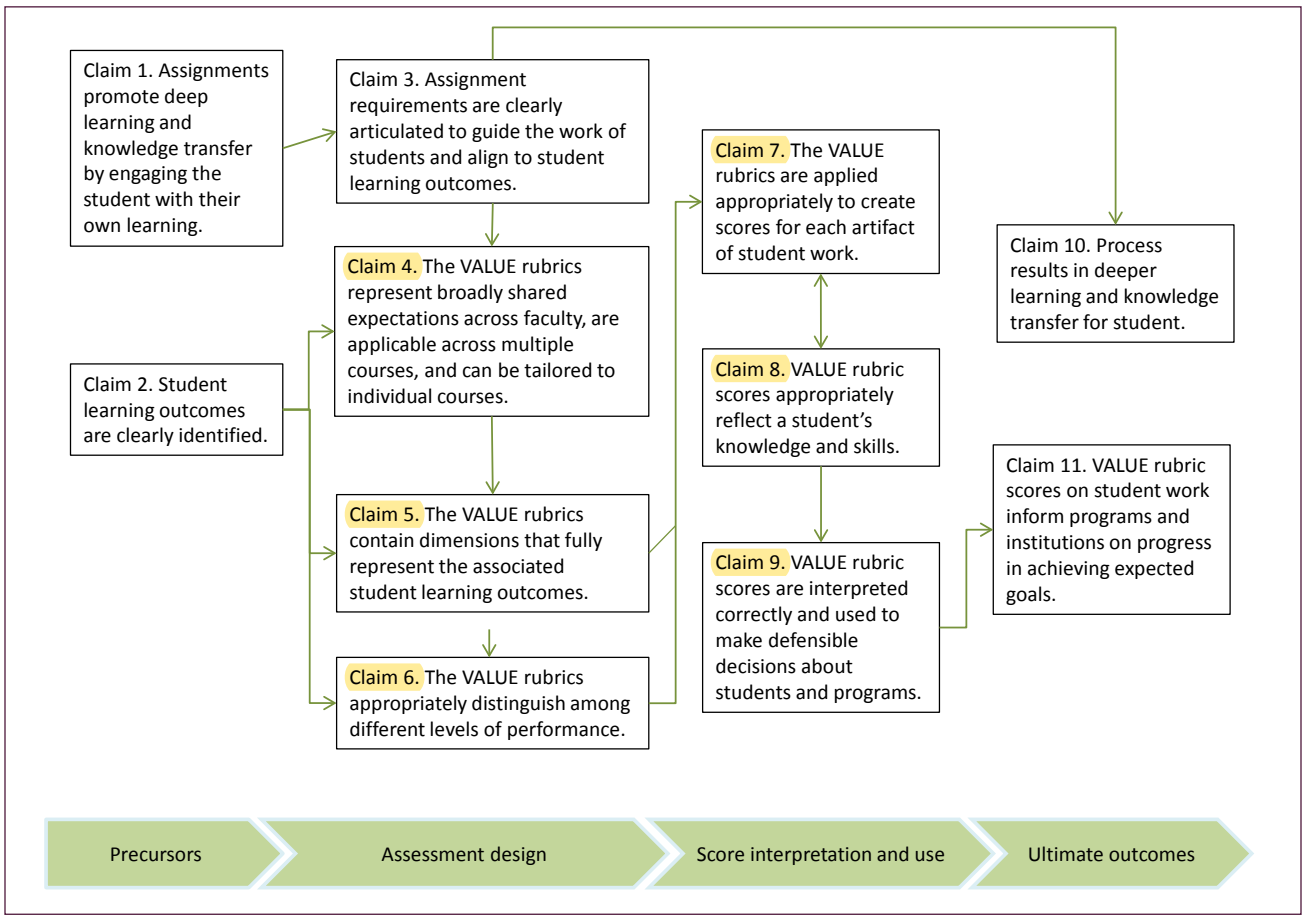
Even though "comparing the validity and reliability of the VALUE process to standardized tests will always be an 'apples to oranges' proposition" (McConnell & Rhodes, 2017, p. 31), it is that very difference—as well as the lack of other comparative models to VALUE—that makes such an exercise of paramount importance. Statistical evidentiary standards from psychological testing and measurement theory may not be a perfect "fit" to this work but leveraging them to help articulate and elucidate the methodological underpinning of VALUE, despite the imperfections, is a necessary and useful component of the work. As such, Perie (2013) drew on validity research in psychological testing and measurement theory (e.g., AERA, APA, NCME, 1999; Kane, 2002) to create a clear plan to evaluate the validity of VALUE, identifying the claims, assumptions, research questions, and evidence needed to satisfy the claims. The resulting interpretive argument (Figure 2) contains 11 claims, covering all aspects of the VALUE approach to assessment, including the ePortfolio process and the rubrics. This interpretive argument is based on the singular goal of VALUE: the improvement of instruction and student learning. As described at the top of each VALUE rubric, "the rubrics are intended for institutional-level use in evaluating and discussing student learning, not for grading." These specifications allowed the rubrics to be useful for colleges and universities to have a common language for discussing student learning outcome attainment. Perie (2013) explicitly cautioned that if institutions engaged in additional applications of the rubrics, such as using them to evaluate a "specific course, professor, or program of study, a different argument would be needed with additional claims" (p. 4). The validity framework presented here assumes the use of the original, unmodified VALUE rubric for assessing students' performance on a learning outcome such as for institution-level, program-level, or course-level assessment. Of course, the rubrics can be, and are, modified for assessment or made more specific to use for grading, but the results derived from these uses should not be used for national comparisons.

It is important to note that for the validity evaluation of VALUE, only claims 4–9 are explicitly addressed, as they focus on the VALUE rubrics themselves rather than elements like assignment/ePortfolio design and institutional factors (e.g., learning outcomes development). That said, it was necessary to refine the interpretive argument to better match the reality of VALUE rubric utilization from local campus implementation to the nationwide scoring efforts, including both the proof-of-concept grant-funded collaborations (e.g., Multi-State Collaborative to Advance Quality Student Learning, Minnesota Collaborative, Great Lakes Colleges Association Collaborative; for details, see McConnell & Rhodes, 2017) conducted from 2014–2017 as well as the outgrowth of that work: the newly instantiated VALUE Institute. Specifically, the interpretive argument had to be refined to reflect that the VALUE data presented here involved the application of the VALUE rubric approach to a sample of individual student work generated by an assignment from a single course and/or an out-of-class, cocurricular experience (e.g., a service learning project) versus a collection of student work over time within an ePortfolio (Rhodes, 2016; Watson, Kuh, Rhodes, Penny Light, & Chen, 2016).

The evidence compiled to support the claims comes from a wide range of sources: (a) the growing body of scholarly literature examining the VALUE rubrics; (b) AAC&U-commissioned research; (c) surveys administered by AAC&U's Office of Quality, Curriculum, and Assessment (OQCA); (d) data generated as part of the grant-funded VALUE collaboratives (2014–2017), including commissioned analyses of the data conducted by consulting methodological experts; and (e) interviews with the key personnel involved in the VALUE rubric design, creation, training, and implementation processes from its inception.

Literature review. The sources within the review of the literature included peer-reviewed journal articles, industry-expert websites, webinars, and occasional white papers (e.g., the National Institute for Learning Outcomes Assessment [NILOA]), along with conference papers, presentations, and communications from

Figure 2. Perie's (2013) interpretive argument for VALUE, revised.



assessment and higher education professionals. It was necessary to consider an array of sources from a broad timeline, including before the creation of VALUE, to comprehensively frame the overall findings. Hence, the literature spans 12 years (2006–2018). By no means exhaustive, the literature review represents a diverse landscape of national and international research on rubrics broadly as well as the specific use of the VALUE rubrics. Though it is beyond the scope of this report to detail the literature review methodology, a brief outline can be found in the Appendix.

AAC&U-commissioned research. In furtherance of its mission "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, 2017b, para. 1), AAC&U regularly commissions research such as member surveys on best practices in undergraduate education (e.g., Hart Research Associates, 2015a, 2015b, 2015c) and surveys of employers to ascertain their perceptions of the preparation of college and university graduates for the workforce (e.g., Hart Research Associates, 2018). Each survey contained questions pertinent to our consideration of the validity of the VALUE approach to assessment.

Surveys administered by OQCA. As the intellectual and administrative home of VALUE, OQCA has administered multiple surveys on the VALUE rubrics and the overall VALUE approach to assessment. For purposes of this report, two surveys in particular are featured: (a) the Institutional Use of the VALUE Rubrics Survey and (b) the 2016–2017 Post-Training and Post-Scoring Surveys from the final year (i.e., *refinement year*) of the VALUE Initiative Collaboratives (Horan, 2018).

Institutional use of the VALUE rubrics survey. In February 2018, AAC&U e-mailed a web-based survey to persons who had accessed one or more of the VALUE rubrics in the past 10 years through the AAC&U website (see www.aacu.org/value-rubrics). At survey closing, 1,448 responses were recorded. The survey aimed to further understand how the VALUE rubrics are used at various institutions, how institutions that use the rubrics view the sufficiency of the rubrics, and how different stakeholders (e.g., administration, faculty, staff) view the utility of the VALUE rubrics. Please note that not all respondents answered all questions, so the number of respondents for questions varied.

2016–2017 refinement year post-training and post-scoring surveys. The summer of 2017 was the end of the three-year VALUE Initiative Collaboratives. The VALUE Initiative involved three grant-funded collaboratives: (a) the SHEEO-facilitated Multi-State Collaborative (see https://www.aacu.org/value/msc) funded by the Gates Foundation, (b) the Minnesota Collaborative funded by the Spencer Foundation, and (c) the GLCA Collaborative funded by the Sherman Fairchild Foundation. Scorers across the country were trained to assess student artifacts using the VALUE rubrics. At the end of scoring, there were 156 scorers for the refinement year. AAC&U administered a survey to trainees and a survey to scorers aimed to better understand issues such as the efficacy of scorer training, the types of assignments that worked well when scoring with the VALUE rubrics, and the degree to which different components of the rubric were sufficient. Note that not all respondents answered all questions, so the number of respondents for questions varied.

VALUE Initiative project-level data and commissioned analyses. Over the course of the three years (2014–2017) of the VALUE Initiative Collaboratives, OQCA compiled a significant amount of data pertinent to the methodological consideration of VALUE, including:

- a database that features more than 20,000 artifacts of scored student work using six of the VALUE rubrics (i.e., Civic Engagement, Critical Thinking, Ethical Reasoning, Intercultural Knowledge and Competence, Quantitative Literacy, and Written Communication), and
- annual Campus Activity Reports from the Minnesota and GLCA Collaboratives delineating how each campus was utilizing the data generated by VALUE scoring as well as information about locally offered faculty development approaches to assignment (re)design and changes in teaching praxis.

Additionally, OQCA contracted with Gary Pike, Professor of Educational Leadership and Policy Studies with the School of Education at Indiana University Bloomington, editor of *The Review of Higher Education,* and former president of the Board of Directors for the Association for Institutional Researchers (AIR), to conduct a generalizability study using the results generated by scoring student artifacts with the Critical Thinking, Quantitative Literacy, and Written Communication VALUE Rubrics. As appropriate, aspects of that study were incorporated as evidence to address aspects of the interpretive argument for VALUE.

Interviews with key personnel. In 2016 and 2017, interviews were conducted with Terrel Rhodes and Wende Morgaine Garrison, who led the VALUE rubric development teams in their original creation. Rhodes continues to be the Executive Director of VALUE as part of his position as Vice President of the Office of Quality, Curriculum, and Assessment at AAC&U. Garrison served as the inaugural Director of VALUE Institute Training and has led training and scoring with the VALUE rubrics for several AAC&U-led initiatives. Rhodes and Garrison organized and facilitated the creation of the VALUE rubrics between 2007 and 2009. Areas of focus within these interviews included understanding how the rubrics were conceived, the processes employed for creating the rubrics, and future trajectories Rhodes and Garrison envisioned for the VALUE rubrics.

In the following sections, each claim and its associated assumptions of the interpretive argument is addressed individually, though it is important to note that the claims are interrelated. Evidence is presented in association with the claim that it most clearly and directly supports, but there is necessarily some overlap between claims. Addressing each claim individually not only allowed for the identification of strengths in the interpretive argument, it helped to reveal gaps as well. As such, certain claims are not as fully supported by evidence as others. While much has been accomplished in the 10 years since VALUE's inception, there are areas ripe for further exploration that would strengthen the argument for VALUE's validity.

PART II

The Interpretive Argument for VALUE: Claims and Assumptions Associated With VALUE Rubric Design and Development

Claim 4: Rubrics Represent Broadly Shared Expectations Across Faculty

4a. The VALUE rubrics were designed with input across faculty from different institutions and disciplines, representing a broad range of experiences and expectations.

4b. The VALUE rubrics can be applied to many courses in a variety of disciplines.

4c. The VALUE rubrics are designed to be easily modified to meet the specifications of a single course.

From the beginning, AAC&U documented the entire VALUE rubric development process, including the names and affiliations of members of the rubric development teams (https://www.aacu.org/development-value-rubrics-rubric-development-teams; e.g., see Appendix B in Hammang, Rhodes, Shulenberger, & Keller, 2010); the names and biographies of the VALUE advisory board (https://www.aacu.org/value/advisory-board); and the e-mails exchanged during the rubric design process. Additionally, a review of the literature supports AAC&U's assertions that the VALUE rubrics can be used to score artifacts of student work from many courses in a variety of disciplines and can be easily modified to meet the unique specifications of a single course.

Assumption 4a: The VALUE Rubrics Were Designed With Input Across Faculty From Different Institutions and Disciplines, Representing a Broad Range of Experiences and Expectations

The decision was made to create a rubric to correspond with each of the Essential Learning Outcomes and the critical knowledge, skills, abilities, and dispositions, which AAC&U identified from the field through its member and employer research endeavors as part of AAC&U's signature initiative Liberal Education and America's Promise (LEAP). In short, the VALUE rubrics were designed to serve as an effective operational tool for measuring the essential learning outcomes (e.g., see Appendix A in Hammang et al., 2010). The first step in the VALUE rubric development process was to call on persons and campuses already familiar with using rubrics for assessment to participate. In 2007, AAC&U chose 12 VALUE Leadership Campuses (https://www.aacu.org/value/leadership-campuses), who shared use of rubrics for assessment and utilization of student ePortfolios to serve in a consultant capacity to the initial 15 (and later 16) VALUE rubric development teams. The development teams were interdisciplinary in nature and included traditional academic faculty from across the arts, humanities, social sciences, natural sciences, and professional programs as well as—when appropriate—assessment, teaching and learning, and cocurricular administrators.

Teams were composed of individuals representing the diversity of AAC&U's membership as well as the diversity of American higher education writ large (e.g., community colleges, public and private four-year institutions). Participation on a VALUE rubric development team was entirely voluntary and was not associated with any form of compensation. The VALUE rubric development teams engaged in a rigorous rubric development process that included reviewing existing campus rubrics, going through at least three rounds of drafting, testing with the 12 VALUE Leadership Campuses, testing with over 100 VALUE Partner campuses (https://www.aacu.org/value/partner-campuses), and redrafting to create the final VALUE rubrics

(Hammang et al., 2010; Rhodes, 2010a). The resulting VALUE rubrics were originally referred to as *meta-rubrics*, as they were framed to consider evidence of learning at the programmatic or institutional level as their unit of analysis, versus more traditional rubrics created to facilitate grading within a single course (Stevens & Levi, 2013). The language of the VALUE rubrics reflects that they are meta-rubrics, with word choices explicitly made to facilitate their use within and across multiple disciplines.

All rubrics consist of the same structure, including: (a) the learning outcome title, (b) the operational definition of the learning outcome, (c) the five or six dimensions of the learning outcome that are explicitly measured by the VALUE rubric, (d) framing language generated by the VALUE rubric development team to provide insight into the guiding principles and/or intended use of each VALUE rubric (e.g., assessing a learning process versus the end product of an activity), (e) a glossary of key terms used in the VALUE rubric, (f) performance levels, and (g) performance descriptors (see https://www.aacu.org/parts-value-rubric). The unmodified VALUE rubrics assume mastery of learning outcomes is achieved with the award of a bachelor's degree (for more information regarding the performance-levels, see Rhodes & Finley, 2013, p. 6).

Assumption 4b: The VALUE Rubrics Can Be Applied to Many Courses in a Variety of Disciplines

A review of the available literature (2006–2018) supports the assertation that the VALUE rubrics broadly encompass the shared skills and knowledge expectations that educators and employers have agreed are necessary for engaged participation in the global economy upon graduation (Adler-Kassner & Harrington, 2010; Barnhisel, Stoddard, & Gorman, 2012; Brown & Rhodes, 2016; Cicek et al., 2015; Eppes, Milanovic, & Sweitzer, 2012; Eynon, Gambino, & Török, 2014a; Hachtmann, Mitchell, Mitchell, & Sasser, 2013; Hammang et al., 2010; Hovland, 2014; Kuh & Ikenberry, 2009; Laguilles, Coughlin, Kelly, & Walters, 2016; Maki, 2015; McConnell, 2013; McNair, 2016; Rhodes, 2008, 2010b, 2011; Rhodes & Finley, 2013; Timmerman, Strickland, Johnson, & Payne, 2011). Universities and colleges varying in size, type, and location have benefited from using rubrics broadly and the VALUE rubrics specifically, demonstrating that VALUE provided meaningful, actionable feedback to faculty while simultaneously meeting the demands of accreditation organizations that seek measurable demonstrations of authentic learning (Carey, 2011; Cicek et al., 2015; Graff, Horvath, Shapiro, & Wood, 2018; Hubert, 2016; Keen, 2009; Kiel, Burclaff, & Johnson, 2015; Oderda et al., 2010; Sullivan, 2015).

There is evidence that by rooting institutional and programmatic assurances in common learning, the VALUE approach to assessment helps to translate evidence of academic quality cohesively and consistently across higher education institutions and stakeholders. Furthermore, while the critical antecedent requirement of defining learning outcomes for institutional and programmatic pedagogy takes considerable time, particularly when led by faculty, the evolving, iterative maturation of these processes has been noted as being a worthwhile exercise (Blaszczynski & Green, 2012; Clark & Eynon, 2011; Coates, 2015; Cooney, 2014; Deardorff, 2011; Escoe, Hall, & Nicholas, 2010; Goff, Potter, Pierre, Carey, & Gullage, 2015; Hatcher, 2011; Hecsh, 2010; Komvies & Smedick, 2012; Mentkowski & Sharkey, 2011; Thyer, 2015). Integrating VALUE from the top-down through the institutional and/or programmatic level, and the bottom-up through individual faculty leveraging the rubrics as part of teaching and learning within a course, can further institutionalize improvement and accountability efforts. This can be achieved by aligning course learning outcomes as embedded assignments with consideration of content, context, students' prior knowledge, and pedagogy, and using the assessment findings to communicate areas of strength across departments and offices (Bergfelt, Brungard, & Rapchak, 2016; McConnell & Doolittle, 2012; Norman, 2017; Stanley, 2016; Sullivan & McConnell, 2017). Faculty, administrators, and staff continue to embed their curricula and assess student learning with dimensions from the VALUE rubrics, thus enabling localized results to be placed in a larger, comparable context within and beyond their own institutions to cultivate external validation of instructional modalities and mastery of

scholarship (Eynon, Gambino, & Török, 2014b; Jankowski, Timmer, Kinzie, & Kuh, 2018; Shea & Turner, 2013; Stupans et al., 2015).

Using the VALUE rubrics on campuses not only inspires dialog about assessment as a means to provide data to accreditors, but also demonstrates the dexterity of the VALUE rubrics for pedagogical innovation (Albertine, 2015; Carter & McConnell, 2018; Graff et al., 2018; Hovland, 2013; Peet, Reynolds-Keefer, & Gurin, 2011). Extant literature also suggests that VALUE rubric-inspired professional development has advanced campus teaching, learning, and assessment praxis. VALUE-related topics addressed by faculty development include (a) addressing discipline-specific needs (Albertine, 2014; Cooney, 2014; Evans, Van Epps, Smith, Matei, & Garcia, 2015; Kruse & Drews, 2013; Stufft, 2013), (b) developing assessment plans (Cicek et al., 2015; Liu, 2017; Reason, 2013; Schlitz et al., 2009; Shada, Cox, & Malik, 2011), and (c) effectively communicating assessment results (Hundley, 2017; Provezis, 2012; Steinke & Fitch, 2011). Additionally, VALUE is credited with inspiring course and assignment (re)design to promote more effective alignment with the VALUE rubrics (Acosta-Morales, McCool, Murphey, & Stephens, 2016; Augsburg, 2012; Baxter, 2013; Belanger, Zou, Mills, Holmes, & Oakleaf, 2015; Bernstein & Greenhoot, 2014; Boersma, Diefenderfer, Dingman, & Madison, 2011; Braid, de Schrynemakers, & Grose, 2011; Cratsley & Slotnick, 2013; Felten, Barnett, & Fuller, 2014; Garritano & Yatcilla, 2014; Gessell, 2015; Gleason et al., 2013; Gray, 2013; Greenhoot & Bernstein, 2011; Jardeleza et al. 2013; Klein-Collins, 2013; Lee, 2012; Oakleaf, 2011; Sorey & DeMarte, 2013). The literature documents how VALUE assists in highlighting curricular strengths and diagnosing weaknesses, therefore informing improvements in teaching and learning within and across programs (Augsburg, 2012; Crist & Robinson, 2015; Holliday et al., 2015; Penny Light, Chen, & Ittelson, 2011; Richards-Schuster, Ruffolo, Nicoll, Distelrath, & Galura, 2014; Williams, 2013).

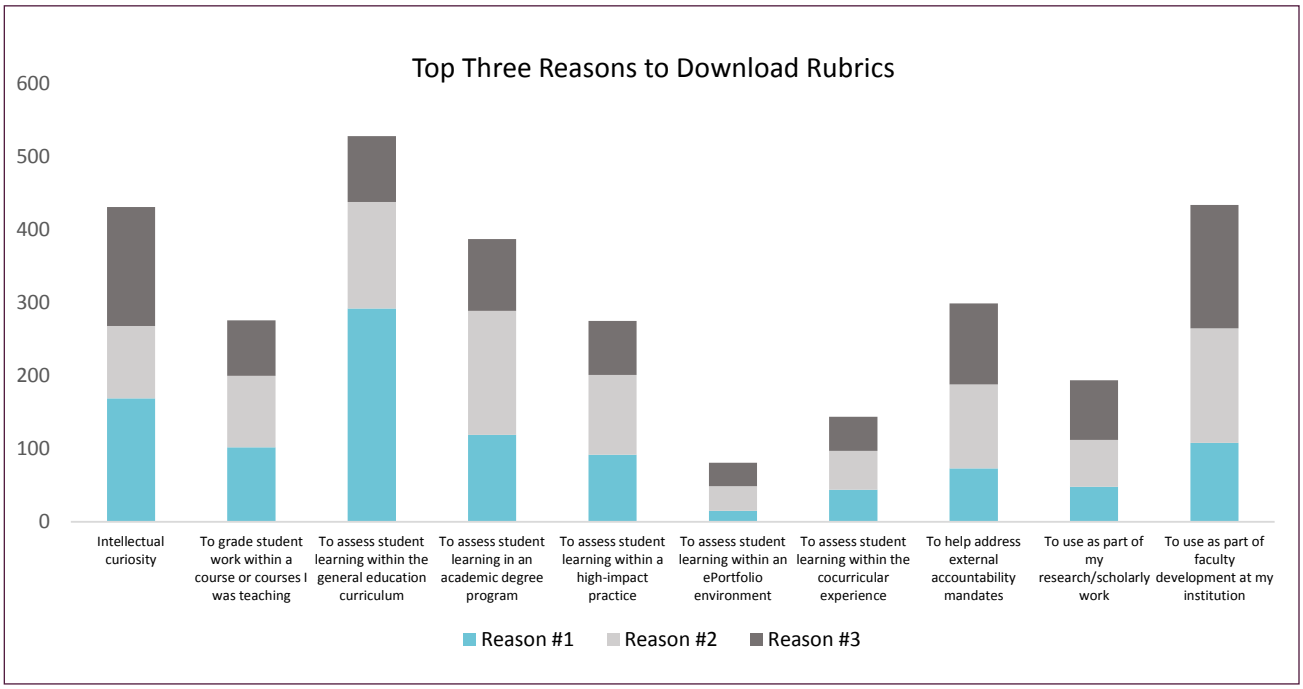
VALUE professional development also appears to promote collaboration across professional roles in college and university learning spaces (i.e., collaboration among faculty, students, institutional researchers, assessment staff, etc.), and provides opportunities to bridge academic disciplines as well as the curricular and the cocurricular (Berg et al., 2014; Carroll, Tchangalova, & Harrington, 2016; Carroll & Traver, 2016; Carter & Purkiss, 2016; Chamberlin, 2018; Cooney, 2014; Graff et al., 2018; Healy, Lancaster, Liddell, & Stewart, 2012; Johnson, Veitch, & Dewiyanti, 2015; Rapchak, Lewis, Motyka, & Balmert, 2015; Woodward, 2015). VALUE rubric assessments transcend the shortcomings typically associated with grading and standardized tests (Banta & Blaich, 2011; Berrett, 2013; Dingman & Madison, 2011; Greenhoot & Bernstein, 2011; Pusecker et al., 2011; Rhodes & Finley, 2013). Concerns over consistently and reliably interpreting grades (due to grade inflation) and static test values can be overcome by utilizing the VALUE rubrics at multiple points in the assessment process (Berrett, 2016).

While the literature does support the assertion that VALUE rubrics can be applied to many courses in different disciplines, it also strongly signals a common area needing improvement: the assignment (re)design process, specifically the creation of explicit prompts containing transparent expectations in the assignment instructions by using the language of the VALUE rubrics. Faculty and administrators find it beneficial to share the rubric ahead of time with students to reiterate expectations. In addition, if rubric language is incorporated into instructional prompts, faculty are equipped with a two-pronged method of justifying and connecting their evaluation to the demonstrated learning submitted by the student (Hodges, 2014; Reddy & Andrade, 2010). This is an aspect of the work that will be addressed in greater detail in subsequent claims (Burke, Ouellette, Miller, Leise, & Utschig, 2012; Darabi & Arrington, 2017; Heinrich, Habron, Johnson, & Goralnik, 2015; Jansen, Mihlbauer, & Mayeshiba, 2016; Kinzie, Helms, & Cole, 2017; Madison, 2012; Overby & Meiman, 2013; Steinberg, 2011; Sullivan & McConnell, 2017).

In addition, results from AAC&U's 2018 Institutional Use of VALUE Rubrics Survey support the assertion that the VALUE rubrics can be applied to many courses in a variety of disciplines. The AAC&U Institutional

Use Survey asked respondents to identify their top three reasons for downloading the VALUE rubrics from a list of 10 possible reasons (Figure 3). The top reasons respondents selected were (1) to assess student learning within the general education curriculum (n = 528), (2) to use as part of faculty development at my institution (n = 434), and (3) intellectual curiosity (n = 431). Respondents who chose to add other reasons included (1) developing rubrics or other assessment materials, (2) learning outcome development, (3) student development, and (4) assessment (Figure 4).

Figure 3. Respondents selected their reasons for downloading the VALUE rubrics in order of first, second, and third.



The variety of reported reasons for downloading the rubrics signals the versatility and broad applicability of the VALUE rubrics across disciplines and uses. Furthermore, the fact that general education assessment is the top reason given for downloading and using the VALUE rubrics is a critical piece of evidence to the claim that the VALUE rubrics can be applied to many courses in a variety of disciplines. By design, general education curricula provide students with—at minimum—exposure to a range of disciplines and fields, often organized around cross-cutting learning outcomes like critical thinking (Carter & McConnell, 2018; Hoy & Wolfe, 2016; Hubert, 2018; McCormack et al., 2012; Penn, 2011a, 2011b; Rather, 2016; Self, Cady, & Olson, 2014; Traver & Nedd, 2018; Tunstall, Matz, & Craig, 2018; Wyman, Waldo, & Doherty, 2016). General education courses are frequently cited as providing an ideal environment for students to cultivate higher-order learning, self-assessment, and reasoning development, which are all aspects of the VALUE rubrics (Anderson & Blair, 2013; Core, 2017; Dingman & Madison, 2011; Duncan & Gleason, 2015; Fitch, Steinke, & Hudson, 2012; Hubert, 2013; Hughes & Jones, 2011; Lee et al., 2018; McClure, Brenchley, & Donahue, 2017; Penny Light, Chen, & Ittelson, 2011; Reason & Hemer, 2014; Roehling, 2018; Simper, 2018; Wozniak, 2012). The VALUE rubrics can serve as a common framework for translating the learning that transpires within courses from different disciplines that have been included in a general education program to satisfy the same requirement(s).

Respondents were also asked about their role/position at their institution and who was primarily using/working with the VALUE rubrics at their institution (Figure 5). These questions are valuable for understanding who is accessing the rubrics and who is using the rubrics. The vast majority of both those downloading

Figure 4. Respondents who selected "other" reasons for downloading the rubric provided open-ended responses. Their responses were coded for themes, which are presented in this graph.

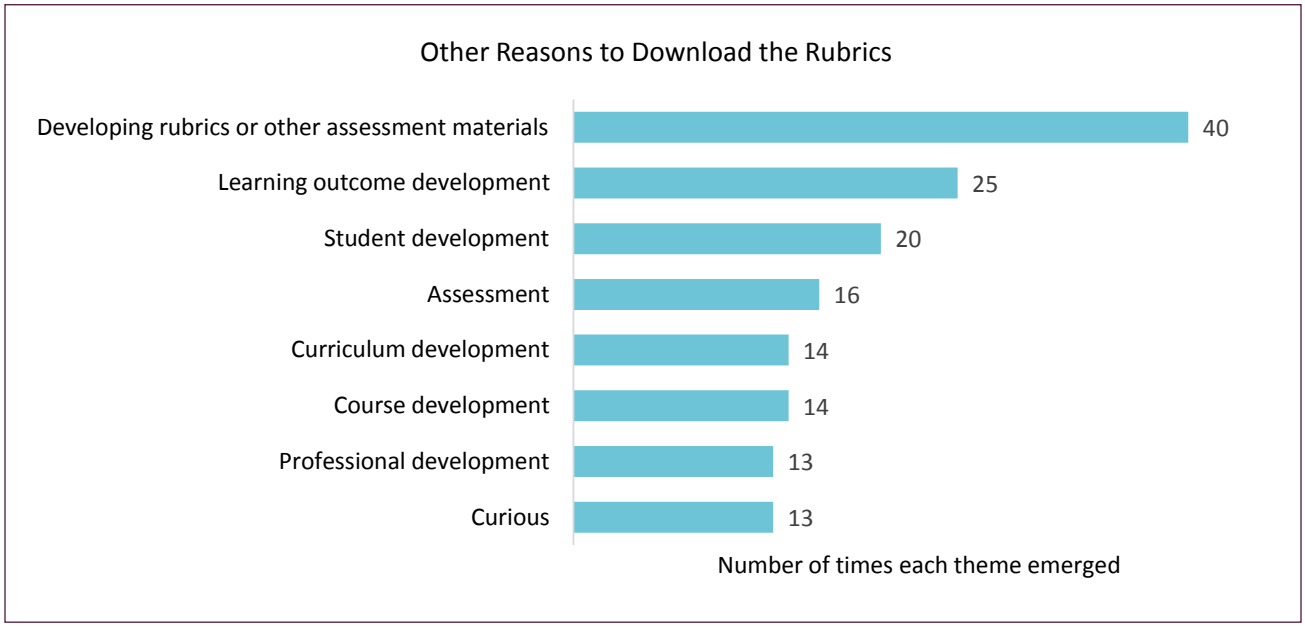
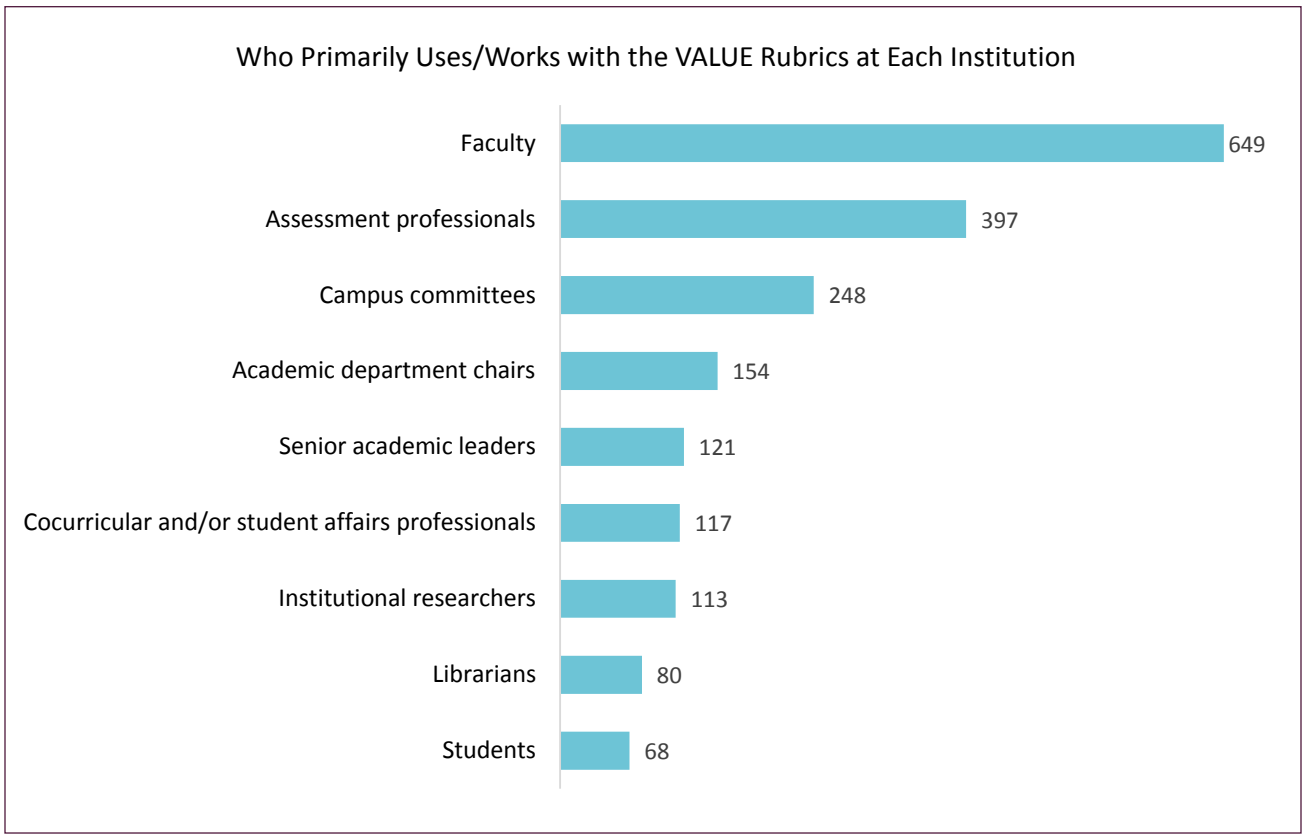


Figure 5. Answers to the question, "Who is primarily using/working with the VALUE rubrics at your institution? Please check all that apply."



the VALUE rubrics directly as well as the identified primary users of the VALUE rubrics on a campus are faculty, signaling the broad applicability of the rubrics across courses and disciplines. Additionally, the varied responses show the utility of the rubrics for a variety of personnel across a college or university campus.

Assumption 4c: The VALUE Rubrics Are Designed to Be Easily Modified to Meet Specifications of a Single Course

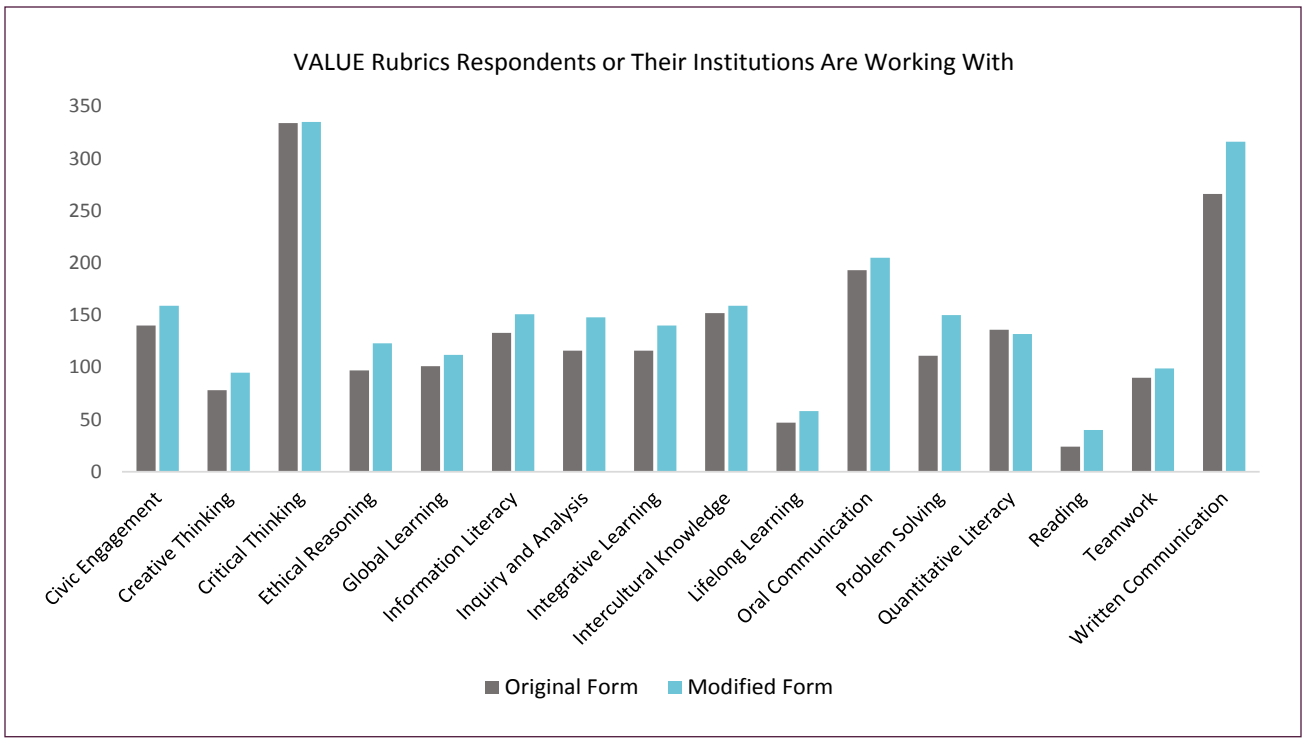
From the moment the VALUE rubrics were released to the higher education community, AAC&U promoted local adoption of the tools by encouraging campuses to modify the rubrics to best meet the needs of their institution's culture and context. The VALUE rubrics themselves are available for download as PDFs and as editable Microsoft Word documents.

A review of the literature provides a sense of the nature and scope of modifications that users have made to the VALUE rubrics. The performance levels and descriptors, number of dimensions, and the associated numeric level values—in original and modified forms—are all noted as easy to use and helpful during assessment. When modified, these rubric elements are collapsed, expanded, and redefined in ways that mirror the distinct intentions of the assessor (please refer to Claim 7 for more information about calibration training and scoring best practices), thus guiding the capacity of the end-user (assessor or student) to interpret the results for local improvements (Osadjan, Jackson, Jallah, & Joel, 2015; Rhodes, Bischof, Milner, Hines, & McNealy, 2014; Weaver et al., 2016; Whitehead, 2015). Examples of VALUE rubric modifications from the literature include (a) reconfiguring the Quantitative Literacy performance levels to better assess learning outcomes in a quantitative reasoning course (Boersma, Diefenderfer, Dingman, & Madison, 2011); (b) redefining performance levels in the Quantitative Literacy and Critical Thinking VALUE Rubrics when used for a comparative study (Hebble & Richards, 2013); (c) reversing performance levels and reformulating descriptors to assess engineering graduate outcomes (Cicek et al., 2015); (d) expanding the dimensions and changing the descriptors in the Information Literacy VALUE Rubric to assess ACRL standards (Kiel et al., 2015); (e) employing factor analysis to incorporate performance descriptors from the Oral Communication VALUE Rubric into another modified rubric (Baryla, Shelley, & Trainor, 2012); (f) assessing work using modified descriptors for the Intercultural Knowledge and Competency, Oral Communication, and Teamwork VALUE Rubrics as part of a study comparing student outcome performance once a curriculum change was made (Bernstein, Strasma, Olwell, & Higbee, 2017); and (g) combining language from the Civic Engagement and Intercultural Knowledge and Competence VALUE Rubrics including altering performance levels to measure social responsibility, as defined by an outside source (Carter & Purkiss, 2016).

Most VALUE rubric references in the literature described making modifications for specific grading of assignments or course-level descriptors for more granular use, but it is important to note that while the VALUE rubrics can be modified for grading purposes, it was never the intent of VALUE to primarily use the rubrics for this purpose (Rhodes & Finley, 2013). As adoption of the VALUE rubrics expanded, a new iteration of rubric modification became central to the promotion of VALUE at "disciplinary, or even classroom levels" (Rhodes, 2011, p. 4; see also Flierl, Bonem, Maybee, & Fundator, 2018; Liu, Frankel, & Roohr, 2014; Mandeville & Stoner, 2015; McConnell & Doolittle, 2012; O'Brien, 2013; Reed, Garcia, Slusser, Konowitz, & Yep, 2017; Riebe & Jackson, 2014). Disciplinary versions of the VALUE rubrics are of increasing interest to classroom faculty as well as disciplinary societies, particularly those who participated in tuning efforts associated with work done on the Degree Qualifications Profile (Lumina Foundation for Education, 2011).

The Institutional Use Survey included several questions related to modifying the VALUE rubrics for specific institutional needs. Figure 6 shows the rubrics that respondents indicated they used in either original or modified form. The most utilized VALUE rubrics in either original or modified form were Critical Thinking, Written Communication, and Oral Communication. The Reading and Foundations and Skills for Lifelong

Figure 6. Responses from respondents indicating which VALUE rubrics they are currently working with in either their original or modified form.



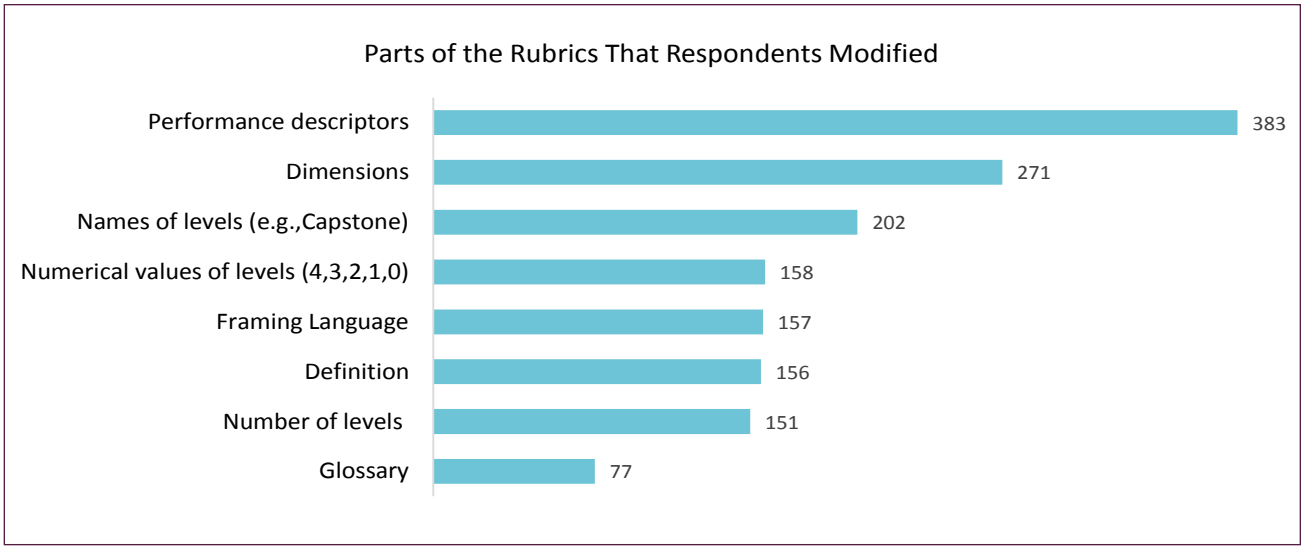
Learning VALUE Rubrics were the least used rubrics, whether in original or modified form. This is interesting to note because during the development of the rubrics, Reading was only added to the body of rubrics at the behest of institutions requesting a tool to assess reading. It is also interesting to note that the Written Communication and Problem Solving VALUE Rubrics differ the most in their use in original form and modified form. These differences may indicate issues or disciplinary preferences with these rubrics, which leads to institutions having to modify them for their own needs.

The Institutional Use Survey asked respondents to articulate their rationale for modifying the VALUE rubrics. The following are example responses to this open-ended question that signal the diversity of the very specific local needs and contexts, which led to VALUE rubric modification: (a) "An instructor needed a graduate student rubric for ethical problem-solving course and wanted to emphasize writing with the critical thinking process. So she adapted the ethical problem solving rubric to her graduate course"; (b) "I modify to connect to my specific course learning outcomes. to increase specificity; to integrate multiple dimensions for various assignments"; (c) "We assess several outcomes at a time, and it's more practical for us to have an integrated rubric"; (d) "A specific rubric for assessing art education graduate students' proposals in the 'Research in Art Education' course was not available. Language from various VALUE rubric dimensions and performance descriptors was adapted and/or combined with content-specific language to create an instrument"; and (e) "They were modified based in intended use and experience. For example, the Inquiry & Analysis Rubric was amended slightly to account for annual report submissions from academic departments. We liked the idea of using the 'same' rubric to apply to faculty work as we were applying to student work."

The Institutional Use Survey also asked respondents specific questions related to their modification of the rubrics, including (a) which parts of the rubric they modified, (b) the degree to which they modified the rubrics, and (c) their justification for modifying the rubrics. Overall, the most frequently modified portion was the performance descriptors, which is unsurprising given the centrality of the VALUE rubric performance

descriptors (Figure 7). Respondents typically cited the need to modify for purposes of grading assignments. Many institutions also cited blending and merging different rubrics together to create more comprehensive rubrics. An interesting modification involved changing the layout of the scores from high to low (e.g., the rubric starts with a score of 4 [*capstone*] at the left) into low to high. This layout was an intentional choice in the development of the VALUE rubrics, as the rubrics were designed to reflect an assets-based versus deficit-model construction (McConnell & Rhodes, 2017). In practice, AAC&U encourages scorers to start the scoring process by focusing on the components that are present, rather than looking for missing/inadequate components of the student work.

Figure 7. Selected responses to the survey question, "Which parts of the rubric have you modified?"



Respondents were also given the opportunity to describe the nature of the modifications that they made to the VALUE rubrics, as well as the degree of the modifications (i.e., the extent to which the revised VALUE rubrics in use deviated from the original). Example responses include, (a) "We have integrated critical thinking and information literacy into oral communication, written communication, and quantitative reasoning"; (b) "Primary use as resource for assignment design"; (c) "used to inform an integrated learning framework"; (d) "This is for assignments: I added in additional levels for grading purposes"; (e) "Used modified rubrics in class as a vehicle for discussion about what is important in written and oral communication. Have not used modified rubrics outside of singular class discussion/assignments"; (f) "We took the skills we want students to express to employers and used descriptions based on those used in each rubric. It helped us take several rubrics and make one we could use in a mock interview. This is used only in the Career Center"; and (g) "in the MSEd program we modified the rubrics to fit a Graduate Level experience where the student has full-time employment outside of the graduate program. The main rubric that is utilized is the Integrative Learning Rubric to help candidates tie their 'real life experiences' with their 'in the classroom (online)' experience".

In summary, those who have downloaded the VALUE rubrics on average see utility in both the original and modified versions for use within a single course. VALUE rubric modifications are made to enhance the applicability of the VALUE rubric to a particular course or assignment, or to facilitate faculty adoption and use. Interestingly, multiple responses indicated that revisions were made to translate a VALUE rubric—designed as a tool to assess undergraduate student learning—into a tool that would be appropriate for post-baccalaureate learning assessment. Again, balancing the ease with which the VALUE rubrics may be modified

and customized to assess learning in a narrowly defined context (e.g., a single assignment or course) with the desirability to maintain the integrity of the VALUE rubric in order to facilitate cross-course, program, or institutional conversations about learning—including but not limited to benchmarking activities—is something that each individual user must consider in their adoption of the VALUE approach to assessment.

Claim 5: The VALUE Rubrics Include Dimensions That Fully Represent the Associated Student Learning Outcomes

5a. For each of the 16 rubrics, the dimensions selected identify the components for that learning outcome.

Determining the validity of the VALUE rubrics entails ensuring the rubrics encompass an inclusive set of student learning outcomes that allow institutions to adequately and fully assess student learning. Assumption 5a asks whether the 16 VALUE rubrics encompass the learning outcomes we should be assessing and, going further, whether the dimensions within those rubrics fully represent the learning outcomes they aim to describe.

Assumption 5a: For Each of the 16 Rubrics, the Dimensions Selected Identify the Components for That Learning Outcome

The decision about which learning outcomes to address within the VALUE rubric creation process started with Terrel Rhodes, AAC&U Vice President in the Office of Quality, Curriculum, and Assessment and Executive Director of VALUE. Immediately before coming to AAC&U, Rhodes served in academic leadership and as a faculty member at Portland State University, where he became familiar with their rubric approach to assessing student learning outcomes using cumulative student portfolios. The expertise Rhodes gained at Portland State not only informed his ability to implement rubric-based assessment of student learning outcomes but also expanded his network of colleagues with expertise in rubric based assessment. Wende Morgaine Garrison, VALUE Initiative Manager from 2008–2010 and current Director of VALUE Institute Training, and Rhodes were colleagues at Portland State. Rhodes recruited Garrison to facilitate the rubric development teams. The expertise of Rhodes and Garrison were key to the efficient and successful development of the VALUE rubrics.

Initially AAC&U planned to create 14 rubrics based on the LEAP Essential Learning Outcomes (https://www.aacu.org/leap/essential-learning-outcomes). As the rubric development process was underway, AAC&U heard from leaders at several institutions that they were primarily concerned with students' reading skills. In response, AAC&U added the Reading VALUE Rubric for a total of 15 original VALUE rubrics. The final VALUE rubric (i.e., Global Learning) was developed later under the auspices of AAC&U's national, multi-project initiative Shared Futures: Global Learning and Social Responsibility (Anderson & Blair, 2013).

The VALUE rubrics were developed by teams of volunteer experts who thoughtfully deliberated to determine each rubric's dimensions and performance descriptors (Anderson & Blair, 2013; Rhodes, 2010a). As discussed earlier in Claim 4, all teams consulted existing campus rubrics to determine the expectations for each learning outcome that were common across institutions (Rhodes, 2010a); each rubric went through at least three rounds of revisions. Before release, the rubrics were tested by faculty in their courses at the VALUE Leadership and Partner Campuses and over 100 other campuses that volunteered to provide feedback to the development teams.

Appropriateness and Completeness of the Dimensions Vis-à-Vis the Learning Outcomes

In 2018, AAC&U asked respondents in the Institutional Use Survey how they first heard about the rubrics. Their responses are shown in Figures 8 and 9. The variety of original reasons for encountering the rubrics, like the variety of positions of those who download the rubrics described in Claim 4, speaks to the versatility and applicability of the rubrics for various reasons and courses. The top two ways respondents first heard about the VALUE rubrics were (a) from colleagues in assessment and/or institutional research at their institutions and (b) from faculty colleagues. The sharing of the VALUE rubrics among colleagues signals evidence of not only traditional notions about face and use validity (e.g., Rhodes & Finley, 2013) but also content validity, as the faculty encouraging others to use the rubrics were experts in their own fields.

The Institutional Use Survey also asked respondents how their campus was using the VALUE rubric(s), either in original form or modified form. The top choices selected were to assess student learning in (a) general education, (b) academic degree programs/majors, and (c) writing intensive experiences (Figure 10). Accompanying open-ended responses highlighted the VALUE rubrics' utility for course and program assessment, student leadership, rubric development, implementing high-impact practices, and faculty development. Respondents were also asked to select from eight possible choices if their institution was engaged with certain practices using the VALUE rubrics. As shown in Figure 11, the most commonly selected modes of engagement were (a) faculty development, (b) assignment (re)design, and (c) conducting VALUE rubric calibration training. As with the other responses, the variety of uses beyond simply using rubrics to assess student learning outcomes shows the applicability of the rubrics for many courses and uses.

As with the questions focused on how respondents first heard about the VALUE rubrics, the range and diversity of the uses of and the practices employed with the VALUE rubrics further demonstrates the face and

Figure 8. Responses to the survey question, "To the best of your recollection, how did you first hear about the VALUE rubrics?"

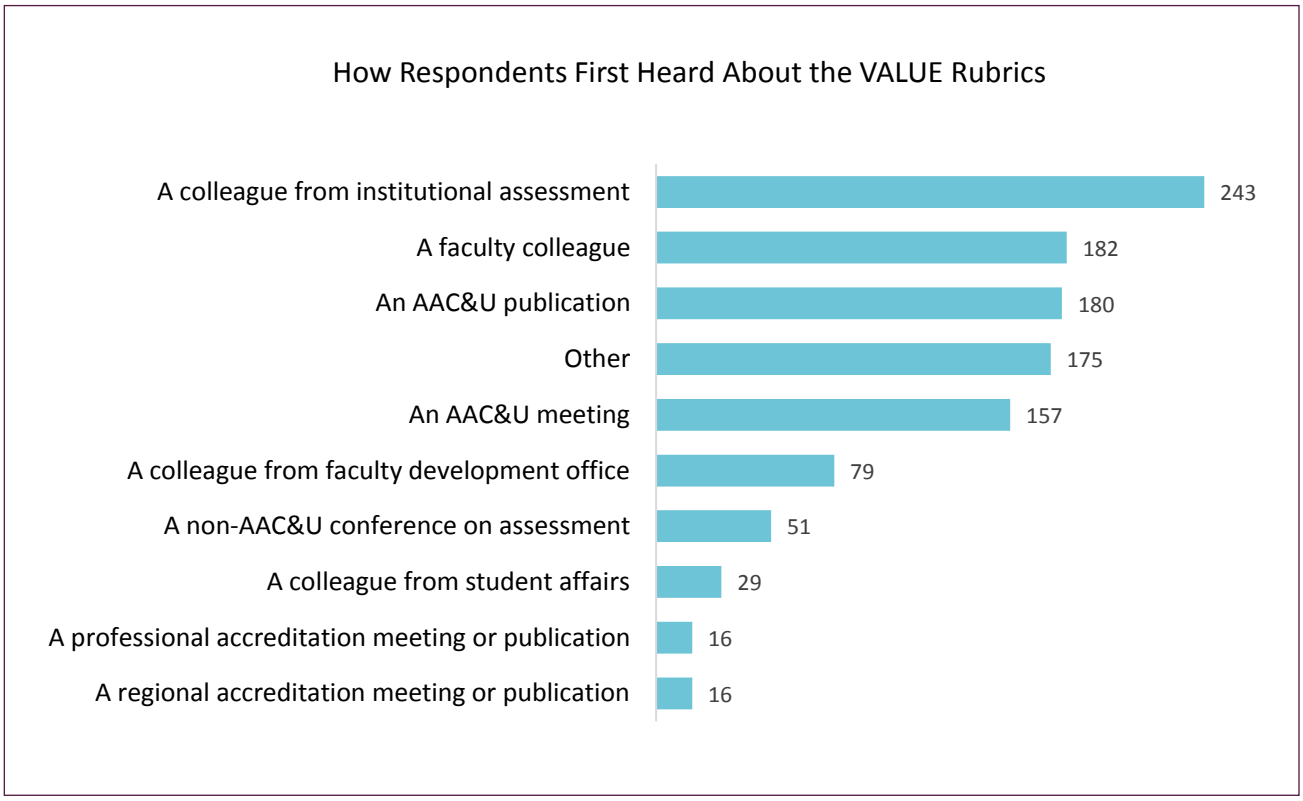


Figure 9. Respondents who selected "other" as how they first heard about the VALUE rubrics. These responses were coded for themes, which are presented in this graph.

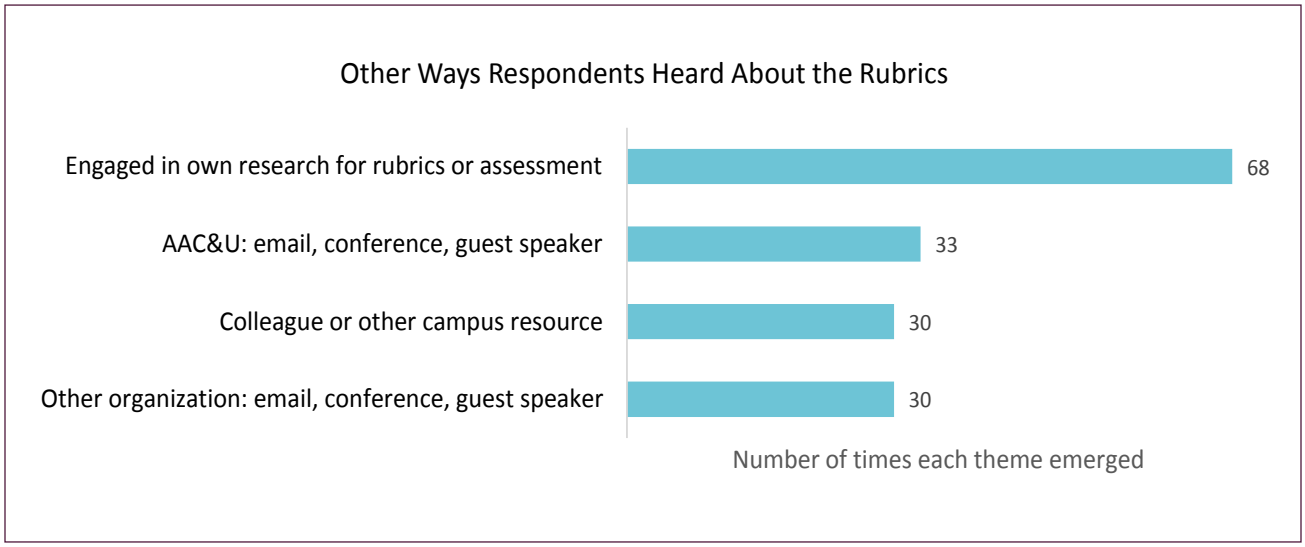


Figure 10. Responses to the survey question about how their institutions are currently using any of the VALUE rubrics (either in their original form or modified for campus needs) to assess student learning.

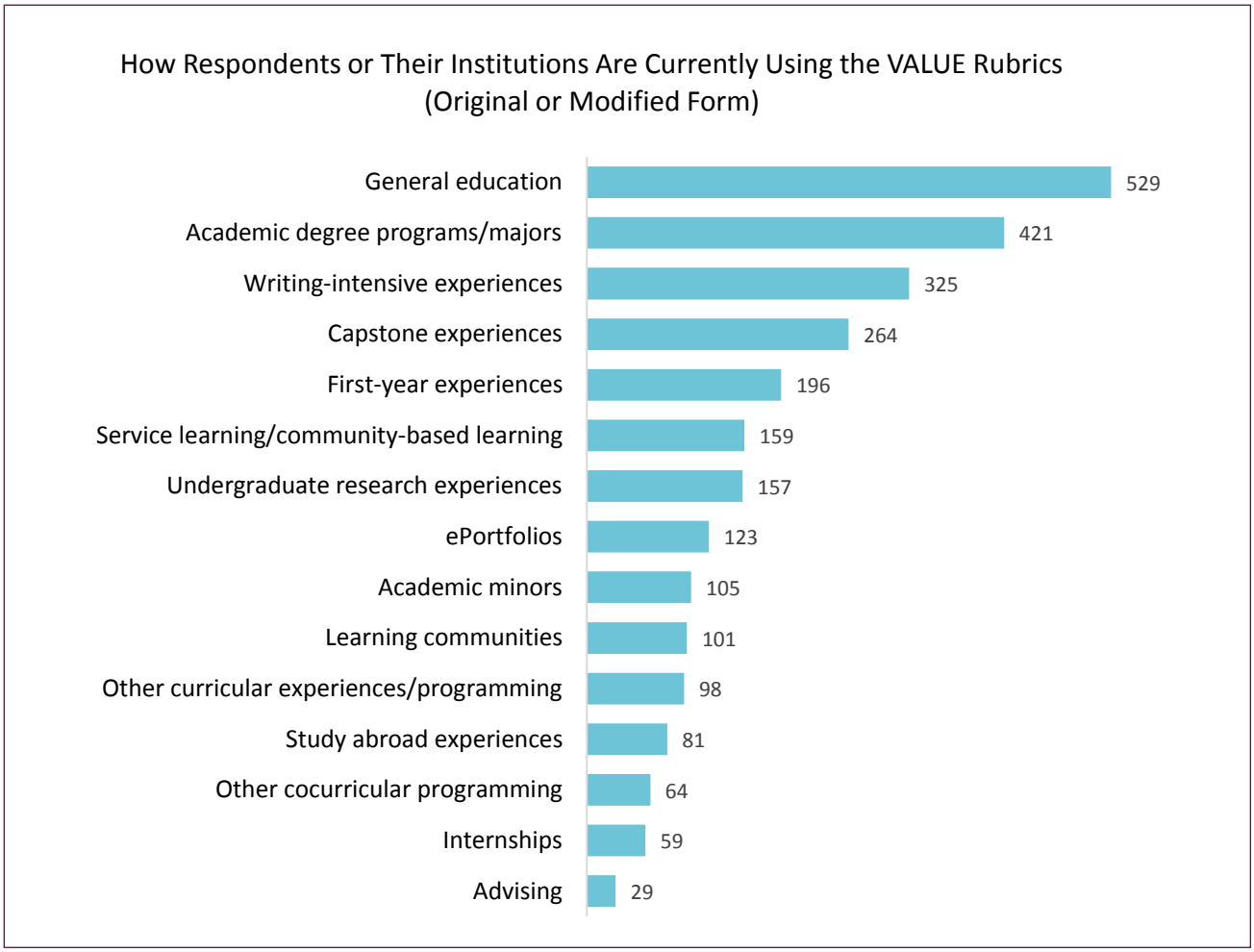
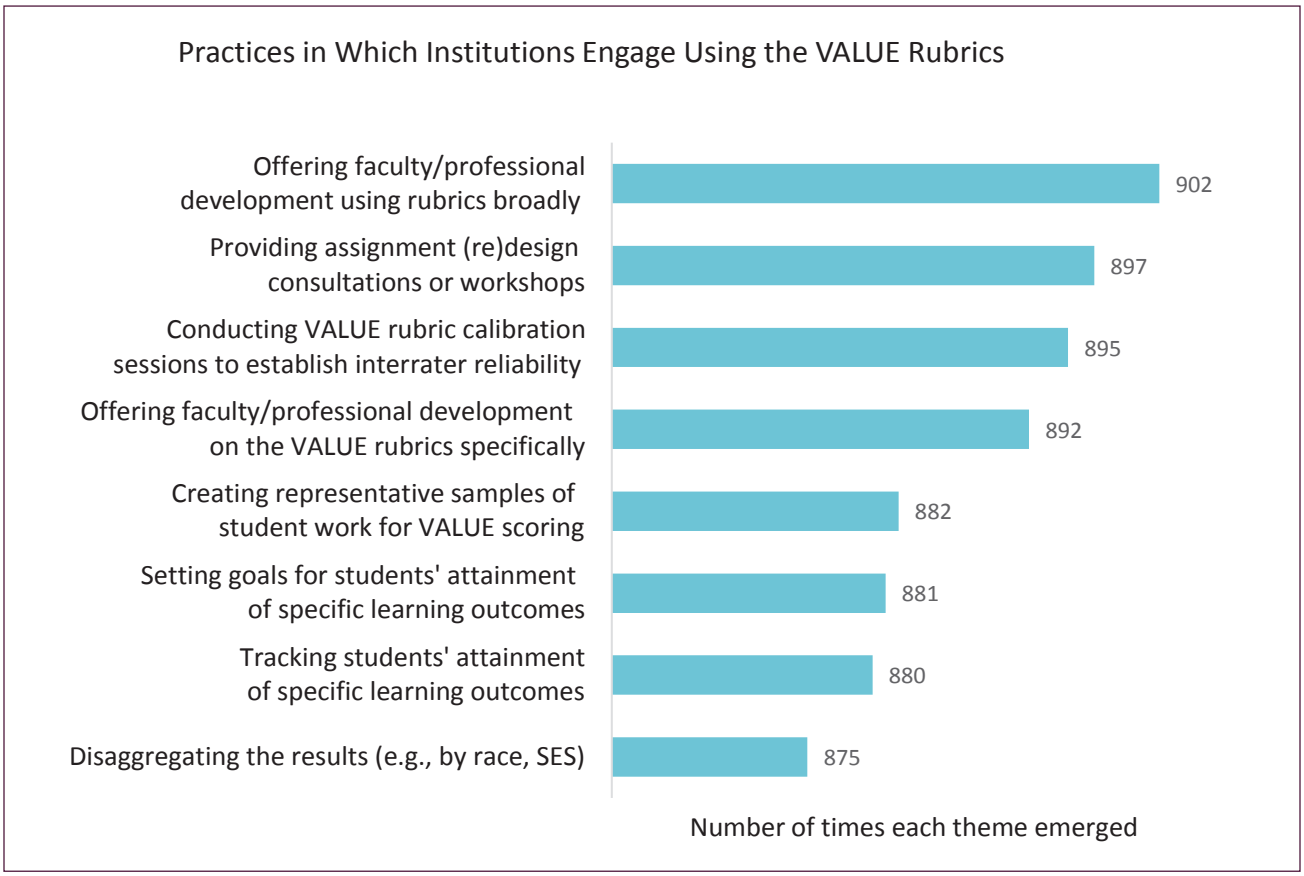


Figure 11. Responses to the survey question, "Does your institution engage in the following practices as part of your work using the VALUE rubrics?"



use validity of the VALUE rubrics. The widespread adoption of the VALUE rubrics, and the feedback provided by those using the VALUE rubrics at their institutions, support the assertion that the 16 learning outcomes selected for VALUE rubrics represent the appropriate student learning outcomes, and that the constituent dimensions encompass the needs that institutions have: curricular, cocurricular, and beyond.

Claim 6: The VALUE Rubrics Appropriately Distinguish Among Different Levels of Performance

6a. The 4-0 VALUE rubric scale appropriately distinguishes across performance levels.

6b. VALUE rubric descriptions clearly articulate differences among the different levels.

6c. The benchmark level, two milestone levels, and one capstone level provide relevant and understandable categories to students and professors.

The three highly interrelated assumptions of Claim 6—that the VALUE scale appropriately distinguishes across performance levels, that the VALUE rubric descriptions of the performance levels clearly articulate the differences among the levels, and that the levels provided are understandable to students and professors—are addressed together in this section.

The VALUE rubric development process included significant attention to the design parameters of the rubrics, such as the number of performance levels (four with zero as an additional scoring option) and number of dimensions (five or six). Interviews with the leaders of the VALUE rubric development process illuminated the conversations that led to the establishment of these VALUE rubric design parameters. Many of the design parameters for the entire suite of VALUE rubrics were established early in the development process, which focused on the creation of the Written Communication and Critical Thinking Rubrics. For example, one of the first drafts of the Written Communication VALUE Rubric had six levels. Community colleges involved in the process pushed back on this, explaining that levels one and two did not have enough nuance to represent the entire community college experience (W. Garrison, personal communication, May 3-5, 2016). Further, distinguishing between levels two and three was a difficult task. It became apparent to the teams that any more than four performance levels for broad learning dimensions would be quite difficult. So, the decision to limit the VALUE rubrics to a smaller number of performance levels was made. The facilitators discussed the possibility of having three levels but knew having an odd number of levels encouraged scorers to gravitate to the mean and score many artifacts at the middle, neutral score.

In addition to the four levels shown on each rubric matrix, scorers have the option to score dimensions as zero. The decision to include the zero option, but not include it on the rubric matrix, arose to provide scorers the option of recording the absence of evidence in the student work on a specific dimension of the rubric or demonstrated achievement not meeting the benchmark level.

Choosing the labels for the rubric levels (*benchmark*, *milestones*, and *capstone*) also resulted from research and discussion around existing rubric labels already in use. For example, the use of language from accreditation standards likely would not resonate with faculty in general. Some descriptors used by campuses with a liberal arts mission would not necessarily resonate with descriptive words for these levels at other institutions such as community colleges or research universities. The decision to use descriptive language that focused on neutral terms for student learning development over time was made to mediate these inevitable differences (W. Garrison, personal communication, May 3–5, 2016). The rubric development teams were especially focused on avoiding negative or pejorative descriptors (e.g., beginner, developing).

The teams also focused on an assets-based approach to the performance descriptors used for scoring student artifacts with the VALUE rubrics. As such, even at the benchmark level, the VALUE rubrics describe positive performance rather than focusing on what the student is not exhibiting and generally avoid the typical practice of rubrics at the time of using an "always, sometimes, never" quantification or a "how often" approach to distinguishing quality between performance levels.

To ascertain whether these design parameters were effective in establishing a scale that appropriately distinguishes across performance levels, descriptors that clearly articulate the differences between the levels, and levels that are understandable to professors, two surveys (i.e., 2018 VALUE Rubric Institutional Use survey, 2017 Refinement Year Post-Training survey) solicited feedback from VALUE rubric users and scorers. An important limitation to the VALUE validity evaluation for Claim 6 (specifically 6c) is the absence of the student perspective (addressed more fully in the overall synthesis of the validity argument). Neither of the surveys discussed as part of this claim can help us determine the extent to which the performance levels of the VALUE rubrics are relevant and understandable to students. (Some faculty, however, did provide their perspectives on this issue, as documented in the discussion of open-ended feedback provided by respondents.) Several of the campuses that tested the rubrics during their development did involve students in providing feedback on the rubric language itself, sharing the information with the development teams as they redrafted the rubrics.

The Institutional Use Survey specifically asked respondents to rate the sufficiency of each component of the rubrics. Responses to these questions are shown in Figure 12. The Institutional Use Survey also provided respondents with the opportunity to elaborate on their sense of the sufficiency of the components of the VALUE rubrics. Some responses to this open-ended question offer further confirmation of the utility of the rubrics as well as suggestions for improving the rubrics themselves and/or developing further resources to enhance campus-level use of the rubrics: (a) "I especially appreciate the clarity of the language used in the rubrics. I think that it is easy for students to understand as well as useful for me as a professor"; (b) "These are exemplary; however, all rubrics, especially ones that are not context-specific such as these, have their limits"; (c) "I think the rubrics could be more student friendly. They are a great starting point for a variety of disciplines and assignments"; and (d) "These are great rubrics. I work in a faculty development/support centre. We recommend them when we work with faculty on course design, and we use them as models for faculty who want to develop their rubrics. We've also used modified versions for purposes other than student evaluation".

Additionally, the 2017 Refinement Year Post-Training Survey asked individuals who scored artifacts of student work that year using either the Civic Engagement, Critical Thinking, Ethical Reasoning, Intercultural Knowledge and Competence, Quantitative Literacy, or Written Communication VALUE Rubric the extent to which each component of the rubric was clear. These responses are shown in Figure 13. The overwhelming majority of respondents agreed that the VALUE rubrics were not only useful for evaluating student work (97%) but that the scoring levels provided sufficient range (97%), the performance descriptors were understandable (82%) and relevant (74%), and the rubric with which they scored student work encompassed the meaning of the learning outcome under consideration (89%). Informed by their experiences using one of the VALUE rubrics as part of the nationwide VALUE Initiative Collaboratives to score artifacts of student work, respondents further provided feedback on rubric dimensions and performance descriptors on specific VALUE rubrics. For example, feedback from those scoring with the Civic Engagement VALUE Rubric indicated that the dimensions Diversity of Communities and Cultures and Civic Action and Reflection were relatively easy to understand. For the Written Communication VALUE Rubric, respondents reported that the Sources and Evidence dimension worked well because the performance descriptors included words such as "credible" and "consistent."

Many scorers also identified specific rubric dimensions that were not working as well (e.g. the Calculation and Assumptions dimensions on the Quantitative Literacy VALUE Rubric). They also identified vague terms (e.g., "accurate," "context") and quantifiers (e.g., "some," "enough") on the rubrics. Scorers also noted that many student artifacts seemed not to provide evidence of learning on various dimensions of the assigned VALUE rubric, suggesting a misalignment between the purpose and parameters of the assignment that generated the student work and the measurement tool (i.e., rubric) being used to assess it for evidence of the learning outcome at hand. In particular, this feedback reinforces the need to consider not only the design and content of the VALUE rubrics themselves but also the importance of assignments designed to develop student learning outcomes, the application of the VALUE rubrics, and interpretation and use of the scores they generate, which is the focus of the next section.

Figure 12. Likert scale responses to the question, "Below, please rate the sufficiency of each component found on the VALUE rubrics."

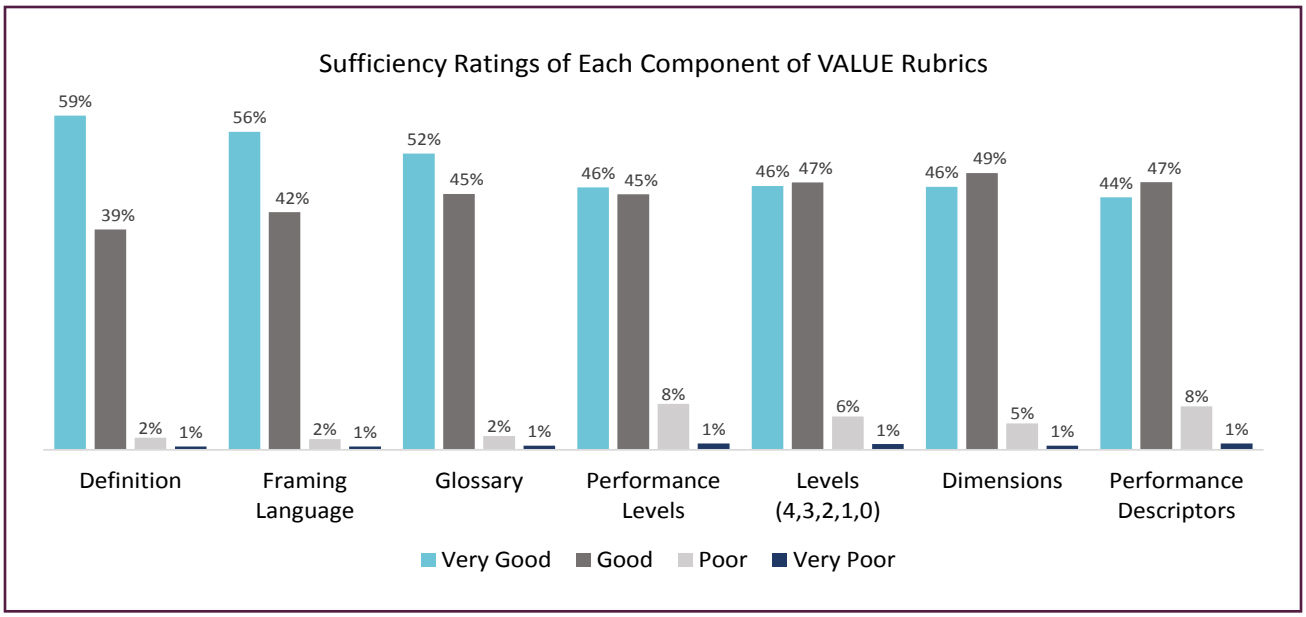
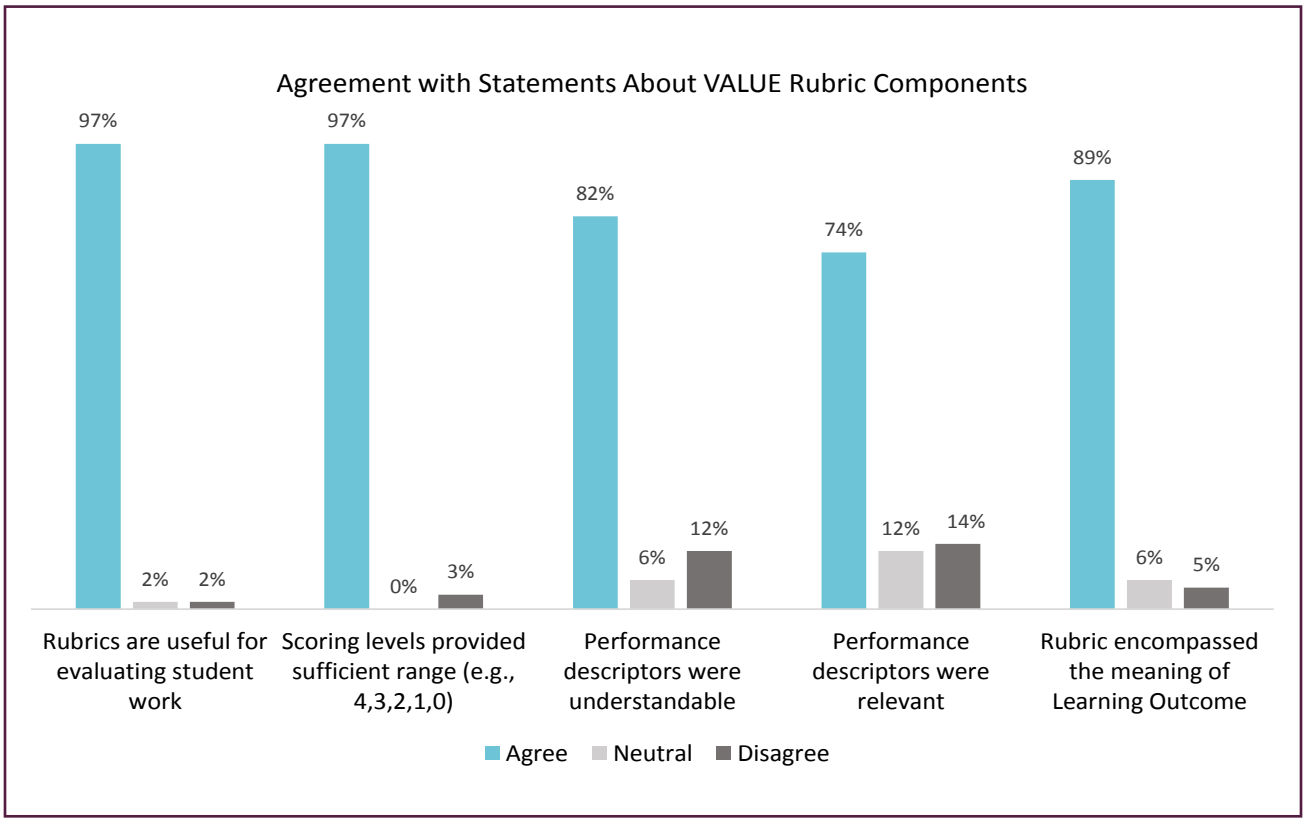


Figure 13. How respondents rated their agreement with statements regarding the clarity of each VALUE rubric component.



PART III

The Interpretive Argument for VALUE: Claims and Assumptions Associated With VALUE Rubric Application, Score Interpretation, and Use

Claim 7: The VALUE Rubrics Are Applied Appropriately to Create Scores for Each Artifact of Student Work

7a. Training fully prepares scorers to apply the VALUE rubrics to artifacts of student work.

7b. VALUE rubrics are applied reliably to artifacts of student work.

7c. VALUE rubrics are applied accurately to artifacts of student work.

Evidence addressing Claim 7 was drawn first from a review of the literature, including AAC&U-produced publications and outside sources, to determine best practices for calibration training with rubrics in general and the VALUE rubrics specifically. Second, those who participated in AAC&U VALUE rubric training and scoring provided valuable feedback regarding the effectiveness of the AAC&U VALUE rubric training. Finally, analyses of project-level data from the VALUE Initiative Collaboratives—including additional analyses conducted by consulting methodologists—helped identify strengths and areas for future work concerning Claim 7.

Assumption 7a: Training Fully Prepares Scorers to Apply the VALUE Rubrics to Artifacts of Student Work

The VALUE Initiative and other rubric assessment approaches have used a number of methods to recruit personnel to participate in calibration training and—ultimately—scoring. In the literature, and mirroring a theme in Claim 4, the notion of one's participation in calibration training as an extension of professional development was frequently discussed. Trainees and scorers were most often volunteers, with some paid a modest stipend for their involvement (Crist & Robinson, 2015; Graff et al., 2018; Siefert, 2011; Sorey & DeMarte, 2013). Typically, trainees and scorers are academic faculty (Carroll et al., 2016; Traver & Nedd, 2018), as was the case with VALUE scoring.

That said, scoring student work is not solely the work of faculty, as librarians, cocurricular professionals, research assistants, and even students have been recruited and trained to engage in rubric scoring of student work (Belanger et al., 2015; Bernstein & Greenhoot, 2014; Cole, Cochran, Troboy, & Roach, 2012; Frank, Simper, & Kaupp, 2017; Golter, Brown, & Van Wie, 2016; Graff et al., 2018; Kiel et al., 2015; Leist, Woolwine, & Bays, 2012; Oakleaf, 2007). Larger recruitment efforts often included the development and use of marketing materials and the creation of incentive programs to bolster participation. Some institutions reported making it a priority to recruit faculty scorers who were not associated with the course(s) from which the artifacts were taken, while others—due to resource constraints—designed scoring systems in which faculty score their own students' work (Crist & Robinson, 2015).

Once personnel are selected to participate in training, rubric calibration training can begin. *Rubric calibration* refers to the assumptions and actions that prepare trainees for formal scoring of student artifacts according

to the specified rubric (Brown & Rhodes, 2016; Finley, 2011; McConnell, 2013). *Scoring practices* refer to the formal execution of the assumptions and actions trainees learn during calibration training to ensure that rubrics are applied reliably and accurately. In other words, calibration training is designed to configure the mindset and activities each scorer is expected to enact so they can provide reliable and accurate scores across multiple artifacts (i.e., *intrarater reliability)* as well as in comparison to the scores that other scorers apply to the same student artifacts (i.e., *interrater reliability)*. While there are examples from the literature of high interrater reliability occurring without training (Bresciani et al., 2009; Finley, 2011), in general interrater reliability is strengthened once calibration training has taken place and weakened when it does not (Burke et al., 2012; Gray & Brown, 2017; Rezaei & Lovorn, 2010).

Training may take place in a variety of formats including traditional face-to-face, online, self-paced, and group settings (Finley, 2011; Raczynski, Cohen, Engelhard, & Lu, 2015). Throughout the literature, there were no significant discrepancies among the formats apart from the often-discussed challenges individuals face while learning in one type of training environment versus another (Oakleaf, 2009; Schlitz et al., 2009). Typically, a lead trainer with previous experience engaging in a similar rubric-based assessment of student learning develops, organizes, and shares documentation to guide the trainees as they build common scorer reasoning practices (Finley, 2011; Holmes & Oakleaf, 2013; Horowitz, 2014; Rhodes & Finley, 2013; Rhodes et al., 2014). Trainers often have multiple years' experience with rubric calibration training and are adept at moderating multiple rounds of discussions (Alschuler, 2016; Bachen et al., 2018; Bodnar & Kadlowec, 2018; Dunbar, Brooks, & Kubicka-Miller, 2006; McKitrick & Barnes, 2012; Turbow & Evener, 2016; Weaver et al., 2016). Another approach to stimulating discussion to help surface scoring patterns and practices between and among scorers is the "speak out loud" method so other trainees can hear a fellow trainee justify how and why they scored a practice artifact (DiPardo, Storms, & Selland, 2011; Joe, Harmes, & Hickerson, 2011; Shapiro, 2012; Trace, Meier, & Janssen, 2016). Others provide trainees with explicit instructions on how to notate evidence within the student artifact that is reflective of a specific dimension and/or performance level from the rubric under consideration (Bodnar & Kadlowec, 2018; Gray, Brown, & Connolly, 2017), and then use that to stimulate conversations around what trainees are seeing and not seeing as evidence of student learning, and how to properly view the artifact in relation to the rubric (Saxton, Belanger, & Becker, 2012).

The end goal of rubric calibration training is for all scorers to agree on the rubric score(s) for the sample student artifacts that comprise the training. The aforementioned techniques are geared toward building consensus and understanding, and calibration is completed when most, if not all, trainees are able to provide scores in a consistent, reliable manner via the consensus building process. This is the approach taken by most locally conducted rubric scoring initiatives and has historically been the approach taken in VALUE rubric calibration. When time and resources permit—including but not limited to having access to previously scored student artifacts and scorers with a depth of experience that would qualify them to serve as expert scorers of student work with a specific rubric—calibration training oriented to guiding trainees to a pre-determined, "expert" score is arguably more robust than training that relies on consensus building alone (Pike, 2018). Tests for determining the calibration of scorers range from the simplest (e.g., percent agreement among scorers) to employing a variety of interrater reliability statistics (interrater reliability will be discussed in greater detail for Assumption 7b). Once trainees are calibrated, the assumption is that they are prepared to officially score the artifacts submitted for the rubric-based assessment.

The AAC&U approach to calibration training has evolved over the years, building upon the best practices identified within the literature to accommodate changes in praxis necessitated by taking VALUE to scale beginning in 2013 with the VALUE Initiative Collaboratives. The "ground rules" for the AAC&U VALUE approach to calibration help orient trainees to the assets-based, growth-oriented mindset behind the design of the VALUE rubrics, reminding them that (a) scoring with the VALUE rubric for assessment is not the same process as grading; (b) they must think globally about the student artifact and the learning outcome at hand, moving beyond their own specific disciplinary lens or content; (c) they should start at the top of the

scale with the assumption that each student artifact has the potential to show evidence of the highest level of performance; and (d) zero is indeed an option as a score (W. Garrison, personal communication, May 3–5, 2016).

From the inception of VALUE, AAC&U has recommended that campuses engage in calibration training. As the VALUE rubrics gained traction, AAC&U staff were invited to institutions to conduct calibration training sessions (T. Rhodes, personal communication, May 6, 2017; W. Garrison, personal communication, May 3–5, 2016) and began to offer workshops at AAC&U's Institutes and conferences, based on the demand from the field (Finley & Rhodes, 2013). In "Using the VALUE Rubrics for Improvement of Learning and Authentic Assessment," Rhodes and Finley (2013) described the process and benefits of calibration training for scoring. These professional development offerings ultimately evolved into formalized calibration training for participants in the Multi-State, Minnesota, and GLCA Collaboratives. For the first two years of the VALUE Initiative Collaboratives, training was conducted in person by experienced VALUE rubric scorers under the project leadership. Training took place over the course of two to three days, in classroom-style rooms with approximately 30–35 individuals in any given training room. The primary mechanism for establishing calibration to a score was consensus building *in situ* within the training session, with pen-and-paper recording of scores to look at percent agreement across raters.

Two significant changes were made to current praxis for AAC&U's nationwide scoring efforts. In the 2016–2017 refinement year, all calibration training was moved to an online environment for cost and scale considerations. Trainees (a) watched and interacted with a series of videos delineating the VALUE scoring process, (b) scored a student artifact, (c) submitted their scores electronically for review, and (d) engaged in discussions over conference calls about the scores moderated by the Training Director. The most recent iteration of calibration training for the 2017–2018 inaugural VALUE Institute involved gaining the permission from state leads to collect, organize, and use anchor artifacts from institutions that submitted student work during the pilot, demonstration, and refinement years. Anchor artifacts were chosen for their ability to exemplify a wide range of scores on various dimensions of the VALUE rubrics. Then, a combination of conference calls, documentation shared via Google Drive, YouTube videos, e-mail correspondence, and Qualtrics surveys were used to train scorers for the newly created VALUE Institute. This also represented AAC&U's first attempt at incorporating the *expert score* approach within calibration training in which experienced VALUE scorers were recruited to score a subset of student work in order to establish the "true" score to which trainees would be calibrated. This expert-scored suite of work representing varying levels of performance became the basis for calibration training. Trainees were considered to be calibrated for scoring if their scores were not more than one level away from the expert score on the majority of the dimensions of the VALUE rubric for the majority of the student artifacts. Trainee scores were also compared with one another in the same manner to ascertain the consensus scores among the trainees apart from the expert scorers.

The Refinement Year Post-Training Survey provided insights into the efficacy of VALUE rubric calibration training. Through several open-ended questions, trainees identified what worked well and what proved to be problematic and offered suggestions for improving the VALUE rubric calibration training (Figures 14, 15, and 16).

Assumptions 7b and 7c: VALUE Rubrics Are Applied Reliably and Accurately to Artifacts of Student Work

Selecting which interrater reliability statistic to use to determine calibration is a contentious topic (e.g., Jonsson & Svingby, 2007; RiCharde, 2008; Szafran, 2017). While some caution against overemphasizing interrater consensus as an end goal in and of itself (RiCharde, 2008), there is utility in exploring interrater reliability, such as using results to inform rubric modifications and to make improvements in training materials (Bachen et al., 2018; Carter & McConnell, 2018; DiPardo et al., 2011; Phillips & McLawhon, 2013; Shapiro, 2012;

Figure 14. Responses to the survey question, "Which parts of the rubric worked well for you during training? Please be as specific as possible, highlighting concepts, vocabulary/word choice, etc., that facilitated your training." The responses were coded for themes, which are presented in this graph.

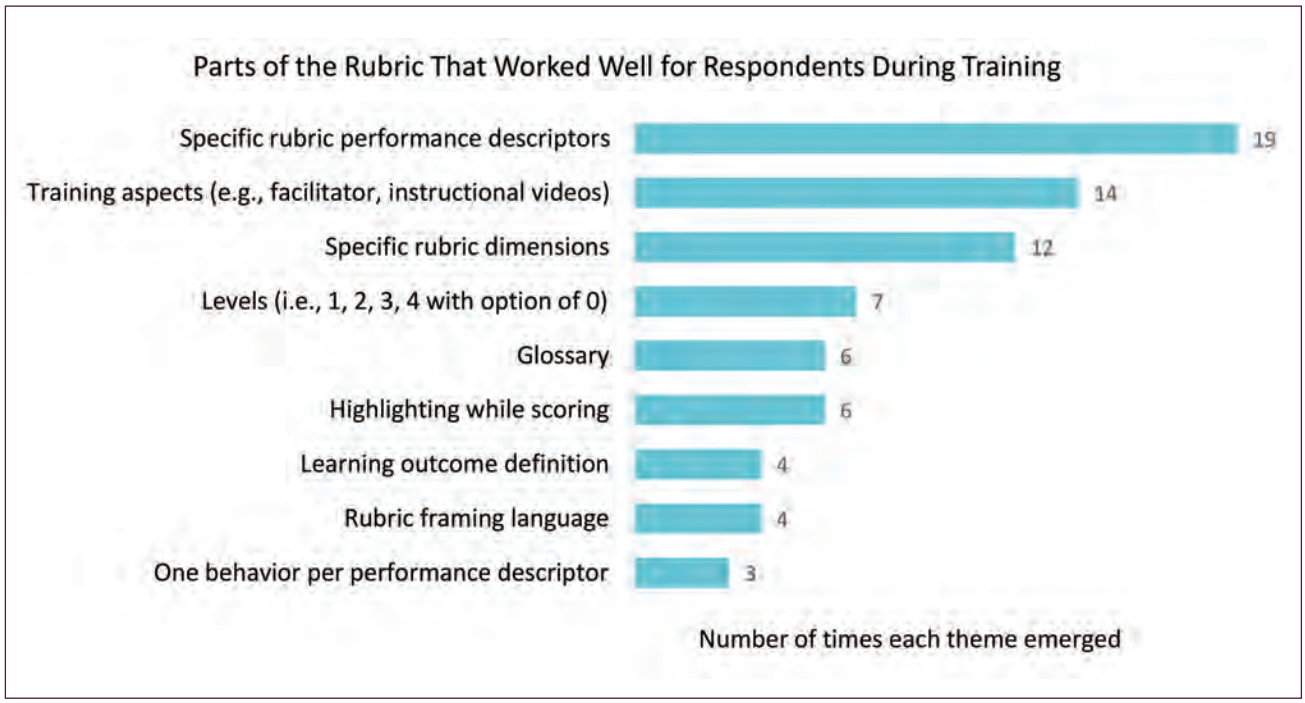


Figure 15. Responses to the survey question, "Which parts of the rubric did not work well for you during training? Please be as specific as possible, highlighting concepts, vocabulary/word choice, etc., that impeded your training." The responses were coded for themes, which are presented in this graph.

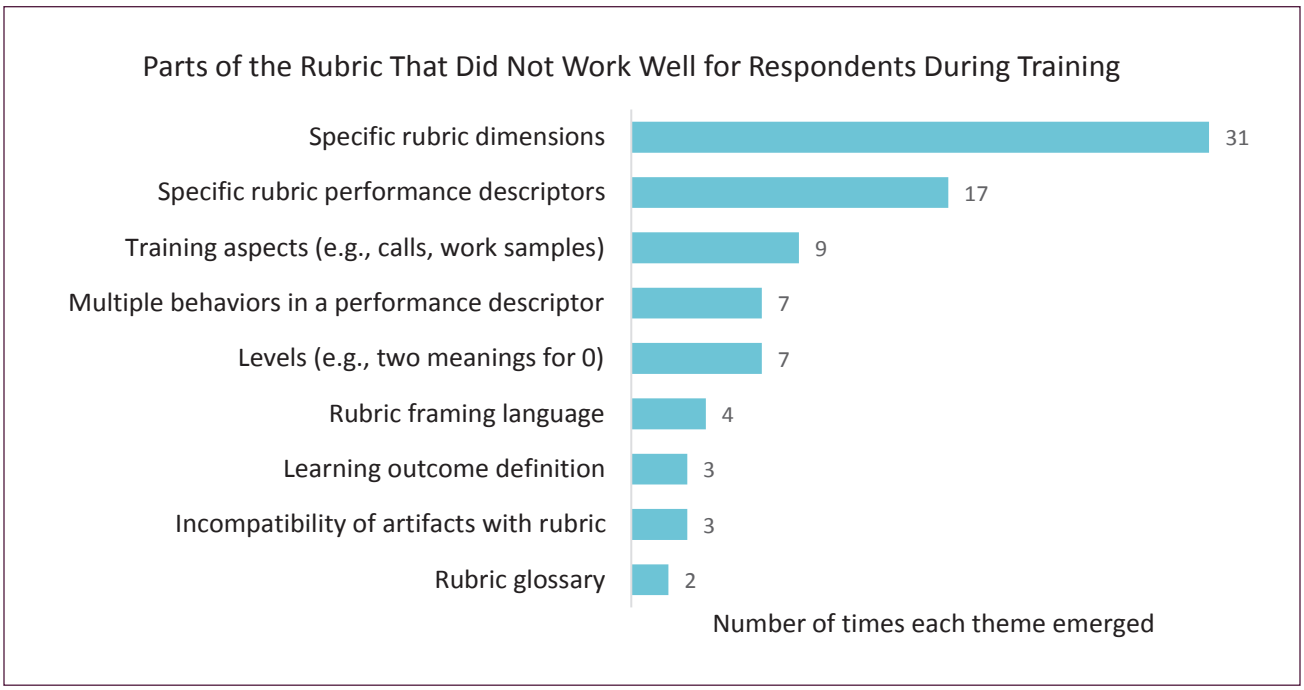
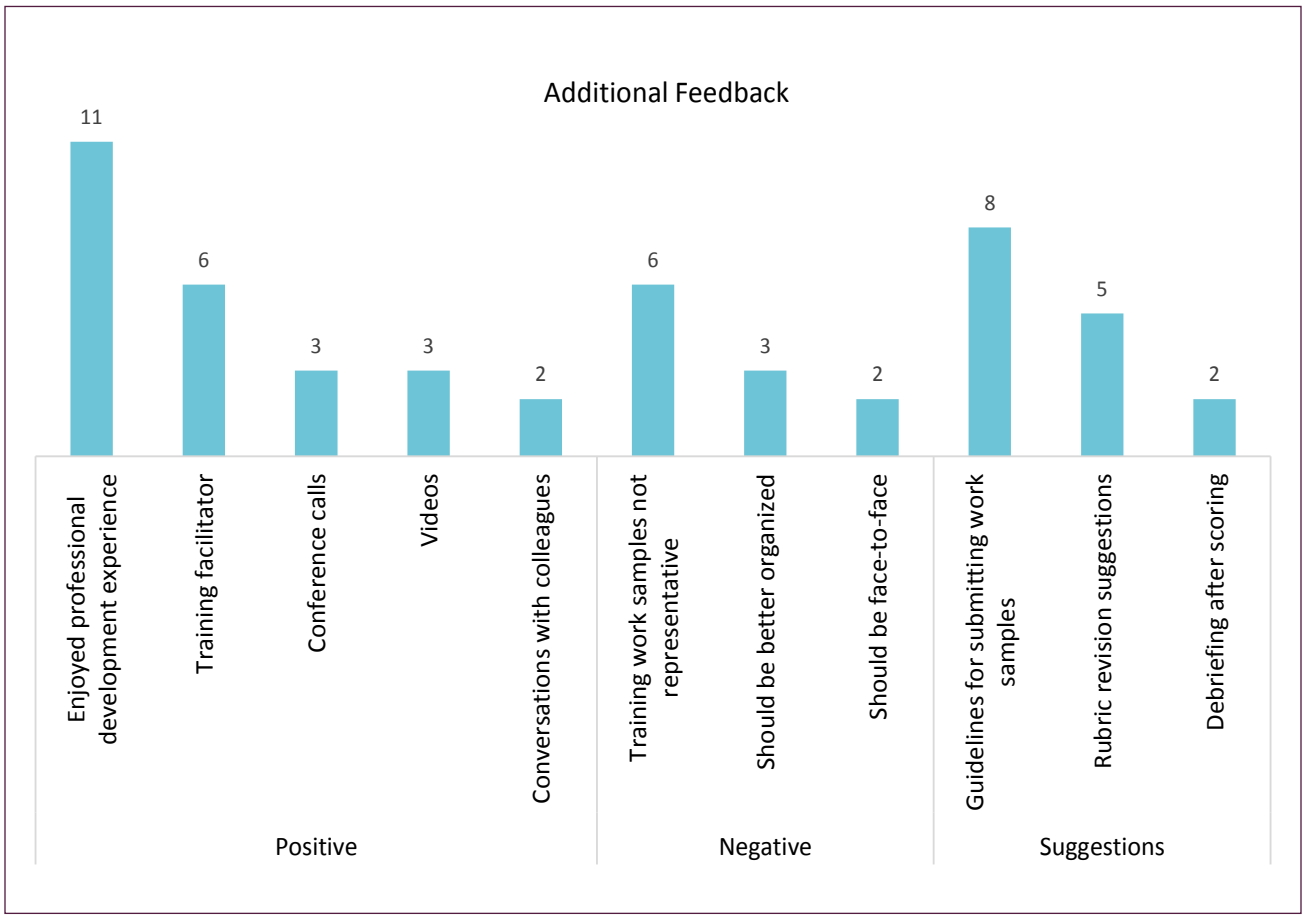


Figure 16. Responses to the survey question, "Finally, please provide us with any additional feedback/comments/questions/concerns that you have as a result of your recent training experience with the VALUE rubric." The question was open ended and coded for themes.



Wang, Engelhard, Raczynski, Song, & Wolfe, 2017). There is no one correct answer regarding which interrater reliability statistic to use, as the choice of test will depend not only upon the number of individuals scoring each student artifact but also the protocols for multiple scoring (e.g., whether scorers are fixed or randomly assigned) and the type of rating system employed (e.g., binary, categorical, ordinal). Even consideration of the reliability of the VALUE rubrics by AAC&U applied different interrater reliability tests (e.g., Finley, 2011; McConnell & Rhodes, 2017).

Determining the accuracy of the application of VALUE rubrics to artifacts of student work is also a complex endeavor. The aforementioned VALUE calibration training and scoring best practices are assumed to be applied using unmodified VALUE rubrics. The reported interrater reliability statistics are also for the unmodified VALUE rubrics. However, as noted throughout the literature, the majority of applications have been in modified formats. Not only were the actual rubrics often modified locally—which implicates changes to the rubric nomenclature and underlying meaning—but calibration training and scoring practices were often changed as well. This is the logical and desirable result of the initial design and purposeful distribution of the VALUE rubrics for local assessment efforts. That said, varied interrater reliability statistics, inconsistent consensus methods, and assorted score calculations contribute to issues of reliability and validity, particularly when scores are used for benchmarking (Steinke & Fitch, 2011). Careful and intentional considerations are needed before assessment results are contextualized within grander contexts (Hathcoat & Penn, 2012). As discussed in Nicholas, Hathcoat, and Brown (2016), assessment strategies will influence whether resultant

scores should be compared or used as a benchmark method and, in cases of significant modification, when they should not. In response, Gary Pike (2016) stated, "lessons to be learned here are that multiple sources of variation can affect institutional-level assessment scores, and that care must be taken to minimize the sources of variation when making comparisons across institutions, over time, and across assessment instruments" (p. 7).

Through the nationwide scoring projects, AAC&U had its first opportunity to look at interrater reliability and dependability (i.e., accuracy) at a larger scale. Each year, approximately 15–20% of student artifacts for the Critical Thinking, Quantitative Literacy, and Written Communication Rubrics were double scored for the explicit purpose of investigating interrater reliability. Working with John Hathcoat, Assistant Assessment Specialist and Assistant Professor of Graduate Psychology at James Madison University, we looked at double scored data from 2015–2016 for those VALUE rubrics using multiple interrater reliability tests, using ordinal weights in each analysis: (a) percent agreement, (b) Cohen's kappa, (c) Brennan-Prediger, and (d) Gwet's AC coefficients. With Cohen's kappa, chance is conceived in the same way as a chi-square test of independence (i.e., raters are assumed to be independent). This value is low across each analysis, but kappa is limited in numerous ways. For example, it tends to be highly influenced by marginal distribution. Brennan-Prediger adjusts for chance by considering the number of categories within the rubric (e.g., 4–0). Gwet's AC adjusts for chance by considering how difficult or easy it is to rate a subject (e.g., hard subjects will have a uniform distribution whereas easy subjects will tend to be placed in the same category). With the exception of Cohen's kappa, the data indicated that the interrater reliability achieved within the nationwide scoring project was moderate to strong overall (Table 1).

Table 1. Interrater reliability for the demonstration year of the VALUE Initiative Collaboratives.

	Weighted % of exact agreement range	Weighted Cohen's kappa range	Weighted Brennan-Prediger range	Weighted Gwet's AC2 range
Critical Thinking	88–89	.26–.34	.57–.62	.64–.70
Quantitative Literacy	84–87	.29–.39	.40–.54	.50–.62
Written Communcation	88–94	.27–.39	.56–.77	.60–.84

Subsequent to our interrater reliability exploration discussed previously, Gary Pike (2018) used double scored student artifact data from the 2015–2017 nationwide collaboratives to conduct the very first evaluation of the dependability of the VALUE scores on the Critical Thinking, Quantitative Literacy, and Written Communication Rubrics using generalizability theory. This analysis revealed that VALUE has not yet achieved the desired levels of dependability traditionally expected of other measures like standardized tests. That said, this work provides us with a baseline from which to improve. Key findings included the following:

- Within all three rubrics, it may be easier to achieve dependable student scores for some dimensions than for others. That said, whether these differences are due to the nature of some of the dimensions, the clarity of the descriptors used in the rubrics, or the training of the raters was beyond the scope of this investigation.
- Given the differences across the dimensions of the Critical Thinking, Quantitative Literacy, and Written Communication Rubrics, it is not surprising that achieving acceptable levels of generalizability is easier for some dimensions than for others.
- More work needs to be done to improve interrater agreement. One possible method of improving interrater agreement is through better training of raters. Another possible avenue for improving interrater agreement is to carefully review the descriptive statements associated with score points on the VALUE

rubrics' dimensions. Dimensions with low levels of generalizability should be a starting point for reviewing and modifying these descriptive statements. A third option for improving the dependability of ratings requires that all student artifacts be evaluated by the same panel of raters.

- Finally, while a change in protocol utilizing a common assignment or prompt would in all likelihood help enhance interrater reliability, such a change would represent a significant—and undesirable—departure from the current VALUE approach to assessment. Instead, there is a need to redouble efforts to focus future work on issues associated with assignment alignment to help achieve greater dependability of the VALUE approach.

Claim 8: VALUE Rubric Scores Appropriately Reflect a Student's Knowledge and Skills

8a. VALUE rubric dimension scores are free from external influences, such as biases of the individual rater.

8b. VALUE rubric dimension scores correlate with other measures of student knowledge and skills.

Perie's (2013) interpretive argument for VALUE surfaces two areas of important consideration in relation to the claim that VALUE rubric scores do indeed appropriately reflect a student's knowledge and skills by asking us to address issues of bias in the VALUE rubric scoring process as well as to investigate if and how VALUE rubric scores correlate to other existing measures of student knowledge and skills (e.g., course grades, GPA, standardized tests). In our consideration of Assumption 8a regarding scorer bias, we provided evidence from the literature and described AAC&U practices designed to mitigate the influence of external factors in the scoring process. We then expanded our consideration from the specific issue of rater bias to a broader discussion of the issue of fairness, providing specific details on AAC&U's efforts to promote fairness in the VALUE approach to assessment.

To address Assumption 8b, we again turned to the literature to inform our understanding of the relationship between VALUE and other measures of learning before discussing the nascent work that AAC&U began in this area through the nationwide VALUE Initiative Collaboratives. We conclude with a brief discussion of the ongoing dynamic tension between maintaining VALUE's inherent flexibility at the local level while simultaneously working to enhance the methodological rigor used at the inter-institutional, nationwide level through the VALUE Institute.

Assumption 8a: VALUE Rubric Dimension Scores Are Free From External Influences Such as Biases of the Individual Rater

All raters bring their own expertise and experiences to the VALUE rubric scoring process. Indeed, VALUE rubric scoring is designed to provide faculty and other college and university educators (e.g., cocurricular professionals, librarians) with the explicit opportunity to leverage their expertise and experience to help elucidate evidence about the quality of student learning as seen through the lens of the VALUE approach. That said, scorers must be cognizant of their own biases that may inappropriately influence the scores they assign for each of the VALUE rubric dimensions.

Not surprisingly, the literature on rubrics and the VALUE rubrics specifically—which were designed as transdisciplinary assessment tools, agnostic to the particularities of any single disciplinary approach—discussed academic discipline as a potential source of bias. Multiple sources describe disciplinary bias broadly as a potential source of error in scoring, while other studies dive into more specific disputes between

disciplines (e.g., differences between writing norms and conventions across different disciplines; Anson, Dannels, Flash, & Gaffney, 2012; Bauerlein, 2012; Carlin et al., 2011; Hathcoat, Penn, Barnes, & Comer, 2016; Oakleaf, 2007; Rezaei & Lovorn, 2010; Stupans et al., 2015). However, it is important to note that the notion of disciplinary bias is context-dependent; a sizable segment of the literature describes how and when the VALUE rubrics have been intentionally revised to better accommodate specific biases of disciplinary language, norms, evidentiary standards, and performance expectations (e.g., Carlin et al., 2011; Cicek et al., 2015; Gleason et al., 2013; Hathcoat & Penn, 2012; Oderda et al., 2010).

Focusing specifically on the issue of rater bias, AAC&U has implemented several scoring protocols designed to mitigate this as a potential source of error. As discussed in Claim 7, AAC&U has established "ground rules" for scoring that are designed to focus scorers' attention on the specifics of the VALUE rubric they are using and each individual student artifact they are rating. Furthermore, AAC&U has implemented several strategies to further moderate the influence of rater bias in the scoring process. In VALUE rubric scoring for the VALUE Initiative Collaboratives and the VALUE Institute, scorers never see the actual assignment that faculty provided to students to generate the artifact being scored. This is done to reinforce in scorers' minds that scoring with a VALUE rubric is indeed different from grading. By not having access to the assignment prompt itself, scorers will not inadvertently focus on how well the student completed the assignment but instead will focus exclusively on identifying evidence of learning within the artifact they are scoring. Additionally, this approach also prevents scorers from evaluating the quality of the assignment prompt, which could in turn influence their scoring. Scorers will not know the course from which the artifact was sampled, though the nature of the student artifact may suggest the discipline or disciplinary category (e.g., arts, humanities, social sciences) of origin. Scorers are also not permitted to know anything about the student behind the artifact: whether the student is from a community college or four-year institution, a private liberal arts college or a public research university, nor anything about the student's demographic characteristics. Finally, at the nationwide scoring level, scorers are never assigned student work from their own course or even their own institution to score. These protocols for reducing bias govern training and scoring for all VALUE Initiatives led by AAC&U and also serve as guidelines or recommendations for local implementation of VALUE. That said, the usual caveat applies: guidelines and ground rules may be modified to better suit the assessment needs, available resources, and culture and context of an individual institution.

The issue of bias, important as it is, represents only one area of concern around the issue of fairness. In assessment of learning, fairness means that every student deserves an equal opportunity to demonstrate what the student has learned. At the programmatic level, where assessment results will potentially influence changes in curricular design or inform pedagogical innovation, fairness toward faculty should also be of concern. While much of AAC&U's focus has been on issues of fairness related to the VALUE rubrics themselves, a burgeoning area of emphasis for our work is assignment selection and assignment design and redesign.

The use of student artifacts generated by robust assignments has the potential to enhance the fairness of the VALUE approach for both students and faculty. Higher quality, well-aligned assignments increase the likelihood that students have the explicit opportunity to demonstrate learning on all dimensions of a VALUE rubric. Faculty have the opportunity to move beyond specious convenience sampling of artifacts for accountability purposes that may or may not have any meaningful connection to the VALUE rubric that will be used to judge the output of those assignments (Pike, 2014). If the rubric being used to measure desired student learning does not in actuality prompt students to demonstrate those dimensions of the learning sought, students are being unfairly judged on their abilities, and faculty (and programs/institutions) unfairly judged on effectiveness and success. Some assignments work well for certain learning outcomes and some do not. For example, assessing critical thinking or written communication skills using a mathematics worksheet would likely not provide enough information to someone determining a student's knowledge and skills for critical thinking dimensions.

AAC&U provides recommendations for selecting and sampling student work for scoring, the first and foremost being that assignments used for assessing student learning outcomes should be signature assignments (Rhodes & Finley, 2013). *Signature assignments*, or signature student work, are a student's best work and can be work used to assess students' cumulative learning at the end of a semester or at graduation. For example, this should not be a test a student did not know about or did not study for. It should be an artifact that allows the student to fully demonstrate knowledge and skills for a learning outcome. In "Using the VALUE Rubrics for Improvement of Learning and Authentic Assessment," Rhodes and Finley (2013) provided several examples of signature assignments for creative thinking.

In the refinement year of the MSC, AAC&U asked trainees and scorers to provide feedback on the assignments they scored and describe the characteristics of the artifacts they felt were the most and least readily assessable (i.e., able to be scored) using the VALUE rubric (i.e., those artifacts that were or were not closely aligned with the rubric). There were several examples of assignment qualities that worked well and did not. We asked for specific examples of types of assignments the scorers would recommend for assessment with the learning outcome they used in scoring. These recommendations are shown in Table 2.

Assumption 8b: VALUE Rubric Dimension Scores Correlate With Other Measures of Student Knowledge and Skills

In student learning outcomes assessment, best practice implicates using more than one assessment tool to measure student learning, be it at the course, program, and/or institutional levels. The goal should not be to find the single perfect measure but instead to create a system for assessment that would draw upon multiple approaches to measuring learning, ideally measures with strengths and weaknesses that complement and counter balance one another for the purpose of triangulating results for a more complete, nuanced picture of learning to emerge. The array of available protocols for authentic assessment take many forms and are reliant on proper training of the assessors (Banta, 2011). Pairing VALUE rubrics with other assessment tools provides insight to administrators, faculty, and staff as they seek to determine whether one assessment tool provides more information about student learning than another. Identifying and interrogating possible correlations between VALUE rubric scores with other resultant measures helps to inform which tools to use as they relate to compliance and funding (Berrett, 2016). Next, we provide a few illustrative examples of VALUE correlational studies.

Bernstein and Greenhoot used the Critical Thinking and Written Communication VALUE Rubrics, along with grades and results from the Collegiate Learning Assessment (CLA), to determine if team-designed courses improved student performance compared to students in a typical course (Bernstein & Greenhoot, 2014; Greenhoot & Bernstein, 2011). Students in both course formats performed no differently on the CLA, however students' rubric scores in both learning outcomes showed improvement for students participating in the team-designed courses. Lastly, the authors observed that rubric scores mirrored students' higher grades after the new course design was implemented.

Frank, Kaupp, Simper, and Scott (2015) directly compared results of the CLA with those from the VALUE rubrics to determine if the scores were correlated. The authors found CLA scores that measured critical thinking and written communication were significantly correlated with the respective VALUE rubrics. They noted that while the CLA provided overarching feedback about student ability, it did not provide any specific evidence that the results were connected to their course. Further, after instructors were interviewed about their experience using the VALUE rubrics, they reported that the VALUE rubrics functioned as an effective tool to aid in designing assessments; thus, through feedback, course instructors became aware of problems related to the course curriculum, particularly that student artifacts did not showcase outcome demonstrations because students were not given the proper prompting to substantiate particular criteria from the VALUE

Table 2. Trainee and Scorer Feedback

VALUE Rubric	Aspects of most readily assessable artifacts	Aspects of least readily assessable artifacts	Recommendations for assignments
Civic Engagement	Experiential learning reflections Aligned with rubric	Series of journal entries Politics or philosophy Public policy papers Summaries of historical events Research papers	Experiential learning reflections Personal reflections Focused, clear prompts aligned with rubric
Critical Thinking	Research papers Topic with multiple perspectives High-quality writing Position papers with conclusions Longer than one page	Reflections No clear purpose No elaboration Lab/science reports Exams/tests	Prompts asking for (at least) two viewpoints Evidence-based research papers with sources Prompt asking student's opinion Clear prompts Longer than one page Position papers that require defending an argument and reaching a conclusion
Ethical Reasoning	Explicit discussion of ethical theories Position papers Aligned with rubric	No elaboration No analysis of ethical reasoning Short-answer questions/ exam questions	Explicit discussion of ethical theories Assignments aligned with rubric Clear ethical issues to address
Intercultural Knowledge and Competence	Experiential learning reflections	Book or movie reviews Fact-based papers (e.g., history, book summaries) Ambiguous terms (e.g., culture, shared understanding)	Reflections Journals Travel blogs Cross-cultural comparisons Experiential learning reflections
Quantitative Literacy	Research papers At least some calculations Science labs Graphs/tables Written work Requires analysis, interpretation, and drawing conclusions	Tests with only calculation Papers with little to no calculation No solution/answer key Worksheets	Requires elaboration and analysis At least some calculation Requires figures, tables, graphs Lab reports Requires writing Shows calculations/ methodology Writing to explain, interpret, and draw conclusions about data and calculations
Written Communication	Essays Research papers High-quality writing Required sources/citations	Short-response answers Writing errors Reflection papers Creative writing pieces Too-narrow prompts Too short Handwritten No sources/citations	Evidence-based research papers Sources/citations Literary essays/analyses Expository/persuasive essays Lab reports Reflections

rubric. Therefore, while CLA and VALUE rubric scores were correlated, instructors report the VALUE rubrics provided more information related to student learning, as well as information to inform course and assignment (re)design.

Other studies have compared VALUE rubrics with other, non-VALUE rubrics to determine whether the non-VALUE method maintained the same learning outcome construct accuracy as the VALUE rubrics (Baryla et al., 2012; Bodnar & Kadlowec, 2018; Boersma et al., 2011). In other words, VALUE rubrics are routinely used as the "standard" to validate other rubrics. For example, Burke et al. (2012) compared scores when raters used the Written Communication VALUE Rubric and another rubric called the Academy of Process Educators Writing Rubric. Both rubrics consisted of five dimensions (called "performance areas") and the authors mapped each dimension to the corresponding dimension on the other rubric. Then, they analyzed rater consistency between the rubrics using a variety of assignments (e.g., research paper and book review). The results indicated that the raters provided significantly different scores on one type of artifact: research paper. For the essay, four of the five performance area pairings indicated a significant difference in the raters' scores. Additionally, three of the five performance-area pairings had significantly different scores for another type of artifact: book review.

Similarly, the City University of New York (CUNY, 2011) captured a more comprehensive look with a multi-campus correlation effort. A task force was created to determine assessment specifications and methodologies that were the most reliable and valid for long-term benchmarking needs which included the ability to "compare across CUNY colleges and between CUNY and other postsecondary institutions" (CUNY, 2011, p. 1). The Critical Thinking Assessment Test, Collegiate Assessment of Academic Proficiency, CLA, and ETS Profile were tested for content, criterion, and generalization validity. The CLA test was the only option deemed comparable in that "tasks require students to demonstrate the higher-order critical thinking and analysis skills reflected in the VALUE rubrics . . . [and] . . . also employs scoring rubrics that are similar in range and scaling to those of the VALUE rubrics" (CUNY, 2011, p. 14). Interestingly, the task force's recommendations for implementing the CLA across CUNY campuses echoed similar concerns about the use of the VALUE rubrics due to limitations related to generalizability. The use of the VALUE rubrics as *the* standard in these studies serves as an example of their face and use validity.

It should be noted that with its Spencer Foundation funding, AAC&U specifically sought to build on this research with its college and university partners in the Minnesota and GLCA Collaboratives (2014-2017). AAC&U developed data collection protocols for the campuses that included the submission of multiple data points for use in such comparative, correlational studies, including (a) national standardized test scores (e.g., student scores on the CLA), (b) GPA, (c) assignment grade (i.e., the grade that the student received on the artifact that was submitted for scoring), and (d) course grade (i.e., the overall grade that the student received for the course as it would appear on the student's transcript).

Unfortunately, there were multiple challenges that prevented this work from being enacted as part of the two collaboratives. First, many of the national standardized tests like the CLA do not provide institutions with individual student scores. Even if such scores were available, variations in timing of administration of tests and the collection of student work artifacts for scoring precluded the collection of paired data. Second, the request for GPA, assignment grades, and course grades were untenable for many institutions. Many lacked the capacity to pull such data in a relational way. For instance, institutional research "offices of one" may not have had sufficient bandwidth to fulfill the request, disparate institutional data systems did not readily communicate with one another, or artifact collection was done primarily through hard copy exchanges with faculty, with completely de-identified student work being submitted for scoring without any record or file linking the anonymous work with other variables of interest.

Finally, the effort suffered from what AAC&U has come to recognize as a *trust deficit* in assessment work. Simply put, the faculty submitting work to the project did not trust that information regarding their students' grades on assignments or courses would not, somehow, be used for purposes they deemed unhelpful or harmful. This is not to suggest that participating faculty withheld data in an attempt to derail this aspect of the work nor to perpetuate the stereotype that all faculty are resistant to student learning outcomes assessment. Instead, we observed that faculty, and by extension their campus-based assessment colleagues, were engaging in the aspects of the work that they found most meaningful—the external scoring of their students' artifacts with the VALUE rubrics and the scoring of other institutions' student work—while exercising an abundance of caution when it came to enabling comparisons or other analyses that may have been at cross purposes with their own. This experience appears to align with recent findings from the National Institute of Learning Outcomes Assessment third national survey of products on assessment (Jankowski, Timmer, Kinzie, & Kuh, 2018, p. 11). Future research will be required to develop a common—but not standardized—method for correlating VALUE rubric scores with these other assessment measures. This effort will be a worthwhile one, as the VALUE Institute strengthens processes for institutional comparisons that involve greater methodological soundness, and AAC&U will continue to review and develop studies that attempt to formulate correlation best practices to fortify this growing body of research work.

Claim 9: VALUE Rubric Scores Are Interpreted Correctly and Used to Make Defensible Decisions About Students and Programs

9a. Professors and department personnel understand how to interpret scores from the VALUE rubrics.

9b. Students understand how to interpret scores from the VALUE rubrics.

9c. Professors make appropriate inferences from the scores about student learning.

9d. Departmental personnel make appropriate inferences from the scores about the program.

9e. Students make appropriate inferences from the scores about their level of understanding.

This section explores whether the *end users* of VALUE (i.e., faculty, academic administrative leadership, and students) correctly interpret and make appropriate inferences from VALUE rubric scores to make defensible decisions that would affect students and/or programs at an institution. We will first address the interrelated supporting assumptions for Claim 9 that deal with faculty and administrators before then turning to the supporting assumptions that deal with students. It is important to note that, while AAC&U can make recommendations regarding the appropriate interpretation and use of VALUE data and provide examples of appropriate institutional decision-making using VALUE rubrics, it is neither possible nor desirable for AAC&U to play a monitoring or "policing" role in campus use of VALUE rubric results.

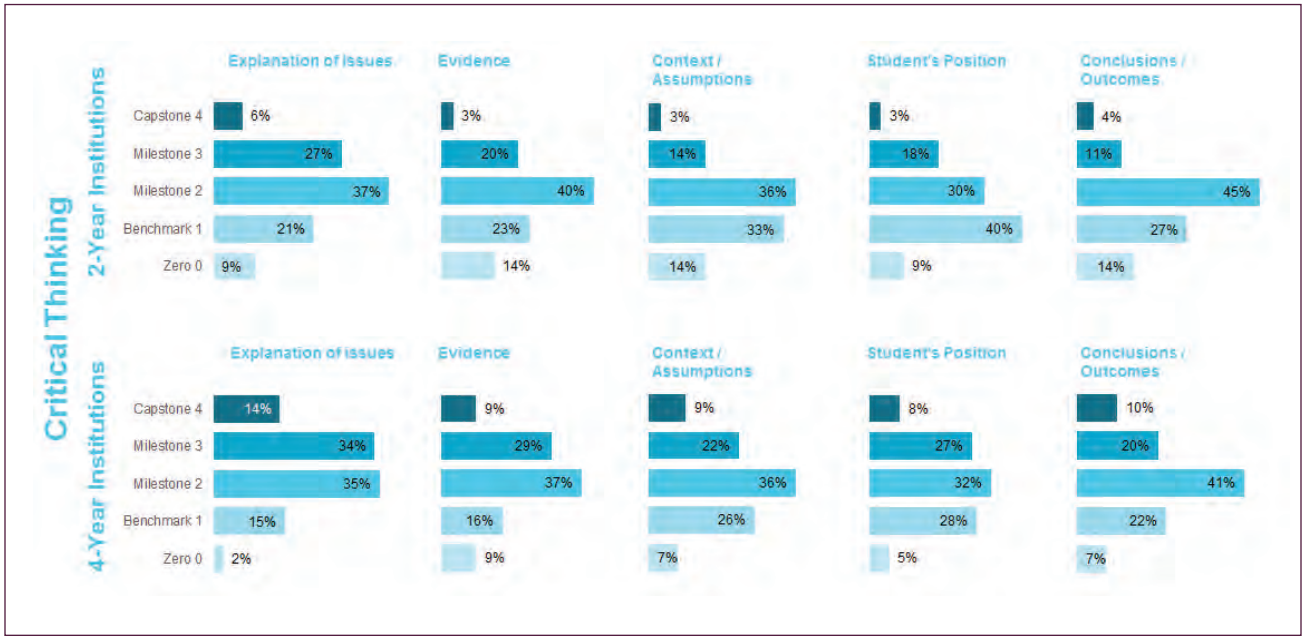
Assumptions 9a, 9c, and 9d: Professors and Department Personnel Understand How to Interpret Scores From the VALUE Rubrics, Make Appropriate Inferences From the Scores about Student Learning, and Make Appropriate Inferences From the Scores About the Program

A robust body of work exists providing educators with guidance on using rubrics writ large as a pedagogical and evaluative tool (e.g., Stevens & Levi, 2013). Over the 10-year existence of the VALUE rubrics, AAC&U

has worked to provide the field with multiple resources specific to the VALUE rubrics to facilitate the appropriate interpretation and use of VALUE-generated evidence of student learning. AAC&U has produced several publications related to using the rubric. The first, "Assessing Outcomes and Improving Achievement: Tips and Tools for Using Rubrics" (Rhodes, 2010a), introduced the rubrics and gave a background to the overarching VALUE initiative and the rubric development process. Rhodes (2010a) explained that the rubrics were intended as "broad, generic, institutional-level rubrics" (p. 2), which can be used as they are or modified for individual institutional purposes. AAC&U does caution that modifying the original rubrics limits the ability to use individual campus results for national comparison, hindering the ability to use those results for accreditation purposes (Rhodes, 2010a). The second publication, "Using the VALUE Rubrics for Improvement of Learning and Authentic Assessment" (Rhodes & Finley, 2013) addressed key elements of, and answers frequently asked questions about, the development and use of the VALUE rubrics. This volume also provided preliminary information about issues related to the validity, reliability, rubric modification, and scorer training, as well as provided specific examples of campus use of the VALUE rubrics, including 12 full case studies (www.aacu.org/value/casestudies). Further campus examples of using the VALUE rubrics have appeared in AAC&U's periodical, *Peer Review* (e.g., Fall/Winter 2011/2012 and Fall 2018 issues).

In addition to sharing the results from the first two years of data collection for the nationwide VALUE initiative, AAC&U put several stakes in the ground regarding the nature and treatment of VALUE data as well as recommendations for analysis and visualization of VALUE rubric results in "On Solid Ground" (McConnell & Rhodes, 2017). For example, AAC&U recommends against providing an average score for each VALUE rubric dimension or an average overall score for the VALUE rubric itself. Instead, AAC&U recommends that institutions report results for each dimension of the rubric as stand-alone bar graphs (see Figure 17 for a sample of this type of display). This allows professors and department personnel to move between the VALUE rubric and a report of the scores to look at the specific performance descriptors of dimensions where students are performing. Rather than having a single critical thinking score, they can see how students are performing on specific aspects of critical thinking. Finally, results from other VALUE research, such as the generalizability study discussed in Claim 7 (Pike, 2018), provide additional information about the appropriate interpretation and use of VALUE rubric data.

Figure 17. VALUE Initiative results for Critical Thinking for the refinement year of the Multi-State Collaborative for students who completed 75% of their credits toward graduation.



Beyond creating guidelines and providing advice, AAC&U solicits feedback from the field to explore how institutions themselves are interpreting and using VALUE data. The amount of activity generated around VALUE score interpretation and use varied from institution to institution and arguably reflected available campus resources. Campuses with a longer history of engaging faculty in assessment activities (with the VALUE rubrics or other measures), campuses that included their centers for teaching and learning in the work of VALUE, and campuses that had robust assessment infrastructures (e.g., larger staff) reported engaging in more activities designed to fully leverage the VALUE rubric results. Additionally, across all three VALUE Initiative Collaboratives, institutions that were more immediately involved in reaffirmation of accreditation efforts reported using the VALUE rubric results to satisfy assessment requirements specific to their region's association. The most sophisticated use of the VALUE rubric results were from campuses who had conducted their own scoring of the work submitted to the nationwide project; those institutions were able to then treat the data generated by nationwide scoring as external validation of the local work that had taken place with the same student artifacts. Above all, engaging with the VALUE rubrics and generating data on student learning stimulated the creation of faculty development opportunities on individual campuses. In the Minnesota and GLCA Collaboratives campus reports and action plans, several institutions reported offering faculty workshops (e.g., through their office of institutional effectiveness or by inviting AAC&U-affiliated experts to come to campus as workshop facilitators). Institutions reported that these workshops were helpful and informative for improving faculty's interpretation of VALUE rubric scores. Participating faculty reported that they were able to pay attention to the skills and knowledge they wanted their students to demonstrate on assignments.

In addition to feedback from the subset of faculty and institutions who participated in the nationwide VALUE Initiative, further evidence of the importance of faculty/professional development as part of the work with the VALUE rubrics comes from AAC&U's Institutional Use Survey. Respondents were asked about practices their institutions engage in as part of their work with the VALUE rubrics. The top three answers chosen were (a) offering faculty/professional development on using rubrics, (b) providing assignment (re)design consultations/workshops, and (c) conducting VALUE rubric calibration sessions (responses to this question are shown in Figure 11). As AAC&U continues to offer development through webinars (https://www.aacu.org/value/webinars), campus visits, and scorer training, it is worthwhile for campuses to consider creating and offering their own professional development opportunities that are designed to address the specific needs of their faculty, staff, and administrators. The Institutional Use Survey also revealed that faculty credit the VALUE approach with helping to initiate positive changes beyond assessing student learning outcomes. Faculty continuously credited their institution's implementation of the VALUE rubrics with inciting their desire to update their courses and assignments. After scoring with the VALUE rubrics, which involves looking at hundreds of student artifacts on the same learning outcome, they reported thinking more deeply about what their students were learning and what they meant when they said they want students to walk away with critical thinking skills or intercultural knowledge and competencies. As a result, AAC&U recommends campuses provide development opportunities to ensure that all faculty, staff, and administrators understand how to interpret VALUE rubric scores.

Admittedly, AAC&U's work in this space has focused on those individuals who have chosen to personally engage with VALUE on their own campuses (as signaled by downloading the VALUE rubrics) and/or participated in an AAC&U-led project. Self-selection and the motivations behind choosing to participate in this work are most certainly one factor at play. That said, campuses must engage in a delicate balancing act between promoting faculty engagement in any assessment effort and mandating that engagement. While there are myriad requirements (e.g., state, regional, federal) concerning the demonstration of the quality of student learning on any one campus, doing so through an approach that provides faculty with the wherewithal to make the assessment activities to meet the mandates meaningful arguably has the best chance of expanding the coalition of colleagues willing to take a deeper look at student learning outcomes assessment.

Assumptions 9b and 9e: Students Understand How to Interpret Scores From the VALUE Rubrics and Make Appropriate Inferences From the Scores About Their Own Level of Understanding

While there is a body of literature that delineates the importance of providing students with practice and feedback broadly (e.g., Ambrose, Bridges, DiPietro, Lovett, & Norman, 2010) and the power of using rubrics to provide students with feedback specifically (Stevens & Levi, 2013) to promote learning, there is a relative dearth of literature as well as AAC&U's research on whether students are able to interpret and use VALUE scores as part of their own learning process. In part, the purpose behind, and unit of analysis of, the VALUE rubrics necessitated the focus on faculty as intended users. As stated previously, the VALUE rubrics were not designed as grading rubrics to facilitate the delivery of feedback to an individual student. Instead, they take a programmatic and/or institutional view of learning: learning that will be adjudicated by faculty and other higher education professionals. Moving forward, consideration of student voice and agency must be explicitly incorporated into the VALUE approach, beginning—but not ending—with future iterations of the VALUE rubrics themselves.

PART IV

The Validity Argument for VALUE: Strengths, Limitations, and Future Work

A review of the evidence presented to address the claims and assumptions of the interpretive argument for evaluating the validity of the VALUE approach to student learning outcomes assessment reveals both existing strengths and current limitations.

At the core of the validity of the VALUE approach we find faculty: faculty as the originators, users, consumers, adopters, adapters, constructive critics, and evangelizers of and for the VALUE rubrics. Without faculty and their fellow educators in the libraries, student affairs offices, centers for teaching and learning, and yes, even in assessment offices, this approach to measuring student learning would not exist. The fact that, 10 years after their release, the VALUE rubrics continue to resonate with the individuals responsible for most directly fostering learning on our campuses—as indicated by their willingness to engage in this work locally as well as contribute to the national VALUE conversation by submitting student artifacts from their own courses for scoring and providing specific, concrete feedback to help improve the VALUE rubrics—is a milestone worth noting, given the number of assessment solutions and "silver bullets" that have come and gone in that same timeframe.

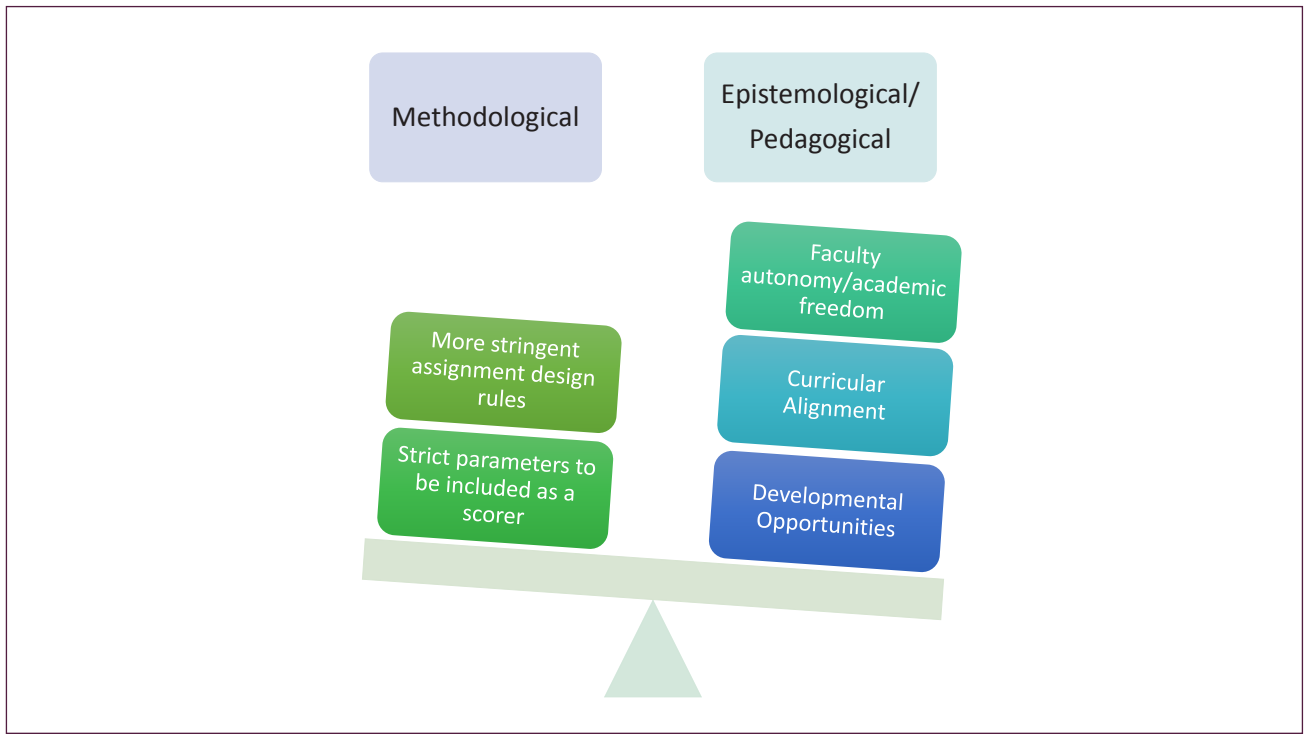
Faculty acceptance of and engagement with the VALUE rubrics is foundational to their validity. From that foundation, further review of the evidence suggests that, in their operationalization of the Essential Learning Outcomes, the VALUE rubrics do indeed provide tools to measure the learning outcomes that educators and employers consistently identify as the knowledge, skills, abilities, and dispositions that college graduates need for success in work and life. Furthermore, each VALUE rubric contains the key dimensions (i.e., the critical constituent constructs) that compose the learning outcome at hand. Our consideration of the literature as well as AAC&U's own research revealed that the VALUE rubrics can indeed be applied to many courses in a variety of disciplines, with or without modification. Evidence also supports the assertion that, writ large, the VALUE rubrics appropriately distinguish among different levels of performance, by providing a scale with sufficient range and clearly articulated performance levels that represent categories faculty find relevant and understandable. Additionally, when trained and calibrated, faculty are able to apply the VALUE rubrics to create scores for each artifact of student work. AAC&U's calibration training protocols and guidance regarding artifact collection and assignment design help to address and promote the fairness of the VALUE approach, including but not limited to addressing issues of bias.

AAC&U has continued its efforts to provide the field with guidance on the interpretation and use of VALUE scores as well, and institutions have taken the initiative to design and implement local complementary faculty development and training opportunities to enhance their own implementation and use of the VALUE rubrics to suit a range of purposes, from pedagogical improvement to addressing accountability mandates. This systematic validity evaluation confirmed that VALUE does indeed represent a powerful approach to assessing student learning, one that is no longer relegated to the sidelines as an "alternative" assessment but, instead, is a widely accepted approach that has strong epistemological, pedagogical, and methodological foundations. That said, the individual VALUE rubrics and the VALUE approach to assessment overall are not perfect. A review of the evidence addressing the claims and assumptions of the interpretive argument reveal that there still is work to be done to (a) enhance the reliability, accuracy, and dependability of VALUE scores; (b) ensure the appropriate interpretation and use of the results generated through VALUE; and (c) address the absence of evidence regarding students' perceptions, interpretations, and uses of the VALUE rubrics.

Addressing these gaps is critical. Each gap implicates myriad possible approaches and solutions, each with its own attendant pluses and minuses. It is necessary to remember how and why VALUE came into being: as an approach to assessment that privileges the learning transpiring within and responding to an institution's curriculum and cocurriculum, using pieces of actual student work, with faculty and other higher education professionals as the arbiters of quality. As such, future improvements to VALUE must not inappropriately privilege methodological litmus tests over VALUE's very identity. For example, while requiring a common prompt to be used for each VALUE rubric would simplify calibration training and likely result in higher interrater reliability, the loss of authenticity, faculty judgment, and pedagogical diversity represents too great a cost. The future work of VALUE will continue to try to strike an appropriate balance between competing concerns, erring on the side of epistemological and pedagogical concerns when necessary (Figure 18). Moving forward, the future of VALUE includes the following:

- Revisiting, revising, and translating the VALUE rubrics. Beginning in 2019, AAC&U will spearhead the revision of all 16 VALUE rubrics. To inform this process, AAC&U will use the results from the existing interrater reliability and generalizability studies and AAC&U surveys, including the review of modified VALUE rubrics submitted to AAC&U from individual faculty and institutions as part of the VALUE Rubric Institutional Use Survey. Additionally, AAC&U will engage faculty and—for the first time—students through focus groups and campus vetting of revised versions of the VALUE rubrics. Another aspect of this work is the creation of translations and guides to the VALUE rubrics for other critical audiences and constituencies (e.g., employers, policy makers, parents, and students). This work has the potential to improve not only the content and design of the VALUE rubrics but also the reliable and accurate application of the VALUE rubrics to student work.
- Improving assignment alignment through assignment (re)design. Drawing on the excellent work of and in partnership with organizations like NILOA, AAC&U will redouble its efforts to help faculty and other higher education professionals find and/or (re)design assignments to ensure alignment between what is asked of students and the standards and criteria by which that work will be judged.
- Enhancing VALUE calibration training. Working with experts in performance-based assessment, AAC&U is revising its VALUE rubric training protocols to move to a more robust and rigorous protocol for training scorers, particularly for those scoring student artifacts as part of the VALUE Institute. Resulting protocol guidelines will be made available for local campus use as well, recognizing that individual institutions may choose to modify the protocols to meet local needs.
- Engaging in cohort-based, longitudinal assessment of student learning using ePortfolios. VALUE will revisit one of its founding assumptions: that practice in assessment requires multiple assessments over time, and that well-planned ePortfolios provide opportunities to utilize college data from multiple assessments across a broad range of learning outcomes and modes for expressing learning, while guiding student learning and building reflective self-assessment capabilities. AAC&U will actively seek out resources to create opportunities for collaboration with colleges and universities to engage in meaningful assessments of learning over time to more fully reveal the true value of undergraduate education.
- Building institutional capacity. Work to date has revealed the diversity of experiences with VALUE across campuses, bringing to the forefront issues of institutional equity when it comes to finding the internal resources to engage in the kind of work the VALUE approach requires. AAC&U will work to not only create additional resources and toolkits designed to help individuals implement and use the VALUE approach to assessment at their institutions but also will actively seek opportunities to support our least-resourced institutions in this important work locally while connecting them to the burgeoning nationwide development of a VALUE-generated landscape of learning.

Figure 18. The VALUE balancing act between competing concerns.



Validity is not a static phenomenon, nor is the VALUE approach to assessment. Future work will undoubtedly shape VALUE's validity, perhaps in unanticipated ways. AAC&U will continue to engage in the interrogation of VALUE and to rely on VALUE's champions and critics (who are often one and the same) to continue pushing to ensure that it lives up to its promise and maintains its focus on the lived learning experiences of students and educators on all of our campuses.

References

Acosta-Morales, O., McCool, E., Murphey, K., & Stephens, M. (2016). Increasing underserved student success through faculty intentionality in problem-centered learning. *Peer Review, 18*(1-2), 12-14. Retrieved from https://www.aacu.org/peerreview/2016/winter-spring/Acosta-Morales

Adler-Kassner, L., & Harrington, S. (2010). Responsibility and composition's future in the twenty-first century: Reframing "accountability." *College Composition and Communication, 62*(1), 73-99.

AERA, APA, NCME. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Albertine, S. (2014). Public health learning—Purposeful, progressive, global by design. *Frontiers in Public Health, 2*, Article 230. doi:10.3389/fpubh.2014.00230

Albertine, S. (2015). Gender equity in higher education: Calling for equitable, integrative, and intergenerational leadership. *Diversity and Democracy, 18*(2), 8-11. Retrieved from https://www.aacu.org/diversitydemocracy/2015/spring/albertine

Alschuler, M. (2016). Faculty inter-rater reliability of a reflective journaling rubric. *Kentucky Journal of Excellence in College Teaching and Learning, 14*, 9-20.

Ambrose, S. A., Bridges, M. W., DiPietro, M., Lovett, M. C., & Norman, M. K. (2010). *How learning works: Seven research-based principles for smart teaching*. San Francisco, CA: Jossey-Bass.

Anderson, C., & Blair, D. (2013). Developing a global learning rubric: Strengthening teaching and improving learning. *Diversity and Democracy, 16*(3). Retrieved from https://www.aacu.org/publications-research/periodicals/developing-global-learning-rubric-strengthening-teaching-and

Anson, C. M., Dannels, D. P., Flash, P., & Gaffney, A. L. H. (2012). Big rubrics and weird genres: The futility of using generic assessment tools across diverse instructional contexts. *Journal of Writing Assessment, 5*(1), 1-17. Retrieved from http://www.journalofwritingassessment.org/article.php?article=57

Association of American Colleges and Universities (AAC&U). (2017). *Mission statement*. Retrieved from https://www.aacu.org/about/mission

Augsburg, T. (2012, May). *From rubric testing to curriculum change: Assessing student learning in a pilot interdisciplinary gateway course*. Paper presented at the Michigan State Conference on Interdisciplinary Teaching and Learning, East Lansing, MI.

Banta, T. (2011). Our primitive art of measurement. *Peer Review, 13/14*(4/1), 35. Retrieved from https://www.aacu.org/publications-research/periodicals/our-primitive-art-measurement

Banta, T., & Blaich, C. (2011). Closing the assessment loop. *Change, 43*(1), 22-27.

Banta, T., & Pike, G. (2007). Revisiting the blind alley of value added. *Assessment Update, 19*(1), 1-4.

Banta, T. W., & Pike, B. R. (2012). *Making the case against—One more time* (Occasional Paper No. 15). Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment.

Barnhisel, G., Stoddard, E., & Gorman, J. (2012). Incorporating process-based writing pedagogy into first year learning communities: strategies and outcomes. *Journal of General Education, 61*(4), 461-487.

Baryla, E., Shelley, G., & Trainor, W. (2012). Transforming rubrics using factor analysis. *Practical Assessment, Research & Evaluation, 17*(4), 1-7.

Bauerlein, M. (2012). Civic literacy. *Academic Questions, 25*(3), 328–333.

Baxter, K. B. (2013). Measuring student learning for interfaith cooperation: The pluralism and worldview engagement rubric. *Journal of College and Character, 14*(3), 259-266.

Belanger, J., Zou, N., Mills, J. R., Holmes, C., & Oakleaf, M. (2015). Project RAILS: Lessons learned about rubric assessment of information literacy skills. *Portal: Libraries and the Academy, 15*(4), 623-644.

Berg, J., Grimm, L. M., Wigmore, D., Cratsley, C. K., Slotnick, R. C., & Taylor, S. (2014). Quality collaborative to assess quantitative reasoning: Adapting the LEAP VALUE rubric and the DQP. *Peer Review, 16*(3), 17-21. Retrieved from https://www.aacu.org/peerreivew/2014/summer/berg

Bergfelt, T. W., Brungard, A. B., & Rapchak, M. E. (2016). What's the VALUE of information literacy? Comparing learning community and non-learning community student learning outcomes. *Learning Communities Research and Practice, 4*(1), 1–17.

Bernstein, D., & Greenhoot, A. (2011). Using VALUE rubrics to evaluate collaborative course design. *Peer Review 13/14*(4/1), 22-26.

Bernstein, D., & Greenhoot, A. F. (2014). Team-designed improvement of writing and critical thinking in large undergraduate courses. *Teaching & Learning Inquiry, 2*(1), 39-61.

Berrett, D. (2013, October). States demand that colleges show how well their students learn. *Chronicle of Higher Education, 60*(9), 1.

Berrett, D. (2016, October). The next great hope for measuring learning. *Chronicle of Higher Education, 63*(8), A30-A33.

Blaszczynski, C., & Green, D. J. (2012). Effective strategies and activities for developing soft skills: Part 1. *Journal of Applied Research for Business Instruction, 10*(1), 1-6.

Bodnar, C. A., & Kadlowec, J. A. (2018). Initial validation of a technical writing rubric for engineering design. *International Journal of Engineering Pedagogy, 8*(1), 81-91. Retrieved from http://online-journals.org/index.php/i-jep/article/view/7728/4828

Boersma, S., Diefenderfer, C., Dingman, S., & Madison, B. (2011). Quantitative reasoning in the contemporary world, 3: Assessing student learning. *Numeracy, Advancing Education in Quantitative Literacy, 4*(2), Article 8.

Braid, B., de Schrynemakers, G. P., & Grose, A. W. (2011). Assessing early integrative learning. *Peer Review, 13/14*(4/1), 12-14. Retrieved from https://www.aacu.org/publications-research/periodicals/assessing-early-integrative-learning

Bresciani, M. J., Oakleaf, M., Kolkhorst, F., Nebeker, C., Barlow, J., Duncan, K., & Hickmott, J. (2009). Examining design and inter-rater reliability of a rubric measuring research quality across multiple disciplines. *Practical Assessment, Research & Evaluation, 14*(12), 1-7.

Brown, G. R., & Rhodes, T. L. (2016). *Assessment practices for advancing transfer student success: Collaborating for educational change*. Washington, DC: Association of American Colleges and Universities.

Burke, K., Ouellette, J., Miller, W., Leise, C., & Utschig, T. (2012). Measuring writing as a representation of disciplinary knowledge. *International Journal of Process Education, 4*(1), 13-27.

Carey, S. J. (2011). From the editor. *Peer Review, 13/14*(4/1), 3. Retrieved from https://www.aacu.org/publications-research/periodicals/editor-93

Carlin, N., Rozmus, C., Spike, J., Willcockson, I., Seifert, W., Chappell, C., . . . Boutwell, B. (2011). The health professional ethics rubric: Practical assessment in ethics education for health professional schools. *Journal of Academic Ethics, 9*(4), 277-290.

Carroll, A. J., Tchangalova, N., & Harrington, E. G. (2016). Flipping one-shot library instruction: Using canvas and Pecha Kucha for peer teaching. *Journal of the Medical Library Association, 104*(2), 125-130.

Carroll, J., & Traver, A. E. (2016). Multicultural education and mass violence: Critical examination by two student groups. *Diversity and Democracy, 19*(2), 28-29.

Carter, D. D., & McConnell, D. K. (2018). *VALUE rubrics as tools for pedagogical innovation* [Webinar]. Retrieved from http://leaptx.org/development-hub/value-rubrics/

Carter, D., & Purkiss, C. (2016). *Social responsibility* [PowerPoint slides]. Retrieved from http://leaptx.org/wp-content/uploads/2016/05/LEAP-Texas-Webinar-Social-Responsibility-August-2016.pdf

Chamberlin, R. (2018, April). *Developing global learning in a home-grown short-term program abroad: A case study in Germany*. Paper presented at the Southeast Coastal Conference on Languages and Literatures, Savannah, GA. Retrieved from https://digitalcommons.georgiasouthern.edu/seccll/2018/2018/41/

Cicek, J. S., Ingram, S., Sepehri, N., Burak, J. P., Labossiere, P., Mann, D., . . . Topping, A. (2015). Rubrics as a vehicle to define the twelve CEAB graduate attributes, determine graduate competencies, and develop a common language for engineering stakeholders. *Proceedings of the Canadian Engineering Education Association, Canada, CEEA14*, 1-7. doi:10.24908/pceea.v0i0.5876

City University of New York (CUNY). (2011). *Report of the CUNY task force on system-wide assessment of undergraduate learning gains* (Assessment Task Force Report 2011). Retrieved from City University of New York website: http://www.ny.edu/about/administration/offices/ue/cue/AssessmentTaskForceReport2011.pdf

Clark, J. E., & Eynon, B. (2011). Measuring student progress with e-portfolios. *Peer Review, 13/14*(4/1), 6-8. Retrieved from https://www.aacu.org/publications-research/periodicals/measuring-student-progress-e-portfolios

Coates, H. (2015). Assessment of learning outcomes. In A. Curaj, L. Matei, R. Pricopie, J. Salmi, & P. Scott (Eds.), *The European higher education area: Between critical reflections and future policies*. (pp. 399-413). Springer, Cham.

Cole, T. L., Cochran, L. F., Troboy, L. K., & Roach, D. W. (2012). Efficiency in assessment: Can trained student interns rate essays as well as faculty members? *International Journal for the Scholarship of Teaching and Learning, 6*(2), Article 6. doi:10.20429/ijsotl.2012.060206

Cooney, E. M. (2014, June). *Adapting the VALUE rubrics to ABET ETAC outcomes A-K*. Paper presented at the 121st ASEE Annual Conference & Exposition, Indianapolis, IN. Retrieved from https://peer.asee.org/20027

Core, R., S. (2017). Assessing global learning in short-term study abroad: Population, environment, and society in Shanghai. *Teaching Sociology, 45*(4), 399-408.

Cratsley, C., & Slotnick, R. C. (2013). Working across institutions to define and assess civic learning. *Diversity and Democracy, 16*(3), 13-15. Retrieved from https://www.aacu.org/publications-research/periodicals/working-across-institutions-define-and-assess-civic-learning

Crist, J. F., & Robinson, K. (2015). Stewards of the word: Employing oral examinations in required theology courses. *Journal of Catholic Higher Education, 34*(2), 235-256.

Darabi, A., & Arrington, T. L. (2017). Designing instruction for critical thinking: A case of a graduate course on evaluation of training. *International Journal of Teaching and Learning in Higher Education, 29*(3), 551-559.

Deardorff, D. K. (2011). Assessing intercultural competence. *New Directions for Institutional Research, 149*, 65-79.

Dingman, S. W., & Madison, B. L. (2011). Twenty-first-century quantitative education: Beyond content. *Peer Review, 13*(3), 15. Retrieved from https://www.aacu.org/publications-research/periodicals/twenty-first-century-quantitative-education-beyond-content

DiPardo, A., Storms, B. A., & Selland, M. (2011). Seeing voices: Assessing writerly stance in the NWP analytic writing continuum. *Assessing Writing, 16*(3), 170-188.

Dunbar, N. E., Brooks, C. F., & Kubicka-Miller, T. (2006). Oral communication skills in higher education: Using a performance-based evaluation rubric to assess communication skills. *Innovative Higher Education, 31*(2), 115-128.

Duncan, W. C., & Gleason, B. L. (2015). Knowledge, abilities, responsibilities: The design of a three-dimensional curriculum framework for modern pharmacy education. *International Journal of Process Education, 7*(1), 53-62.

Eppes, T. A., Milanovic, I., & Sweitzer, F. (2012). Towards liberal education assessment in engineering and technology programs. *Journal of College Teaching and Learning, 9*(3), 171-178.

Escoe, G., Hall, W., & Nicholas, M. (2010). *Assessing learning outcomes at the University of Cincinnati: Comparing rubric assessments to standardized tests*. Retrieved from https://www.aacu.org/campus-model/assessing-learning-outcomes-university-cincinnati-comparing-rubric-assessments

Evans, J. J., Van Epps, A. S., Smith, M. T., Matei, S. A., & Garcia, E. (2015, June). *A transdisciplinary approach for developing effective communication skills in a first-year STEM seminar.* Paper presented at the 2015 ASEE Annual Conference & Exposition, Seattle, WA. doi:10.18260/p.23468

Eynon, B., Gambino, L. M., & Török, J. (2014a). Completion, quality, and change: The difference e-portfolios make. *Peer Review, 16*(3), 8-13. Retrieved from https://www.aacu.org/publications-research/periodicals/completion-quality-and-change-difference-e-portfolios-make

Eynon, B., Gambino, L., & Török, J. (2014b). *Outcomes assessment and institutional learning*. Retrieved from http://c2l.mcnrc.org/wp-content/uploads/sites/8/2014/01/Outcomes_Assessment1.pdf

Felten, P., Barnett, B., & Fuller, J. (2014). Supporting religious pluralism at Elon University. *Diversity and Democracy, 17*(4), 23-24. Retrieved from https://www.aacu.org/diversitydemocracy/2014/fall/felten-barnett-fuller

Finley, A. P. (2011). How reliable are the VALUE rubrics? *Peer Review, 13/14*(4/1), 31-34. Retrieved from https://www.aacu.org/publications-research/periodicals/how-reliable-are-value-rubrics

Finley, A., & Rhodes, T. (2013, July). *The VALUE rubrics calibration session: How to start working with rubrics to assess student learning*. Paper presented at the 2013 Institute on Integrative Learning and the Departments: Faculty Leadership for the 21st Century, Portland, OR.

Fitch, P., Steinke, P., & Hudson, T. (2012). Research and theoretical perspectives on cognitive outcomes of service learning. In P. H. Clayton, R. G. Bringle, & J. A. Hatcher (Eds.), *Research on service learning: Conceptual frameworks and assessment* (pp. 1-35). Sterling, VA: Stylus.

Flierl, M., Bonem, E., Maybee, C., & Fundator, R. (2018). Information literacy supporting student motivation and performance: Course-level analyses. *Library and Information Science Research, 40*(1), 30-37.

Frank, B. M., Kaupp, J. A., Simper, N., & Scott, J. (2015). *Multi-method longitudinal assessment of transferrable intellectual learning outcomes*. Paper presented at the 122nd ASEE Annual Conference and Exposition: Making Value for Society, Seattle, WA. Retrieved from https://www.asee.org/public/conferences/56/papers/13653/view

Frank, B., Simper, N., & Kaupp, J. (2017). Formative feedback and scaffolding for developing complex problem solving and modelling outcomes. *European Journal of Engineering Education, 43*(4), 552-568. doi:10.1080/03043797.2017.1299692

Garritano, J. R., & Yatcilla, J. (2014). Participatory design of the active learning center: A combined classroom and library building. In N. F. Foster (Ed.), *Participatory design in academic libraries: New reports and findings* (pp. 88-99). Washington, DC: Council on Library and Information Resources.

Gessell, D. A. (2015). Beyond magical realism: Teaching creative thinking using Gabriel García Márquez's "the handsomest drowned man in the world." *Eureka Studies in Teaching Short Fiction, 11/12*, 84-94.

Gleason, B., Gaebelein, C., Grice, G., Crannage, A., Weck, M., Hurd, P., . . . Duncan, W. (2013). Assessment of students' critical-thinking and problem-solving abilities across a 6-year doctor of pharmacy program. *American Journal of Pharmaceutical Education, 77*(8), Article 166. doi:10.5688/ajpe778166

Goff, L., Potter, M. K., Pierre, E., Carey, T., & Gullage, A. (2015). *Learning outcomes assessment: A practitioner's handbook*. Ontario, Canada: University of Windsor Scholarship at UWindsor, Centre for Teaching and Learning. Retrieved from https://scholar.uwindsor.ca/cgi/viewcontent.cgi?article=1005&context=ctlreports

Golter, P., Brown, G., & Van Wie, B. (2016). Preliminary validation of a critical thinking rubric for fluid mechanics and heat transfer. *Journal of STEM Education, 17*(4), 19-28.

Graff, N., Horvath, F., Shapiro, D., & Wood, S. (2018, January). *Using integrated VALUE rubrics to engage all faculty in reflective assessment that improves teaching and learning*. Paper presented at the 2018 AAC&U Annual Meeting: Can Higher Education Recapture the Elusive American Dream?, Washington, DC. Retrieved from https://csumb.edu/tla/aacu-2018-csumb-faculty-engagement

Gray, J. S., & Brown, M. A. (2017, December). *Operationalizing value rubric dimensions: Case studies on calibration activities to enhance inter-rater agreement*. Paper presented at the SACSCOC Annual Meeting, Dallas, TX. Retrieved from http://www.sacscoc.org/2017AMHo/CS%20Handouts/CS-56_Gray.pdf

Gray, J. S., Brown, M. A., & Connolly, J. P. (2017). Examining construct validity of the quantitative literacy VALUE rubric in college-level STEM assignments. *Research & Practice in Assessment, 12,* 20-31.

Gray, P. (2013, June). *Developing assessment rubrics in project-based courses: Four case studies*. Paper presented at the 9th International CDIO Conference, Cambridge, Massachusetts. Retrieved from http://ww.cdio.org/files/document/file/W5A4_Gray_006.pdf

Greenhoot, A., & Bernstein, D. (2011). Using VALUE rubrics to evaluate collaborative course design. *Peer Review, 13/14*(4/1), 22-26. Retrieved from https://www.aacu.org/publications-research/periodicals/using-value-rubrics-evaluate-collaborative-course-design

Hachtmann, F., Mitchell, N., Mitchell, B., & Sasser, S. (2013, June). *International advertising education in a digital world: Achieving global competency*. Paper presented at the 2013 Global Conference of the American Academy of Advertising, Honolulu, Hawaii. Retrieved from http://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1072&context=journalismfacpub

Hammang, J. M., Rhodes, T. L., Shulenberger, D., & Keller, C. (2010). *Rising to the challenge: Meaningful assessment of student learning.* Washington, DC: Association of American Colleges and Universities.

Hart Research Associates. (2015a). *Bringing equity and quality learning together: Institutional priorities for tracking and advancing underserved students' success.* Washington, DC: AAC&U and Hart Research Associates. Retrieved from https://www.aacu.org/about/2015-membersurvey

Hart Research Associates. (2015b). *Recent trends in general education designs, learning outcomes, and teaching approaches.* Washington, DC: AAC&U and Hart Research Associates. Retrieved from https://www.aacu.org/about/2015-membersurvey

Hart Research Associates. (2015c). *Trends in learning outcomes assessment: Key findings from a survey among administrators at AAC&U member institutions.* Washington, DC: AAC&U and Hart Research Associates. Retrieved from https://www.aacu.org/about/2015-membersurvey

Hart Research Associates. (2018). *Fulfilling the American dream: Liberal education and the future of work. Selected findings from online surveys of business executives and hiring managers.* Washington, DC: AAC&U and Hart Research Associates. Retrieved from https://www.aacu.org/sites/default/files/files/LEAP/2018EmployerResearchReport.pdf

Hatcher, J. A. (2011). Assessing civic knowledge and engagement. *New Directions for Institutional Research, 149,* 81-92. doi:10.1002/ir.382

Hathcoat, J., & Penn, J. (2012). Generalizability of student writing across multiple tasks: A challenge for authentic assessment. *Research & Practice in Assessment, 7,* 16–28.

Hathcoat, J., Penn, J., Barnes, L., & Comer, J. (2016). A second dystopia in education: Validity issues in authentic assessment practices. *Research in Higher Education, 57*(7), 892-912.

Healy, M. A., Lancaster, J. M., Liddell, D. L., & Stewart, D. L. (2012). The role of the campus professional as a moral mentor. *New Directions for Student Services, 139,* 83-92. doi:10.1002/ss.20025

Hebble, A., & Richards, M. (2013). Assessing student critical thinking skills for online quantitative courses. In S. Nenninger (Ed.), *Fifth Annual General Business Conference Proceedings: Vol. 4* (pp. 214-224). Huntsville, TX: Sam Houston State University, College of Business Administration. Retrieved from http://mkwright.org/files/GB_Proceedings_2013-1.pdf

Hecsh, J. (2010). Resource-friendly reform in general education. *Diversity and Democracy, 13*(3), 16-17.

Heinrich, W. F., Habron, G. B., Johnson, H. L., & Goralnik, L. (2015). Critical thinking assessment across four sustainability-related experiential learning settings. *Journal of Experiential Education, 38*(4), 373-393.

Hodges, L. C. (2014). Demystify learning expectations to address grade inflation. *College Teaching, 62*(2), 45-46.

Holliday, W., Dance, B., Davis, E., Fagerheim, B., Hedrich, A., Lundstrom, K., & Martin, P. (2015). An information literacy snapshot: Authentic assessment across the curriculum. *College & Research Libraries, 76*(2), 170-187.

Holmes, C., & Oakleaf, M. (2013). The official (and unofficial) rules for norming rubrics successfully. *Journal of Academic Librarianship, 39*(6), 599-602.

Horan, E.M. (2018). *The use of the VALUE rubrics in higher education: Selected findings from online surveys of faculty, administrators, and higher education professionals.* Retrieved from https://www.aacu.org/value/research/use-surveys

Horowitz, S. M. (2014). Hands-on learning in special collections: A pilot assessment project. *Journal of Archival Organization, 12*(3/4), 216-229.

Hovland, K. (2013). Equitable designs for global learning. *Diversity and Democracy, 16*(2), 10-12.

Hovland, K. (2014). What can global learners do? *Diversity and Democracy, 17*(2), 8-11. Retrieved from https://www.aacu.org/diversitydemocracy/2014/spring/hovland

Hoy, A., & Wolfe, K. (2016). High-impact learning for self and society: Community-engaged signature work. *Diversity and Democracy, 19*(4), 4-8.

Hubert, D. (2013). *Assessing general education outcomes with ePortfolios at SLCC*. Retrieved from http://slcc.mcnrc.org/oa-story/

Hubert, D. (2016). ePortfolios, assessment, and general education transformation. *Peer Review, 18*(3), 25-28. Retrieved from https://www.aacu.org/peerreview/2016/summer/Hubert

Hubert, D. (2018, June). *Signature assignments and reflection in general education*. PowerPoint presented at the Institute on General Education and Assessment, Salt Lake City, UT. Retrieved from https://www.aacu.org/sites/default/files/files/Signature%20Assignments%20and%20Reflection%20in%20General%20Education.pdf

Hughes, R. L., & Jones, S. K. (2011). Developing and assessing college student teamwork skills. *New Directions for Institutional Research, 149*, 53-64. doi:10.1002/ir.380

Jankowski, N. A., Timmer, J. D., Kinzie, J., & Kuh, G. D. (2018). *Assessment that matters: Trending towards practices that document authentic student learning*. Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment. Retrieved from http://www.learningoutcomesassessment.org/documents/NILOA2018SurveyReport.pdf

Jansen, K. R., Mihlbauer, L., & Mayeshiba, M. (2016). *The value of critical thinking: Designing assignments that evaluate critical thinking using AAC&U VALUE rubrics: A pilot study.* Milwaukee, WI: University of Wisconsin. Retrieved from https://ce.uwex.edu/wp-content/uploads/2016/02/aacn2016-FINALtoPRINTreduced.pdf

Jardeleza, S., Cognato, A., Gottfried, M., Kimbirauskas, R., Libarkin, J., Olson, R., . . . Thomas, S. (2013). The value of community building: One center's story of how the VALUE rubrics provided common ground. *Liberal Education, 99*(3), 52-57. Retrieved from https://www.aacu.org/publications-research/periodicals/value-community-building-one-centers-story-how-value-rubrics

Joe, J. N., Harmes, J. C., & Hickerson, C.A. (2011). Using verbal reports to explore rater perceptual processes in scoring: A mixed methods application to oral communication assessment. *Assessment in Education: Principles, Policy & Practice, 18*(3), 239-258. doi:10.1080/0969594X.2011.577408

Johnson, S., Veitch, S., & Dewiyanti, S. (2015). A framework to embed communication skills across the curriculum: A design-based research approach. *Journal of University Teaching & Learning Practice, 12*(4), Article 6. Retrieved from https://ro.uow.edu.au/jutlp/vol12/iss4/6/

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review, 2*(2), 130-144.

Kane, M. T. (2002). Validating high-stakes testing programs. *Educational Measurement, 21*(1), 31-41. doi:10.1111/j.1745-3992.2002.tb00083.x

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17-64). New York, NY: American Council on Education/Macmillan.

Keen, C. (2009). New efforts to assess civic outcomes. *Journal of College and Character, 10*(7). doi:10.2202/1940-1639.1689

Kiel, S., Burclaff, N., & Johnson, C. (2015). Learning by doing: Developing a baseline information literacy assessment. *Portal: Libraries and the Academy, 15*(4), 747-766.

Kinzie, J., Helms, R., & Cole, J. (2017). A glimpse of global learning: assessing student experiences and institutional commitments. *Liberal Education, 103*(2), 30-37. Retrieved from https://www.aacu.org/liberaleducation/2017/spring/kinzie_helms_cole

Klein-Collins, R. (2013). *Sharpening our focus on learning: The rise of competency-based approaches to degree completion* (Occasional Paper No. 20). Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment. Retrieved from http://learningoutcomesassessment.org/occasionalpapertwenty.html

Komvies, S. R., & Smedick, W. (2012). Using standards to develop student learning outcomes. *New Directions for Student Services, 140*, 77-88. doi:10.1002/ss

Kruse, G., & Drews, D. (2013). Using performance tasks to improve quantitative reasoning in an introductory mathematics course. *International Journal for the Scholarship of Teaching and Learning, 7*(2), Article 19. doi:10.20429/ijsotl.2013.070219

Kuh, G., & Ikenberry, S. (2009). *More than you think, less than we need: Learning outcomes assessment in American higher education*. Champaign, IL: National Institute for Learning Outcomes Assessment. Retrieved from http://www.learningoutcomeassessment.org/documents/niloafullreportfinal2.pdf

Laguilles, J. S., Coughlin, M. A., Kelly, H. A., & Walters, A. M. (2016). Postgraduate outcomes in American higher education. *New Directions for Institutional Research, 169*, 11-23. doi:10.1002/ir.20166

Lederman, D. (2007, September 27). Meeting of the minds. *Inside Higher Ed*. Retrieved from https://www.insidehighered.com/news/2007/09/27/meeting-minds

Lee, J. J., Townend, P. A., Boersma, J. M., Tippett, T. P., Clements, C. M., & Hicks, R. A. (2018). Promoting community engagement: A campus-wide approach to applied learning. *Partnerships, 9*(1), 31-42.

Lee, V. S. (2012). What is inquiry-guided learning? *New Directions for Teaching and Learning, 129,* 5-14.

Leist, C. W., Woolwine, M. A., & Bays, C. L. (2012). The effects of using a critical thinking scoring rubric to assess undergraduate students' reading skills. *Journal of College Reading and Learning, 43*(1), 31-58.

Liu, O. L. (2017). Ten years after the Spellings Commission: From accountability to internal improvement. *Educational measurement: Issues and practice.*1-8. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/emip.12139/full

Liu, O. L., Frankel, L., & Roohr, K. C. (2014). *Assessing critical thinking in higher education: Current state and directions for next-generation assessment* (ETS Research Report No. RR-14-10). Retrieved from the Educational Testing Services: doi:10.1002/ets2.12009

Lumina Foundation for Education. (2011). *Degree qualifications profile*. Indianapolis, IN: Lumina Foundation for Education. Retrieved from http://www.luminafoundation.org/resources/dqp

Madison, B. (2012). If only math majors could write. *Numeracy, 5*(1), Article 6. Retrieved from https://scholarcommons.usf.edu/numeracy/vol5/iss1/art6/

Maki, P. L. (2015). *Assessment that works: A national call, a twenty-first-century response*. Washington, DC: Association of American Colleges and Universities.

Mandeville, D., & Stoner, M. (2015). Assessing the effect of problem-based learning on undergraduate student learning in biomechanics. *Journal of College Science Teaching, 45*(1), 66-75.

McClure Brenchley, K. J., & Donahue, L. M. (2017). Stress reduction in a high stress population: A service-learning project. *Journal of Social and Political Psychology, 5*(2), 463-476.

McConnell, K. D. (2013). Rubrics as catalysts for collaboration: A modest proposal. *European Journal of Higher Education, 3*(1), 74-88.

McConnell, K. D., & Doolittle, P. E. (2012). Classroom-level assessment: Aligning pedagogical practices to enhance student learning. In C. Secolsky & B. D. Denison (Eds.), *Handbook on measurement, assessment, and evaluation in higher education* (ch. 2, pp. 15-30). New York, NY: Routledge.

McConnell, K. D., & Rhodes, T. L. (2017). *On solid ground*. Washington, DC: Association of American Colleges and Universities.

McCormack, J., Williams, B., Noren, K., Beyerlein, S., Cordon, D., & Morgan, J. (2012). Development and usability of a generic quantitative problem-solving rubric for student learning. *International Journal of Process Education, 4*(1), 29-41.

McKitrick, S. A., & Barnes, S. M. (2012). Assessment of critical thinking: An evolutionary approach. *Journal of Assessment and Institutional Effectiveness, 2*(1), 1-29.

McNair, T. B. (2016). The time is now: Committing to equity and inclusive excellence. *Diversity and Democracy, 19*(1), 4-7. Retrieved from https://www.aacu.org/diversitydemocracy/2016/winter/mcnair

Mentkowski, M., & Sharkey, S. (2011). How we know it when we see it: Conceptualizing and assessing integrative and applied learning-in-use. *New Directions for Institutional Research, 149*, 93-107.

Nicholas, M. C., Hathcoat, J. D., & Brown, B. L. (2016). The multi-state collaborative: A preliminary examination of convergent validation evidence. *Assessment Update, 28*(3), 5-14.

Norman, C. R. (2017). *Students' performance on institutional learning outcomes*. St. Cloud, MN: St. Cloud State University. Retrieved from http://repository.stcloudstate.edu/hied_etds

Oakleaf, M. (2007). Using rubrics to collect evidence for decision-making: What do librarians need to learn? *Evidence Based Library and Information Practice, 2*(3), 27-42.

Oakleaf, M. (2009). Using rubrics to assess information literacy: An examination of methodology and interrater reliability. *Journal of the American Society for Information Science and Technology, 60*(5), 969-983.

Oakleaf, M. (2011). Staying on track with rubric assessment: Five institutions investigate information literacy learning. *Peer Review, 13/14*(4/1), 18-21.

O'Brien, K. (2013). *Deakin learning futures agenda 2020: Stage 2: Assessment and learning design*. Victoria, Australia: Deakin Learning Futures. Retrieved from https://www.deakin.edu.au/__data/assets/pdf_file/0004/19327/clo-rubric-aqf7.pdf

Oderda, G. M., Zavod, R. M., Carter, J. T., Early, J. L., Joyner, P. U., Kirschenbaum, H., . . . Plaza, C. M. (2010). An environmental scan on the status of critical thinking and problem solving skills in colleges/schools of pharmacy: Report of the 2009-2010 academic affairs standing committee. *American Journal of Pharmaceutical Education, 74*(10), Article S6. doi:10.5688/aj7410S6

Osadjan, J., Jackson, C., Jallah, J., & Joel, T. (2015). *Handbook for improved assessment of student learning outcomes in short-term faculty-led study abroad.* Chicago, IL: Roosevelt University. Retrieved from http://sites.miis.edu/courtneyannejackson/files/2015/02/Jackson-Joel-Jallah_-Final-Design-and-Assessment-Project.pdf

Overby, L., & Meiman, M. (2013). *Assessing student learning in undergraduate research*. Retrieved from http://ud.mcnrc.org/ref-practice/

Peet, M. R., Reynolds-Keefer, L., & Gurin, P. (2011). Fostering integrative knowledge and lifelong learning. *Peer Review, 13/14*(4/1), 15-17. Retrieved from https://www.aacu.org/publications-research/periodicals/fostering-integrative-knowledge-and-lifelong-learning

Penn, J. D. (2011a). Future directions for assessing complex general education student learning outcomes. *New Directions for Institutional Research, 14)* 109-117. doi:10.1002/ir.384

Penn, J. D. (2011b). The case for assessing complex general education student learning outcomes. *New Directions for Institutional Research, 149*, 5-14. doi:10.1002/ir.376

Penny Light, T., Chen, H., & Ittelson, J. C. (2011). Using ePortfolios to support assessment. In *Documenting Learning with ePortfolios* (pp. 95–104). San Francisco, CA: Jossey-Bass.

Perie, M. (2013). *Developing a validity argument for the VALUE rubrics*. Unpublished manuscript.

Phillips, L., & McLawhon, R. (2013). Assessing intercultural competence. *Diversity and Democracy, 16*(3), 16-17.

Pike, G. (2014). Assessment measures: Course embedded assessment designs - lessons from the MSC. *Assessment Update: Progress, Trends, and Practices in Higher Education, 26*(6), 7–8, 10. doi:10.1002/au.

Pike, G. (2016). Sources of variation in evaluations of student artifacts: A comment on research by Nicholas, Hathcoat, and Brown. *Assessment Update, 28*(3), 7-13. doi:10.1002/au.30059

Pike, G. (2018). *The dependability of student-, institution-, and state-level scores for critical thinking, written communication, and quantitative literacy: Final report to AAC&U*. Unpublished manuscript.

Provezis, S. (2012). *LaGuardia Community College: Weaving assessment into the institutional fabric.* Retrieved from http://www.learningoutcomeassessment.org/documents/LaGuardiaCC.pdf

Pusecker, K. L., Torres, M. R., Crawford, I., Levia, D., Lehman, D., & Copic, G. (2011). Increasing the validity of outcomes assessment. *Peer Review, 13/14*(4/1), 27-30. Retrieved from https://www.aacu.org/publications-research/periodicals/increasing-validity-outcomes-assessment

Raczynski, K. R., Cohen, A. S., Engelhard, G., & Lu, Z. (2015). Comparing the effectiveness of self-paced and collaborative frame-of-reference training on rater accuracy in a large-scale writing assessment. *Journal of Educational Measurement, 52*(3), 301-318.

Rapchak, M. E., Lewis, L. A., Motyka, J. K., & Balmert, M. (2015). Information literacy and adult learners: Using authentic assessment to determine skill gaps. *Adult Learning, 26*(4), 135–142.

Rather, K. F. (2016). Learning to LEAP. *Assessment Update*, *28*(1), 11-14.

Reason, R. D., (2013). Creating and assessing campus climates that support personal and social responsibility. *Liberal Education, 99*(1), 38-43. Retrieved from https://www.aacu.org/publications-research/periodicals/creating-and-assessing-campus-climates-support-personal-and-social

Reason, R. D., & Hemer, K. (2014). *Civic learning and engagement: A review of the literature on civic learning, assessment, and instruments*. Retrieved from https://www.aacu.org/sites/default/files/files/qc/CivicLearningLiteratureReviewRev1-26-15.pdf

Reddy, Y. M., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education, 35*(4), 435–448.

Reed, C.-R., Garcia, L. I., Slusser, M. M., Konowitz, S., & Yep, J. (2017). Linking essential learning outcomes and interprofessional collaborative practice competency in health science undergraduates. *Journal of the Scholarship of Teaching and Learning, 17*(1), 15-23.

Rezaei, A. R., & Lovorn, M. (2010). Reliability and validity of rubrics for assessment through writing. *Assessing Writing, 15*, 18–39.

Rhodes, T. (2008). VALUE: Valid assessment of learning in undergraduate education. *New Directions for Institutional Research, S1,* 59-70. doi:10.1002/ir.262

Rhodes, T. L. (2010a). *Assessing outcomes and improving achievement: Tips and tools for using rubrics*. Washington, DC: Association of American Colleges and Universities.

Rhodes, T. (2010b). Since we seem to agree, why are the outcomes so difficult to achieve? *New Directions for Teaching and Learning*, *121*, 13-21. doi:10.1002/tl.384

Rhodes, T. L. (2010c). Valid assessment of learning in undergraduate education. In *Rising to the challenge: Meaningful assessment of student learning* (pp. 16-25). Washington, DC: Association of American Colleges and Universities.

Rhodes, T. (2011). Emerging evidence on using rubrics. *Peer Review, 13/14*(4/1), 4-5. Retrieved from https://www.aacu.org/publications-research/periodicals/emerging-evidence-using-rubrics

Rhodes, T. L. (2016). The VALUE of assessment: Transforming the culture of learning. *Change: The Magazine of Higher Learning, 48(5)*, 36-43.

Rhodes, T., Bischof, M., Milner, J., Hines, S., & McNealy, T. (2014, January). *The value of quality degrees*. Paper presented at the Association of American Colleges and Universities Annual Meeting, Washington, DC.

Rhodes, T. L., & Finley, A. (2013). *Using the VALUE rubrics for improvement of learning and authentic assessment.* Washington, DC: Association of American Colleges and Universities.

RiCharde, R. S. (2008). The humanities versus interrater reliability. *Assessment Update*, *20*(4), 1-16.

Richards-Schuster, K., Ruffolo, M. C., Nicoll, K. L., Distelrath, C., & Galura, J. A. (2014). Using ePortfolios to assess program goals, integrative learning, and civic engagement: A case example. *International Journal of ePortfolio, 4*(2), 133-141.

Riebe, L., & Jackson, D. (2014). The use of rubrics in benchmarking and assessing employability skills. *Journal of Management Education, 38*(3), 319-344.

Roehling, P. V. (2018). *Flipping the college classroom: An evidence-based guide*. Cham, Switzerland: Springer International. doi:10.007/978-3-319-6939277

Saxton, E., Belanger, S., & Becker, W. (2012). The Critical Thinking Analytic Rubric (CTAR): Investigating intra-rater and inter-rater reliability of a scoring mechanism for critical thinking performance assessments. *Assessing Writing, 17*(4), 251-270.

Schlitz, S. A., O'Connor, M., Pang, Y., Stryker, D., Markell, S., Krupp, E., . . . Redfern, A. K. (2009). Developing a culture of assessment through a faculty learning community: A case study. *International Journal of Teaching and Learning in Higher Education, 21*(1), 133-147.

Self, C., Cady, D., & Olson, M. (2014). Connecting democratic engagement and global learning in general education. *Diversity and Democracy, 17*(3), 14-16.

Shada, A., Cox, R., & Malik, S. (2011). Growing a new culture of assessment: Planting ePortfolios in the metro academies program. *International Journal of ePortfolio, 1*(1), 71-83.

Shapiro, D. F. (2012). Collaborative faculty assessment of service-learning student work to improve student and faculty learning and course design. *Michigan Journal of Community Service Learning, 19*(1), 44–57.

Shea, P., & Turner, B. (2013). *The interstate passport initiative: A new transfer framework* [Webinar]. Retrieved from https://interstatepassport.wiche.edu/webinars

Siefert, L. (2011). Assessing general education learning outcomes. *Peer Review, 13/14*(4/1), 9-11. Retrieved from https://www.aacu.org/publications-research/periodicals/assessing-general-education-learning-outcomes

Simper, N. (2018). Rubric authoring tool for supporting the development and assessment of cognitive skills in higher education. *Teaching and Learning Inquiry, 6*(1), 10-24.

Simper, N., Kaupp, J., Frank, B., & Scott, J. (2016). Development of the transferable learning orientations tool: Providing metacognitive opportunities and meaningful feedback for students and instructors. *Assessment and Evaluation in Higher Education, 41*(8), 1159-1175.

Sorey, K. C., & DeMarte, D. T. (2013). Assessment of general education: Adapting the AAC&U value rubrics. *Peer Review, 15*(2), 27-28. Retrieved from https://www.aacu.org/publications-research/periodicals/assessment-general-education-adapting-aacu-value-rubrics

Stanley, T. A. (2016). Ringing true: Applied civic learning at Emory & Henry College. *Diversity and Democracy, 19*(4), 20-21.

Steinberg, K. S. (2011). Assessing civic mindedness. *Diversity and Democracy, 14*(3), 12-14. Retrieved from https://www.aacu.org/publications-research/periodicals/assessing-civic-mindedness

Steinke, P., & Fitch, P. (2011). Outcome assessment from the perspective of psychological science: The TAIM approach. *New Directions for Institutional Research, 149,* 15-26.

Stevens, D. D., & Levi, A. J. (2013). *Introduction to rubrics: An assessment tool to save grading time, convey effective feedback, and promote student learning*. Sterling, VA: Stylus.

Stufft, M. (2013). Putting collaboration front and center: Assessment strategies for theatre departments. *Theatre Topics, 23*(1), 53-67.

Stupans, I., McAllister, S., Clifford, R., Hughes, J., Krass, I., March, G., . . . Woulfe, J. (2015). Nationwide collaborative development of learning outcomes and exemplar standards for Australian pharmacy programmes. *International Journal of Pharmacy Practice, 23*(4), 283-291.

Sullivan, D. F. (2015). *The VALUE breakthrough: Getting the assessment of student learning in college right*. Washington, DC: Association of American Colleges and Universities.

Sullivan, D. F., & McConnell, K. D. (2017). Big progress in authentic assessment, but by itself not enough. *Change, 49(1)*, 14-25.

Szafran, R. F. (2017). The miscalculation of interrater reliability: A case study involving the AAC&U VALUE rubrics. *Practical Assessment, Research and Evaluation, 22*(11), 1-7.

Thyer, E. (2015). *Critical thinking.* Retrieved from http://www.deakin.edu.au/__data/assets/pdf_file/0012/51222/critical-thinking.pdf

Timmerman, B. E. C., Strickland, D. C., Johnson, R. L., & Payne, J. R. (2011). Development of a "universal" rubric for assessing undergraduates' scientific reasoning skills using scientific writing. *Assessment and Evaluation in Higher Education, 36*(5), 509-547.

Trace, J., Meier, V., & Janssen, G. (2016). "I can see that": Developing shared rubric category interpretations through score negotiation. *Assessing Writing, 30,* 32–43.

Traver, A. E., & Nedd, R. (2018). Incorporating campus-based cultural resources into humanities courses. *Liberal Education, 104*(1), 6. Retrieved from https://www.aacu.org/liberaleducation/2018/winter/traver_nedd

Tunstall, S., Matz, R., & Craig, J. (2018). Quantitative literacy courses as a space for fusing literacies. *Journal of General Education, 65*(3/4), 178.

Turbow, D. J., & Evener, J. (2016). Norming a VALUE rubric to assess graduate information literacy skills. *Journal of the Medical Library Association, 104*(3), 209-214.

U.S. Department of Education (USDOE), Commission on the Future of Higher Education. (2006). *A test of leadership: Charting the future of U.S. higher education.* Retrieved from https://www2.ed.gov/about/bdscomm/list/hiedfuture/index.html

U.S. Department of Education (USDOE), Office of Postsecondary Education. (n.d.). *Fund for the Improvement of Postsecondary Education (FIPSE).* Retrieved from https://www2.ed.gov/about/offices/list/ope/fipse/index.html

U.S. Department of Education (USDOE), Office of Postsecondary Education. (2007). *FY 2007 comprehensive program—New grant awards by organization.* Retrieved from https://www2.ed.gov/programs/fipsecomp/cp-awards-2007.pdf

Wang, J., Engelhard, G., Raczynski, K., Song, T., & Wolfe, E. W. (2017). Evaluating rater accuracy and perception for integrated writing assessments using a mixed-methods approach. *Assessing Writing, 33,* 36-47.

Watson, C. E., Kuh, G. D., Rhodes, T., Penny Light, T., & Chen, H. (2016). Editorial: ePortfolios – the eleventh high impact practice. *International Journal of EPortfolio, 6*(2), 65–69.

Weaver, K. F., Morales, V., Nelson, M., Weaver, P. F., Toledo, A., & Godde, K. (2016). The benefits of peer review and a multi-semester capstone writing series on inquiry and analysis skills in an undergraduate thesis. *CBE: Life Sciences Education, 15*(4), 1-9.

Welsh, J. (2017). How do pros at large, decentralized institutions manage outcomes data? *Intersection of Assessment and Learning*, *Fall,* 38–41. Retrieved from https://cdn.ymaws.com/www.aalhe.org/resource/resmgr/docs/Int/AAHLE_Fall_2017_Intersecton.pdf

Whitehead, D. M. (2015). Global service learning: Addressing the big challenges. *Diversity and Democracy, 18*(3), 8-13. Retrieved from https://www.aacu.org/diversitydemocracy/2015/summer/whitehead

Williams, T. R. (2013). Examine your LENS: A tool for interpreting cultural differences. *Frontiers, 22*, 148-165.

Woodward, K. M. (2015). Information literacy in competency-based education: Reflections on the flex option at the University of Wisconsin-Milwaukee. *College and Research Libraries News, 76*(3), 118-121.

Wozniak, N. M. (2012). Enhancing inquiry, evidence-based reflection, and integrative learning with the lifelong ePortfolio process: The implementation of integrative ePortfolios at Stony Brook University. *Journal of Educational Technology Systems, 41*(3), 209–230.

Wyman, S. M., Waldo, J. T., & Doherty, D. (2016). Methods and models for museum learning at the Samuel Dorsky Museum of Art. *Journal for Learning Through the Arts, 12*(1). doi:10.21977/D912127457

Appendix

The literature review synthesized how, and in what ways, individuals and institutions engage with the AAC&U VALUE rubrics. For the purposes of this review, "engage" referred to sources that 1) documented the extent of use, modification, or omission of the VALUE rubrics during pedagogical development and assessment, 2) described calibration and scoring practices for the VALUE rubrics, or 3) tested the validity, reliability, and/or generalizability of the VALUE rubrics through empirical study. "Individuals" and "institutions" referred to those affiliated with higher education. While not an exhaustive review, the authors' aim was to situate VALUE rubric engagement and test findings within the larger body of rubric research by selecting sources about general and VALUE-specific rubrics that were published over a broad timeline to ensure comprehensiveness (2006–2018). A systematized literature review allowed for variations in engagement and testing to be deeply explored and provided preliminary explanations as to why, for example, the initial intent for the VALUE rubrics does not mirror most use cases (e.g., ePortfolio pedagogy and assessment is the exception, not the rule).

Marianne Perie's working paper for AAC&U, *Developing a Validity Argument for the VALUE Rubrics* (2013), and the aforementioned interpretive argument-based requirements to evaluating validity (McConnell, Horan, Zimmerman, & Rhodes, pp. 2–6) served as a framework for source inclusion. Perie's claims 4–9 provided thematic guidance and the assumptions served as sub-theme domains for investigation (2013, pp. 8–11). This method enabled the authors to detect past and current engagement and testing with the VALUE rubrics as well as identify areas for future research and improvement of the VALUE approach to assessment.

The review of sources was conducted as an environmental scan between May 2018 and August 2018 via keyword and phrase search using EBSCOHost, GALILEO Scholar, ERIC Institute of Education Sciences, and Google Scholar. Initial VALUE-specific sources included those produced by AAC&U and VALUE partners (e.g., development campuses), and the authors used those publication citations to expand their inquiry. In addition to the inclusion framework described above, the credibility and assortment of sources was limited to peer-reviewed journal articles (hard and electronic modes), industry-expert websites, webinars, and occasional white papers (e.g., NILOA), along with conference papers, presentations, and communications from assessment and higher education professionals.

Over 200 sources were organized and coded in NVivo software by claim and sub-claim themes via in-text analysis. Themes included VALUE rubric use, modification, and reasons for not using the rubrics; professional development on how to apply the VALUE rubrics; calibration and scoring methods; as well as VALUE rubric measure correlations. The authors refined the literature for inclusion in this publication based on relevance (i.e., those most relevant to the validity framework). While concise, future literature reviews should be conducted in other areas, such as K–12 education use, multimodal assessment applications, alternative assessment tools inspired by the VALUE rubrics, or concerns related to accessibility.

About the Authors

Kathryne (Kate) Drezek McConnell is the Assistant Vice President for Research and Assessment at AAC&U, where she focuses on using course-embedded assessment to improve teaching and learning while also addressing accountability and accreditation requirements. She brings extensive campus-based experience in assessment, evaluation, general education reform, and faculty development to this work. She holds a PhD in Educational Psychology from Virginia Tech, where she also served as affiliate faculty, teaching courses on cognition, information processing, and college teaching.

Erin M. Horan is the Postdoctoral Research Analyst in the Office of Quality, Curriculum, and Assessment at AAC&U. Her research at AAC&U focuses on the assessment of student learning outcomes in higher education, specifically related to the validity of assessment techniques that utilize authentic student work. She teaches courses in educational psychology and applied research methods in education at the University of Georgia. Erin holds a PhD from the University of Georgia in Educational Psychology, Applied Cognition and Development.

Bethany Zimmerman is the Research Associate in AAC&U's Office of Quality, Curriculum, and Assessment, where she coordinates project activities related to the VALUE Institute and Summer Institute on General Education and Assessment. In collaboration with Erin Horan, she has written research on VALUE and LEAP initiatives. She holds a MA in Political Science and a BA in Interdisciplinary Studies from Virginia Tech, where she taught courses on Global Economy and Politics and on Nations and Nationalities, focusing on primordialism, reification, and the nation-state.

Terrel L. Rhodes is Vice President for the Office of Quality, Curriculum, and Assessment at AAC&U, where he focuses on the quality of undergraduate education, access, general education, and assessment of student learning. He is also the Executive Director of the Valid Assessment of Learning in Undergraduate Education (VALUE) project and director of the annual AAC&U Institute on General Education and Assessment. He holds a PhD in political science from the University of North Carolina at Chapel Hill. Before moving into national higher education work, he was a faculty member for 25 years.

About AAC&U

The Association of American Colleges and Universities (AAC&U) is the leading national association dedicated to advancing the vitality and public standing of liberal education by making quality and equity the foundations for excellence in undergraduate education in service to democracy. Its members are committed to extending the advantages of a liberal education to all students, regardless of academic specialization or intended career. Founded in 1915, AAC&U now comprises 1,400 member institutions—including accredited public and private colleges, community colleges, research universities, and comprehensive universities of every type and size.

AAC&U functions as a catalyst and facilitator, forging links among presidents, administrators, faculty, and staff engaged in institutional and curricular planning. Through a broad range of activities, AAC&U reinforces the collective commitment to liberal education at the national, local, and global levels. Its high-quality programs, publications, research, meetings, institutes, public outreach efforts, and campus-based projects help individual institutions ensure that the quality of student learning is central to their work as they evolve to meet new economic and social challenges.

Information about AAC&U can be found at www.aacu.org.



ISBN 978-0-9961404-7-8
90000
9 780996 140478

Association of American Colleges and Universities

1818 R St. NW, Washington, DC 20009
www.aacu.org