Ten Core Competencies Assessment Scales (CCAS): Validity Evidence for Measures of Core Competencies in Liberal Arts Education

Ascan Koerner, Judith Anderson, Colin McFadden, Kyle Nickodem, Danielle N. Dupuis, Mohammed A. A. Abulela

University of Minnesota

Abstract

It is generally accepted that college education in general, and liberal arts education in particular, helps students develop important career and world readiness competencies (Boyer 2030 Commission, 2023). There is, however, no well established instrument supporting a scalable approach to the reliable and valid assessment of students' development of these competencies. In a series of eight studies involving data from more than 55,000 undergraduate students, we report on the development of an assessment instrument for ten core career and world readiness competencies, from theoretical conceptualization to demonstrating predictive validity of the instrument. Specifically, we collected four sources of validity evidence that support score interpretations and uses: (a) validity evidence based on the assessment content, (b) validity evidence based on response processes, (c) validity evidence based on internal structure, and (d) validity evidence based on association with other variables. Stated differently, we evaluated the psychometric properties of the assessment scales, demonstrated their factor structures, score reliability, and measurement invariance across demographic groups. We then present empirical evidence that shows scores obtained from the scales are not only reliable, but also have convergent, discriminant, and predictive validity. We conclude that score interpretations from the ten Core Competency Assessment Scales developed as part of the RATE™ (reflect, articulate, translate, evaluate) tool are a reliable, valid, and scalable solution for institutions interested in assessing their students' competency development to aid student learning and document institutional effectiveness.

Introduction

Liberal arts education has been the bedrock of higher education in the United States since the first colleges and universities opened their doors in the 17th and 18th century, in what were then British and Spanish colonies. Despite this long and proud tradition, in recent years, there has been an ongoing public debate over the continued value of a liberal arts education, the return on investment for students, and the relevance of liberal arts in a digital future that many critics see as dominated by science and technology. At the same time, increasing polarization of the political discourse has placed liberal arts in the middle of a culture war between progressive and conservative political forces. Our purpose here, however, is not to rehash the arguments for and against the liberal arts. Rather, we are persuaded that liberal arts education (in many spaces general education) is valuable because it leads students to develop a set of competencies giving them distinct advantages in their future professional, public, and private lives. We concede, however, that currently there is no easy and scalable way to assess students' development of these competencies. As a consequence, the value of a liberal arts education is difficult to demonstrate empirically.

In this paper, we address this lacuna by describing the development of 10 easily administered assessment scales that measure students' development of 10 core competencies inherent to the liberal arts (Koerner, 2017). These assessments were developed as part of a comprehensive learning process and creation of a tool for liberal arts students (RATE™: reflect, articulate, translate, evaluate) that aids them with developing core competencies. RATE™ was developed by faculty and staff of the College of Liberal Arts at the University of Minnesota as part of the College's Career Readiness Initiative, the centerpiece of its strategic plan unveiled in 2014. First, we describe the process of defining and conceptualizing the ten competencies. Then, we illustrate the rigorous, eight-year process of development and validation of the assessments that represent the evaluation portion of the award-winning RATE™ tool, its psychometric properties and studies showing its congruent, discriminate, and predictive validity that combine to represent overwhelming quantitative evidence that strongly suggest these ten assessment scales are reliable and valid measures of the core competencies.

Purpose and Overview of RATE™

The process of identifying the core competencies and developing reliable measures of them originated with a Career Readiness Initiative in the College of Liberal Arts (CLA) at the University of Minnesota. The initiative centered around identifying the competencies students develop through their liberal arts education and best prepare them for post-graduate life and career success and developing pedagogy for teaching these competencies more explicitly to our students. The latter is achieved mainly by training faculty and students to use the metacognitive process of reflection, articulation, translation, and evaluation (RATE™) in regard to specific competencies they develop in their liberal arts education. This learning process is supported by a set of learning strategies and tools, including the innovative RATE™ tool, a web-based application that prompts students to engage in the four step process of reflecting, articulating, translating, and evaluating their competency development. RATE™ helps students foster and understand their development of core competencies by guiding them through the process of reflecting on and articulating how they have developed a particular core competency and how they can translate it outside the context of their courses, for example, at work. In its final step, evaluation, it allows students to self-assess the extent to which they have developed a particular competency by completing the core competency assessment scale.

A secondary function of RATE™ is to use student assessments of their competency development to inform faculty and departments about how their courses and curricula contribute to students' development of their core competencies. The data also allows the College and departments to be informed about and monitor students' development of the competencies over time. Based on this data, faculty may make adjustments to their courses to better facilitate student competency development, while departments and administrators at the college-level can identify areas for student support via advising and career coaching, for example. The competencies also serve as the default student learning outcomes for CLA, which means that core competency assessment data collected via assigning students to complete a RATE™ can be used to support student learning outcomes assessment. Supporting these goals does not require the assessments to be completed as part of the RATE™ process. Rather, the assessments can be used independently for assessment of competency development only.

Conceptualization and Development of the Core Competencies & Assessment Scales

There have been two interrelated processes in the development of the competencies and their assessments that both began in Fall of 2015. The first process was the conceptual development of the core competencies and the second process was about the development of assessments of the competencies. After the first process resulted in definitions and operationalization of ten core competencies, and the second process resulted in initial versions of the assessments, the RATE™ tool was deployed and used by faculty and students starting in Fall 2017. Since this initial deployment, the RATE™ tool, including instructions and the prompts moving students through the RATE™

process, have evolved together with the ten assessments. Specifically, the ongoing development of the assessments resulted in several changes described below leading to updates and continuous re-evaluation of RATE™. This means that the ongoing student feedback we have received has been critical to the development and refinement of the tool and that this student data has driven much of the assessments' development. In addition, students have also directly participated in the development of the assessments through participation in "think aloud" sessions, usability lab evaluations, and course evaluations. Their feedback influenced the ongoing refinement of the language and usability of the RATE™ tool in general, and of the assessments in particular, including the structure and content of the scale items. In total, data from over 55,000 student participants has informed the development of the assessments. This frequent and authentic engagement of students in this process guarantees the assessments are student-centered and meaningful to respondents and users in a way that assessments designed entirely or mostly by practitioners or researchers seldom are. This approach has also surfaced a variety of issues that would have otherwise gone entirely unnoticed. For example, the initial version of the tool presented students with a choice for evaluating "curricular" or "co-curricular" experiences. During usability testing, we observed students expressing confusion over the difference between curricular and co-curricular activities. Based on this feedback, we changed the language prompting students to select between "course" and "non-course" experiences. Opportunities like these allowed us to learn from students and apply their feedback to improve the tool which has been critical for its development as a reliable and valid measure.

Develop Definitions and Initial Assessments

Defining Competencies

The ten core competencies are outcomes of a liberal arts education and constitute the core of career readiness as defined by the College of Liberal Arts (CLA) Career Readiness Initiative. These competencies were identified as the result of an extensive literature review that involved scholarly and popular media sources such as journal and magazine articles, op-ed pieces, websites, and social media posts. In addition, we interviewed University of Minnesota faculty, as well as, employers who recruit and hire CLA students to determine qualities they see as key to a liberal arts education and career success. Results indicated competencies such as critical thinking, communication, and problem solving, among others, were valued by both faculty and employers, and that mastery of competencies was perceived as *more* important than completing specific undergraduate majors.

To conduct the review, in 2016, the authors assembled a team of faculty and staff from across the CLA under the leadership of Career Readiness Faculty Director and primary author, Ascan Koerner, to discuss and define major concepts related to career readiness of liberal arts students. This team engaged in research of secondary sources and identified a variety of existing competency frameworks, including, but not limited to, the National Association of Colleges and Employers Career Readiness Competencies (NACE, 2015); the Association of American Colleges and Universities national initiative VALUE (Valid Assessment of Learning in Undergraduate Education) Rubrics (AAC&U, 2009); the Association of American Medical Colleges fifteen core competencies for entering medical students (AAMC, 2016); and the Measuring College Learning Essential Competencies from the Social Science Research Council (Arum, Roksa, & Cook, 2016). They also reviewed two surveys conducted on behalf of the AAC&U by Hart Research Associates (2013, 2015) along with the University of Minnesota's Student Development Outcomes and Student Learning Outcomes.

The research team reviewed these various competency frameworks for conceptual overlap and alignment by comparing and contrasting similar concepts and their definitions side by side to evaluate their similarities and differences. Despite the apparent differences between these organizations, there was a surprising agreement among the different competency frameworks (see Table 1). Critical thinking, communication, problem solving, engaging diversity, teamwork & leadership, ethical reasoning, and civic engagement were present in all frameworks, as was some form of career management/professionalism, with additional competencies being present in only some of the frameworks. Overall, however, there seems to be a broad consensus about the content and value for career readiness of liberal arts education.

Table 1. Alignment between CLA Core Competencies, AAC&U Value Rubrics, and NACE Career Competencies.

CLA Core Competencies	**AAC&U VALUE Rubrics** (updated)	**NACE Career Readiness Competencies** (updated version from 2020)
Applied Problem Solving	Problem solving	
Analytical & Critical Thinking	Critical Thinking & Inquiry and Analysis Reading	Critical Thinking
Engaged Diversity	Intercultural Knowledge and Competence & | Global Learning	Equity & Inclusion
Ethical Reasoning & Decision Making	Ethical Reasoning	
Oral & Written Communication	Written Communication & Oral Communication	Communication
Innovation & Creativity	Creative Thinking	
Active Citizenship & Community Engagement	Civic Engagement - Local and Global	
Career Management	Foundations and Skills for Lifelong Learning Integrative Learning	Career & Self-Development & Professionalism
Teamwork & Leadership	Teamwork and problem solving	Leadership & Teamwork
Digital Literacy	Information Literacy, Quantitative Literacy	Technology

The next step in the conceptualization process was to arrive at a final set of competencies and to define them in a way that reflected student experiences and that could be operationalized. In this process, we aimed for exhaustive and mutually exclusive definitions, a task that was complicated by the fact that there is conceptual overlap between competencies. For example, good problem solving requires some mastery of critical thinking, and leadership requires some mastery of communication. We also aimed our definitions to be meaningful to undergraduate liberal arts students and their academic experiences, who we centered as the primary audience for the definitions. In contrast, the AAC&U VALUE rubrics were written for faculty, and therefore the language is more geared towards academic disciplines, while NACE's definitions are more geared toward employers. Finally, the result of the comprehensive review identified ten core competencies, nine "academic" competencies inherent in the liberal arts and career management as the tenth competency. We then articulated definitions for each competency (Table 2). Definitions were based on labels and conceptualizations identified by our review and written in collaboration with CLA faculty with expertise related to the competency.

Table 2. Core Competencies for Career Readiness with Short Definitions and Examples of Operationalizations.

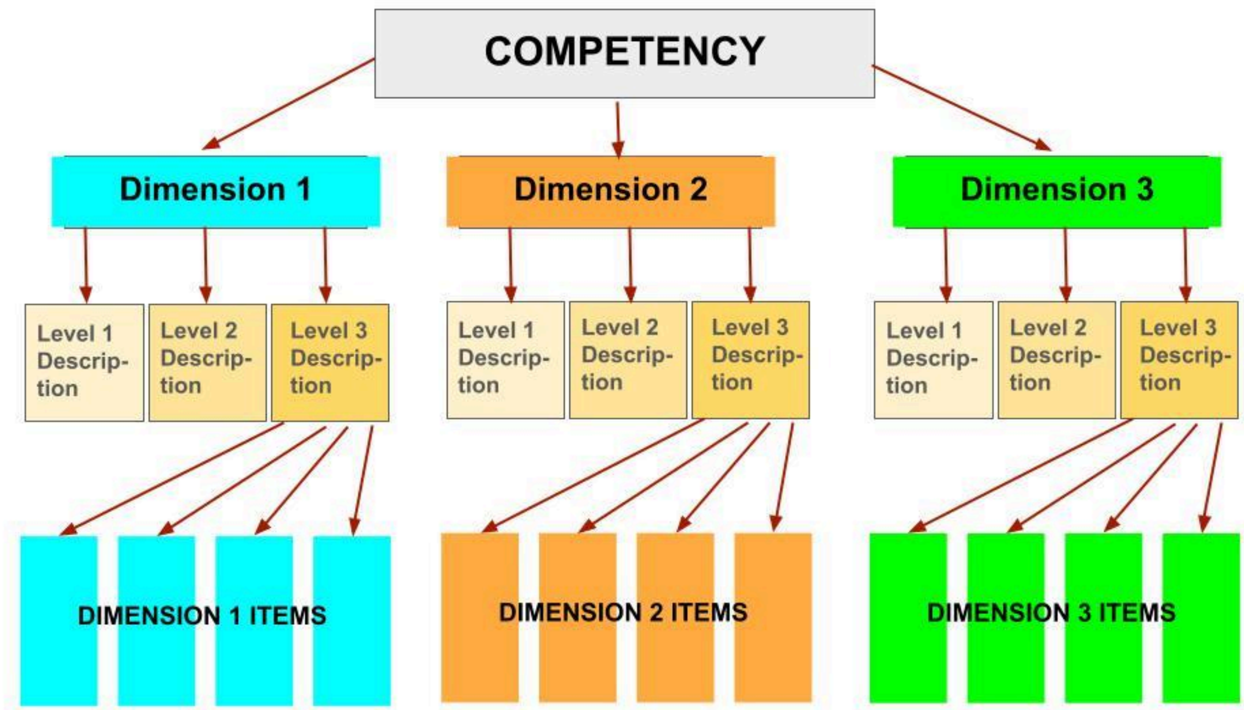
Core Competency Labels and Definitions
Active Citizenship & Community Engagement develops a consciousness about one's potential contributions and roles in the many communities one inhabits, in person and online, and takes action accordingly. Those competent in Active Citizenship & Community Engagement: ● Actively engage with the communities in which they are involved. ● Build awareness of how communities impact individuals, and how, in turn, an individual impacts, serves, and shapes communities. ● Evolve their awareness of culture and power in community dynamics.
Analytical & Critical Thinking comprehensively explores issues, ideas, knowledge, evidence, and values before accepting or formulating an opinion or conclusion. Those competent in Analytical & Critical Thinking: ● Recognize there may be more than one valid point of view. ● Evaluate an issue or problem based on multiple perspectives, while accounting for personal biases. ● Identify when information is missing or if there is a problem, prior to coming to conclusions and making decisions.
Applied Problem Solving is the process of designing, evaluating, and implementing a workable strategy to achieve a goal. Those competent in Applied Problem Solving: ● Recognize constraints. ● Generate a set of alternative courses of action. ● Evaluate alternatives using a set of criteria. ● Select and implement the most effective solution. ● Monitor the actual outcomes of that solution.
Career Management is the active engagement in the process of exploring possible careers, gaining meaningful experience, and building skills that help one excel after college and lead to employment or other successful post-graduation outcomes. Those competent in Career Management: ● Understand their values, interests, identity, personality, skills, strengths, and Core Career Competencies. ● Are able to articulate how those characteristics, combined with and shaped by a liberal arts education, lead to career success.
Digital Literacy leverages knowledge of information and communications technology and media literacies, and utilizes the interpersonal skills necessary to succeed in a digital space. Those competent in Digital Literacy: ● Assess sources of information. ● Use technologies responsibly. ● Adapt tools to new purposes. ● Keep up with the evolving technology landscape.
Engaging Diversity cultivates awareness of one's own identity and cultural background and that of others through an exploration of domains of diversity, which may include: race, ethnicity, country of origin, sexual orientation, ability, class, gender, age, spirituality, etc. This requires an understanding of historical and social contexts and a willingness to confront perspectives of dominant cultural narratives and ideologies, locally, nationally, or globally. Those competent in Engaging Diversity: ● Understand how culture affects perceptions, attitudes, values, and behaviors. ● Recognize how social structures and systems create and perpetuate inequities, resulting in social and economic marginalization and limited opportunities. ● Commit to the fundamental principles of freedom of thought and expression, equality, respect for others, diversity, and social justice; and to participate in society as conscious global citizens. ● Are able to navigate an increasingly complex and diverse world by appreciating and adopting multiple cultural perspectives or worldviews.

Ethical Reasoning & Decision Making recognizes ethical issues arising in a variety of settings or social contexts, reflects on the ethical concerns that pertain to the issue, and chooses a course of action based on these reflections. Those competent in Ethical Reasoning & Decision Making: ● Assess their own personal and moral values and perspectives as well as those of other stakeholders. ● Integrate these values and perspectives into an ethical framework for decision making. ● Consider intentions and the short- and long-term consequences of actions and the ethical principles that apply in the situation before making decisions.
Innovation & Creativity generates new, varied, and unique ideas, and makes connections between previously unrelated ideas. Those competent in Innovation & Creativity: ● Challenge existing paradigms and propose alternatives without being constrained by established approaches or anticipated responses of others. ● Employ their knowledge, skills, abilities, and sense of originality. ● Have a willingness to take risks and overcome internal struggle to expose their creative self in order to bring forward new work or ideas.
Oral & Written Communication intentionally engages with an audience to inform, persuade, or entertain. Those competent in Oral & Written Communication: ● Consider relationships with the audience and the social and political context in which one communicates, as well as the needs, goals, and motivations of all involved. ● Have proficiency in, knowledge of, and competence with the means of communication (including relevant language and technical skills). ● Ensure that communication is functional and clear.
Teamwork & Leadership builds and maintains collaborative relationships based on the needs, abilities, and goals of each member of a group. Those competent in Teamwork & Leadership: ● Understand their own roles and responsibilities within a group, and how they may change in differing situations. ● Are able to influence others without necessarily holding a formal position of authority, and have the willingness to take action. ● Leverage the strengths of the group to achieve a shared vision or objective. ● Effectively acknowledge and manage conflict toward solutions.

Initial Operationalization and Level Descriptors

The first step in the development of the assessment scales for the core competencies was to operationalize the definitions. To this end, for each definition we identified underlying dimensions that we considered essential for each of the competencies (Figure 1). We then described these dimensions in terms of the observable practices, behaviors, values, or abilities indicative of the designated competency, resulting in 4-6 dimensions for each definition of a competency, except for career management, which is defined in terms of a wide range of distinct behaviors, resulting in eight dimensions that align closely with the CLA Career Services student learning outcomes. The behaviors that operationalize the competencies were then described at three levels of proficiency representing a basic trajectory of student growth (Level 1 = Novice, Level 2 = Intermediate, Level 3 = Advanced), resulting in descriptors for all dimensions of the competencies at the three levels of proficiency to clearly identify which behaviors indicate what level of competency development. In the last step, we wrote several assessment items operationalizing each dimension of each competency in the form of "can-do" statements, initially using a 5-point Likert-type response format (1 = Strongly Disagree to 5 = Strongly Agree). The items were written to reflect the highest level of proficiency (i.e., Level 3 = Advanced), so that greater agreement with the items would indicate greater mastery of the competency. Generally, we aimed to write about five items for each dimension, for an average of 25 items per competency (Range = 22 - 36). These items described concrete behaviors, practices, values, and abilities for each dimension. In addition, we also wrote "external validation" items for each dimension which are items that, rather than indicating the respondents' own behavior, are behaviors by others that indicate their perception of the competency of the respondent. For example, for teamwork and leadership, an external validation item would be, "When working in groups, others usually ask me to assume the leadership role." Care was taken to ensure respondents could imagine themselves across disciplinary settings for all items, as well as in co-curricular experiences, such as learning abroad, a volunteer experience, a job fair, or research.

Figure 1. Architecture: Competency, dimensions, levels descriptors, and items



Validity Testing: Content and Response Options Reviews and Psychometric Analyses

To support the 10 competencies score interpretation and uses, we collected four sources of validity evidence namely, (a) validity evidence based on the assessment content, (b) validity evidence based on response processes, (c) validity evidence based on internal structure, and (d) validity evidence based on association with other variables (convergent, discriminant, and predictive; American Educational research Association et al., 2014). We also estimated score reliability for each of the 10 assessments, which is also considered validity evidence. Additionally, we tested measurement invariance (MI) to ensure that the 10 assessments function equally across student demographic variables (e.g., sex). How we collected each source of validity evidence, the psychometric procedures undertaken, and relevant results are described in more detail below.

Validity Evidence Based on the Assessment Content (First Review and Iteration of Assessments)

Validity evidence based on the assessment content includes ensuring domain representation (i.e., the item pool assesses the intended competencies) and domain relevance (i.e., item pool does not include redundant or irrelevant items), among others (Sireci, 1998; Sireci & Faulkner-Bond, 2014). To collect this source of validity evidence, we subjected the first version of the definitions and assessment scales to additional rounds of content review in Spring of 2017. Specifically, we invited a number of stakeholders and content experts to make suggestions to clarify, correct errors, add missing information, or address any potential biases in the definitions or items. Invited stakeholders included members of CLA Career Services, Academic Advising, and CLA Career Readiness team. Additionally, we sought input and insight from University of Minnesota faculty and staff from the departments and offices of Design, Housing, and Apparel; Philosophy; Communication Studies; Public Engagement; Diversity, Equity, and Access; Technology and Information Systems; Educational Innovation; and Global Programs and Strategic Alliance. We also invited community members including service learning experts, coaches, professional artists and musicians, teachers from Saint Paul and Minneapolis Public Schools and the MacPhail Center for Music (Minneapolis). Additional stakeholders included faculty external to the University of Minnesota (from McNally Smith College of Music, Augsburg University, Bard College, & Vassar College). CLA student focus groups gave additional input on the form and content of the competency definitions and assessment items. In this first review, given the current context of enrollment in higher education in the United States in the 21st century, we paid additional attention to the needs of multilingual learners whose first language is not English and specifically CLA students where in 2016 almost ten percent of undergraduates students spoke English as a second language.

In response to the qualitative reviews of the first version of definitions and assessments, the core research team edited the first versions of the definitions and assessment items and conducted a close reading of the newly revised definitions and assessments to verify clear correspondence between definitions, dimensions, and assessment items and to ensure each dimension was explicitly represented by assessment items and that item content expressed the highest level of proficiency for a given competency. After this review and revision process was complete, we started our quantitative investigation of the ten assessments for score reliability and validity of inferences based on those scores.

Validity Evidence Based on Internal Structure

To ensure that the assessment items load on their intended competencies, we conducted exploratory and confirmatory analyses. In more detail, we conducted principal component analysis (PCA) followed by another PCA study to ensure replications. We also estimated score reliability in the two analyses. To test the stability of the obtained factor structure, we cross-validated the resulting factor structure using the K-fold cross-validation approach, which splits the sample into k-folds and performs exploratory and confirmatory factor analyses (more information is provided in the “Factor structure stability” section under study 7).

Principal Component Analysis (Study 1)

Revised assessment items were tested in Spring 2017 in a pilot study of 2,197 CLA undergraduate students, who were incentivized to participate in the study through entry into a drawing for one of three iPads Mini for completing an online Qualtrics survey. Of the respondents, 1,357 (66%) identified as female, 633 (32%) as male, 33 (1.6%) as gender queer, and 91 (4.3%) did not report their gender. Their average age was 20 years old (Range = 18-27); 19% were first year students, 23% were sophomores, 26% juniors, 30% seniors, and 2% were non-degree seeking or graduate students. Most students (68%) started their studies at the University, while 32% were transfer students. These numbers very closely mirrored the demographics of the student population as a whole, suggesting a representative sample.

Participants were asked to complete assessments for three randomly assigned competencies, each of which consisted of between 22 and 36 items. Each scale was presented to about 700 participants. After listwise deletion of cases with missing data, we were left with data from 2,134 students (97% of the sample) and an average of 640 (Range = 570-663) responses to each scale. In SPSS, we performed PCA with oblimin rotation for each competency scale to reduce the complexity and length of the competency scales and to check which items load on each competency scale. For each scale, poor performing items were successively eliminated based on low (<.50) or ambiguous component weights (i.e., item with > .20 weight on two or more components). Because the constructs were defined with several underlying dimensions in mind, we required each scale to contain at least three components, resulting in eight scales with three components, and two scales with four components (digital literacy and career management, respectively). These components, on average, accounted for 64% of total item variance for each of the ten competency scales. In the end, this PCA yielded a total of 123 items (Mean = 12.3 per competency; Range = 10-20) from the initial total of 272 items.

Table 3. Scale Summary Results from Initial Pilot Study

Competency	n	Mean (SD)	# of components	alpha
Active Citizenship & Community Engagement	640	4.30 (0.82)	3	.82
Analytical & Critical Thinking	663	4.10 (1.03)	3	.77
Active Problem Solving	650	4.17 (0.87)	3	.78
Career Management	628	4.34 (0.78)	4	.84
Digital Literacy	570	4.26 (0.82)	4	.77
Engaging Diversity	661	4.39 (0.82)	3	.83
Ethical Reasoning & Decision Making	660	4.45 (0.72)	3	.80
Innovation & Creativity	657	4.08 (0.85)	3	.78
Oral & Written Communication	647	4.27 (0.81)	3	.75
Teamwork & Leadership	627	4.10 (0.78)	3	.78

Replication Analyses (Study 2)

The 123 retained items that constituted the 10 assessment scales were incorporated into RATE™, which was then deployed for all ten competencies and made available to all CLA students starting in Fall 2017. Student responses from Fall 2017 (n = 1057) were used to conduct a replication of study 1 for the ten scales and to further refine the retained items. The scales were completed by an average of 106 students per competency (Range = 70-131). The scales overall were reliable with an average α =.88 (Range: .74-.96; see Table 4). As a result of these analyses, weaker performing items were inspected and edited for clarity, however, no item was deleted. RATE™ was updated for the Spring 2018 semester and the newly revised items for the 10 assessment scales were deployed. Student responses from Spring 2018 (n = 2827) were used to conduct additional PCA to further refine the scales and scale items. The updated scales were completed by an average of 383 students per competency (Range: 221-829). The scales overall again were reliable with an average to α = .88 (Range: .84 - .94). The largest change in results was the improved reliability of the Teamwork & Leadership competency, which improved from α = .74 to .86. (Table 5).

Table 4. Fall 2017 Scale Summaries ($N = 1057$)

Competency	n	Mean (SD)	alpha
Active Citizenship & Community Engagement	98	4.30 (0.82)	.90
Analytical & Critical Thinking	121	4.10 (1.03)	.93
Applied Problem Solving	107	4.17 (0.87)	.88
Career Management	100	4.34 (0.78)	.92
Digital Literacy	103	4.26 (0.82)	.84
Engaging Diversity	109	4.39 (0.82)	.96
Ethical Reasoning & Decision Making	105	4.45 (0.72)	.89
Innovation & Creativity	70	4.08 (0.85)	.88
Oral & Written Communication	113	4.27 (0.81)	.83
Teamwork & Leadership	131	4.10 (0.78)	.74

Table 5. Spring 2018 Scale Summaries ($N = 2827$)

Competency	n	Mean (SD)	alpha
Active Citizenship & Community Engagement	303	4.33 (.73)	.87
Analytical & Critical Thinking	829	3.99 (.91)	.86
Applied Problem Solving	271	4.20 (.78)	.86
Career Management	265	4.09 (.96)	.93
Digital Literacy	221	4.16 (.82)	.82
Engaging Diversity	255	4.32 (.79)	.95
Ethical Reasoning & Decision Making	243	4.23 (.85)	.91
Innovation & Creativity	335	4.04 (.88)	.86
Oral & Written Communication	625	4.09 (.88)	.84
Teamwork & Leadership	480	4.12 (.82)	.86

Validity Evidence based on Response Processes: Response Option Revisions & Items Refinement (Study 3)

To collect validity evidence based on response processes, we conducted individual think-aloud sessions with 28 undergraduate students (6-8 students per competency) in Spring 2019. This source of validity evidence has been recommended to ensure that the assessment items are interpreted by respondents as intended and the number of response options is appropriately selected, so that respondents can express their actual standing on the measured competencies (Padilla & Benítez, 2014). In the think-aloud sessions, students evaluated the scales for 2-3 competencies and were instructed to read the items aloud and to verbalize their thinking about the items and how they would respond to them. The extensive notes from these sessions were then evaluated by the research team for emerging themes that may indicate problems students had with responding to the assessment items. Overall, the students interpreted the items in line with their intended meanings. After being prompted, a few students made suggestions for minor revisions for some of the items to improve readability. Most importantly, however, the think-aloud sessions revealed that some students thought that their responses would be more discriminating if they were given a wider range of response options and would respond in terms of frequency of, rather than agreement with, behaviors that characterize mastery of competency. This observation led us to conduct an additional analysis of the responses collected between Fall 2017 - Spring 2019 (N = 15,917), which revealed that the two highest responses (Agree, Strongly Agree) accounted for about 75% of all responses (Table 6), potentially creating ceiling effects and thus were unlikely to fully capture the variability in levels of the competencies. In response to this finding and the students' comments about frequency versus agreement, we changed the response options to a 1 - 7 scale with anchors of 1 = Never, 4 = Half of the time, and 7 = Always.

Table 6. Summary of Item Analysis with Agreement Responses

Competency	Response Frequencies					# items	α	*R*
	SD	D	N	A	SA			
Active Citizenship & Community Engagement	2%	3%	11%	35%	49%	11	.93	.55
Analytical & Critical Thinking	2%	7%	18%	39%	33%	12	.89	.42
Applied Problem Solving	2%	4%	13%	42%	40%	11	.91	.48
Career Management	2%	5%	14%	36%	42%	15	.94	.54
Digital Literacy	3%	5%	15%	35%	43%	11	.90	.47
Engaging Diversity	1%	3%	10%	33%	52%	20	.96	.55
Ethical Reasoning & Decision Making	2%	4%	12%	37%	45%	11	.91	.50
Innovation & Creativity	1%	5%	17%	39%	38%	11	.89	.42
Oral & Written Communication	2%	6%	15%	38%	39%	11	.89	.43
Teamwork & Leadership	1%	4%	13%	41%	41%	10	.91	.50

Note. SD = Strongly Disagree, D = Disagree, N = Neutral, A = Agree, SA = Strongly Agree, *R* = Item-Total correlation

To test whether the change in response choice had the desired effect of increasing the range of responses without negatively affecting reliability, responses from the revised items were collected during the 2019-2020 academic year (N = 11,909). Sample sizes across competencies varied, ranging from 430 for Active Citizenship & Community Engagement up to 3,219 for Analytical & Critical Thinking. Classical test theory item analysis showed continued high levels of internal consistency reliability and similar item-total correlations for all 10 competencies with the 7-point scale as compared to the 5-point scale (Table 7). The response frequencies, however, continued to be similarly negatively skewed, suggesting a ceiling effect with students selecting the top two categories over 70% of the time. Based on these results, we determined that the 7-point scale was not superior to the 5-point scale and decided to change the response options for a final time for Spring 2021 to five options with a label for each point: Rarely, Sometimes, Often, Very Often, and Always.

To test whether this decision was justifiable, we repeated the classical test theory item analysis on the student responses from Spring 2021 (Table 8). Although most responses were still on the high end of the scale, the response options of Rarely, Sometimes, Often, Very Often, and Always produced greater variability than previous iterations. Additionally, internal consistency reliability, as measured by Cronbach's alpha, and average item-total correlations remained high.

Table 7. Summary of Item Analysis With 7-Point Frequency Response Options

	Response Frequencies								
Competency	1	2	3	4	5	6	7	α	R
Active Citizenship & Community Engagement	0%	1%	2%	5%	12%	31%	49%	.86	.37
Analytical & Critical Thinking	1%	2%	4%	12%	20%	32%	29%	.87	.36
Applied Problem Solving	0%	1%	2%	9%	18%	35%	34%	.89	.44
Career Management	1%	1%	3%	8%	17%	29%	40%	.93	.48
Digital Literacy	2%	2%	3%	6%	15%	30%	42%	.87	.42
Engaging Diversity	1%	1%	3%	8%	16%	31%	41%	.94	.46
Ethical Reasoning & Decision Making	1%	3%	4%	11%	19%	31%	31%	.81	.30
Innovation & Creativity	0%	2%	5%	11%	22%	34%	26%	.86	.37
Oral & Written Communication	2%	3%	4%	9%	17%	31%	35%	.83	.34
Teamwork & Leadership	0%	1%	2%	8%	18%	35%	36%	.83	.33

Notes. The 7 response options had anchors of 1 = Never, 4 = Half of the time, and 7 = Always. R = Item-Total Correlation

Table 8. Summary of Item Analysis With 5-Point Frequency Response Options

	Response Frequencies						
Competency	1	2	3	4	5	α	R
Active Citizenship & Community Engagement	2%	5%	15%	37%	40%	.91	.48
Analytical & Critical Thinking	2%	12%	23%	35%	28%	.89	.40
Applied Problem Solving	1%	7%	16%	32%	34%	.90	.33
Career Management	3%	9%	19%	35%	34%	.94	.48
Digital Literacy	3%	8%	17%	32%	40%	.88	.41
Engaging Diversity	1%	6%	18%	33%	42%	.95	.48
Ethical Reasoning & Decision Making	4%	15%	23%	33%	26%	.87	.38
Innovation & Creativity	1%	8%	25%	40%	26%	.94	.30
Oral & Written Communication	4%	11%	20%	33%	31%	.86	.37
Teamwork & Leadership	1%	8%	20%	39%	32%	.87	.41

Validity Evidence Based on Association with Other Variables

We collected three forms of validity evidence based on associations with other variables including convergent (Studies 4 & 5), discriminant (Study 6), and predictive (Study 8) evidence. Each of the three sources of evidence is described in more detail below.

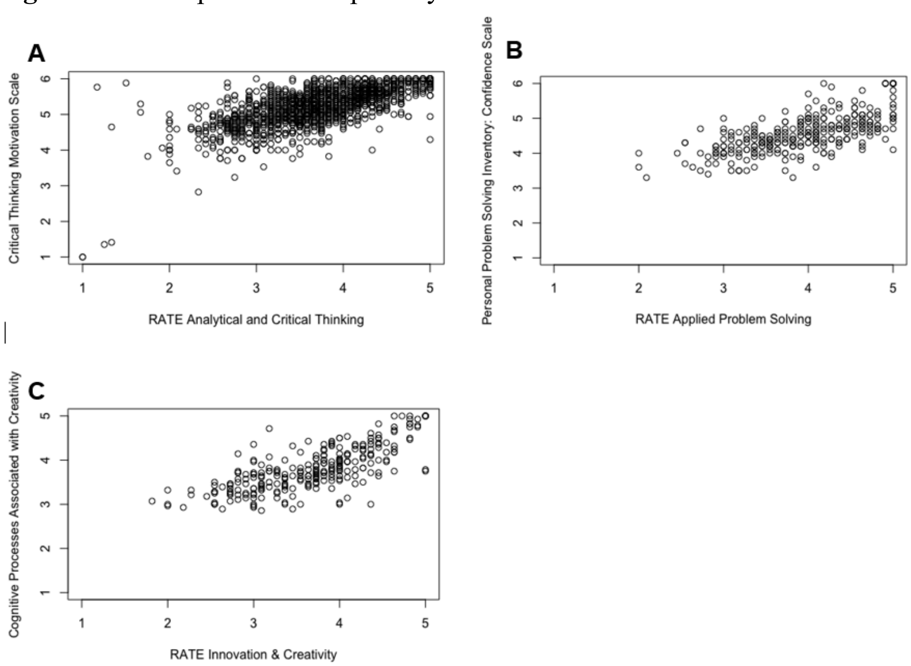
Convergent Validity Evidence (Studies 4 & 5)

To start the process of testing for convergent validity, we conducted an extensive literature search in Spring 2019 to identify suitable measures for pursuing a convergent validity study. Criterion measures were selected if they measured similar constructs to the competencies, where published, and had acceptable evidence of score reliability and validity. Unfortunately, there are no established and readily available measures for many of our competencies, so we were limited in the number of external measures we could use for convergent validity testing. In total, we conducted six studies to test for the convergent validity evidence of three of the ten competency assessment scales. CLA students who completed assessments for the three competencies were randomly selected to participate in the studies and, after consenting to participate, were presented with both the relevant assessment scale as well as the criterion measures. Their responses were then correlated to assess the extent of agreement between the two measures. Greater correlations between measures signify great agreement, and large positive correlations between the assessment scale and the criterion measure are evidence of validity of the assessment scale.

In Fall 2020, the Critical Thinking Motivation Scale (Valenzuela, Nieto, & Saiz, 2011) was administered to 177 students who also took the Analytical & Critical Thinking competency assessment (ACT) scale. The correlation between scores on the two measures was $r = .68$. A scatterplot of scores is presented in Figure 2A. Likewise, the Applied Problem Solving competency (APS) scale and the Personal Problem Solving Inventory: Confidence Scale (Heppner & Peterson, 1982) was administered to 72 CLA undergraduate students in Fall 2020, with scores yielding a correlation of $r = .69$ (see Figure 2B). Lastly, the Innovation & Creativity competency (IC) scale and the Cognitive Processes Associated with Creativity (Miller, 1982) was administered to 76 other CLA undergraduate students in Fall 2020, which resulted in a correlation of $r = .73$ (see Figure 2C). The consistent and strong positive correlations between the three competency assessment scales and the external measures provide strong evidence for the criterion validity of these three assessment scales.

Because we changed the response scale from seven to five scale points starting in Spring 2021 (see above), the correlational studies for the three competencies were repeated in Spring 2021. For ACT ($n = 292$) we observed a correlation of $r = .69$. For APS ($n = 101$) we obtained a correlation of $r = .65$, and for IC ($n = 90$) results showed a correlation of $r = .69$. The consistently high correlations between the competency assessment scales and these established measures provide strong evidence that the assessment scales for the three investigated competencies for which we found established measures, that is, Analytical & Critical Thinking, Applied Problem Solving, and Innovation & Creativity, have high convergent validity evidence, further demonstrating the validity evidence of core competency assessments.

Figure 2. Scatterplots of Competency Scores With Criterion Measure Scores



Discriminant Validity Evidence (Study 6)

Whereas the convergent validity evidence is established by observing positive correlations between a measure and other established measures of the same construct, discriminant validity evidence is established by showing that a measure is not correlated to variables the theoretical construct is not related to. This is particularly relevant in regard to variables that are otherwise prominent in a population and that are suspected to potentially bias the results, such as demographics like race or sex. To examine the association between demographic variables that are generally considered to be predictive of student outcomes and competency assessment scales, we conducted chi-squared analyses to test for group differences using the data of 16,557 undergraduate students who completed a competency assessment between Fall 2018 and Spring 2020 semesters. To conduct this test, we first transformed the assessment scores to three levels: 1 = low; 2 = medium; 3 = high. We then compared the distribution of the levels for each competency for different groups defined by three demographic variables (sex, race, and first-generation status). Results are presented in Tables 9-11. For each of the three demographic variables, a family-wise Type I error rate of $\alpha < .10$ was set, which led to an $\alpha = .001$ for each individual comparison. Results showed no association with race or first generation status for any of the competencies. For sex, results showed an association between sex and the levels for the following competencies: Active Citizenship & Community Engagement, Career Management, Engaging Diversity, and Teamwork & Leadership. In all four cases, students who identified as female had higher scores than males.

Table 9. Association between Competency Level and Sex by Competency

Competency	rho^2	F	df1	df2	p-value	n
Active Citizenship & Community Engagement	0.01	11.24	1	953	0.00	955
Analytical & Critical Thinking	0.00	1.53	1	4,690	0.22	4,692
Applied Problem Solving	0.00	5.22	1	1,130	0.02	1,132
Career Management	0.02	16.65	1	1,054	0.00	1,056
Digital Literacy	0.00	0.29	1	719	0.59	721
Engaging Diversity	0.01	9.25	1	816	0.00	818
Ethical Reasoning & Decision Making	0.00	3.30	1	839	0.07	841
Innovation & Creativity	0.00	0.00	1	1,137	0.96	1,139
Oral & Written Communication	0.00	6.88	1	2,928	0.01	2,930
Teamwork & Leadership	0.01	15.39	1	2,235	0.00	2,237

Table 10. Association between Competency Level and Race by Competency

Competency	rho^2	F	df1	df2	p-value	n
Active Citizenship & Community Engagement	0.00	0.33	8	922	0.95	931
Analytical & Critical Thinking	0.00	0.57	8	4,601	0.81	4,610
Applied Problem Solving	0.02	2.14	8	1,102	0.03	1,111
Career Management	0.00	0.50	8	1,038	0.86	1,047
Digital Literacy	0.00	0.16	8	704	1.00	713
Engaging Diversity	0.01	1.00	8	800	0.44	809
Ethical Reasoning & Decision Making	0.01	0.63	8	820	0.75	829
Innovation & Creativity	0.00	0.62	8	1,116	0.76	1,125
Oral & Written Communication	0.00	1.75	8	2,883	0.08	2,892
Teamwork & Leadership	0.00	0.98	8	2,200	0.45	2,209

Table 11. Association between Level and First-Generation Status by Competency

Competency	rho^2	F	df1	df2	p-value	n
Active Citizenship & Community Engagement	0.00	0.23	1	953	0.63	955
Analytical & Critical Thinking	0.00	0.04	1	4,697	0.85	4,699
Applied Problem Solving	0.00	0.24	1	1,131	0.62	1,133
Career Management	0.00	0.80	1	1,056	0.37	1,058
Digital Literacy	0.00	0.79	1	720	0.37	722
Engaging Diversity	0.00	1.12	1	818	0.29	820
Ethical Reasoning & Decision Making	0.00	0.12	1	840	0.73	842
Innovation & Creativity	0.00	1.16	1	1,138	0.28	1,140
Oral & Written Communication	0.00	0.82	1	2,929	0.37	2,931
Teamwork & Leadership	0.00	0.13	1	2,238	0.72	2,240

Measurement Invariance (Study 7)

Measurement Invariance (MI) refers to the extent to which the items measure a competency equivalently across groups (Svetina et al., 2020). Evidence of MI is particularly important when scores are used to make group comparisons. We investigated MI of the ten competency assessments using multigroup confirmatory factor analysis (MGCFA). A total of 24,937 assessments were completed by 13,203 students between June 2021 and February 2024. Most students (n = 7348, 55.7%) completed a single assessment, although 15 students completed ten or more assessments. Many students (n = 2391, 18.1%) completed assessments for the same competency multiple times. Students typically completed one or more assessments for a single competency (n = 7965, 60.3%), although 3254 (24.6%) students did so for two competencies, 1321 (10.0%) for three competencies, and 663 (5.0%) students completed one or more assessments for four or more competencies. For the MI investigation we used students' first completed assessment for any given competency, which means we deleted 3,627 assessments from our data for a final N = 21,310 assessments.

Factor Structure Stability

Prior to conducting MI, we investigated the factor structure of each competency, which refers to the number of latent factors measured by the items and the pattern of item-factor associations within the competency. For each competency, we first used the kfa package (Nickodem & Halpin, 2022) in R (R Core Team, 2023) to perform k-fold cross-validation of the possible factor structures. In this approach, the full sample was randomly split into k = 3 subsamples. This reveals the stability of confirmatory factor analysis (CFA) model across the k subsamples, and thus, provides validity evidence for the generalizability of the optimal factor structure. The competencies were hypothesized to be 1-factor structures (i.e., all items loading onto a single factor representing the competency). Following kfa, we estimated the full sample fit statistics of the optimal model. The comparative fit index (*CFI*), root mean square error of approximation (*RMSEA*) and standardized root mean squared residual (*SRMR*) were used to evaluate model fit. Typical guidelines indicating good model fit are $CFI > .95$, $RMSEA < .08$, and $SRMR < .06$; however, these were established with continuous variables (Hu & Bentler, 1999). Consistent guidelines with ordinal variables have not been established. Given the ordinal item responses (i.e., Rarely, Sometimes, Often, Very Often, and Always), all models used weighted least squares estimation with the mean and variance scaled chi-square test statistic and associated fit indices. There was no missing data for assessment item responses.

For eight of the ten competencies, the 1-factor structure was the only stable factor structure across all 3 subsamples in the k-fold cross-validation. As shown in Table 12, the 1-factor structure also showed good fit to the full sample for these competencies. For Applied Problem Solving and Digital Literacy, in addition to the 1-factor structure, a 2-factor structure was also stable across all 3 subsamples, and also showed improved model fit. For Applied Problem Solving, the second factor consisted of only three items and did not provide unique information above and beyond the 1-factor structure in subsequent analyses (e.g., predictive validity). Thus, Applied Problem Solving should be treated as a unidimensional construct. Conversely, the model fit in Table 12 and subsequent analyses suggest that Digital Literacy should be treated as a two dimensional construct measuring 1) students ability to discern the quality of information (7 items) and 2) their comfort using technology (4 items).

Table 12. Confirmatory Factor Analysis Model Fit

Model	*n*	χ^2	*df*	*p<*	*CFI*	*RMSEA*	*90% CI*	*SRMR*
ACCE	613	207.59	44	*.001*	.98	.08	[.07, .09]	.04
ANCT*	5020	2318.69	53	*.001*	.95	.09	[.09, .10]	.04
APPS – 1-factor	1296	792.81	44	*.001*	.94	.11	[.11, .12]	.05
APPS – 2-factor	1296	274.73	43	*.001*	.98	.06	[.06, .07]	.03
CRMG	1403	1431.96	104	*.001*	.95	.10	[.09, .10]	.05
DGLI – 1-factor	773	945.76	44	*.001*	.89	.16	[.15, .17]	.10
DGLI – 2-factor	773	241.67	43	*.001*	.98	.08	[.07, .09]	.05
ENDV	2089	3581.48	170	*.001*	.94	.10	[.10, .10]	.05
ERDM*	747	330.24	43	*.001*	.96	.09	[.09, .10]	.05
INCR	1240	254.84	43	*.001*	.98	.06	[.06, .07]	.03
OWCM	5630	1551.74	44	*.001*	.96	.08	[.07, .08]	.04
TKLP	2499	1018.30	35	*.001*	.95	.11	[.10, .11]	.05

Notes. ACCE = Active Citizenship & Community Engagement, ANCT = Analytic & Critical Thinking, APPS = Applied Problem Solving, CRMG = Career Management, DGLI = Digital Literacy, ENDV = Engaging Diversity, ERDM = Ethical Reasoning & Decision Making, INCR = Innovation & Creativity, OWCM = Oral & Written Communication, TKLP = Teamwork & Leadership. *One residual correlation was added between two items to improve model fit.

Multigroup Confirmatory Factor Analysis

We then investigated MI by academic level, sex, and race/ethnicity using MGCFA with the stable factor structure found for each competency. Based on a power analysis, MI by sex only included students who identified as female or male. The race/ethnic and academic level groups were included when the sample size of the group was sufficiently large, which varied by competency. We did not combine race/ethnic groups with small sample sizes as the subsequent heterogeneity of the groups leads to uninterpretable results.

Given that all assessment items are ordinal, we used the MI model identification and testing procedures suggested by Wu and Estabrook (2016) and demonstrated by Svetina and colleagues (2020). The model constraints were compared using the chi-square difference test, change in fit indices (e.g., *CFI*), and $RMSEA_D$. The latter has been shown to be more appropriately sensitive to model fit than the oversensitive chi-square difference test or under sensitive change in fit indices (Beribisky & Hancock, 2023; Savalei et al., 2023). Specifically, the 90% confidence interval for $RMSEA_D$ is used as a test of not-close-fit with the null hypothesis that the two models were not equal (H0: > .10). Thus, when the upper bound of the 90% confidence interval was < .10, we concluded that measurement invariance held.

Item thresholds refers to the levels of the latent competency whereby a student would, for instance, respond to an item with Sometimes instead of Rarely or Often instead of Sometimes. Item loadings refers to the strength of the association between an item and the latent competency it is measuring. Table 13 shows that when item thresholds and loadings are constrained to be equal across academic level, the fit of the MGCFA model remains suitable, as evidenced by the upper bound of the 90% confidence interval of $RMSEA_D$ being < .10 for all competencies. Even the chi-square difference test, which is known to be oversensitive to trivial misfit when sample size is large (> 250), was nonsignificant (p > .05) for most competencies. This provides evidence of MI across academic level for the ten competencies measured by the assessment scales. Likewise, results reported in Tables 14 and 15, respectively, show that MI held across sex and race/ethnicity, respectively, for all ten competencies. In other words, the assessment scales measure the ten competencies similarly across academic level, sex, and race/ethnicity. This MI demonstrates that students' self-assessments are not affected by their sex, race, ethnicity, or class standing, thereby providing strong evidence of the discriminant validity of the instrument.

Table 13. MI Results by Academic Level

Competency	Constraint Added	n	χ^2	df	p	$RMSEA_D$	$90\%CI$
ACCE	Thresholds	613	53.5	54	.49	.000	[.000, .000]
	Loadings	613	50.6	60	.80	.000	[.000, .000]
ANCT	Thresholds	5020	170.1	168	.44	.004	[.000, .019]
	Loadings	5020	87.4	77	.20	.015	[.000, .028]
APPS	Thresholds	1296	69.6	77	.71	.000	[.000, .000]
	Loadings	1296	58.6	70	.83	.000	[.000, .000]
CRMG	Thresholds	1403	128.2	112	.14	.029	[.000, .050]
	Loadings	1403	104.2	105	.51	.000	[.000, .000]
DGLI	Thresholds	773	91.0	63	.01	.068	[.033, .097]
	Loadings	773	48.8	63	.91	.000	[.000, .000]
ENDV	Thresholds	2089	146.4	140	.34	.013	[.000, .033]
	Loadings	2089	147.5	133	.18	.020	[.000, .037]
ERDM	Thresholds	747	69.2	66	.37	.021	[.000, .062]
	Loadings	747	63.8	60	.34	.024	[.000, .065]
INCR	Thresholds	1240	101.7	77	.03	.046	[.015, .068]
	Loadings	1240	91.8	70	.04	.045	[.009, .068]
OWCM	Thresholds	5630	171.1	147	.09	.015	[.000, .024]
	Loadings	5630	73.6	70	.36	.009	[.000, .024]
TKLP	Thresholds	2499	79.3	63	.08	.029	[.000, .047]
	Loadings	2499	58.8	63	.63	.000	[.000, .000]

Notes. The 8 groups were compared (e.g., first semester freshman, second semester sophomore). ACCE = Active Citizenship & Community Engagement, ANCT = Analytic & Critical Thinking, APPS = Applied Problem Solving, CRMG = Career Management, DGLI = Digital Literacy, ENDV = Engaging Diversity, ERDM = Ethical Reasoning & Decision Making, INCR = Innovation & Creativity, OWCM = Oral & Written Communication, TKLP = Teamwork & Leadership.

Table 14. MI Results by Sex

Competency	Constraint Added	n	χ^2	df	p	$RMSEA_D$	$90\%CI$
ACCE	Thresholds	611	5.9	9	.75	.000	[.000, .000]
	Loadings	611	9.3	10	.50	.000	[.000, .000]
ANCT	Thresholds	5004	29.4	24	.21	.009	[.000, .020]
	Loadings	5004	38.6	11	.00	.032	[.021, .043]
APPS	Thresholds	1292	9.3	11	.59	.000	[.000, .000]
	Loadings	1292	11.7	10	.31	.016	[.000, .047]
CRMG	Thresholds	1401	21.1	16	.17	.021	[.000, .044]
	Loadings	1401	12.7	15	.63	.000	[.000, .000]
DGLI	Thresholds	772	7.6	9	.57	.000	[.000, .000]
	Loadings	772	7.0	9	.64	.000	[.000, .000]
ENDV	Thresholds	2085	22.7	20	.30	.011	[.000, .030]
	Loadings	2085	11.8	19	.90	.000	[.000, .000]
ERDM	Thresholds	743	10.5	11	.49	.000	[.000, .000]
	Loadings	743	5.9	10	.83	.000	[.000, .000]
INCR	Thresholds	1235	11.3	11	.42	.006	[.000, .043]
	Loadings	1235	18.7	10	.05	.037	[.006, .064]
OWCM	Thresholds	5616	49.7	21	.00	.022	[.014, .030]
	Loadings	5616	13.0	10	.22	.010	[.000, .024]
TKLP	Thresholds	2495	9.6	9	.38	.007	[.000, .033]
	Loadings	2495	13.5	9	.14	.020	[.000, .041]

Notes. Two groups were compared (Male, Female); ACCE = Active Citizenship & Community Engagement, ANCT = Analytic & Critical Thinking, APPS = Applied Problem Solving, CRMG = Career Management, DGLI = Digital Literacy, ENDV = Engaging Diversity, ERDM = Ethical Reasoning & Decision Making, INCR = Innovation & Creativity, OWCM = Oral & Written Communication, TKLP = Teamwork & Leadership.

Table 15. MI Results by Race/Ethnicity

Competency	Constraint Added	*n*	χ^2	*df*	*p*	$RMSEA_D$	*90%CI*
ACCE*	Thresholds	546	20.3	18	.32	.026	[.000, .073]
	Loadings	546	22.6	20	.31	.027	[.000, .071]
ANCT	Thresholds	4860	74.3	72	.40	.005	[.000, .018]
	Loadings	4860	38.5	33	.23	.012	[.000, .025]
APPS	Thresholds	1246	36.2	33	.32	.018	[.000, .046]
	Loadings	1246	34.1	30	.28	.021	[.000, .049]
CRMG	Thresholds	1359	40.9	45	.64	.000	[.000, .000]
	Loadings	1359	53.5	45	.18	.024	[.000, .045]
DGLI*	Thresholds	693	12.0	18	.85	.000	[.000, .000]
	Loadings	693	10.0	18	.93	.000	[.000, .000]
ENDV	Thresholds	2015	55.0	60	.66	.000	[.000, .000]
	Loadings	2015	54.5	57	.57	.000	[.000, .000]
ERDM*	Thresholds	677	18. 1	22	.70	.000	[.000, .000]
	Loadings	677	27.4	20	.12	.041	[.000, .075]
INCR	Thresholds	1195	42.8	33	.12	.031	[.000, .056]
	Loadings	1195	29.8	30	.48	.000	[.000, .000]
OWCM	Thresholds	5448	73. 4	63	.17	.011	[.000, .020]
	Loadings	5448	49.6	30	.01	.022	[.010, .032]
TKLP	Thresholds	2421	37.1	27	.09	.025	[.000, .043]
	Loadings	2421	36.4	27	.11	.024	[.000, .042]

Notes. Groups were Black, White, Asian, and Hispanic unless indicated by * (Black, White, and Asian); ACCE = Active Citizenship & Community Engagement, ANCT = Analytic & Critical Thinking, APPS = Applied Problem Solving, CRMG = Career Management, DGLI = Digital Literacy, ENDV = Engaging Diversity, ERDM = Ethical Reasoning & Decision Making, INCR = Innovation & Creativity, OWCM = Oral & Written Communication, TKLP = Teamwork & Leadership.

Predictive Validity Evidence (Study 8)

Predictive validity evidence is established if the competency scores are correlated with outcomes measured at a later time. For the predictive analysis, we used competency assessments to predict students' responses to the Post-Graduation Survey (PGS), a survey administered by the University of Minnesota that is made available to every graduating senior of the University upon completion of their baccalaureate degree. For this study, we used PGS responses from Spring 2018 to Spring 2022 as dependent variables and students' RATE™ competency scores–computed as the mean item response with each competency–as the independent variables. Consequently, only a portion of students who completed a RATE™ had also graduated and completed the PGS (*n* = 671). A total of 18 variables from the PGS were correlated with RATE™ scores and whether or not (1/0) a student completed a RATE™ for a competency. The 18 PGS variables are grouped in three areas: Primary Status, Competency Level, and Career Development. For primary status, three dichotomous variables were created from the PGS item, "What is the primary status after graduation?" Responses of Employed, Continuing Education, and Employed or Continuing Education constituted the three dichotomous variables with each coded as 1 = Yes and 0 = No. In the area of competency level, for the ten competencies along with "Communicating in a foreign language" (11 items), graduates were asked, "To what extent has your undergraduate education helped you develop competence in…" Response options were 1 = Not at all, 2 = To a small extent, 3 = To a moderate extent, 4 = To a large extent, and 5 = To a very large extent. Lastly, in the career development area graduates were asked whether they agreed or disagreed with four statements (1 = Strongly disagree, 2 = Disagree, 3 = Neither disagree nor agree, 4 = Agree, 5 = Strongly agree):

1. I have had conversations with faculty in the College of Liberal Arts about how I am developing my Core Competencies.
2. The College of Liberal Arts has supported me in my career development.
3. My experience as a student in the liberal arts has prepared me to succeed after graduation.
4. I can explain how my undergraduate academic and co-curricular experiences prepared me for my career.

Given that the 18 PGS variables were dichotomous or ordinal, we computed polychoric correlations using the lavaan package (Rosseel, 2012) in R (R Core Team, 2023).

Sample sizes for the polychoric correlations between RATE™ scores and PGS responses are presented in Table 16 and the correlations are in Table 17. Among the notable findings regarding future employment status were that assessment scale scores on Active Citizenship & Community Engagement and Ethical Reasoning & Decision Making were positively correlated with gaining employment immediately following graduation. Regarding future plans for graduate education, students with higher scores on Active Citizenship & Community Engagement, Career Management, and the Information Quality factor of the Digital Literacy competency were more likely to pursue graduate education following graduation than their peers, whereas students with higher scores in Innovation & Creativity were less likely to pursue graduate education. In regard to students' self-reported competency development in the PGS, with the exception of Ethical Reasoning & Decision Making, students' scores on assessment scales for a particular competency were positively correlated with attributing growth across most competencies to their undergraduate education. For instance, students with higher RATE™ scores on Career Management also reported that their undergraduate education to a larger extent helped develop their competency not only in Career Management (r = .43), but also the other nine competencies (r = .29 – 45).

With respect to other career related outcomes, results showed that students with higher RATE™ scores in Ethical Reasoning & Decision Making, Innovation & Creativity, and Teamwork & Leadership were more likely to agree that their experiences in the liberal arts prepared them for success after graduation. Similarly, students with higher RATE™ scores in Career Management, Digital Literacy-Technology Use, Oral & Written Communication, and Teamwork & Leadership were more likely to agree that they could explain how their undergraduate academic and co-curricular experiences prepared them for their career. In addition, students with higher RATE™ scores in Innovation & Creativity and Teamwork & Leadership were more likely to agree that CLA supported their career development. Finally, students with higher RATE™ scores in seven competencies were more likely to have conversations with faculty about Career Competency development while students with higher RATE™ scores in Active Citizenship & Community Engagement were less likely to have those conversations with faculty (see Table 17).

Table 16. Sample Sizes for the Polychoric Correlations Between RATE™ Score and Post-Graduation Survey Question Area

Competency	Primary Status	Competency Level	Career Development
Analytic & Critical Thinking	214	90	90
Applied Problem Solving	79	35	35
Ethical Reasoning & Decision Making	46	23	23
Innovation & Creativity	76	30	30
Digital Literacy	87	45	45
Engaging Diversity	124	53	53
Active Citizenship & Community Engagement	43	16	16
Teamwork & Leadership	161	65	65
Oral & Written Communication	205	86	86
Career Management	168	74	75

Table 17. Polychoric Correlation of Unidimensional RATE™ scores with Post-Graduate Survey (PGS) Outcome Items

Outcome	ACCE	ANCT	APPS	CRMG	DGLI-IQ	DGLI-TU	ENDV	ERDM	INCR	OWCM	TKLP
Employed	**.66**	.20	0.07	.03	0.07	-.14	.15	**.42**	.00	-.20	.20
Continuing Education	**.25**	-.06	0.14	**.43**	**0.21**	0.09	.12	.15	-**.47**	.17	.13
Employed/Cont. Ed.	**.74**	**.26**	0.19	**.31**	0.18	-.09	**.26**	**.61**	-**.30**	-.11	**.31**
PGS ACCE	**.25**	**.21**	**0.21**	**.29**	0.14	**0.21**	**.25**	.01	**.26**	**.21**	**.27**
PGS ANCT	**.39**	.11	**0.38**	**.39**	0.13	**0.27**	.18	.18	**.21**	**.36**	**.40**
PGS APPS	**.35**	**.29**	**0.21**	**.45**	0.14	**0.28**	**.30**	.08	**.25**	**.27**	**.33**
PGS CMFL	.03	**.21**	-.08	.07	0.02	0.05	.07	.02	**.22**	.08	.14
PGS CRMG	**.56**	**.38**	0.18	**.43**	-.08	-.02	**.25**	**.22**	.17	**.27**	**.36**
PGS DGLI	**.39**	**.30**	**0.26**	**.35**	0.06	0.15	**.38**	**.28**	**.34**	**.27**	**.25**
PGS ENDV	.15	**.24**	**0.29**	**.35**	0.15	0.17	**.21**	.17	**.36**	**.29**	**.29**
PGS ERDM	**.31**	**.31**	**0.38**	**.41**	0.00	0.16	**.25**	.02	**.23**	**.40**	**.32**
PGS INCR	**.40**	**.21**	0.12	**.34**	0.00	0.12	**.41**	.07	**.40**	**.28**	**.28**
PGS OWCM	**.55**	**.28**	0.17	**.34**	0.15	**0.24**	**.31**	.15	**.31**	**.38**	**.45**
PGS TKLP	**.53**	**.30**	0.14	**.31**	**0.31**	**0.32**	**.39**	.10	**.32**	**.29**	.20
Career Conversations	-**.32**	.16	**0.32**	**.34**	0.05	**0.34**	-.04	**.31**	**.38**	**.29**	**.37**
Career Support	.02	.07	0.05	.19	-.12	-.12	-.19	-.15	**.52**	.06	**.26**
Post Grad Success	.05	.16	0.16	.19	0.08	0.12	.05	**.21**	**.48**	.12	**.34**
Explain Preparation	.04	.20	0.14	**.26**	0.12	**0.22**	.09	.18	.18	**.29**	**.42**

Note. ACCE = Active Citizenship & Community Engagement, ANCT = Analytic & Critical Thinking, APPS = Applied Problem Solving, CMFL = Communicating in a Foreign Language, CRMG = Career Management, DGLI = Digital Literacy, IQ = Information Quality, TU = Technology Use, ENDV = Engaging Diversity, ERDM = Ethical Reasoning & Decision Making, INCR = Innovation & Creativity, OWCM = Oral & Written Communication, TKLP = Teamwork & Leadership. **Bold** indicates correlation > |.2|.

When comparing students who completed a RATE™ in a particular competency to other students who completed at least one RATE™ but not in that competency, those who completed a RATE™ in (n = 671 for primary status variables; n = 287 for competency level variables; n = 288 for career development variables):

- Analytic & Critical Thinking (r = .20) and Engaging Diversity (r = .21) were more likely to be employed whereas those who completed a RATE™ in Applied Problem Solving (r = -.33) and Teamwork & Leadership (r = -.22) were less likely to be employed
- Active Citizenship & Community Engagement (r = .27), Digital Literacy (r = .24), Ethical Reasoning & Decision Making (r = .21), and Teamwork & Leadership (r = .25) reported that their undergraduate experience to a higher extent helped develop their competency in those areas.

Conclusion

In a series of eight studies involving well over 55,000 students, we constructed and rigorously collected appropriate validity evidence for ten assessment scales that measure students' development of core competencies. We defined the ten competencies in theoretical concepts that reflect a broadly shared perception of the competencies inherent in higher education. We then, with the help of subject matter and professional experts, operationalized the competencies in terms of observable behaviors that students can easily self report. We established the underlying factor structure of these assessments as unidimensional and demonstrated across several samples that the assessment scales produce reliable and reproducible factor structures and scores. We also demonstrated these qualities in confirmatory factor analysis and also showed that the scales are unbiased and their scores are not influenced by demographic characteristics of respondents. In addition to face validity, which we established in our careful construction of the theoretical concepts and their operationalization, we also established the concurrent validity evidence of those scales when instruments were available, and the discriminant validity evidence of all ten assessment scales. Finally, by correlating RATE™ assessment scores to students' postgraduate outcomes, we also demonstrated the predictive validity evidence of the assessment scales. Taken together, these studies provide overwhelming evidence that the quality of the Core Competency Assessment Scales have reliable, valid, easily scalable score-based inferences and ideally suited to aid colleges and universities in assessing how their students develop core career and world readiness competencies.

References

American Association of Medical Colleges. (April, 2016). *The premed competencies for entering medical students*. Students & Residents. https://students-residents.aamc.org/real-stories-demonstrating-premed-competencies/premed-competencies-entering-medical-students

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Arum, R., Roksa, J., & Cook, A. (Eds.). (2016). *Improving quality in American higher education: Learning outcomes and assessments for the 21st century*. John Wiley & Sons.

Association of American Colleges and Universities. (2009). *Valid Assessment of Learning in Undergraduate Education (VALUE)*. Author. https://www.aacu.org/initiatives/value

Hart Research Associates (2015). *Falling Short? College Learning and Career Success: Selected Findings from Online Surveys of Employers and College Students Conducted*. Association of American Colleges and Universities. https://www.aacu.org/research/falling-short-college-learning-and-career-success

Hart Research Associates (2013). *It Takes More Than a Major: Employer Priorities for College Learning and Student Success: An Online Survey Among Employers.* Association of American Colleges and Universities. https://www.aacu.org/research/it-takes-more-than-a-major-employer-priorities-for-college-learning-and-student-success

National Association of Colleges and Employers. (2015). Career Readiness for the New College Graduate: A Definition and Competencies.

National Association of Colleges and Employers. "*What is Career Readiness?*" (April 17, 2024) https://www.naceweb.org/career-readiness/competencies/career-readiness-defined