Innovative Higher Education
https://doi.org/10.1007/s10755-026-09910-8

Building a Learning Evidence Ecosystem: A Partnership Model for Institutional Assessment

Megan Rodgers Good[1] · **Laura A. Lambert**[2] · **Joseph M. Kush**[3] · **Bethany Blackstone**[4]

Received: 9 July 2025 / Accepted: 22 April 2026

Abstract

Beyond the financial return on investment metrics that dominate public discourse, institutions need evidence systems aligned with their core mission: student learning. This study reports on a pilot implementation of a longitudinal learning evidence ecosystem designed to track developmental gains across the undergraduate experience. Leveraging a campus-wide data collection infrastructure with response rates exceeding 95%, we examined information literacy development using matched assessments from students' first and final years. Findings revealed modest but statistically significant gains, consistent with national benchmarks. More notably, students who entered the Honors College in their first semester demonstrated significantly stronger sophomore-year performance than later entrants, suggesting that early, structured academic engagement can yield disproportionate developmental benefits. Our collaborative model brings together assessment practitioners and academic leaders with curricular authority—such as the Dean of the Honors College—enabling exploratory, institutionally relevant assessment research. The findings illustrate how institutions can design learning evidence ecosystems that go beyond compliance, support evidence-informed decision-making, and offer a deeper understanding of educational value. We conclude with design principles for implementing similar models in other institutional contexts.

Keywords Longitudinal assessment · Information literacy · Honors education · Educational improvement

✉ Megan Rodgers Good
good2mr@jmu.edu

[1] Center for Assessment and Research Studies & Department of Graduate Psychology, James Madison University, MSC 6806, 298 Port Republic Road, Harrisonburg, VA 22807, USA

[2] Department of Integrated Science and Technology, James Madison University, Harrisonburg, VA, USA

[3] IXL Learning, San Mateo, CA, USA

[4] Honors College & Department of Political Science, James Madison University, Harrisonburg, VA, USA

Published online: 14 May 2026

Building a Learning Evidence Ecosystem: A Partnership Model for Institutional Assessment

The rising cost of undergraduate education continues to fuel public skepticism about the value of higher education. Outstanding U.S. student loan debt reached $1.77 trillion in 2025 (Hanson, 2025) as more students rely on borrowing to finance their education. Against this backdrop, a 2024 Pew survey found that only 22 percent of adults believe college is worth the cost if loans are required (Fry et al., 2024). Gallup data echo this concern: public confidence in higher education has declined sharply over the past decade, falling from 57 percent in 2015 to 36 percent in 2024 of adults expressing a great deal or quite a lot of confidence, while the share expressing very little or no confidence tripled from 10 to 32 percent during that period (Jones, 2024).

This skepticism has led many higher education advocates and critics to emphasize a single metric to characterize the value of college: graduates' earning potential. Georgetown's Center on Education and the Workforce, for instance, now ranks 4,600 institutions solely by projected lifetime Return on Investment (ROI; Georgetown University Center on Education & the Workforce, 2025). Importantly, ROI is typically operationalized solely as a function of financial considerations—costs of attendance, debt incurred, and future earnings. These metrics have a role to play in conversations about the value of higher education, but they fundamentally overlook higher education's core service: the facilitation of sustained intellectual development (e.g., critical thinking, information literacy, ethical reasoning, and civic engagement; see Rhodes & Finley, 2013). If we truly wish to "measure what matters" (Doerr, 2018; Grove, 1983), the sector must complement analyses of economic outcomes with evidence of student learning and growth.

ROI metrics should inform, not dominate, conversations about higher education's value. What is needed is not to replace financial indicators but to complement them with evidence systems capable of capturing learning – the sector's defining service. While external audiences rarely demand evidence about learning, institutions committed to their educational mission need evidence systems that operate at the institutional level and can adapt to emerging competencies (e.g., AI literacy) to inform concrete improvements.

Ultimately, debates about the value of higher education must consider whether and how institutions foster the knowledge, skills, and dispositions they claim as outcomes (Lederman, 2026). Yet the prevailing infrastructure and culture of assessment are not well-suited to answer these questions (Eubanks, 2023). This manuscript identifies the barriers that prevent most institutions from generating campus-wide learning evidence and offers a novel learning evidence ecosystem model designed to address these persistent gaps. To illustrate this argument, we present evidence from a pilot implementation of a campus-wide learning evidence ecosystem at a public research university. This pilot, focused on information literacy development and honors programming, demonstrates how institutions can empirically examine educational value using longitudinal learning data.

Challenges to Documenting Learning Gains at Scale

Higher education critics rarely consider what—or how much—students learn during their collegiate experience, and studies empirically documenting learning improvement in higher education remain rare (Banta & Blaich, 2011). This review synthesizes work on routine assessment practices, measurement challenges, and national efforts to create shared evaluation standards. It then positions our learning evidence ecosystem approach as a model for building collaborative partnerships to evaluate learning and generate actionable assessment data for institutions.

Fragmented Assessment in an Accreditation-Driven Culture

Institutional accreditors require every academic program (e.g., Psychology, BS) to articulate student-learning outcomes (SLOs) and demonstrate that assessment results inform improvement, as required by the Southern Association of Colleges and Schools Commission on Colleges (SACSCOC, 2023, Standard 8.2.A). Program faculty typically develop learning outcomes, create discipline-specific measures (e.g., rubrics for capstone projects, locally written exams, juried performances), aggregate results, and document curricular changes. Traditional assessment thus follows a confirmatory cycle: establish SLOs, collect evidence, judge adequacy, implement changes, and file a report (Fulcher et al., 2024).

Staffing and budget constraints reinforce this compliance-focused approach, leaving limited capacity for exploratory questions that could yield actionable insights (e.g., Do first-generation students develop differently across academic domains? How do high-impact practices interact with individual characteristics?). Such questions seldom surface under frameworks designed for verification rather than inquiry (Ewell, 2009; Good, 2023).

Moreover, assessment is fragmented by design: accreditor standards are written and reviewed at the program (and often general education) level, not as an integrated, institution-wide evidence system (Smith et al., 2015). A benefit of this predefined, program-level approach is stronger faculty ownership because faculty author outcomes and control curricular changes. Yet the resulting evidence is siloed, and local measures are rarely psychometrically validated (Eubanks, 2021), limiting their use for questions about educational value.

Recognizing the limitations of locally developed assessments, two national initiatives have attempted to provide broadly applicable measurement standards. The Collegiate Learning Assessment Plus (CLA+), a standardized performance-task examination, enables longitudinal "value-added" estimates across participating campuses. Multi-institution studies report modest but measurable gains in broad cognitive skills over the course of college, with improvements typically observed after two years and continuing through graduation (Arum & Roksa, 2011; Zahner & Steedle, 2015).

The Association of American Colleges & Universities' VALUE rubrics offer a faculty-scored alternative for outcomes such as critical thinking, written

communication, and information literacy (Rhodes & Finley, 2013). Faculty members trained to apply the VALUE Scoring Collaborative's rubrics score student work from across institutions (VALUE Scoring Collaborative, n.d.). However, because each campus submits different artifact samples each cycle, the data cannot track students or programs longitudinally. Further, the use of different artifacts presents significant measurement error (Hathcoat, 2018).

While the CLA+ and VALUE projects demonstrate cross-institutional measurement feasibility and support improvement efforts, their external administration and fixed assessment cycles limit institutional flexibility. Institutions cannot readily modify constructs to align with evolving priorities, integrate findings with campus-specific interventions and student characteristics, or pursue exploratory questions using comprehensive longitudinal data.

What remains largely absent from higher education is systematic, institution-wide infrastructure that institutions control for examining learning over time. Public stakeholders focus on ROI and graduation rates (Fry et al., 2024; Jones, 2024). Accreditors mandate program-level learning assessment (SACSCOC, 2023). National initiatives like VALUE and CLA+ provide important benchmarking frameworks. Yet few institutions have developed internal capacity to generate comprehensive longitudinal learning data and flexibly adapt assessed constructs to emerging competencies (such as AI literacy). Fewer still can integrate learning outcomes with institutional data to support exploratory inquiry into what works, for whom, and under what conditions. In short, these national measures weren't built for the institution-level, flexible use case we target here. These limitations point to the need for a fundamentally different infrastructure—one that collects high-quality data at the institutional level, supports exploratory analyses, and embeds those with authority to act on results. Our learning evidence ecosystem represents one such model.

Challenges to Measuring Learning

Learning presents inherently greater quantification challenges than post-graduation wages. Course-embedded measures engage motivated students but yield narrow, context-specific evidence. Campus-wide administrations broaden coverage yet, in low-stakes testing environments, often suffer from student disengagement and reduced effort. Test-taking motivation strongly predicts score validity (Wise & DeMars, 2005), while ceiling effects can mask growth when students begin college with moderate proficiency levels (American Educational Research Association et al., 2014).

Even well-designed instruments frequently produce scale scores that non-specialists struggle to interpret meaningfully. What does a "72" in critical thinking indicate about workplace readiness or civic engagement capacity? These interpretation challenges reinforce public preference for intuitive ROI calculations over abstract learning indicators, despite the latter's potential significance for understanding educational impact. Compounding these measurement challenges, even high-quality assessment data often fails to generate meaningful institutional change (Fulcher et al., 2014).

Yet even when measurement challenges are overcome and institutions produce high-quality data, another barrier often emerges: acting on that evidence. As Fulcher & Prendergast (2021) note, improvement efforts depend on having the authority and vision to act on assessment evidence. At the program level, faculty typically possess this authority and can directly modify curriculum or pedagogy. However, as assessment scales upward—to general education or institution-wide outcomes—authority becomes increasingly distributed across many faculty and units, making coordinated improvement challenging. These structural realities underscore the value of partnerships with units that hold both implementation authority and interest in learning research, such as honors or transfer services.

The Assessment-to-Action Gap

Assessment's fundamental purpose is not to satisfy accreditors, but to assemble evidence that can guide improvement, thereby enhancing learning and increasing credential value. But meaningful change requires champions with sufficient authority to align curriculum, faculty development, and resource allocation (Smith et al., 2018). Because most academic authority resides at the program level, few institutional actors possess both incentives and capacity to coordinate change at scale. This structural limitation points toward a critical need: assessment approaches that transcend traditional program boundaries while embedding decision-makers directly in the research and improvement efforts, as seen in emerging models like Auburn University's Insight Lab (Auburn University, n.d.). Without such integration, even high-quality assessment data may fail to generate the institutional changes necessary to enhance student learning.

Designing Assessment for Inquiry and Institutional Improvement

The literature reveals persistent limitations constraining current assessment approaches: fragmented program-level evidence collection, psychometric uncertainty in local measures, modest documented learning gains, predominantly confirmatory reporting cultures, and diffuse authority structures that impede systematic change. If institutions are to systematically examine learning as an ongoing object of inquiry and use that evidence to inform improvement, they need fundamentally different assessment infrastructure—infrastructure that can:

- capture longitudinal data at multiple points throughout students' academic journeys
- support exploratory analyses addressing equity concerns and program design questions
- embed institutional decision-makers directly within research processes to ensure findings inform practice
- accommodate diverse constructs reflecting stakeholder priorities and emerging competencies

The learning evidence ecosystem approach described in subsequent sections addresses these infrastructure needs. By combining routine, comprehensive data collection with a partnership model that embeds institutional decision-makers directly in research processes, it offers one viable approach toward 'measuring what matters' in higher education. Rather than competing with ROI metrics, this approach provides complementary evidence that can enrich institutional decision-making and enable more nuanced understanding of educational value. The following section describes how we operationalized these design principles through a longitudinal learning evidence ecosystem implemented at our university.

The Learning Evidence Ecosystem Approach

We implemented the learning evidence ecosystem at a public research university to generate credible, longitudinal learning data that inform action. While inquiry-oriented approaches to assessment have been advocated at the course and program level (Hutchings, 2010), the current ecosystem extends this logic to the institutional level, reconceptualizing existing assessment infrastructure as an integrated platform for exploratory research and continuous improvement rather than compliance (Ewell, 2009). The approach is built on four critical elements: (1) infrastructure for institution-wide data collection that (2) generates repeated measures from students over time, and (3) is sufficiently flexible to accommodate questions of interest to (4) campus partners with demonstrated interest and ability to act on results.

Infrastructure for Data Collection and Assessment

The system requires all students to complete assessment batteries at three critical junctures: the beginning of the first year, the mid-career mark (45–70 credits), and at graduation, extending the two-timepoint model described in Pastor et al. (2019). When students complete their degrees within typical timeframes, their responses can be matched longitudinally across all three collection points, enabling cross-sectional and developmental analyses. The assessment batteries combine validated learning outcomes instruments with survey instruments capturing high-impact practice engagement, demographic characteristics, and personality dimensions using the Big Five inventory (BFI-2-S; Soto & John, 2017).

This infrastructure builds upon over thirty years of required assessment administration for entering first-year and mid-career students at our institution, originally implemented in the late 1980s with registration holds for non-compliance. The senior data collection effort was incorporated in 2024. Compliance remains robust through established enforcement mechanisms: students cannot register for subsequent semesters without completing early assessments, and graduating seniors cannot receive diplomas without participating in the final battery. This approach consistently achieves 96–98% response rates across all time points, effectively creating near-census institutional data rather than the convenience samples typical of voluntary assessment efforts.

Building Partnerships to Learn About Learning

While the current ecosystem is built on structures and processes originally designed to generate assessment data for General Education compliance reporting, the current ecosystem represents a fundamental reconceptualization of this existing capacity. We now position these collection points as opportunities for systematic exploration of student development patterns, program effectiveness, and institutional impact. Unlike traditional program-specific assessment, this approach generates institution-wide data that can accommodate diverse constructs of stakeholder interest while maintaining the flexibility to modify instruments regularly based on emerging priorities and research questions. Recognizing that data collection alone does not guarantee improvement and that meaningful change requires champions with the authority to act on evidence (Fulcher & Prendergast, 2021), the ecosystem embeds institutional partners directly within research teams from question development through implementation. This partnership model addresses the authority diffusion problem that often prevents assessment findings from generating meaningful change in higher education settings.

Our pilot partnership emerged when the newly appointed Honors College Dean approached the institutional assessment support team seeking evidence about honors programming effectiveness. Participation in the Honors College is optional, and there are multiple paths into and through the Honors curriculum. Most students apply for admission to the Honors College as incoming first-year or transfer students, but others apply and join partway through their undergraduate career. A group of incoming first-year scholarship recipients are invited to join the Honors College without applying.

Unlike traditional academic units, administrative units like honors colleges, transfer services, and academic advising possess both incentives to demonstrate impact and authority to implement programmatic changes based on research findings. The Dean expressed explicit openness to findings regardless of whether they supported or challenged current practices—a research-friendly stance that contrasts sharply with defensive compliance cultures. Importantly, she could also identify ways the program could be modified in response to findings prior to data collection.

The resulting research team includes the Honors Dean, two Assessment Center faculty members, a doctoral student, a psychology faculty member with personality trait expertise, and an institutional research representative. This collaborative structure developed organically around shared interest in rigorous exploration rather than predetermined outcomes. Team members jointly developed research questions before data collection began, ensuring that investigations would address genuine programmatic concerns rather than post-hoc fishing expeditions.

Addressing Persistent Assessment Limitations

This ecosystem approach directly responds to the five limitations identified in our literature review:

1. Fragmented program-level evidence: The system generates institution-wide, required participation data rather than isolated program samples, enabling analyses that cross traditional academic boundaries and examine campus-wide patterns.
2. Psychometric uncertainty: Measures to be given at the institutional level must have psychometric information available to improve confidence in findings while maintaining flexibility to address context-specific questions.
3. Modest documented learning gains: Rather than seeking large effect sizes across the institution, the approach enables targeted exploration of specific constructs, populations, and programs where meaningful change might be more readily detected and understood.
4. Confirmatory reporting cultures: The partnership model explicitly prioritizes exploratory questions that can combine learning outcomes with personality, demographic, and engagement variables—investigations that traditional compliance frameworks typically ignore.
5. Diffuse authority structures: By embedding administrative partners with implementation authority directly within research teams, findings can translate more readily into programmatic modifications and institutional improvements.

Implementation and Scope

The ecosystem's flexibility allows intentional selection of assessment instruments that align with evolving institutional priorities and partner needs. For the first pilot of senior data collection, we selected a battery of instruments previously administered to this cohort as entering students and sophomores (for those graduating in four years). The battery included an information literacy assessment aligned with a general education learning outcome that had been administered repeatedly over many years and demonstrated strong psychometric properties. Future cycles may emphasize different competency areas as new partnerships develop, and institutional needs evolve.

The partnership with the Honors College represents the first application of the broader learning evidence ecosystem. It demonstrates how the model supports exploratory research questions that are both institutionally relevant and programmatically actionable, providing a template for future partnerships. We have already initiated a second partnership with transfer student services. The approach appears most suitable for administrative units with clear programming goals, implementation authority, and genuine commitment to evidence-based improvement—characteristics that distinguish promising partners from those seeking only confirmatory validation of existing practices. However, resource constraints may limit our ability to extend the partnership model to multiple units simultaneously.

The Current Study

This study was conducted at a public R2 (high research activity) university in the mid-Atlantic region. The institution enrolls approximately 21,000 undergraduate and 1,700 graduate students. The in-state to out-of-state ratio is ~ 75:25%, and about 1%

of the student body identified as international. Approximately 26% of students identified as students of color.

The Honors College serves approximately 6% of undergraduates across all majors and is demographically similar to the overall student body in terms of race and in-state/out-of-state status. Male students are slightly underrepresented in the Honors College relative to their proportion in the broader undergraduate population. The Honors College admits students both before matriculation and through later entry points, creating a mix of early and mid-career participants. Honors students pursue one of three pathways: a Minor Only Pathway, a Senior Portfolio Project (SPP) Pathway, or a Departmental Capstone Pathway. All Honors graduates complete a 19-credit curriculum that combines a first-year seminar, interdisciplinary coursework, and experiential learning requirements. The optional capstone pathways culminate in either a reflective ePortfolio and synthesis paper (SPP) or an independent, faculty-mentored research or creative project completed "with distinction" in the student's major field. Roughly half of Honors students complete one of these capstone options.

To demonstrate the ecosystem's capacity for generating actionable evidence, we report findings from our inaugural partnership with the Honors College focusing on information literacy development. Information literacy represents a foundational competency emphasized in a general education asynchronous module and should be particularly salient for those honors students who elect to complete a departmental capstone. This makes it an appropriate construct for examining both institutional learning patterns and program-specific effects.

This pilot study addresses two primary research questions developed collaboratively by our research team:

1. Do students show improvement in their information literacy scores from first-year to senior year?
2. What is the effect of participating in honors programming on students' information literacy scores?

These questions exemplify the ecosystem's exploratory potential: the first examines institutional-level learning gains using longitudinal data unavailable through traditional assessment approaches, while the second investigates program effectiveness through comparisons enabled by campus-wide data collection. Together, they demonstrate how the partnership model can generate evidence relevant to both institutional stakeholders and specific programmatic units.

Method

Participants

This study drew on a near census cohort of undergraduates at a public R2 university in the mid-Atlantic, following one entering class longitudinally. Students were

tested at three time points throughout their undergraduate careers: first as incoming first-year students in Fall 2021 ($n = 3{,}557$), again as sophomores in Spring 2023 ($n = 2{,}888$), and finally as seniors in the 2024–25 academic year ($n = 3{,}617$). To examine information literacy growth, we retained only students who had complete data, provided informed consent for research use, and were at least 18 years old. The university's Institutional Review Board approved the protocol (IRB-FY25-785).

Data Collection

Our learning evidence ecosystem systematically assesses student learning at three critical time points through the undergraduate experience. Entering first-year students complete assessments during a two-week window preceding the first-class period of their initial semester. Mid-career students (45–70 credit hours) participate during a single designated "Assessment Day" in February. Graduating seniors complete their final assessment battery during a six-week window at the beginning of their final semester, whether fall or spring.

All assessments are administered remotely and are unproctored. Students receive multiple reminder communications throughout each testing window, as needed. As described previously, compliance is maintained through established enforcement mechanisms: students cannot register for subsequent semesters without completing early assessments; graduating seniors may have their degree conferred, but they cannot receive physical diplomas until they fulfill the final assessment requirements. Across the three administrations, our data collection protocol typically yields response rates of 96–98%, including the pilot of the senior data collection structure. For this study, every participant completed one cognitive test and two attitudinal instruments, each detailed in the following section.

Measures

Information Literacy

The information literacy assessment ("InfoCore") is a 30-item, multiple-choice instrument that measures foundational information literacy competencies. This instrument was developed internally by content experts and reviewed by assessment professionals. Reliability estimates in the form of Cronbach's coefficient alpha were 0.75 for the first-year administration, 0.84 for the sophomore administration, and 0.81 for the senior administration. All three values are considered acceptable for this use-case, though the first-year values could be higher (Raykov & Marcoulides, 2011). Information literacy is a general education outcome at our institution, and this cohort completed an asynchronous first-year module with embedded quizzes that introduced the knowledge and skills later tested by the InfoCore. Many majors—especially those with a research emphasis—reinforce these skills in upper-division coursework. Items are scored dichotomously (correct = 1, incorrect = 0); thus, higher total scores reflect stronger information literacy proficiency.

Big Five Inventory-2 Short

The Big Five Inventory-2 Short (BFI-2-S; Soto & John, 2017) is a 30-item reduced form of the Big Five Inventory-2 (Goldberg, 1993) used to assess five personality domains, including: Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness. Factor analytic work by Soto and John (2017) demonstrated the BFI-2-S to produce highly reliable and valid scores for each of the five domains, with coefficient alpha values ranging from 0.73–0.84. Each dimension of the BFI-2-S is measured by five Likert-scale items ranging from 1 (Strongly Disagree) to 5 (Strongly Agree). Total scores for each dimension were calculated as the sum of item scores, resulting in a range of 5 to 25 for each dimension.

Student Opinion Survey

The Student Opinion Scale (SOS; Sundre & Moore, 2002) is an attitudinal measure that gauges test-taking motivations and perceived assessment importance. Administered after a test or test battery, the instrument reports two subscale scores: effort and importance. The effort subscale uses five Likert-type items ranging from 1 (Strongly Disagree) to 5 (Strongly Agree) to capture students' momentary effort levels (Sundre & Finney, 2002). For the current study we analyzed only the effort subscale; higher mean scores indicate greater examinee effort during the assessment session. Scores from this measure are considered reliable, with coefficient alpha values consistently in the 0.80s (Sundre & Moore, 2002).

Filtering Because the information literacy test was administered in a low-stakes setting, examinee motivation poses a threat to score validity (Sundre & Kitsantas, 2004; Wise & DeMars, 2005). We therefore applied a filter commonly used with the SOS. Only students whose five-item Effort subscale summed to 15 or higher—the midpoint that represents "neutral effort" on a 5-point Likert scale—were retained (Swerdzewski et al., 2011). Students falling below that threshold were presumed to have invested insufficient effort to yield confidence in their scores. Note that this had implications for differing sample sizes across the three assessment time points. Although the total number of respondents was similar across all time points, a greater number of students were removed from analyses due to low motivation at the second time point (i.e., sophomores in Spring 2023).

Statistical Analyses

RQ1: Learning Gains in Information Literacy

To obtain a baseline estimate of growth, we first conducted a repeated measures *t*-test comparing mean InfoCore scores from the first-year and senior administrations. This test yields the raw mean difference in information-literacy performance across the cohort but does not adjust for 1) students' self-reported test-taking effort

at each time point or 2) individual variation in starting proficiency. Consequently, subsequent models incorporated these factors to provide a more nuanced estimate of learning gains.

To estimate learning gains while simultaneously controlling for test-taking effort and starting proficiency, we complemented the repeated measures *t*-test with a multiple regression model. Senior year InfoCore scores served as the outcome, predicted by first-year InfoCore score and SOS ratings at both administrations:

$$InfoCore_{Senior} = \beta_0 + \beta_1 InfoCore_{FirstYear} + \beta_2 Effort_{FirstYear} + \beta_3 Effort_{Senior} + \epsilon \quad (1)$$

To aid in interpretability, first-year information literacy scores were centered around the mean ($M = 68.9$).

RQ2: Effect of Honors Programming on Information Literacy

To understand the impact of honors-specific programming on information literacy gains, a series of quasi-experimental analyses were conducted. Importantly, because the decision to participate in an honors-based curriculum is not randomly assigned (rather, students self-select to participate in honors programming), a naïve comparison between the two groups (honors students vs. not honors students) would likely result in biased parameter estimates. Propensity score methods can be used to correct for likely confounding by producing treatment (honors) and control (non-honors) groups with balanced covariate distributions, ultimately allowing for unbiased treatment effect estimates (Rosenbaum & Rubin, 1983).

A total of three treatments of interest were considered, including 1) ever honors, 2) first-semester honors, and 3) honors pathway. More specifically, the ever honors treatment variable was defined as 0 for students who never participated in honors-based programming and 1 for students who participated in honors-based programming for at least one semester. The first-semester honors treatment variable was defined as 0 for students who did not participate in honors during their first semester and 1 for those who participated in honors during their first semester. Importantly, both students who never participated in honors-programming and those who began participating in honors programming after their first semester (e.g., those who joined sophomore year) were included in this control group. Finally, the honors pathway treatment variable was defined as 1 for students who completed the Honors College's departmental capstone pathway and 0 for other students. Similarly, the control group included both students who never participated in honors-programming and those who did participate in honors programming but with either a minor or portfolio pathway.

To account for possible confounding, a rich set of pre-treatment covariates were included in estimation of the propensity scores, including: gender (1 = female, 0 = male), race (1 = white, 0 = non-white), underrepresented status, defined as whether a student is from a locality with low Associate and Baccalaureate attainment rates (1 = underrepresented, 0 = non-underrepresented), in-state status (1 = in-state, 0 = out-of-state), total transfer credits ($M = 7.1$, $SD = 10.5$), and each of the Big Five domains; Extraversion ($M = 20.3$, $SD = 3.9$), Agreeableness ($M = 22.8$,

SD = 3.4), Conscientiousness (*M* = 21.5, *SD* = 3.9), Neuroticism (*M* = 16.5, *SD* = 4.0), and Openness (*M* = 22.0, *SD* = 3.7). The race variable was originally captured using seven categories (American Indian/Alaska Native, Asian, Black/African American, Hispanic/Latino, Native Hawaiian/Other Pacific Islander, White, and Not Specified), with students able to select multiple identities. However, due to within-group sparseness and complete separation across subgroups, the race variable was dichotomized, resulting in *n* = 1350 White students (80%), *n* = 345 non-White students (20%). This is relatively in-line with the institution's larger enrollment characteristics, with approximately 26% students of color. Only students with complete data on the covariates were included; see Tables 1 and 2 for demographic information and descriptive statistics.

Propensity score weighting was conducted using the WeightIt package (Greifer, 2025) in R version 4.5.0 (R Core Team, 2025). Specifically, energy weighting (Huling & Mak, 2024) was conducted to estimate the average treatment effect on the treated (ATT), which can be conceptualized as the effect of honors programming among those who participated in honors. Covariate balance was assessed by ensuring all covariates and their two-way interactions achieved a standardized mean difference less than 0.10 (Austin & Mamdani, 2006). Finally, propensity score weights were incorporated in an outcome model, in addition to all pre-treatment covariates, to estimate the effect of the honors treatment on information literacy, controlling for first-year information literacy score and effort. For the ever honors and honors pathway treatment models, the outcome of interest was senior-year information literacy (i.e., distal impact), while for the first-semester honors treatment model, the outcome of interest was sophomore-year information literacy (i.e., immediate impact). By re-incorporating the covariates as predictors in the weighted outcome model, this approach is considered doubly-robust; so long as either the propensity score model

Table 1 Descriptive statistics for categorical variables by honors engagement

Variable	Ever Honors				First-Semester Honors				Honors Pathway			
	Control		Treatment		Control		Treatment		Control		Treatment	
	n	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
Gender	1,543	91.0	152	9.0	927	91.8	83	8.2	1,662	98.1	33	1.9
Male	563	36.5	46	30.3	292	31.5	27	32.5	601	36.2	8	24.2
Female	980	63.5	106	69.7	635	68.5	56	67.5	1061	63.8	25	75.8
Race												
Non-White	313	20.3	32	21.1	170	18.3	17	20.5	338	20.3	7	21.2
White	1230	79.7	120	78.9	757	81.7	66	79.5	1324	79.7	26	78.8
Underrepresented status												
Non-underrepresented	1515	98.2	150	98.7	917	98.9	83	100.0	1632	98.2	33	100
Underrepresented	28	1.8	2	1.3	10	1.1	0	0.0	30	1.8	0	0
In-state status												
Out-of-state	342	22.2	30	19.7	198	21.4	18	21.7	365	22.0	7	21.2
In-state	1201	77.8	122	80.3	729	78.6	65	78.3	1297	78.0	26	78.8

Table 2 Descriptive statistics for continuous variables by honors engagement

Variable	Ever Honors				First-Semester Honors				Honors Pathway			
	Control		Treatment		Control		Treatment		Control		Treatment	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Total Transfer Credits	6.87	10.22	9.30	13.00	6.26	7.16	5.30	7.00	7.08	10.57	7.12	7.48
BFI-2-S												
Extraversion	20.26	3.86	21.02	4.22	20.22	3.93	21.02	4.05	20.31	3.90	20.88	4.03
Agreeableness	22.71	3.40	23.28	3.73	22.98	3.47	23.17	3.76	22.73	3.43	24.06	3.21
Conscientiousness	21.45	3.89	21.93	4.17	21.66	3.95	21.64	4.20	21.46	3.92	23.27	3.25
Neuroticism	16.49	3.97	16.03	3.82	16.60	4.08	16.11	4.13	16.44	3.97	16.64	3.36
Openness	21.90	3.73	22.72	3.37	21.86	3.82	22.73	3.57	21.97	3.73	22.45	2.29
Effort												
First-year	19.96	2.75	20.45	2.72	20.18	2.73	20.61	2.44	19.99	2.75	20.82	2.64
Sophomore-year	19.05	3.55	20.45	2.72	19.62	2.92	19.96	2.81	19.03	3.56	20.35	2.64
Senior-year	19.40	2.90	19.89	2.78	19.24	3.36	19.66	3.66	19.44	2.90	19.79	2.36
Information Literacy												
First-year	68.20	13.86	75.88	8.99	69.62	12.58	75.66	7.98	68.74	13.72	76.46	7.86
Sophomore-year	66.67	17.81	75.33	13.15	67.67	17.22	75.90	12.53	67.24	17.63	78.99	11.52
Senior-year	71.10	15.38	76.84	13.86	71.72	15.05	77.03	11.78	71.49	15.36	77.98	11.84

BFI-2-S Big Five Inventory-2-Short version

and/or the outcome model is correctly specified, unbiased treatment effects may be obtained (Bang & Robins, 2005).

Results

RQ1: Learning Gains in Information Literacy

After retaining only those who completed information literacy assessments at freshman and senior year with more than neutral effort, data from $n = 1{,}700$ students were used for analyses. Results from the repeated measures *t*-test indicated a significant improvement in students' information literacy score over the course of their time at the university. Seniors performed an average of 2.72 percentage points better than they did their first year ($t(1699) = 7.07$, $p < .001$).

However, this approach does not account for effort at either time point, nor control for baseline information literacy. As shown in Table 3, a multiple regression model that included first-year scores and effort at both time points accounted for approximately 25% of the variance in senior-year information literacy scores ($F(3, 1696) = 191.8$, $p < .001$), offering a more robust alternative to the repeated measures *t*-test. More specifically, first-year score was a significant positive predictor of senior score, with students scoring on average 0.4 percentage points better at the senior administration for every point increase in first-year score. Examining effort at both the first-year and senior time points, results indicated that only effort on the senior administration was a significant positive predictor. Specifically, for every one-point increase in senior effort, senior information literacy scores increased an average of 1.66 percentage points.

While these results are statistically significant, they are less impactful from a practical standpoint. On average, senior scores improved 2.7 percentage points which is equivalent to slightly less than answering one additional question correctly.

RQ2: Effect of Honors Programming on Information Literacy

To address the second research question, the initial model tested whether any participation in honors programming predicted students' senior-year information-literacy scores. After retaining only those who completed information literacy

Table 3 Average senior InfoCore score after controlling for effort and first-year InfoCore score

				95% CI for *b*	
	b (SE)	*t*	*p*	LL	UL
Intercept	42.69 (1.68)	25.339	< 0.001	39.39	45.99
First year Effort	−0.21 (0.12)	−1.695	0.090	−0.46	0.03
Senior Effort	1.66 (0.12)	14.225	< 0.001	1.43	1.89
First year InfoCore	0.40 (0.02)	16.640	< 0.001	0.36	0.45

$n = 1{,}700$, $R^2 = 0.25$. *CI* Confidence interval, *LL* Lower limit, *UL* Upper limit

assessments at freshman and senior year with more than neutral effort, $n = 152$ students (9.0%) comprised the ever honors treatment group, while $n = 1{,}543$ (91.0%) comprised the control group. Energy weighting resulted in meaningfully balanced groups with respect to all covariates and their two-way interactions, with an average absolute value standardized mean difference of 0.004 (max = 0.034). Results from the doubly-robust weighted outcome model indicated treatment had a significant, positive effect on senior-year information literacy scores ($\delta = 2.41$, $p = .029$). That is, participating in honors programming for at least one semester at any point during the undergraduate experience resulted in senior-year information literacy scores that were 2.41 points higher than those not partaking in honors programming.

The second model considered the impact of first-semester honors participation on the immediate outcome of information literacy scores sophomore year. After retaining only those who completed information literacy assessments at freshman and sophomore year with more than neutral effort, $n = 83$ students (8.2%) comprised the ever honors treatment group, while $n = 927$ (91.8%) comprised the control group. Energy weighting resulted in meaningfully balanced groups with respect to all covariates and their two-way interactions, with an average absolute value standardized mean difference of 0.004 (max = 0.019). Results from the doubly-robust weighted outcome model indicated treatment had a significant, positive effect on sophomore-year information literacy scores ($\delta = 3.68$, $p = .023$). That is, participating in honors programming during the first semester resulted in sophomore-year information literacy scores that were 3.68 points higher than those from students who did not participate in honors during the first semester.

The third and final model considered the impact of completing an honors departmental capstone project on the distal outcome of information literacy scores senior year. After retaining only those who completed information literacy assessments at freshman and senior year with more than neutral effort, $n = 33$ students (1.9%) comprised the treatment group, while $n = 1{,}662$ (98.1%) comprised the control group. Energy weighting resulted in meaningfully balanced groups with respect to all covariates and their two-way interactions, with an average absolute value standardized mean difference of 0.005 (max = 0.054). Results from the doubly-robust weighted outcome model indicated treatment had no effect on senior-year information literacy scores ($\delta = 1.58$, $p = .421$), see Table 4. That is, completing an honors-based capstone project was not meaningfully related to senior-year information literacy scores.

Taken together, these analyses illustrate both the promise and limitations of an institutional learning evidence ecosystem. The system produced longitudinal, psychometrically credible findings about learning gains and program effects—results rarely obtainable through conventional assessment structures. Yet, as the modest effect sizes suggest, the true value of the ecosystem lies in its ability to enable systematic inquiry and improvement rather than in any single metric.

Table 4 Effect of honors programming on information literacy

	δ		*SE*	*p*
Model 1				
Ever Honors	2.41	*	1.10	.029
Model 2				
First-semester Honors	3.68	*	1.61	.023
Model 3				
Honors-based Capstone	1.58		1.96	.421

δcorresponds to average treatment effect on the treated (ATT) using propensity score weights. *$p < .05$

Discussion

Modest Gains in Learning

These findings—though modest in scale—illustrate how credible evidence of student learning can enrich conversations about educational value. Demonstrating learning growth, even on a single construct such as information literacy, offers a counterpoint to purely economic conceptions of institutional worth. Students showed an average improvement of 2.7 percentage points between their first year and graduation—a small but statistically significant gain, slightly smaller than prior research using instruments such as the CLA+ (Arum & Roksa, 2011; Zahner & Steedle, 2015). While these gains may appear modest in absolute terms, they reflect a broader pattern observed across higher education: learning occurs, but institutional-level effects tend to be incremental rather than dramatic. This pattern is perhaps unsurprising given the breadth and complexity of the four-year undergraduate experience, where concentrated growth is easier to detect within specific programs than across diverse educational pathways.

More encouragingly, our analysis revealed that participation in honors programming, particularly when initiated early, was associated with higher performance. Students who enrolled in honors during their first semester demonstrated significant improvements in sophomore-year information literacy that exceeded the gains observed in the broader sample. This finding underscores the value of early, structured academic programming and suggests that students may benefit from enhanced scaffolding during the foundational stages of their college experience, when academic habits and approaches to information evaluation are still forming.

Unexpectedly, we found no significant association between departmental capstone completion and information literacy performance. This null finding was unexpected, given the recognized value of undergraduate research as a high-impact practice and the clear connection to information literacy. However, this result becomes more interpretable when considered within our institutional context. The honors capstone experience at our institution is deliberately decentralized, with significant variation in structure, mentorship intensity, and performance expectations across disciplines. This finding exemplifies how our partnership model enables programmatic refinement—the Honors Dean can now act on evidence suggesting that

departmental capstones, while valuable, may benefit from supplemental scaffolding of information literacy competencies.

While disciplinary autonomy is preserved, this variability may limit the capstone's impact on information literacy competencies. These findings suggest that future honors curriculum enhancements could benefit from integrating common programming designed to complement the capstone experience across disciplines – an example of how assessment findings can inform curricular redesign when decision-makers are embedded in the research process. More broadly, even modest evidence of developmental growth demonstrates that rigorous assessment can yield actionable insights about student learning.

Study Limitations and Design Challenges

These findings must be interpreted within the context of several methodological constraints that reflect broader challenges in longitudinal learning assessment (e.g., controlling for prior ability and isolating curricular effects). Capturing authentic learning across broad constructs over extended timeframes remains inherently difficult, as evidenced by the strong predictive power of early proficiency scores on senior-year performance. This pattern—where students who begin with higher skills tend to maintain their advantage—raises important questions about ceiling effects and the enduring influence of pre-college preparation that may limit observable growth.

The low-stakes assessment environment presents additional motivational challenges. Although we employed effort-based filtering to address this concern, residual variation in student engagement likely attenuated the magnitude of observable score gains. Furthermore, beyond a brief asynchronous first-year module, students had no common information literacy interventions throughout their undergraduate experience. While we anticipated that major coursework would provide some reinforcement of these skills, the extent and quality of such reinforcement remains unknown and likely varied considerably across programs.

An additional consideration regarding timeframe to degree warrants discussion. In the current analyses, only those students who were seniors in the 2024–25 academic year after beginning as first-year students in Fall 2021 were included. That is, students who took shorter or longer to degree completion were not included in the analyses. While this timeframe appropriately captures the typical career trajectory for a majority of students at our institution (i.e., a 4-year graduation rate of approximately 67%), there may be important student subgroups who were disproportionately excluded. Research has consistently documented longer timeframes to degree completion for first-generation, low-income, and working students (Engle & Tinto, 2008). As a result, it is important to interpret the current findings within the context of a four-year graduation pathway, while recognizing that findings may differ for students outside this timeframe.

These constraints represent more than methodological limitations—they constitute fundamental design challenges that should be addressed in future assessment research. Robustly documenting learning gains requires not only sophisticated psychometric approaches and thoughtful analytical controls, but also educational

interventions with sufficient intensity and coherence to produce measurable effects (Finney et al., 2021). While these design challenges are substantial, they do not undermine the value of our more rigorous assessment model. Instead, they underscore the complexity of measuring learning and the need for assessment ecosystems that pair methodological precision with educational intensity.

Toward a More Rigorous and Action-Oriented Assessment Model

Despite these challenges and the modest nature of our observed gains, we find substantial reason for optimism about the potential for more rigorous assessment research. Our institutional ecosystem enables a methodologically sophisticated approach to studying learning that prioritizes longitudinal evidence, incorporates contextual variables such as personality and engagement, and employs advanced causal inference strategies like propensity score matching—analytical tools rarely deployed in assessment studies of this scope and scale.

This shift toward rigorous inquiry represents more than methodological advancement; it signals a fundamental reconceptualization of assessment's institutional role. Rather than treating assessment as an obligation to satisfy external accreditors, our model positions assessment infrastructure as a platform for exploratory research into questions of genuine institutional importance. This reorientation addresses a persistent cultural challenge in higher education: the tendency to assume that existing practices are effective without subjecting them to systematic scrutiny. Institutions often operate on faith that their programs work, that their curricula develop intended competencies, and that their investments in student success yield measurable returns. Our approach challenges this comfortable assumption by bringing rigorous analyses of learning to bear on programmatic decisions, enabling institutions to invest strategically in areas with documented successes while innovating in domains where expected learning gains cannot be demonstrated.

Building Institutional Capacity and Exploratory Assessment

A central insight from our implementation is that meaningful exploratory inquiry requires creating institutional capacity—not by adding resources, but by strategically reallocating existing ones. To make this partnership model possible, our internal assessment office had to reorganize how we allocated resources previously dedicated to accreditation compliance activities. This reallocation freed capacity to support the kind of exploratory research we describe here, shifting from producing routine compliance documentation toward collaborative investigation of questions that matter for practice.

Equally critical was the strategic selection of our partner. We chose to work with the Honors College Dean because she brought positive energy to the collaboration, demonstrated genuine openness to evidence-based program revision, and expressed explicit willingness to accept null findings if they emerged—a research-friendly stance that contrasts sharply with partners seeking only confirmatory validation of existing practices. This willingness to be surprised by data, rather than simply

affirmed by it, proved essential to the partnership's success. More broadly, this partnership represents a cultural shift in how our institution conceptualizes assessment itself. Assessment has long functioned primarily in a confirmatory mode: programs articulate learning outcomes, collect evidence, demonstrate adequacy, and file reports. Our ecosystem approach extends assessment beyond its traditional confirmatory role, introducing an exploratory dimension that enables experimentation and inquiry into student development, program effectiveness, and educational impact.

The research team approach emerged organically from this reorientation, bringing together assessment practitioners, institutional research, disciplinary faculty, and administrative decision-makers in collaborative inquiry. This team structure has proven energizing for all involved, because it is an intellectual enterprise with direct practical implications.

Practical Implications: From Evidence to Action

The partnership model embedded within our ecosystem approach addresses a fundamental implementation challenge that has plagued assessment efforts for decades: the gap between data collection and meaningful action. While there are examples of learning improvement at the degree program level (Lending et al., 2018), scaling such change is difficult (Fulcher & Prendergast, 2021).

Because administrative units like honors colleges, academic advising, and student success programs possess both motivation to demonstrate impact and authority to implement programmatic changes, embedding their leaders directly within assessment research teams creates natural pathways from findings to practice. Our Honors Dean's immediate engagement with the finding about early participation's disproportionate impact exemplifies this dynamic. Rather than receiving a report months after data analysis and treating findings as retrospective information, she participated in research design, interpretation, and application planning throughout the process. This integration broadens assessment's institutional role: alongside its ongoing reporting obligations, it now serves as a collaborative inquiry process that informs and improves educational practice at scale.

The ecosystem's flexibility represents another practical advantage. Our institution can readily integrate emerging constructs of interest and adapt assessment cycles to address evolving educational priorities. While information literacy provided a logical starting point given its central role in general education and the availability of psychometrically sound instruments, future investigations may examine more contemporarily relevant domains such as Artificial Intelligence literacy.

Future research might also incorporate qualitative data to understand emergent findings more deeply. For instance, student, faculty, and administrator interviews could provide unique insights into how stakeholders value and approach learning. Combining quantitative results—such as our finding about early honors participation—with qualitative investigations into students' experiences of that programming could illuminate mechanisms that explain observed patterns. Such mixed-methods approaches might help capture different dimensions of student development that quantitative measures alone cannot fully reveal, particularly regarding students'

perceptions of their own growth, the meaning they make of educational experiences, and the ways institutional culture shapes their engagement with learning opportunities.

Because our data approach census-level coverage, findings can be disaggregated and shared with various campus stakeholders—including academic deans, departmental unit heads, and student success coordinators—to identify equity gaps, understand differential program effectiveness across student populations, and guide strategic investments in educational infrastructure.

Implementation Considerations: Scaling the Model to Other Contexts

The learning evidence ecosystem described here builds upon over thirty years of established assessment infrastructure at our institution, with participation requirements enforced through registration holds and diploma withholding. This foundation—while enabling the high response rates essential to our approach—may not exist at institutions where assessment culture is less mature or where such enforcement mechanisms would face significant resistance.

Institutions seeking to implement similar models should consider beginning with modest ambitions that can be scaled over time. Rather than attempting to launch a comprehensive longitudinal system immediately, campus leaders might:

1. select a high-priority institutional learning outcome that is both measurable and meaningful to multiple stakeholder groups;
2. assemble a research team that includes a curricular decision-maker alongside assessment practitioners to ensure findings can inform practice;
3. identify an administrative partner whose unit has clear programming goals, implementation authority, and genuine commitment to evidence-based improvement—characteristics that distinguish promising collaborators from those seeking only confirmatory validation of existing practices;
4. begin with two-time point data collection (e.g., entering and graduating students) before expanding to mid-career assessments once initial processes are established; and
5. incorporate complementary data sources—such as student satisfaction and first-destination career outcomes—so that learning evidence appears alongside other positive institutional indicators.

Framing the ecosystem this way may make participation more appealing to senior administrators and enhance the perceived value of learning evidence within broader institutional narratives. Beyond implementation logistics, sustaining an evidence ecosystem depends on institutional culture—how learning data are interpreted, discussed, and acted upon.

Questions about how this ecosystem cultivates a culture of assessment and how institutions can scale such efforts extend beyond any single partnership. Culture change occurs gradually as stakeholders see evidence being used to inform genuine programmatic improvements rather than to fulfill compliance requirements. Each

successful collaboration reinforces assessment's credibility and normalizes its role as a resource for improving educational effectiveness. Sustaining this shift requires leadership that values evidence-based decision-making, investment in assessment capacity, and recognition of contributions to learning improvement. When these conditions are in place, assessment evolves from a peripheral compliance activity into a core institutional function that advances educational quality.

Reclaiming Assessment's Educational Promise While Acknowledging its Limitations

This work represents a fundamental shift toward treating assessment as rigorous educational inquiry rather than compliance theater. While our findings reveal the complexity of measuring learning—modest gains, null effects, and the persistent influence of prior preparation—they also demonstrate the value of sustained, methodologically sophisticated investigation. The real promise lies not in any single study's results, but in building institutional capacity for ongoing, evidence-based reflection on educational practice.

This work reframes assessment as sustained institutional inquiry into how, when, and for whom learning occurs. Demonstrating educational value requires acknowledging that meaningful insights often emerge from modest or null findings and acting on what evidence reveals. Though the gains reported here are limited, they illustrate the type of inquiry needed to understand and improve learning outcomes authentically. The learning evidence ecosystem offers a path forward—preserving accountability functions while transforming assessment into a research engine that generates actionable insights. In doing so, institutions can move beyond compliance toward evidence-based reflection and change, strengthening both educational quality and public confidence in higher education's purpose.

In an era when higher education faces legitimate questions about value and effectiveness, institutions cannot rely solely on economic metrics or reputation to justify their mission. By building assessment systems that document authentic learning, involve decision-makers in interpreting results, and share findings, colleges can restore public confidence while strengthening their own educational core. Such transparency does not diminish higher education's promise—it fulfills it, demonstrating that student learning remains its defining value.

Authors Contributions All authors contributed to the study's conception and design. Data cleaning, analysis, and results interpretation were led by Joseph M. Kush and Laura A. Lambert. Megan Rodgers Good wrote the literature review and discussion sections. Bethany Blackstone contributed the honors program content and provided comprehensive editorial feedback and refinement. All authors reviewed and approved the final manuscript.

Funding The authors received no financial support from any organization for the research, authorship, or publication of this study.

Data Availability Data available pending approval through James Madison University.

Code Availability Code available upon request.

Declarations

Ethical Approval This study utilized retrospective institutional data collected at James Madison University for the purposes of program improvement and evaluation. The research protocol was reviewed by the James Madison University Institutional Review Board and determined to be exempt from further review under federal guidelines. IRB approval was granted under protocol number IRB-FY25-785.

Consent for Human Participants This study was approved by the James Madison University Institutional Review Board (IRB-FY25-785) and determined to be exempt. All data used in this research were originally collected for institutional improvement purposes. Participation in the data collection was part of standard university processes, and students were explicitly given the option to consent to the use of their responses for research purposes. Only data from students who provided consent were included in this study.

Conflict of interest The authors declare that they have no financial or non-financial conflicts of interest related to the content of this manuscript. No funding sources or affiliations have influenced the study's design, execution, interpretation, or reporting.

References

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

Arum, R., & Roksa, J. (2011). *Academically adrift: Limited learning on college campuses.* University of Chicago Press.

Auburn University. (n.d.). *Insight Lab*. Office of the provost and academic affairs. https://www.auburn.edu/academic/provost/academic-insight/insight-lab/

Austin, P. C., & Mamdani, M. M. (2006). A comparison of propensity score methods: A case-study estimating the effectiveness of post-AMI statin use. *Statistics in Medicine, 25*(12), 2084–2106. https://doi.org/10.1002/sim.2328

Bang, H., & Robins, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics, 61*(4), 962–972. https://doi.org/10.1111/j.1541-0420.2005.00377.x

Banta, T. W., & Blaich, C. (2011). Closing the assessment loop. *Change: The Magazine of Higher Learning, 43*(1), 22–29. https://doi.org/10.1080/00091383.2011.538642

Doerr, J. (2018). *Measure what matters: How Google, Bono, and the Gates Foundation rock the world with OKRs*. Penguin.

Engle, J., & Tinto, V. (2008). Moving beyond access: College success for low-income, first-generation students. The Pell Institute for the Study of Opportunity in Higher Education. https://files.eric.ed.gov/fulltext/ED504448.pdf

Eubanks, D. (2021). Assessing for student success. *Intersection: A Journal at the Intersection of Assessment and Learning, 2*(2), 1–11. https://doi.org/10.61669/001c.21272

Eubanks, D. (2023). Part 1: Assessment standards are broken. *Assessment Update, 35*(6), 1–3.

Ewell, P. T. (2009). Assessment, accountability, and improvement: Revisiting the tension (NILOA Occasional Paper No. 1). National Institute for Learning Outcomes Assessment.

Finney, S. J., Wells, J. B., & Henning, G. W. (2021). The need for program theory and implementation fidelity in assessment practice and standards (Occasional Paper No. 51). National Institute for Learning Outcomes Assessment.

Fry, R., Braga, D., & Parker, K. (2024, May 23). Is college worth it? Pew Research Center. https://www.pewresearch.org/social-trends/2024/05/23/is-college-worth-it-2/

Fulcher, K. H., & Prendergast, C. O. (2021). *Improving student learning at scale: A how-to guide for higher education.* Stylus.

Fulcher, K. H., Good, M. R., Coleman, C. M., & Smith, K. L. (2014, December). *A simple model for learning improvement: Weigh pig, feed pig, weigh pig* (Occasional Paper No. 23). National Institute for Learning Outcomes Assessment.

Fulcher, K. H., Good, M. R., & Sanchez, E. R. H. (2024). *Assessment 101 in higher education: The fundamentals and how to apply them*. Routledge. https://doi.org/10.4324/9781032730264

Georgetown University Center on Education and the Workforce. (2025). *Ranking 4,600 colleges by ROI.* https://cew.georgetown.edu/cew-reports/roi2025/

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist, 48*(1), 26–34. https://doi.org/10.1037/0003-066X.48.1.26

Good, M. R. (2023). Part 2: For the sake of educational quality—Let's minimize bureaucracy and maximize impact. *Assessment Update, 35*(6), 4–5. https://doi.org/10.1002/au.30368

Greifer, N. (2025). *WeightIt: Weighting for covariate balance in observational studies.* (Version 1.4.0.9002) [R package]. https://ngreifer.github.io/WeightIt/

Grove, A. S. (1983). *High output management*. Random House.

Hanson, M. (2025, March 16). Student loan debt statistics. Education data initiative. https://educationdata.org/student-loan-debt-statistics

Hathcoat, J. D. (2018). The role of assignments in the multi-state collaborative: Lessons learned from a master chef. *Peer Review, 20*(4), 7–10.

Huling, J. D., & Mak, S. (2024). Energy balancing of covariate distributions. *Journal of Causal Inference, 12*(1), 1–22. https://doi.org/10.1515/jci-2022-0029

Hutchings, P. (2010). *Opening doors to faculty involvement in assessment* (NILOA Occasional Paper No. 4). National Institute for Learning Outcomes Assessment.

Jones, J. (2024, July 8). *U.S. confidence in higher education now closely divided.* Gallup.com. https://news.gallup.com/poll/646880/confidence-higher-education-closely-divided.aspx

Lederman, D. (2026, January 20). Better defining and measuring higher ed's value. *Inside Higher Ed*. https://www.insidehighered.com/opinion/columns/tough-love/2026/01/20/better-defining-and-measuring-higher-eds-value

Lending, D., Fulcher, K. H., Ezell, J. D., May, J. L., & Dillon, T. W. (2018). Example of a program-level learning improvement report. *Research & Practice in Assessment, 13*, 34–50.

Pastor, D. A., Foelber, K. J., Jacovidis, J. N., Fulcher, K. H., Sauder, D. C., & Love, P. D. (2019). University-wide assessment days: The James Madison University model. *The AIR Professional File, 144*, 1–13.

R Core Team. (2025). *R: A language and environment for statistical computing* (Version 4.5.0) [Computer software]. R Foundation for Statistical Computing. https://www.r-project.org/

Raykov, T., & Marcoulides, G. A. (2011). *Introduction to psychometric theory*. Routledge/Taylor & Francis Group.

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

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41–55. https://doi.org/10.1093/biomet/70.1.41

Smith, K. L., Good, M. R., Sanchez, E. H., & Fulcher, K. H. (2015). Communication is key: Unpacking "use of assessment results to improve student learning." *Research & Practice in Assessment, 10*, 15–29.

Smith, K. L., Good, M. R., & Jankowski, N. (2018). Considerations and resources for the learning improvement facilitator. *Research & Practice in Assessment, 13*, 20–26.

Soto, C. J., & John, O. P. (2017). Short and extra-short forms of the Big Five Inventory-2: The BFI-2-S and BFI-2-XS. *Journal of Research in Personality, 68*, 69–81. https://doi.org/10.1016/j.jrp.2017.02.004

Southern Association of Colleges and Schools Commission on Colleges. (2023). *The principles of accreditation: Foundations for quality enhancement* (2023 ed.). https://sacscoc.org/app/uploads/2023/01/2023-POA.pdf

Sundre, D. L., & Finney, S. J. (2002, April 1–5). *Enhancing the validity and value of learning assessment: Furthering the development of a motivation scale* [Paper presentation]. American Educational Research Association Annual Meeting, New Orleans, LA, United States.

Sundre, D. L., & Kitsantas, A. (2004). An exploration of the psychology of the examinee: Can examinee self-regulation and test-taking motivation predict consequential and non-consequential test performance? *Contemporary Educational Psychology, 29*(1), 6–26. https://doi.org/10.1016/S0361-476X(02)00063-2

Sundre, D. L., & Moore, D. L. (2002). The Student Opinion Scale: A measure of examinee motivation. *Assessment Update, 14*(1), 8–9.

Swerdzewski, P. J., Harmes, J. C., & Finney, S. J. (2011). Two approaches for identifying low-motivated students in a low-stakes assessment context. *Applied Measurement in Education, 24*(2), 162–188. https://doi.org/10.1080/08957347.2011.555217

VALUE Scoring Collaborative. (n.d.). *VALUE scoring collaborative.* Association of American Colleges & Universities. https://www.aacu.org/value/value-scoring-collaborative

Wise, S. L., & DeMars, C. E. (2005). Low examinee motivation in low-stakes assessment: Problems and potential solutions. *Educational Assessment, 10*(1), 1–17. https://doi.org/10.1207/s15326977ea1001_1

Zahner, D., & Steedle, J. T. (2015, April). *Comparing longitudinal and cross-sectional school effect estimates in postsecondary education* [Paper presentation]. National Council on Measurement in Education, Chicago, IL, United States. https://files.eric.ed.gov/fulltext/ED582107.pdf

Megan Rodgers Good is Executive Director of the Center for Assessment and Research Studies and Assistant Professor in Graduate Psychology at James Madison University. She earned her Ph.D. in Assessment and Measurement from James Madison University and previously served as Auburn University's inaugural Director of Academic Assessment and SACSCOC liaison. Her work focuses on building sustainable systems for institutional improvement, with research interests in learning improvement, accountability, and the value of higher education.

Laura A. Lambert is an A&P faculty member and assessment coordinator in the department of Integrated Science and Technology at James Madison University. She earned her Ph.D. In Assessment and Measurement from James Madison University. Her research interests include methodological challenges in the field of undergraduate research experiences, learning improvement, and equity in assessment practices.

Joseph M. Kush is a psychometrician at IXL learning. He completed his B.S. in psychology from Syracuse University and his Ph.D. in research, statistics and evaluation from the University of Virginia. His primary research interests focus on latent variable modeling, measurement invariance, multilevel modeling, and propensity score methods.

Bethany Blackstone is the Dean of the Honors College and a professor of political science at James Madison University. She earned her doctorate in political science from Emory University, and her research focuses on judicial and congressional politics and the policymaking role of U.S. courts. Her work in honors education centers on high-impact undergraduate learning experiences.