

# Do STEM Programs Provide Authentic Assessments? Exploring Their Prevalence and Opportunities for Growth

## Aradaryn Marsh, Althea Hagan, Lisa Benson

Clemson University, United States of America

**Abstract:** Authentic assessments in undergraduate STEM courses promote higher-order thinking and real-world relevance and are increasingly recognized for supporting student learning and inclusivity. This study examines the prevalence of authentic assessments– characterized by realism, cognitive challenge, and evaluative judgment– within an undergraduate natural resource (NR) department at a research-intensive (R1), predominantly white institution (PWI) in the United States (U.S.) Instructors of NR courses completed a survey to score the authenticity of their assessments, providing data on the design and implementation. Results revealed that while 42% of courses (n=26) met high authenticity standards, traditional exams remained dominant. Only four high-authenticity scoring courses used authentic assessments in place of traditional exams. These findings highlight the need for broader integration of authentic assessment practices to foster inclusivity and prepare students for real-world challenges. This manuscript also offers insights for future research and implications for faculty development, policy, and assessment practices.

Keywords: Authentic Assessment; Higher Education; Inclusive Learning; Racially Marginalized Students. DOI: <u>https://doi.org/10.31756/jrsmte.4111SI</u>

## Introduction

## Context and Importance of Authentic Assessment in STEM

Science, Technology, Engineering, and Mathematics (STEM) education is crucial in preparing students for real-world challenges upon graduation. However, traditional assessment practices—dominated by standardized and memory-based evaluations—often fail to measure essential skills such as critical thinking, problem-solving, and application of knowledge in professional contexts (Biggs & Tang, 2011; Shepard, 2000). Recent educational initiatives, including those from the National Science Foundation (NSF), emphasize the need for equitable and inclusive approaches to education that better support diverse student populations (National Science Foundation, n.d.). "Authentic assessments"—evaluations that require students to apply their knowledge in real-world contexts, engage in higher-order thinking, and reflect on their performance—offer a promising alternative that moves beyond memorization toward inclusivity (Shepard, 2000; Villarroel et al., 2017; Schultz et al., 2022).

Despite growing advocacy for improved assessment practices, traditional exams remain the dominant form of evaluation in STEM disciplines (Schultz et al., 2022). These conventional methods often privilege students with strong test-taking abilities while disadvantaging those from racially marginalized (RM) backgrounds, who may face additional barriers in standardized testing environments (Supovitz & Brennan, 1997; Penfield & Lee, 2010). Assessment practices that do not account for diverse cultural and experiential knowledge can contribute to disparities in student performance and retention rates (Solano-Flores & Nelson-Barber, 2001). The natural resource (NR) discipline, in particular, has historically struggled with racial underrepresentation, and the persistence of exclusionary assessment practices may further limit the participation and success of RM students (Sharik et al., 2015; Polka, 2019; Rodriguez & Peterson, 2020).

## Addressing the Research Gap

While studies have explored the benefits of authentic assessment across various disciplines, limited research has examined its prevalence within NR programs at research-intensive (R1), predominantly white institutions (PWIs). Furthermore, there is a need to investigate whether instructors' perceptions of their assessments align with objective measures of authenticity. Understanding these dynamics is crucial for advancing inclusive assessment practices and fostering more equitable learning environments.

This study addresses this gap by investigating the extent to which authentic assessments are implemented within an undergraduate NR department at an R1 PWI in the United States. Using a survey developed from Hobbins et al.'s (2021) rubric for authentic assessment, this research seeks to:

- 1. Examine the prevalence of authentic assessments within NR courses.
- Compare instructor self-reported assessments with researcher-assigned authenticity scores to assess potential discrepancies.
- 3. Highlight implications for faculty development and institutional policy to promote inclusive assessment practices.

This research aims to inform the integration of authentic assessment practices in higher education, enriching the learning experiences of diverse student populations and ensuring that educational practices align with the evolving needs of students and the professional world. This study aims to address the following research question: What is the prevalence of authentic assessment within an undergraduate natural resource (NR) department at a research-intensive (R1) predominantly white institution (PWI) in the United States (U.S.)?

While the resulting survey is a valuable tool, its true significance lies in what it can reveal about current assessment methods and how that information can be applied in research and practice. By providing educators, administrators, and researchers with a standardized way to evaluate assessment authenticity, this work opens the door for more consistent and meaningful comparisons across disciplines and institutions. This research helps lay the groundwork for enhancing assessment strategies to better prepare students for real-world challenges and contribute to more inclusive and equitable educational environments.

#### Authentic Assessment: A Framework for Evaluation

The essence of authentic assessment, which serves as a pedagogical framework for this study, falls within three overarching dimensions: realism, cognitive challenge, and evaluative judgment (Villarroel et al., 2017). Hobbins et al. (2021) provide a definition rooted within these three dimensions: "Authentic assessment refers to a formally evaluated assessment activity which engages students with problems or important questions that are relevant to everyday life beyond the classroom; prompts students to use higher levels of thinking to extend knowledge and thinking, while also providing an opportunity to enhance self-regulated learning by engaging with grading criteria and providing and receiving feedback" (p. 1262).

Realism describes the alignment of questions or tasks in the classroom with tasks one may face in one's professional life (Hobbins et al., 2021). This dimension is especially relevant in performance-based tasks where students demonstrate their knowledge in a way that represents performance found in the workplace (Villarroel et al., 2017; Saye, 2013; Palmer, 2004).

Cognitive challenge refers to assessments requiring higher-order cognitive skills, as defined by Bloom's taxonomy (Anderson et al., 2001). The transfer of knowledge necessary for cognitive challenge goes hand in hand with the previously mentioned dimension of realism, in that students need to be able to practice skills required of them beyond a traditional exam at a university. In other words, successfully memorizing information for a decontextualized exam does not indicate how well that same individual can utilize that knowledge when needed in the real world (Villarroel, 2017; Bloxham & Boyd, 2007).

Evaluative judgment refers to the need for students to judge the quality of their work and performance. Two ways exist to accomplish this (Villarroel et al., 2017; Hobbins et al., 2021). First, students should judge the quality of their work based on criteria (typically through provided rubrics), and second, they should provide and receive feedback on their work (Hobbins et al., 2021).

#### Workforce Preparation and Student Well-Being

The current workforce faces a significant challenge as recent graduates often lack the necessary skills and adaptability to meet the demands of a professional environment. Employers express concerns about graduates' problem-solving abilities, adaptability, and communication skills (Singh & Ramly, 2014; Villarroel et al., 2017). This issue contributes to heightened stress among graduates, who feel unprepared as they enter the workforce (Villarroel et al., 2017). Authentic assessments can help bridge this gap by aligning student learning experiences with professional expectations (Hobbins et al., 2021).

Additionally, traditional, high-stakes assessments contribute to significant test anxiety, particularly when framed as evaluative tools with severe consequences (von der Embse et al., 2017). Students in high-stakes, summative courses often describe these exams as "weed-out" mechanisms prioritizing stress over learning (Kenyon, 2023). In contrast, assessments framed as "learning scenarios" or "activities" result in lower levels of test anxiety and improved performance (Keppell & Carless, 2006; Durning et al., 2016). For example, although presentation-based assessments can still induce anxiety, they provide students with valuable workplace preparation, improving confidence in professional communication (Jones et al., 2020).

Surface learning, often involved in passing traditional assessments, is no longer helpful in a world with information a click away (Lynam & Cachia, 2017). Educators must prepare students to be employees who can problem-solve, make decisions, source credible information, communicate, and work well in a team; it is no longer enough to be knowledgeable in their field (Lynam & Cachia, 2017). Educators can create dynamic and engaging environments that

foster critical thinking, problem-solving skills, and creativity by moving away from traditional assessment practices and embracing authenticity in practice.

## **Paper Structure**

This manuscript is organized to present findings on the prevalence of authentic assessments within an undergraduate NR department. The introduction provided background on the significance of authentic assessments in promoting inclusive learning. The methods section describes the survey development and implementation, detailing the process of gathering instructor data. The results section presents key findings on assessment practices within the department, followed by a discussion of these findings in the context of existing literature. Finally, the conclusion outlines the implications for faculty development and future research directions.

## Methods

#### **Study Design and Conceptual Framework**

This work is inspired by Hobbins et al. (2021), who documented authentic assessment within an undergraduate health science curriculum using a rubric aligned with Villarroel et al.'s (2017) core dimensions of authentic assessment. This study builds upon their work by transforming the rubric into a survey format tailored to systematically assess instructors' assessment practices.

### **Participant Sampling and Context**

All instructors teaching undergraduate NR courses at a R1 PWI were invited to participate (N=50). A total of 26 instructors responded, yielding a 52% response rate. While this convenience sample limits broad generalizability, it offers a meaningful snapshot of departmental assessment practices. The sample size was appropriate for descriptive analysis of ordinal survey data, providing reliable medians, modes, and frequency distributions (Bartlett, Kotrlik, & Higgins, 2001; VanVoorhis & Morgan, 2007). An estimated margin of error of approximately 10% at a 95% confidence level further supports the reliability of the descriptive findings (Cochran, 1977).

#### Survey Development Process

Survey development incorporated cognitive testing and iterative refinement to ensure clarity and alignment with research goals (Willis & Artino, 2013). Experts in education research and faculty development reviewed early drafts, leading to improvements such as reordering questions, revising language for clarity, and enhancing inclusivity. Additional usability features, including a progress bar and navigation buttons, were added to support participant experience.

#### Survey Instrument Components:

The final survey included the following sections:

- Consent and Course Identification: Instructors began by reviewing an informed consent statement and identifying the course they reported on. They were asked to complete the survey separately for each course they taught.
- 2. Assessment Focus: Instructors were prompted to select the single assessment type most influential in determining final grades (e.g., midterm, final exam, project).
- 3. Assessment Characteristics:
  - a. Multiple-choice questions aimed to determine the dominant form(s) of participation required for the identified assessment (e.g., multiple choice, presentation, or essay), the cognitive skill(s) engaged by the assessment (e.g., memory skills, application, and/or analytical skills), and the methods of feedback used for the assessment.
  - b. Likert-scale questions asked instructors to rate the extent to which the assessment engaged students with problems relevant to their future professional lives and the opportunities provided for students to judge their own performance.
- Demographic Information: Instructors reported their current position, age group, racial or ethnic background, and gender identity.
- 5. Supporting Documentation: Instructors were given the option to upload files of their course syllabus and the assessment in question to provide additional context.

## **Data Analysis**

Survey data were analyzed to determine the prevalence of authentic assessment methods within the department. Given the limited response rate (covering 26 out of 50 courses), there was insufficient data to perform statistical comparisons. Instead, descriptive statistics, including frequencies and percentages, were used to summarize key findings.

The primary researcher scored each assessment based on survey responses and documents uploaded by instructors. This scoring process used the rubric originally by Hobbins et al. (2021), which was used in survey development. The researcher's scores were necessary for standardizing the results, enabling a systematic comparison and evaluation of assessment practices across courses. These standardized researcher-assigned scores were analyzed descriptively to identify trends and patterns in authenticity across the curriculum, which are summarized in the following section.

#### Results

Before interpreting the results of this analysis, it is essential to acknowledge that the dataset's relatively small sample size (n=26) limits broader generalizability. Given the ordinal nature of Likert-scale data, the study reports medians, modes, and frequency distributions rather than means and standard deviations, ensuring appropriate treatment of ordinal-level data (Jamieson, 2004). While self-reported scores offer valuable insight into instructors' perceptions of their assessments, they may be influenced by individual biases or varying interpretations of authenticity criteria. Therefore, the researcher-assigned scores, based on a standardized rubric, provide a more consistent measure of

authenticity and are the primary focus of analysis. **Figure 2** below highlights the discrepancies between the two scores (self-reported by instructors and researcher-assigned). By focusing on the researcher's scores, we aim to minimize subjectivity and enhance the findings' validity to reflect the authenticity level accurately.

Also worth noting is a possible participation bias inherent in the dataset, as those who are particularly passionate about providing unique and meaningful assessments—potentially resulting in higher authenticity scores—may have been more likely to respond to the survey invitation. This suggests that the survey responses may disproportionately represent individuals already committed to authentic assessment practices. As a result, researchers may infer that the prevalence of such practices among those who did not participate is likely lower.

## **Response Rate**

Out of the 50 instructors invited to participate in the survey, responses were collected for 31 courses, with 26 responses providing sufficient information for analysis. This resulted in a response rate of 52%. Given the focus on descriptive rather than inferential statistics, the sample size was deemed appropriate for identifying trends in assessment practices within the department.

#### **Prevalence of Authentic Assessments**

A percentile-based approach was used to determine the prevalence of authentic assessments. The 75th percentile cutoff (scores of 9 to 12 on the 0–12 scale) was used to classify an assessment as high in authenticity, ensuring that only assessments meeting a substantial portion of the rubric's criteria were categorized as authentic. The decision to use this cutoff also aligns with standard practices in educational assessment (Cohen, 1988; McMillan &Foley, 2011), where scores in the top quartile are often associated with high levels of achievement or quality.

Among the 26 responses included in the analysis, 42.3% (n=11) of assessments fell within the high authenticity range (refer to Figure 3). Despite these scores, traditional exams remained the predominant assessment type. Among the 11 high-authenticity courses, only four *replaced* traditional exams with authentic assessments, as evidenced by the instructors' uploaded syllabi. The remaining seven required students to also take traditional assessments.

## **Descriptive Statistics**

The researcher-assigned scores were analyzed using appropriate measures for ordinal data (see Table 1). The median score of 8.0 and mode of 9.0 suggest that many courses incorporated some degree of authentic assessment, though variability remained high. While most scores aligned between self-reported and researcher-assigned scores, there were notable instances where the researcher's assessment was more critical, leading to lower scores, with only two instances of the researcher-assigned score being higher than self-reported scores (refer to Figure 2).

#### **Demographic Trends**

The dataset includes five distinct age groups: "25-34", "35-44", "45-54", "55-64", and "65 or older". The most common age group is "35-44," with 12 respondents. In terms of racial and ethnic background, five were identified:

"White," "Prefer not to say," "Hispanic or Latino," "Hispanic or Latino & White," and "Asian." The majority of participants identified as "White" (20 out of 26). Additionally, most respondents self-identified as male (18 out of 26), while the other eight self-identified as female.

## Table 1

Summary of Descriptive Statistics for Researcher-Assigned Assessment Scores

Statistic	Score
Median	8.0
Mode	9.0
Minimum	0.0
Maximum	11.0

## Figure 2

Discrepancies between researcher-assigned and instructor-reported scores highlight tendencies to over- or underreport authenticity



*Note*: This scatter plot shows the frequency of discrepancies between researcher-assigned and instructors' self-reported authenticity scores. The x-axis represents the difference in points between the two scores, while the y-axis shows how often each discrepancy occurred.



## **Figure 3** Distribution of Instructor Self-Reported and Researcher-Assigned Authenticity Scores

*Note*: This bar chart compares the frequency distribution of authenticity scores as reported by instructors (blue bars) and as assigned by researchers (yellow bars). The 75th percentile cutoff is indicated by the dashed red line at a score of 9. Scores ranging from 9 to 12 are considered "authentic," with these assessments being categorized as such based on their evaluation scores.

#### **Gender-Based Discrepancies in Scoring**

Over-reporting authenticity was observed in four male instructors and three female instructors. Under-reporting authenticity was less common, with only one male and one female instructor reporting scores lower than those assigned by the primary researcher. Although these discrepancies are not directly related to the primary research question, they are crucial to recognize as they highlight the potential for over- or under-reporting tendencies in self-reported data. This calls for the development of survey instruments that can mitigate bias and capture more accurate self-assessments.

## Discussion

### Limitations

While this study provides important insights into assessment practices within an NR department, several limitations must be considered. First, the relatively small sample size (n=26) and the study's confinement to a single department at an R1 and PWI limit the generalizability of the findings. Departmental culture, including attitudes toward assessment, diversity initiatives, and instructional support, likely influenced the prevalence of authentic assessments observed and may differ significantly across institutions or disciplines. It's also possible that the dataset only partially captures the diversity of assessment practices across the entire department, especially since participants who are more engaged with authentic assessment may have been more likely to respond, introducing potential participation bias. Additionally, while the majority of female instructors in the department responded to the survey (8 out of 12), the overall gender imbalance in the department's faculty (12 female, 23 male) limited our ability to draw meaningful conclusions about gender-based differences in assessment practices.

This study relied on Likert-scale responses representing ordinal, not continuous data. While many studies treat Likertscale responses as interval data, doing so assumes equal distances between response options, which is not always valid (Jamieson, 2004). We addressed this by using median, mode, and categorical groupings rather than mean and standard deviation. These methods limit the depth of insight into experiences with authentic assessments. Qualitative data, such as interviews or focus groups, could provide a deeper understanding of the challenges and successes in implementing these practices.

Standardized, researcher-assigned scores may introduce subjectivity due to the limitations of the survey design. The survey's structure may have prevented participants from providing the detail necessary for a fully accurate assessment, potentially affecting the researcher-assigned scores. As a result, the rubric-based scores assigned by researchers to the instructors' descriptions of assessments may have led to under- or over-estimating certain assessments' authenticity. While the survey was developed based on an established rubric, it may have limitations in fully capturing the complexity of assessments. Some nuances of assessment design and implementation may not be easily conveyed through the survey's multiple-choice and Likert-scale questions, further limiting the depth of the findings.

During the analysis of the survey responses, it was observed that for the two evaluative judgment questions—namely, "What opportunities do you provide for students to judge their own performance on your assessment?" and "What forms of feedback do you provide to students for this assessment?"—the researcher-assigned scores were often consistent with the self-reported scores. This consistency arose primarily due to the researchers' reliance on available survey data and uploaded documents to verify the accuracy of the instructors' responses. Given the subjective nature of these questions and the limited detail in the survey responses, it was challenging to assign an objective score that may have differed from an instructor's assessment. Any significant deviation would require more robust evidence or more transparent documentation than available.

#### **Persistence of Traditional Exams**

This study highlights a persistent gap between authentic assessment design and implementation, particularly the continued dominance of traditional exams. Using a percentile-based cutoff to define "authentic" assessments effectively identified courses that met a high standard of assessment authenticity. Even among courses with high authenticity scores, most still incorporated traditional exams alongside more authentic methods. This finding highlights a disconnect—while many assessments were designed with authentic elements, traditional exams continue to dominate, suggesting that a total shift away from conventional assessment methods has not yet been realized. This illustrates the ongoing challenge of integrating authentic assessments as the primary evaluative tools in higher education.

## **Conclusions and Implications**

This study examined the prevalence of authentic assessments within an NR department at an R1 and PWI. By leveraging a survey developed from the Hobbins et al. (2021) rubric, this research provides insights into whether authentic assessment practices are being adopted and where traditional assessments still dominate. Approximately 42.3% of the assessments analyzed met high standards of authenticity. However, many of these courses still relied on traditional exams, illustrating a gap between authentic assessment design and implementation in place of traditional methods. This research emphasizes the need for a deeper integration of authentic assessments in higher education, as such assessments foster more inclusive and equitable learning environments and better prepare students for real-world challenges.

The survey tool uncovers the current assessment practices and offers a foundation for improving future assessments across institutions. By consistently measuring assessment authenticity, educational institutions can make informed decisions, enhance curriculum development, and better support instructors in transitioning to more meaningful, student-centered assessment methods.

## **Implications for Theory, Practice, and Policy**

The findings from this study suggest that while there is a growing emphasis on authentic assessments in higher education, the persistence of traditional assessment methods indicates a potential disconnect between theory and practice. From a theoretical perspective, this highlights the need to better understand the factors supporting or hindering the adoption of authentic assessments. From a practical perspective, educational institutions may need to provide more support and resources to instructors to help them transition from traditional exams to more authentic assessment methods. For instance, resources such as the Authentic Assessment Toolbox (Mueller, n.d.) or the University of Illinois Center for the Advancement of Teaching Excellence (Messier, 2022) provide practical guidance and examples of authentic assessments across disciplines.

Institutional and department support in the form of tenure, promotion, and retention guidelines would also be beneficial. Policymakers should consider developing guidelines or incentives to encourage the broader use of authentic assessments, potentially leading to more meaningful and applicable student learning experiences.

The implications of these findings are also significant for RM students. Given that traditional assessments have been shown to disproportionately disadvantage RM students (Supovitz & Brennan, 1997), the continued reliance on these methods may perpetuate existing inequities in educational spaces. Authentic assessments, which often allow for more culturally relevant and inclusive tasks (Penfield & Lee, 2010), offer a pathway to mitigate these disparities by providing RM students with more opportunities to succeed and demonstrate their competencies in ways that resonate with their lived experiences. By offering this flexibility in how students engage with course material, they can draw on their various cultural capitals. This aligns with frameworks like Community Cultural Wealth (Yosso, 2005), highlighting the strengths RM students bring to academic spaces. By validating and integrating these forms of cultural

capital into assessment, authentic assessments not only foster academic success but also promote students' confidence and agency. When RM students are empowered to use their unique perspectives and skills, they counteract the deficit model that often frames their academic performance in traditional assessments (Yosso, 2005). By adopting these practices more widely, institutions can help close the achievement gaps that persist for RM students and contribute to more equitable educational outcomes across diverse student populations.

## **Future Research**

Building on the findings of this study, future research will involve student interviews and the application of a novel framework in STEM courses, aiming to explore the intersection of authentic assessment, racial equity, and student learning experiences. This work aims to deepen our understanding of how authentic assessments can help reduce academic disparities between racially marginalized (RM) students and their peers (Supovitz & Brennan, 1997). Incorporating student voices will also allow for triangulation of the current study's findings on assessment prevalence, offering a more comprehensive understanding of assessment dynamics from both instructor and student perspectives.

Other avenues could involve a more in-depth investigation into the barriers preventing the widespread adoption of authentic assessments despite their recognized value. Qualitative studies could provide insights into faculty challenges and the support they need to implement these practices effectively. Research could explore students' perspectives on authentic assessments, examining their experiences, perceptions, and concerns. Given that assessments reflect what instructors value (Kenyon, 2023), future research could explore understanding how students engage with authentic assessments and their impact on student motivation and learning to provide insights for refining assessment practices which help students build connections between what they are learning and what they will need to know in the future. Further research could also examine the impact of authentic assessments on student learning outcomes, comparing them to traditional exams to evaluate their effectiveness in promoting deeper learning and critical thinking skills. Finally, longitudinal studies could track the adoption of authentic assessments over time, assessing whether educational policies or interventions influence their uptake across different institutions.

## References

- Anderson, L. W., Krathwohl, D. R., & Bloom, B. S. (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's Taxonomy of educational objectives. Longman.
- Bartlett, J. E., Kotrlik, J. W., & Higgins, C. C. (2001). Organizational Research: Determining appropriate sample size in survey research. *Information Technology, Learning, and Performance Journal*, 19(1), 43–50.
- Biggs, J., & Tang, C. (2011). Teaching for Quality Learning at University (4th ed.). Open University Press.
- Bloxham, S., & Boyd, P. (2007). *Developing effective assessment in higher education: a practical guide*. McGraw-Hill Education (UK).

Cochran, W. G. (1977). Sampling techniques (3rd ed.). John Wiley & Sons.

- Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). L. Erlbaum Associates.
- Darling-Hammond, L., (1994). Performance-based assessment and educational equity, *Harvard Educational Review*, 64(1): 5–31. https://doi.org/10.17763/haer.64.1.j57n353226536276
- Durning, S. J., Dong, T., Ratcliffe, T., Schuwirth, L., Artino, A. R., Boulet, J. R., & Eva, K. (2016). Comparing open-book and closed-book examinations: a systematic review. *Academic Medicine*, 91(4), 583-599. https://doi.org/10.1097/ACM.000000000000977
- Fowler, F. J. (2013). Survey research methods (5th ed.). Sage.
- García, G. E., & Pearson, P. D. (1994). Chapter 8: Assessment and Diversity. Review of Research in Education, 20(1), 337-391. <u>https://doi.org/10.3102/0091732X020001337</u>
- Gay, G. (2000). Culturally responsive teaching: theory, research, and practice. Teachers College Press.
- Haynes, N., Jacobson, S. K., & Wald, D. M. (2015). A life-cycle analysis of minority underrepresentation in natural resource fields. *Wildlife Society Bulletin*, 39(2), 228–238. https://doi.org/10.1002/wsb.525
- Hobbins, J., Kerrigan, B., Farjam, N., Fisher, A., Houston, E., & Ritchie, K. (2021). Does a classroom-based curriculum offer authentic assessments? A strategy to uncover their prevalence and incorporate opportunities for authenticity. *Assessment & Evaluation in Higher Education*, 47(8), 1259–1273. https://doi.org/10.1080/02602938.2021.2009439
- Jones, E., Priestley, M., Brewster, L., Wilbraham, S. J., Hughes, G., & Spanner, L. (2020). Student wellbeing and assessment in higher education: the balancing act. Assessment & Evaluation in Higher Education, 46(3), 438– 450. https://doi.org/10.1080/02602938.2020.1782344
- Kenyon, C. (2023). Assessing What We Value: Interactions between Student Perceptions of Assessments in the Calculus Classroom and Their Future-Oriented Motivation (Doctoral dissertation). Retrieved from https://open.clemson.edu/all\_dissertations/3275
- Keppell, M., & Carless, D. (2006). Learning-oriented assessment: A technology-based case study. Assessment in Education: Principles, Policy, and Practice, 13(2), 179–191. https://doi.org/10.1080/09695940600703944
- Ladson-Billings, G. (1999). Preparing teachers for diverse student populations: a critical race theory perspective. *Review of Research in Education*, 24: 211 247. https://doi.org/10.3102/0091732X024001211
- Lynam, S., & Cachia, M. (2017). Students' perceptions of the role of assessments at higher education. *Assessment & Evaluation in Higher Education*, 43(2), 223–234. https://doi.org/10.1080/02602938.2017.1329928
- Mcmillan, J. & Foley, J. (2011). Reporting and Discussing Effect Size: Still the Road Less Traveled? *Practical Assessment, Research and Evaluation. 16.*
- Messier, N. (2022). "Authentic Assessments". Center for the Advancement of Teaching Excellence at the University of Illinois Chicago. Retrieved December 10, 2024, from https://teaching.uic.edu/cate-teaching-guides/assessment-grading-practices/authentic-assessments/

- Mueller, J. (n.d.). Authentic task examples. Authentic Assessment Toolbox. Retrieved December 10, 2024, from https://jonfmueller.com/toolbox/examples/authentictaskexamples.htm
- National Science Foundation. (n.d.). Broadening participation in science, technology, engineering & mathematics fact sheet. U.S. National Science Foundation. Retrieved from https://www.nsf.gov/od/broadeningparticipation/bp\_portfolio\_dynamic.jsp
- Palmer, S. (2004). Authenticity in Assessment: reflecting undergraduate study and professional practice. *European Journal of Engineering Education*, 29(2), 193–202.
- Penfield, R. D., & Lee, O. (2010). Test-based accountability: Potential benefits and pitfalls of science assessment with student diversity. *Journal of Research in Science Teaching*, 47(1), 6–24. https://doi.org/10.1002/TEA.20307
- Polka, E. (2019, July 13). The Green Ceiling. Public Health Post. Retrieved from https://www.publichealthpost.org/databyte/the-green-ceiling/
- Rodriguez, S. L. & M. N. Peterson. 2020. Human dimensions of wildlife management. In Silvy (Eds.), *The Wildlife Techniques Manual: Volume 2 management (pp.39-58)*. Johns Hopkins University Press. Baltimore, Maryland, USA.
- Saye, J., & Social Studies Inquiry Research Collaborative (SSIRC). (2013). Authentic pedagogy: Its presence in social studies classrooms and relationship to student performance on state-mandated tests. *Theory & Research in Social Education*, 41(1), 89–132. https://doi.org/10.1080/00933104.2013.756785
- Schultz, M., Young, K., Gunning, T., & Harvey, M. L. (2022). Defining and measuring authentic assessment: a case study in the context of tertiary science. *Assessment & Evaluation in Higher Education*, 47(1), 77–94. https://doi.org/10.1080/02602938.2021.1887811
- Sharik, T.L. & Frisk, S.L. (2011). Student perspectives on enrolling in undergraduate forestry degree programs in the United States. *Journal of Natural Resources & Life Sciences Education*, 40(1), 160–166. https://doi.org/10.4195/jnrlse.2010.0018u.
- Sharik, T.L., Lilieholm, R.J., Lindquist, W., & Richardson, W.W. (2015). Undergraduate enrollment in natural resource programs in the United States: Trends, drivers, and implications for the future of natural resource professions. *Journal of Forestry*, 113(6),538–551. https://doi.org/10.5849/jof.14-146.
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4–14. https://doi.org/10.3102/0013189x029007004
- Simmons, W., & Resnick, L. (1993). Assessment as the catalyst of school reform. *Educational Leadership*, *50*(5), 11–15. Association for Supervision and Curriculum Development.

- Singh, P., R. Thambusamy, & M. Ramly. (2014). "Fit or Unfit? Perspectives of Employers and University Instructors of Graduates' Generic Skills." *Social and Behavioral Sciences 123*: pp. 315–324. doi:10.1016/j.sbspro.2014.01.1429.
- Solano-Flores, G., & Nelson-Barber, S. (2001). On the cultural validity of science assessments. *Journal of Research in Science Teaching*, 38(5), 553–573. https://doi.org/10.1002/TEA.1018
- Supovitz, J. A., & Brennan, R. T. (1997). Mirror, mirror on the wall, which is the fairest test of all? An examination of the equitability of portfolio assessment relative to standardized tests. *Harvard Educational Review*, 67(3), 472–502. https://doi.org/10.17763/HAER.67.3.66V546274672G154
- Villarroel, V., Bloxham, S., Bruna, D., Bruna, C., & Herrera-Seda, C. (2017). Authentic assessment: Creating a blueprint for course design. Assessment & Evaluation in Higher Education, 43(5), 840–854. https://doi.org/10.1080/02602938.2017.1412396
- VanVoorhis, C. R., & Morgan, B. L. (2007). Understanding power and rules of thumb for determining sample sizes. *Tutorials in Quantitative Methods for Psychology*, *3*(2), 43–50. https://doi.org/10.20982/tqmp.03.2.p043
- von der Embse, N., Jester, D., Roy, D., & Post, J. (2018). Test anxiety effects, predictors, and correlates: A 30-year meta-analytic review. *Journal of Affective Disorders*, 227, 483–493. https://doi.org/10.1016/j.jad.2017.11.048
- Whitcomb, K. M., & Singh, C. (2021). Underrepresented minority students receive lower grades and have higher rates of attrition across STEM disciplines: A sign of inequity? *International Journal of Science Education*, 43(7), 1054–1089. https://doi.org/10.1080/09500693.2021.1900623
- Willis, G. B., & Artino, A. R., Jr (2013). What Do Our Respondents Think We're Asking? Using Cognitive Interviewing to Improve Medical Education Surveys. *Journal of graduate medical education*, 5(3), 353-356. https://doi.org/10.4300/JGME-D-13-00154.1
- Yosso, Tara J. (2005). "Whose Culture Has Capital? A Critical Race Theory Discussion of Community Cultural Wealth." *Race, Ethnicity and Education*, 8(1), 69–91. https://doi.org/10.1080/1361332052000341006

## **Appendix- Authentic Assessment Survey**

Welcome to the assessment survey!

We want to assure you that your responses will be kept confidential. This data will be used solely for research purposes and will not be shared in a way that could personally identify you.

If you have received a recruitment email mentioning multiple courses, we kindly ask that you complete the survey separately for each course. Don't worry, there will be a question to identify the course each time. Your feedback for

each course is greatly appreciated!

There are a few things we'd like you to keep in mind while taking this survey:

1. We ask that you consider the *single* assessment type that carries the most weight in determining a student's overall grade in your course. This may be a midterm, final exam, or project.

2. Please evaluate and provide responses specifically related to the chosen assessment *only*. Avoid generalizing your responses to include other assessments or aspects of the class.

3. Please respond honestly, there are no right or wrong answers here. It may be helpful to have your assessment and syllabus file open while taking this survey.

Your feedback is valuable and we are very thankful for your time in taking this survey!

Q1 Please enter your first and last name.

(**Note:** This information will only be used to identify who has not yet responded. All survey results will be deidentified and analyzed in an aggregated manner to ensure confidentiality.)

Q2 Please select the course code for which you will be taking this survey:

Q3 You selected 'other course not listed here'. Please enter the course code and name for which you are taking this survey:

Approximately how many students are typically enrolled in this course?

Q4 Per the instructions, now is the time to identify the *single* assessment type that carries the most weight in determining a student's overall grade in your course.

This assessment is considered this course's\_\_\_\_\_. o Midterm Exam (1) o Final Exam (2)

o Project (4)

o Other (Please specify) (3)

Q5 \_\_\_\_\_% of a student's final grade comes from this assessment type.

(Please enter	a number between	1 1 and 100.)
---------------	------------------	---------------

Q6 What are the dominant forms of participation that are required for this assessment type? (Please select up to two)

- $\Box \qquad \text{Multiple Choice (1)}$
- $\Box$  Matching (2)
- □ Fill-in-the-blank/Short answer (3)
- $\Box$  Presentation (4)
- $\Box$  Essay (5)
- □ Other (please describe) (6) \_\_\_\_\_

Q7 To what extent does this assessment engage students with problems that are relevant to their future professional life beyond the classroom?

o Minimally; a gap remains between classroom and real-world context (1)

o Moderately; it begins to bridge the gap between classroom and real-world context (2)

o Highly; it bridges the gap between classroom and real-world context (3)

Q8 What cognitive skills does this assessment require students to utilize?

Memory Skills: students identify and/or provide info or facts; recognition or understanding (associated verbs: identify, describe, summarize, define, recount, explain) (1)

Application and Analytical Skills: requires the unpacking and organization of information in multiple sources, types or relationships; requires a response to a hypothetical situation (associated verbs: compare /contrast, relate, interpret, integrate) (2)

Transfer skills: requires students to design or put elements together to form a coherent whole and/or make an original product (associated verbs: judge, decide, critique, suggest, design, create, innovate). (3)

Q9 What opportunities do you provide for students to judge their *own* performance on your assessment? o Students are provided with criteria published to Canvas AND latent criteria prior to taking the assessment (e.g., a follow-up to initial instructions via discussion or email, graded exemplars, or assessment criteria that are co-created with students). (2) o Latent criteria are not only provided and published to Canvas prior to taking the assessment but students are engaged with such criteria pre- &/or post-completion. (e.g., critiques/critical reflections, self and/or peer assessment using those criteria) (3)

o Students are provided with explicit criteria published on Canvas prior to taking the assessment. (e.g., instructions, rubric, grading scheme) (1)

o No opportunities are provided for students to judge their own performance. (0)

Q10 What forms of feedback do you provide to students for this assessment?

o Students are only given a grade (0)

o Content-specific feedback and/or a grade is provided. (1)

o Content-specific feedback AND generic skill (communication, leadership, integrity, creativity, attention to detail, etc.) feedback are provided. (2)

o Content-specific feedback AND generic skill feedback are provided in response to multiple iterations of a similar assignment (whether graded or ungraded) (i.e., drafts, practice tasks.) (3)

Q11 Please upload a copy of the syllabus for the course.

(Note: PDF, DOC, or DOCX file type preferred)

Q12 Please upload a copy of the assessment for which you've taken this survey (PDF, DOC, or DOCX file types preferred).

Should you need to upload multiple files, you can create a single PDF or upload a compressed .zip file with all attachments.

(**Note:** We prioritize the integrity of your assessment and preventing any cheating or unauthorized access. Your file will be securely stored in a password-protected Box account with limited access. Your test materials will remain confidential and inaccessible to unauthorized parties.)

Q13 Please indicate your current position at the university:

o Professor (1)

o Associate Professor (2)

- o Assistant Professor (3)
- o Senior Lecturer (7)
- o Principal Lecturer (8)
- o Lecturer (4)
- o Adjunct faculty (5)
- o Other (please specify) (6) \_\_\_\_

Q14 What is your age group?

o 25-34	(1)	
o 35-44	(2)	
o 45-54	(3)	
o 55-64	(4)	
o 65 or	older (5)	
o Prefer not to say (6)		
Q15 WI	nat is your racial or ethnic background? (Please select all that apply.)	
	American Indian or Alaska Native (1)	
	Asian (2)	
	Black or African American (3)	
	Hispanic or Latino (4)	
	Native Hawaiian or Other Pacific Islander (5)	
	White (6)	
	Prefer not to say (8)	
	Other (please specify) (7)	

Q16 What is your gender identity? (Please select all that apply.)

- $\Box$  Male (1)
- $\Box$  Female (2)
- □ Non-binary (3)
- Genderqueer (4)
- $\Box$  Transgender (5)
- Gender non-conforming (6)
- $\Box$  Prefer not to say (7)
- □ I use a different term (please specify) (8) \_\_\_\_\_

Q17 Many of your recruitment emails request that you take this survey for **more than one course.** Would you like to submit another response now?

o Sure! Submit and start the survey over for a different course. (1)

o I will take it again later through the original link. (2)

o N/A- I only need to take it for one course. (3)

#### **Corresponding Author Contact Information:**

Author name: Aradaryn Marsh

Department: Engineering and Science Education

University, Country: Clemson University, United States of America

Email:aradars@clemson.edu

**Please Cite:** Marsh, A., Hagan, A. & Benson, L. (2025). Do STEM Programs Provide Authentic Assessments? Exploring Their Prevalence and Opportunities for Growth. *Journal of Research in Science, Mathematics and Technology Education*, 8(SI), 263-280. DOI: https://doi.org/10.31756/jrsmte.4111SI

**Copyright**: This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Conflict of Interest: The authors declare no conflict of interest.

**Publisher's Note**: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

**Data Availability Statement**: The data supporting the findings of this study are available from the corresponding author upon reasonable request.

**Ethics Statement:** This study was reviewed and approved by the Clemson University Institutional Review Board (IRB2023-051). Informed consent was obtained from all participants.

**Author Contributions**: A.M. led survey design, data analysis, and manuscript drafting. A.H. and L.B. provided guidance on study design and critical revisions. All authors read and approved the final manuscript.

Received: December 14, 2024 • Accepted: May 30, 2025