Journal of Research in Science, Mathematics and Technology Education
https://jrsmte.com/

Higher Education STEM Faculty Views on Collaborative Assessment and Group Testing

Allan Jay Dinglasan & Jennifer L. Weible

Central Michigan University, United States of America

Abstract: Transitioning to online learning led to an increased reliance on collaborative assessment and group testing in higher education STEM courses. While research has shown that such assessments enhance student engagement, critical thinking, and confidence, their adoption remains inconsistent. This study explores STEM faculty perceptions of collaborative assessment, focusing on its legitimacy, benefits, and implementation barriers. Through a mixed-methods approach, surveys and semi-structured interviews were conducted with STEM faculty across North America. Thematic analysis revealed three key findings: (1) faculty recognized the pedagogical value of collaborative testing in fostering student confidence and engagement, (2) concerns about implementation, including academic integrity and student accountability, were seen as manageable with proper safeguards, and (3) resistance to collaborative assessment was often rooted in entrenched pedagogical beliefs and institutional inertia. While faculty acknowledged its potential benefits, some remained hesitant due to concerns about grading fairness, student workload distribution, and long-term knowledge retention. The study highlights the need for a paradigm shift in STEM education, emphasizing collaborative pedagogy as a viable alternative to traditional assessments. Addressing faculty concerns through professional development, institutional support, and data-driven discussions on efficacy may help normalize collaborative assessment practices. These findings contribute to ongoing discussions about assessment innovation in higher education and provide insights into how faculty attitudes shape instructional choices.

Keywords: Collaborative assessment, group testing, STEM education, faculty perceptions, higher education, pedagogy reform
DOI: https://doi.org/10.31756/jrsmte.4113SI

Introduction

Following the pivot to online learning in 2020, faculty at educational institutions worldwide transitioned to less traditional means of assessment and evaluation. This included faculty teaching courses in science, technology, engineering, and mathematics (STEM) fields, more traditionally online at the university level, (Wilson, 2020; Dietrich et al., 2020, Huang, 2020) These more non-traditional evaluation methods included collaborative assessment, also known as group testing, For this paper, we define collaborative assessment as “an approach to assessment whereby students work together in pairs or groups, allowing them to benefit from their peers’ knowledge and teacher feedback in the same activity” (Thaanyane & Jita, 2024, pg 495). This broad definition holds even across theoretical courses and large undergraduate sections. As faculty and students returned to campuses, the use of collaborative assessment lessened in both willingness to implement and consistent use with the easing of pandemic restrictions. Although previous research found benefits to collaborative assessment use, students working collaboratively in groups on tests or projects, as it allows for a robust discussion of answers and for other forms of learning to occur beyond what a single test can accrue (Newton, 2019), its use in higher education has yet to be normalized and lacks research on the level of faculty “buy-in” that exists.

Although much research on collaborative education and group testing has examined student outcomes, academic integrity, and implementation particularly in K12 settings, higher education faculty implementing collaborative assessment has not been deeply studied. In addition, with the lack of research on implementation across higher education, this study seeks to examine instructors’ lived experiences using collaborative testing as an alternative assessment and evaluation tool and practice in higher education STEM courses. The goal of this study is therefore to explore observed benefits of collaborative assessment and develop best practices for implementation based on faculty's experiences to support others seeking to use this pedagogical practice.

Literature Review

Research has shown that what students and instructors believe to be true about collaborative learning are rooted in preconceived notions and experiences in working as part of various group settings. (Hillyard, Gillespie & Littig, 2010) The perceptions of students are connected to instructors' views of group work and their perceptions of efficacy for collaborative pedagogy. With collaborative laboratory work often featured predominately in STEM courses and work, faculty perceptions of collaborative work are crucial to understand as it impacts students' espoused beliefs about how chemistry and science works (Smith & Sepulveda, 2018). Hamnett et al., (2018) found that group work was essential for post graduate science students in terms of confidence building and familiarization of each other's experiences.

Faculty management of group settings is pivotal in fostering positive attitudes towards group work; research has found that skills developed by students in group work have proven to be useful. Students who felt supported in their work viewed group work more positively and thought it benefited them greatly compared to those who were in an unsupported work environment (Natoli et al., 2014). Additionally, with those participating in unsupported group activities felt they benefited from enhanced organizational skills. When researchers compared business and non-business students' views about collaborative work that included work distribution and fairness in grading, business students expressed more concerns about the equitable distribution of work and sought more freedom over group dynamics compared to non-business students who wanted more structure (Grzimek et al., 2020). They concluded that those who equated group work with attaining life or workplace skills held more positive attitudes towards collaboration (Grzimek et al.).

Research has also found issues with collaborative work. Students have indicated that it was frustrating to work in teams. Some members may take the work less seriously. Higher achieving students indicate concerns on being grouped with their less capable counterparts (Grzimek et al., 2014) Also, social loafing, where individuals work harder when by themselves rather than when placed in groups to work collaboratively, is ubiquitous in workplaces, classrooms and laboratories (Karau & Williams, 1993) as well as within online and face-to-face learning environments (Piezon, 2011). Concerns about equitable collaboration in a dysfunctional environment where one individual tends to dominate are present (Theobald et al., 2017), but research has indicated that positive attitudes towards group work can be nurtured when instructors support "emerging leaders" who work towards minimizing free riders in their courses (Rudawaska, 2017). Additionally, creating a better environment for facilitating teamwork can counter negative attitudes held by students. Taking steps beyond instructor effort, it has also been posited that campus initiatives and interdepartmental coordination at the institutional level are crucial to getting students to appreciate the benefits of collaborative learning (Hillyard et al., 2010).

While benefits of group work are evident, research indicates that from the students' perspective, collaborative learning tells a twofold story of challenge and support (Delucchi, 2006). Students viewed group work positively but do not prefer it over independent work (Marks & O'Connor, 2013). They value lectures but collaborative activities are seen as better exam preparation tools (Machemer & Crawford, 2007). While researchers found students to be more engaged in their learning when collaborating (Ellis, 2016), they also found them to be experiencing negative emotions and a reluctance to share (Medaille & Usinger, 2018). Neber et al. (2001) found that quiet students found group work challenging also indicated that collaborative learning situations could have more benefit for gifted learners.

Concerns surrounding the efficacy of collaborative learning it difficult to address or implement on a practical level. Johnston and Miles (2005) examined peer assessments as evaluation tools in group work and found that students gave their own contributions higher ratings than the contributions of their colleagues. While some students identified their free-riding teammates and the lack of accountability in collaborative learning settings, they also challenged their instructors whom they saw as unmotivated to properly manage group dynamics (Marks & O'Connor, 2013).

Despite mixed reviews, research has indicated how students benefit from collaborative learning. When team-based learning was used in medical school, students gained critical thinking skills and advantages from independent learning and sharing what they learned, (Zgheib et al., 2016). Even when group activities do not end up as intended, this "failure experience" provided students learning opportunities not found in most classrooms (Lam, 2019). Many of tasks such as generating problem questions and solutions are arguably what are expected of competent STEM students. However, despite benefits, the practicality of assessing individual performance in a group setting has historically created resistance from students, parents (King & Behnke, 2005), and many educators.

Collaborative Assessment and "Pedagogical Dissatisfaction"

The sense of "pedagogical dissatisfaction" prevalent among faculty forced to teach online created enormous pressure to change the way science courses were taught in remote learning settings (Sansom, 2020). Researchers found professors were affected by personal factors and from structure and culture of their institutions, they adapted quickly to using platforms provided, minimizing any changes to their courses. However, they struggled with assessment, holding on to diverging pedagogical beliefs and the possible effects on students (Rupnow, LaDue, James & Bergan-Roller, 2020). Sansom (2020) alluded to a "deep pedagogical dissatisfaction" that resulted when pandemic pressures led to instructors making pedagogical decisions that may not coincide with their personal beliefs about teaching and learning. Even though historically educators have been resilient amidst rapid change, having learned to adapt practices used by previous instructors to add to their own, fueled a drive for continuous improvement (Kusch, 2016). Instructors quickly recalibrated curriculum delivery while trying to maintain continuity and ensure high quality content. Despite the apparent readiness of some, most faculty had a steep learning curve with few resources, resulting in new strategies employed for lectures, tutorials, project groups, laboratory work and assessments (Dietrich et al., 2020).Components that utilized technology were not difficult to integrate into the course, i.e., online chemistry homework, but pedagogy for performing assessments often was significantly different from past practices (Clark et al., 2020). Research indicates the importance of authentic assessment activities and assessments that promote self-regulation among student learners. Additionally, it emphasizes academic integrity concerns and challenges within the faculty ranks on implementing online discussions and collaboration, not just in introductory chemistry courses (Raje & Stitzel, 2020) but in other STEM courses as well. In response, many faculty experimented with alternative assessment tools and practices during the shift to online learning, particularly collaborative testing and assessment.

Collaborative Testing Defined

Meseke, Nafziger, and Meseke (2010) defined collaborative learning as students in pairs or groups completing assigned tasks in the classroom. Collaborative assessment is a group test but also involves using the use of the exam beyond evaluation purposes for learning (Muir & Tracey, 1999).

One form of collaborative assessment is designed in multiple parts The "pyramid" form involves implementing it in two stages where individual students complete their work before their learning group, with their exam grades being a hybrid of the scores for their individual and group components (Zipp, 2007). Siegel et al. (2015) defines group testing as a strategy where student learners work in groups while completing an assessment or evaluation piece. Two-stage testing however involves students taking the test on their own first before working on the same assessment again in small groups that allow them to discuss answers in a setting beyond what individual pen and paper tests can provide (Newton, 2019).

Collaborative Testing's Benefits to Students

Collaborative testing has shown increased benefits and improved outcomes to students in the biomedical sciences (Khong & Tanner, 2020), physics (Jang, 2017), physiology (Giuliodori et al, 2008), biology and life sciences (Leight, 2012; Cooke et al., 2019), earth and ocean sciences (Gilley, 2014), anatomy (Green et al., 2015), nursing (Rivaz et al., 2015; Hickey, 2006), introductory geology (Knierim et al., 2015), astronomy (Miller, 2019) and mathematics (Kinnear, 2020). Research found that collaborative testing enhanced student performance (Giuliodori et al, 2008; Knierim et al., 2015) with students scoring higher on the group portion of the exam (Miller, 2019). This system consistently produced significantly higher test scores (Bloom, 2009) on both multiple choice and long answer question formats (Newton et al., 2019), and compared to traditional testing methods, showed increases in test performance and afforded students with critical thinking skills to aid in problem solving (Rivaz et al., 2015). However, these gains were also dependent on saturation of concepts, testing complex ideas, and lessening retention as time passed (Miller, 2019).

Implementing successful collaborative assessments involves more than having the correct answers propagate among the students in the course (Jang, 2017) but also an acknowledgement of the need to foster a positive social environment to keep it sustainable (McConnell, 2002). Faculty surveyed by Hickey (2006) noted that most students were not as argumentative during the test review process as they had previously been, after having gone through collaborative testing. The study concluded that the collaborative testing process provided a mechanism for students to ask questions and listen to alternative perspectives. He found that discussions inherent in group testing assisted students in ensuring their questions are adequately answered which made arguments with faculty more unlikely than not (Hickey, 2006).

Collaborative testing affords benefits to students of all learning abilities, with both low and high performers showing improvement in learning outcomes (Khong & Tanner, 2020). In a study where all students in the course collaborated on an exam, even the strongest members of a team had greater rates of correct answers (Jang, 2017).While group scores generally tended to be higher than individual scores on collaborative assessments, the higher scores helped those students who did not do as well (Giuliodori et al., 2008; Gilley, 2014; Knierim et al., 2015). Green (2015) pointed to evidence that showed lower performing students benefiting from team grades even without the affordances of active learning, but this was posited as possibly being due to a lack of strategic guidance on the part of the instructor. Evidence showed how students retained course content better (Cortright et al., 2003; Bloom, 2009) when assessments are done collaboratively (Leight, 2012). Previous studies also pointed to how collaborative testing offers retention benefits across students in different performance categories, when looking at the period in between the pre-test and the post-test but not when taken over a shorter time (Cooke et al., 2019). However, a study by Meseke et al., (2010) pointed to increased short term performance compared to traditional testers but no evidence of learning and increased retention was found.

Scant Research on Faculty Perceptions

Research on faculty perceptions of collaborative testing is limited. Siegel (2015) surveyed six faculty; they thought students learned through their process, were supportive of collaborative testing and found it useful for students with varying skills (Siegel, 2015). Concerns about "unintended collaboration" (Lewis, 2020), as well as grade inflation (Bloom, 2009) were reduced when a weighted grading system was implemented. A second study (Hickey,2006) (n=4) found faculty believed that collaborative testing helped student understanding and critical thinking without compromising study habits and preparedness. There were mixed results concerning impact on text anxiety behavior, the ideal group size, or whether students studied more because of a collaborative testing format, but faculty thought that the amount of time provided was sufficient. (Hickey, 2006)

Strongly held views can evolve whenever innovative teaching practices are introduced. A case study explored how students shift from opposing to supporting group exams in an undergraduate calculus class (Dobie & MacArthur, 2022). Dobie and MacArthur's (2022) research raised questions about how faculty can gain student support for more novel teaching methods as well as challenged beliefs about math pedagogy. This research will explore whether STEM instructors can experience the same change in attitude as students.

Social Constructivism and Collaborative Assessment

The shift to emergency remote learning during the pandemic forced educators to rethink assessment and evaluation practices. Bowen and Phelps (1997) argued that STEM curriculum changes focused on content revision and not assessment, pointing to how chemistry students are learning new topics like molecular modeling or statistical analysis but are still tested with traditional pen-and-paper exams. Along with a call for more interactive methods, this conceptual model sought a re-examination of thinking on how students are evaluated, whereas pedagogy and content evolve, so does assessment. Collaborative assessment, where students work together on a test or quiz, fits this framework for assessment of learning and shows the interconnectedness of assessment, content, and instruction.

Collaborative pedagogy is rooted in social constructivist ideas, particularly within Johnson and Johnson's (2015) social interdependence theory where people's learning comes from interactions with others and the environment (Johnson & Johnson, 2015; Mishra, 2014). Social constructivism is evident in teaching methods like group work, peer reviews, and training with simulations in non-STEM fields (Giuliodori et al., 2008). Collaborative assessment brings social constructivism into science courses. When students use collaborative testing, they have the chance to share ideas and build on what they already know. As teachers try to create more engaging learning environments where students work together and create knowledge (American Association for the Advancement of Science, 2011; National Research Council, 2012), new ways to include collaboration in assessments can emerge. Although collaborative assessment is rooted in social constructivism, this study seeks to explore the teachers' perceptions and lived experiences when using collaborative assessment with students without examining the student learning potentially taking place.

Methods

Research Questions

Specifically, the study will explore the following questions:

1. How do higher education STEM faculty perceive the authenticity and legitimacy of collaborative assessment and group testing?
2. What benefits and drawbacks do STEM faculty associate with collaborative assessment?
3. What barriers do STEM faculty identify in implementing collaborative assessment?

Setting and Participants

Purposeful sampling was used for this study. Purposeful sampling assumes that the investigator wants to discover, understand and gain insight and therefore must select a sample from which the most can be learned. (Merriam & Tisdell, 2017). A screening survey was used to recruit potential interviewees based on their previous experience implementing collaborative assessment and group testing in their courses. Of the seven respondents who replied to targeted email to faculty teaching STEM classes at higher learning institutions to schedule a virtual interview, five of the professors were able to be scheduled. The interviews took place between April and May 2023. Table 1 shows the background data of the participants and the interviews.

Table 1
Background data of participants

	Katie	Gerald	Louise	Reanne	Adley
Gender	Female	Male	Female	Female	Male
Subjects taught	Geosciences	Biology	Chemistry	Statistics	Physiology
Interview length	45 min	51 min	31 min	30 min	36 min
Institution Location	Pacific Northwest USA	Midwest USA	Southwest USA	South USA	Central Canada

These faculty teach STEM courses using a collaborative assessment protocol in both face to face and online environments. Three participants were female and two were male from biology, chemistry, geosciences, physiology and statistics. They teach at 4-year colleges and universities in the USA and Canada.

Data Collection

Interviews followed a semi-structured protocol where a virtual conferencing app was used to record responses in keeping with best practices for observation of tone and body language (Opdenakker, 2006), having observed how the use of a telephone versus face-to-face interviews in previous studies made no difference in the outcome (Sturges & Hanrahan, 2004). The interviews were transcribed using the conferencing app closed captions and then verified by one researcher.

Limitations

This study focused on the lived experiences of five STEM faculty members who utilized some form of collaborative assessment in their courses. This small sample size limits the ability to generalize to a larger or broader population.

Analysis

Thematic analysis was utilized to draw out emerging patterns using a process of eliciting codes from interview transcripts, then analyzing and grouping into themes (Creswell, 2014). Merriam and Tisdell (2017) describe the constant comparative method where codes are continually compared with one another in a process that triggers more elaborate codes from those found initially. These methods were used to deeply analyze the transcripts of the interviews.

Coding Process

The time frame allocated for the study allowed for a more formalized member check process. The formal member check entailed providing participants with an explanation of the analysis and ascertaining whether they perceived that the analysis was accurate based on the information shared during the interviews. Since none of the participants disputed any parts of the analysis presented in the member check, the interview transcripts were able to be more reliably subjected to coding through thematic analysis. Had there been any disputed parts of the interview transcripts, the researcher and participant(s) would have followed-up with informal correspondence to make the necessary clarifications.

Initial coding happened parallel with transcribing the interviews to find emerging themes and see if the questions elicited sufficiently rich data. The constant comparative method was used to compare data with codes, and codes with codes, so that it would result in sorting of the initial codes into more elaborate codes (Merriam & Tisdell, 2017).

Because of the limited number of interviews, a manual analysis (without the use of qualitative analysis software) was conducted using the following procedures: (a) read each transcript fully and identify key words and phrases that stand out, (b) highlight text using the different colors for each main idea that emerged. These were assigned in the code book of tentative ideas for topics and noticeable themes and patterns, (c) use recurring ideas within and across each interview to establish a list of emerging themes and (d) examine relationships among ideas and themes to create an overarching theme or theoretical concept. Codes developed from this analysis (and emerging codes that connected to literature) are listed in Table 2 below.

The researcher recorded the interview and transcribed it, followed by reading three times as an initial analysis for each interview. The first read allowed the researcher to get an overview of participants' sentiments as well as greater familiarization of the data. A second read of the transcripts provided an opportunity for the researcher to focus on broad themes aligned with the three research questions while a final read delineated additional topics within the research question themes. A second researcher separately read the transcript and examined codes before discussing the codes to ensure fidelity of code application. Discrepancies were discussed until consensus was reached. Codes were discussed and defined before coding the remaining interviews. These discussions served as checkpoints to ensure that bias was being filtered out as the coding was taking place. The first researcher coded the entire dataset with the second checking approximately 20% for fidelity.

Table 2
Initial codes and subcodes developed for collaborative assessment and group testing

Efficacy perceptions	Potential barriers to implementing	Academic integrity concerns
More authentic assessment piece	Lack of knowledge for proper implementation	Frustration
Ease of implementation	Equity issues	Acceptance of inevitability
Preparation for real life teamwork	Time constraints	Encouraging of cheating and "unintended collaboration"
Heightened student conceptual understanding	Teaching philosophy for STEM unaligned with CA & GT	Does not encourage individual accountability
Similar outcomes to traditional assessment	Lack of support from colleagues and administration	
	Lack of technological tools/know-how for proper implementation	

While going through and highlighting the pertinent data on the transcripts, the researcher, mindful of the more fine-tuned codes that could come out of the transcripts, made necessary amendments to the code book. The codes previously generated were simplified to a minimum, considering the preliminary information gathered. The data were copied into Microsoft Excel for focused coding where themes that summarized groups of codes were created. The following "Results and Discussion" section is organized according to these emerging themes.

Results and Discussion

From the thematic analysis, three themes emerged. In addition, several subthemes were found that provide a more nuanced understanding of perceptions of collaborative assessment.

1. Efficacy of collaborative assessment and group testing
2. Concerns and barriers to implementation can be easily mitigated
3. Need for a paradigm shift towards collaborative pedagogy.

Efficacy of Collaborative Assessment and Group Testing

The theme of efficacy was found to have several subthemes: improved self-esteem and self-confidence and increased motivation. Faculty felt that students participating in collaborative assessment were more motivated to participate in class as well as self-reporting more confidence as a learner.

Improved Self-esteem and Self-confidence

Faculty shared a common appreciation for the heightened levels of self-confidence and self-esteem as students became more familiar with collaborative assessment and group testing practices. The value is further enhanced when faculty see positive effects on students' perceived abilities to do well in STEM courses. Katie (geosciences) felt group assessments benefited her students; they were more at ease taking chances and became more confident with their responses, which is helpful even if students do not get the correct answer.

> "It doesn't matter what discipline you're in, especially in the STEM fields, the success of a student is based upon their confidence, and this is giving (them) opportunities to become confident with the collaboration..., in coming up with an idealized answer and, it doesn't have to be the right answer." (Katie)

Adley (physiology) pointed to how group testing in his courses allowed for "stand-out" students to emerge, those who during the collaborative assessment process developed the confidence to advocate for their own chosen response, even when it appears not to be the consensus answer. Gerald (biology) concurred about the positive effects that group testing had on his own students.

Increased Motivation

Faculty noticed that increased motivation for students to succeed was observable in the STEM courses where collaborative assessments were implemented. Students were more motivated in the group assessments when the instructions were being more explicitly explained to them (Katie) and were more likely to share what they know to the group. Louise noticed that her general and organic chemistry students were *"happier"* with how the group exams turned out, often wanting a repeat of the group exams soon after having done it.

Positive psychological effects aside, instructors' confidence in the efficacy of collaborative assessments largely appears in how this practice encourages students to be confident on their own abilities to do well in STEM. Instructors perceived such practices as allowing for a more enriching opportunity for students to teach and learn from each other (Katie) while remaining confident that *"on average,"* it benefits even the most knowledgeable of them (Adley). While having a safe environment for students during group testing remains important for these instructors, they believe that the essential value of the collaboration lies in having students feel like they are genuine scientists in training. Adley (physiology) posited that collaborative assessment often afforded students with an initial point to nurture the "spark of an idea" when "sometimes...they just need a little prompting…and ...when other students and the course can give that to them...they can get going." In Katie's (geosciences) experience, sometimes this boost of confidence can truly be a game changer:

> "When I was in high school, I remember...having an instructor make me feel like I didn't know anything and I felt like I was dumb and, ... (I want to) make sure that no one in my classroom ever feels like they

can't get it…. I had one student say to me…. 'I have never felt smart in a science classroom before until I took this class'." (Katie)

These faculty perceptions coincided with previous findings regarding the efficacies of collaborative assessment and group testing in eliciting greater confidence among students. While students in one study saw this negatively and were reluctant to share, (Medaille & Usinger, 2018), it did confirm heightened positive emotions. These perceptions aligned with previous research touting group work as providing science students with opportunities to build their confidence (Hamnett et al., 2018) and in a more engaging learning environment. (Ellis, 2016) Faculty perceptions of an increase in the confidence, self-esteem and motivation of their students are also in line with previous accounts of the benefits afforded students when exposed to "failure experiences", especially when the activities do not lead to their intended outcomes (Zgheib et al., 2016; Lam, 2019).

A Valuable Pedagogical Tool for STEM

Faculty alluded to studies pointing to collaborative assessment being a more effective pedagogical tool. Gerald thought that collaborative assessment was the closest that higher education had to teaching effective collaboration, considering that "there is no way to develop those soft skills in a traditional lecture." Reanne (statistics) reiterated that it is beneficial since it teaches students to question their peers and ensure that they have time to do so. Katie posited that being able to collaborate effectively is a rare practical skill that can be taught, knowing the reality that students will likely never be working in isolation after graduation. She added that collaborative assessment practices allow students the opportunity to network and mentor each other while allowing instructors to do the same. Faculty also pointed to being able to teach students how to justify their answers to their colleagues (Louise), a skill worth teaching students (Gerald).

Reanne valued how implementing collaborative assessments gave her the ability to teach difficult concepts in the course while also benefiting those who are not necessarily strong math students. Louise saw the same in her chemistry courses, where group exams became more of an "assessment *for* learning" versus traditional "assessment *of* learning" events and alluded to the paradigm shift involved, going from a "cram and dump and forget about it" [scenario] into a powerful learning experience."

Barriers to Implementation Can Be Easily Mitigated

Implementation Barriers as Non-Insurmountable Challenges

Professors interviewed did not view collaborative assessment through rose-tinted glasses but instead were confident that whatever barriers they saw to proper implementation of these practices were not insurmountable. Most expressed disappointment that despite having been around for a long time, collaborative assessment practices are still not widely accepted, with its many different aspects remaining to be explored. Reanne commented that the small number of her colleagues who may have heard about collaborative assessment are still in the dark about properly implementing it. She added that while this may be a minor challenge, it is nothing that more research (to spread knowledge of its efficacies) and more materials (to expose faculty to the same) cannot alleviate. Gerald pointed that it is not just higher education faculty that would benefit from embracing collaborative assessment practices. The absence of any pushback from students when group testing was implemented in his department, he said, came because of faculty getting students to "buy-in" to the practice.

The muted optimism expressed by the faculty interviewed regarding the benefits of collaborative assessment outweighing its drawbacks, come from their own experiences. While most of the professors are "hopeful" that more of their colleagues would try it in the future, Adley is optimistic that a "data driven conversation" on the benefits of group testing could convince skeptical colleagues to give it another look based on what students are saying about it.

> "I think it's interesting even just looking at the (Discord) server itself as the exam is taking place and watching the kinds of conversations that are taking place. I think that alone would perhaps amaze some of my colleagues... (how)... you have this rare insight into what's going on in the heads of your students while they're running their exam'…." (Adley)

Academic Integrity Issues Mitigated

The professors corroborated with the limited research on faculty concerns about academic integrity in group testing environments (Lewis, 2020; Bloom, 2009) but suggested that it is not an issue if a serious effort is made to minimize it. Furthermore, they exhibited a keen awareness of academic integrity in the classroom and use multiple methods to mitigate it. An emphasis on the individual accountability already built-in to collaborative assessments, as well as a large percentage of students' grades being based on individual work was key for many of this mitigative approach.

Instructors faced these academic integrity issues with realistic expectations; it is not easily ameliorated. However, they felt that effort helped moderate instances of perceived cheating. In Katie's experience, trusting students is something she learned to value in her efforts to maintain academic integrity in her courses. While acknowledging that cheating may not be as rampant in her discipline as it is in other areas, she spoke of the emotional toll that this can sometimes take. Reanne fears that artificial intelligence (AI) has made the situation worse. She focuses on keeping her students engaged saying that academic integrity issues arise when certain group members become disengaged.

Need for a Paradigm Shift

The professors cited psychological and ideological differences, along with a lack of investment in better teaching methods and a passive attitude towards evidence of successful implementation as driving a "pedagogical inertia" phenomenon that has made it harder to be normalized and legitimized. This is in line with earlier findings that cited faculty as continuing to hold diverging beliefs about teaching and students. (Rupnow, LaDue, James & Bergan-Roller, 2020).

Katie acknowledged that similar to students having different learning preferences, faculty also have varying teaching styles but noted that faculty without training in pedagogy often were unwilling to change. She added that how faculty teach can be traced back to their own professional journey.

> "A lot of people still teach how they were taught because they were successful, right? They wouldn't be here...unless they were successful in their learning and so they figured it out…that that's their comfort level." (Katie)

The pedagogical inertia that results is attributable to this philosophical outlook. Adley agreed with his colleague that faculty teaching the way they were taught, is indeed a strong influence in their behavior and motivation. His experience with group testing led to his call for a paradigm shift as far as training students to not just be competent individuals but also expert collaborators. While Adley believes that most faculty feel a need to ensure students are independently knowledgeable about the course material as well as pressure to keep doing individual assessments to ensure their competence, he also says that only when his colleagues abandon this individualistic "ideology" and focus more on "building a more knowledgeable community" will more of them embrace collaborative assessment practices. Part of the skepticism, he said, is that his colleagues "envision that the students who worked collaboratively, will on the average, be less knowledgeable going forward because they just inherently feel there's a lot of parasitism." Short of faculty buying in to these ideas and an ideological shift occurring, Adley sees a return by his colleagues to "pre-pandemic mentality" as making it harder to convince them to get onboard with collaborative assessment.

> "I think we're in two very different ideological or pedagogical camps and I think it's just the traditional camp has the momentum behind it that … it's hard to sway colleagues to come over and see what I'm doing." (Adley)

Gerald thinks that when faculty refuse to do collaborative assessments, they were never early adoptors of any practice. He also acknowledged that it is sound practice to change the way we do things as teachers especially when a better way for students to learn can be derived out of making such a change.

> "We just know this is a better way to learn…. Moving to collaborative assessments will not make you a better instructor. It won't make you a worse instructor either. The reason you do it is because it's better for the student." (Gerald)

For Katie, that pedagogical inertia manifested itself with her colleagues who were not invested in the teaching aspect of their jobs. While she recognized that teaching is not the primary responsibility of most tenured faculty, the lackluster enthusiasm she saw for implementing collaborative assessment was often overshadowed by colleagues who were "just checking boxes and not truly doing it, just going through the motions." Katie advocates for a more thorough understanding of collaboration as being the goal of the assignment or exam, harkening back to the "assessment for learning" framework.

Gerald felt fortunate that the investment that his colleagues and the entire department made towards implementing collaborative assessment had been paying dividends knowing most faculty in his department practice some form of collaboration in their assessments. He acknowledged the challenge for colleagues to "buy in" to this practice, knowing that they had limited time to implement it properly.

It does not help the cause of those advocating for an ideological shift when, according to faculty interviewed, professors insist on having access to data on whether collaborative assessment practices are efficacious but passively ignoring it whenever it shows positive outcomes. Adley's colleagues wanted "hard evidence" of group exams not eroding the individual performance of students as well as long range data on knowledge retention, to convince themselves to get on board. He is often disappointed when colleagues remain dismissive of the practice despite evidence pointing to its efficacy and steer the conversation towards less data-driven arguments to justify their non-committal. Gerald shared that disappointment with faculty who remain laid back towards implementing collaborative assessment despite research giving the practice high marks for increasing learner outcomes.

> "The data is clear, (so) it's almost malpractice not to change the way you teach, right? Because if you know that teaching that way is gonna help students succeed better and you refuse to do that?" (Gerald)

Best Practices Identified from this Analysis

- *Use collaborative assessment to supplement traditional pen-and-paper evaluations.* There are many topics in STEM disciplines that require higher levels of analysis in tests and exams where students would greatly benefit from a collective effort in their responses.
- *Higher order thinking questions go well with collaborative assessments and group testing.* Instructors can more reasonably expect a higher quality response to a test question requiring greater depth and breadth in problem solving, knowing the collective effort made is a product of multiple perspectives.
- *Design collaborative assessments with individual accountability in mind.* Ensuring that a student's grade in any group examination is an aggregate of individual and collective efforts, with the individual component weighted much heavier relative to the group component, promotes greater accountability and mitigates the effects of loafing or poor group dynamics.
- *Integrate technology in collaborative assessments.* Using collaboration apps and software like Google Docs and Discord not only facilitates for better communication and discussion among students but it also serves as a "paper trail" for the instructor to better assess the level of participation and collaboration.

Conclusions and Implications

This study confirmed the affordances that collaborative assessment and group testing that previous research has outlined. Professors added that collaborative assessment gives students experience before entering the workforce indicating that collaborative assessment is a valuable pedagogical tool for STEM to teach skills that would not have otherwise been available to students. Faculty interviewed suggested that barriers to effective implementation are not insurmountable. They remained realistic about their proposed solutions resolving most problems.

Academic integrity issues, while viewed as implementation hurdles, were also seen as not insurmountable. Faculty reiterated that individual accountability measures built into collaborative assessment helped mitigate the issue. They cited philosophical and ideological barriers as significant to what they see as a lack of enthusiasm for collaborative and assessment practices.

While the limited number of participants in this study do not speak for all higher education STEM educators, it presents broad implications for the future of collaborative assessment. The use of collaborative assessment practices has been a consequence of reform efforts to integrate essential aspects of the scientific enterprise into STEM pedagogy. As the fulcrum of change in any curriculum effort, higher education instructors are crucial players in the success or failure of pedagogical initiatives. How college faculty view collaborative assessment practices crucial for a future where collaborative assessments are the norm rather than the exception in college STEM classrooms.

This need places a heavier burden on current STEM educators to be more cognizant of the science behind successful pivots in best practices to impact the students moving into STEM careers. This much needed paradigm shift involves a simultaneous ideological move forward towards an academic environment where more collaboration is valued, A more serious effort to move beyond discussing the benefits and drawbacks of collaborative assessment and implementing a data-driven approach to enhance benefits is a first step. Advocating for more resources to support implementation and greater assistance to faculty willing to embrace collaborative assessment logically follows.

Faculty also noted the need for an open mindset to combat academic integrity issues that threaten to compromise implementation and acceptance of collaborative assessment in higher education as well as pedagogical reform efforts at all levels. As technology and society evolve, adopting a more mindful and purposeful ideology would fortify faculty to combat pedagogical inertia before it sets in, as well as the stagnation in STEM education that is an omnipresent threat to maintaining best practices.

References

American Association for the Advancement of Science. (2011, April). *Vision and change in undergraduate biology education: A call to action*. Washington, D.C.

Bloom, D. (2009). Collaborative test taking: Benefits for learning and retention. *College Teaching*, *57*(4), 216–220. https://doi.org/10.1080/87567550903218646

Bowen, C. W., & Phelps, A. J. (1997). Demonstration-Based Cooperative Testing in General Chemistry: A Broader Assessment-of-Learning Technique. *Journal of Chemical Education*, *74*(6), 715. https://doi.org/10.1021/ed074p715

Clark, T. M., Callam, C. S., Paul, N. M., Stoltzfus, M. W., & Turner, D. (2020). Testing in the time of covid-19: A sudden transition to unproctored online exams. *Journal of Chemical Education*, *97*(9), 3413–3417. https://doi.org/10.1021/acs.jchemed.0c00546

Cooke, J. E., Weir, L., & Clarkston, B. (2019). Retention following two-stage collaborative exams depends on timing and student performance. *CBE—Life Sciences Education*, *18*(2). https://doi.org/10.1187/cbe.17-07-0137

Cortright, R. N., Collins, H. L., Rodenbaugh, D. W., & DiCarlo, S. E. (2003). Student retention of course content is improved by collaborative-group testing. *Advances in Physiology Education*, *27*(3), 102–108. https://doi.org/10.1152/advan.00041.2002

Creswell, J. W. (2014). Research Design: Qualitative, Quantitative and Mixed Methods Approaches (4th ed.). Thousand Oaks, CA: Sage.

Delucchi, M. (2006). The Efficacy of Collaborative Learning Groups in an Undergraduate Statistics Course. *College Teaching*, *54*, 244–248. https://doi.org/10.3200/CTCH.54.2.244-248

Dietrich, N., Kentheswaran, K., Ahmadi, A., Teychené, J., Bessière, Y., Alfenore, S., Laborie, S., Bastoul, D., Loubière, K., Guigui, C., Sperandio, M., Barna, L., Paul, E., Cabassud, C., Liné, A., & Hébrard, G. (2020). Attempts, successes, and failures of distance learning in the time of covid-19. *Journal of Chemical Education*, *97*(9), 2448–2457. https://doi.org/10.1021/acs.jchemed.0c00717

Dobie, T. E., & MacArthur, K. (2022). Exploring Shifts in Student Attitudes Toward Group Exams in College Calculus: The Case of Dane. *International Journal of Research in Undergraduate Mathematics Education*, *8*(1), 149–175. https://doi.org/10.1007/s40753-021-00148-7

Ellis, R. A. (2016). Students' approaches to groupwork in a blended course, associations with perceptions of the online environment and academic achievement – when is learning engaged? *Education and Information Technologies*, *21*(5), 1095–1112. https://doi.org/10.1007/s10639-014-9370-4

Gilley, B., & Clarkston, B. (2014). Research and teaching: Collaborative testing: Evidence of learning in a controlled in-class study of undergraduate students. *Journal of College Science Teaching*, *043*(03). https://doi.org/10.2505/4/jcst14_043_03_83

Giuliodori, M. J., Lujan, H., & DiCarlo, S. E. (2008). Collaborative group testing benefits high performing as well as low performing students. *The FASEB Journal*, *22*(S1). https://doi.org/10.1096/fasebj.22.1_supplement.766.4

Green, R. A., Cates, T., White, L., & Farchione, D. (2015). Do collaborative practical tests encourage student-centered active learning of gross anatomy? *Anatomical Sciences Education*, *9*(3), 231–237. https://doi.org/10.1002/ase.1564

Grzimek, V., Kinnamon, E., & Marks, M. B. (2020). Attitudes about classroom group work: How are they impacted by students' past experiences and major? *Journal of Education for Business*, *95*(7), 439–450. https://doi.org/10.1080/08832323.2019.1699770

Grzimek, V., Marks, M. B., & Kinnamon, E. (2014). Do Differences in GPA Impact Attitudes About Group Work? A Comparison of Business and Non-Business Majors. *Journal of Education for Business*, *89*(5), 263–273. https://doi.org/10.1080/08832323.2013.872591

Hamnett, H. J., McKie, A. E., & Morrison, C. (2018). Postgraduate students' attitudes towards group work: Experiences within a forensic chemistry programme. *Chemistry Education Research and Practice*, *19*(4), 1240–1252. https://doi.org/10.1039/c8rp00126j

Hickey, B. L. (2006). Lessons learned from collaborative testing. *Nurse Educator*, *31*(2), 88–91. https://doi.org/10.1097/00006223-200603000-00012

Hillyard, C., Gillespie, D., & Littig, P. (2010). University students' attitudes about learning in small groups after frequent participation. *Active Learning in Higher Education*, *11*(1), 9–20. https://doi.org/10.1177/1469787409355867

Huang, J. (2020). Successes and Challenges: Online teaching and learning of chemistry in higher education in China in the time of covid-19. *Journal of Chemical Education, 97*(9), 2810-2814. doi:10.1021/acs.jchemed.0c00671

Jang, H., Lasry, N., Miller, K., & Mazur, E. (2017). Collaborative exams: Cheating? Or learning? *American Journal of Physics*, *85*(3), 223–227. https://doi.org/10.1119/1.4974744

Johnson, D. W., & Johnson, R. T. (2015). *Cooperation and Competition*. https://doi.org/10.1016/B978-0-08-097086-8.24051-8

Johnston, L., & Miles, L. (2005). Assessing contributions to group assignments. *Assessment & Evaluation in Higher Education*, *29*(6), 751–768. https://doi.org/10.1080/0260293042000227272

Karau, S. J., & Williams, K. D. (1993). Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology*, *65*(4), 681–706. http://dx.doi.org.cmich.idm.oclc.org/10.1037/0022-3514.65.4.681

Khong, M. L., & Tanner, J. A. (2020). A collaborative two-stage examination in BIOMEDICAL sciences: Positive impact on feedback and peer collaboration. *Biochemistry and Molecular Biology Education*, *49*(1), 69–79. https://doi.org/10.1002/bmb.21392

King, P. E., & Behnke, R. R. (2005). Problems Associated with Evaluating Student Performance in Groups. *College Teaching*, *53*(2), 57–61. https://doi.org/10.3200/CTCH.53.2.57-61

Kinnear, G. (2020). Two-stage collaborative exams have little impact on subsequent exam performance in undergraduate mathematics. *International Journal of Research in Undergraduate Mathematics Education*, *7*(1), 33–60. https://doi.org/10.1007/s40753-020-00121-w

Knierim, K., Turner, H., & Davis, R. K. (2015). Two-stage exams improve student learning in an introductory geology course: Logistics, attendance, and grades. *Journal of Geoscience Education*, *63*(2), 157–164. https://doi.org/10.5408/14-051.1

Kusch, J. (2016). Learning experiences of university biology faculty: A qualitative pilot study. *Journal of College Science Teaching, 045*(06). https://doi:10.2505/4/jcst16_045_06_24

Lam, R. (2019). What students do when encountering failure in collaborative tasks. *Npj Science of Learning*, *4*(1), 6. https://doi.org/10.1038/s41539-019-0045-1

Leight, H., Saunders, C., Calkins, R., & Withers, M. (2012). Collaborative testing improves performance but not content retention in a large-enrolment introductory biology class. *CBE—Life Sciences Education*, *11*(4), 392–401. https://doi.org/10.1187/cbe.12-04-0048

Lewis, S. E. (2020). Chemistry assessments through the sudden implementation of online instruction. *Journal of Chemical Education*, *97*(9), 3418–3422. https://doi.org/10.1021/acs.jchemed.0c00697

Machemer, P. L., & Crawford, P. (2007). Student perceptions of active learning in a large cross-disciplinary classroom. *Active Learning in Higher Education*, *8*(1), 9–30. https://doi.org/10.1177/1469787407074008

Marks, M. B., & O'Connor, A. H. (2013). Understanding Students' Attitudes About Group Work: What Does This Suggest for Instructors of Business? *Journal of Education for Business*, *88*(3), 147–158. https://doi.org/10.1080/08832323.2012.664579

McConnell, D. (2002). The experience of collaborative assessment in e-learning. *Studies in Continuing Education*, *24*(1), 73–92. https://doi.org/10.1080/01580370220130459

Medaille, A., & Usinger, J. (2018). "That's Going to Be the Hardest Thing for Me": Tensions Experienced by Quiet Students during Collaborative Learning Situations. *Educational Studies*, 1–18. https://doi.org/10.1080/03055698.2018.1555456

Merriam, S. B., & Tisdell, E. J. (2017). *Qualitative research: A guide to design and implementation*. Langara College.

Meseke, C. A., Nafziger, R., & Meseke, J. K. (2010). Student attitudes, satisfaction, and learning in a collaborative testing environment. *Journal of Chiropractic Education*, *24*(1), 19–29. https://doi.org/10.7899/1042-5055-24.1.19

Miller, S. T. J. (2019). *What is the impact of collaborative exams on learning and attitudes in introductory astronomy classes?* Journal of Astronomy & Earth Sciences Education. Retrieved September 20, 2021, from https://eric.ed.gov/?id=EJ1228738.

Mishra, R. K. (2014). Social Constructivism and Teaching of Social Science. *Journal of Social Studies Education Research*, *5*. https://doi.org/10.17499/jsser.22283

Muir, S. P., & Tracy, D. M. (1999). Collaborative essay testing: Just try it! *College Teaching*, *47*(1), 33–35. https://doi.org/10.1080/87567559909596077

National Academies of Sciences, Engineering, and Medicine. 2012. A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas. Washington, DC: The National Academies Press. https://doi.org/10.17226/13165.

Natoli, R., Jackling, B., & Seelanatha, L. (2014). The impact of instructor's group management strategies on students' attitudes to group work and generic skill development. *Pedagogies*, *9*(2), 116–132. https://doi.org/10.1080/1554480X.2014.912519

Neber, H., Finsterwald, M., & Urban, N. (2001). Cooperative Learning with Gifted and High-achieving Students: A review and meta-analyses of 12 studies. *High Ability Studies*, *12*(2), 199–214. https://doi.org/10.1080/13598130120084339

Newton, G., Ritchie, K., Albabish, W., Gilley, B., Rajakaruna, R., & Kulak, V. (2019). Research and teaching: Two-stage (collaborative) testing in science teaching: Does it improve grades on short-answer questions and retention of material? *Journal of College Science Teaching*, *048*(04). https://doi.org/10.2505/4/jcst19_048_04_64

Opdenakker, R. J. G. (2006). Advantages and disadvantages of four interview techniques in qualitative research. In *Forum Qualitative Sozialforschung= Forum: Qualitative Social Research* (Vol. 7, No. 4, pp. art-11). Institut fur Klinische Sychologie and Gemeindesychologie.

Piezon, S. L. (2011). Social loafing and free riding in online learning groups [Ph.D., The Florida State University]. In *ProQuest Dissertations and Theses*. https://www.proquest.com/docview/897114500/abstract/D120FE02B6D740C7PQ/1

Raje, S., & Stitzel, S. (2020). Strategies for effective assessments while ensuring academic integrity in general chemistry courses during COVID-19. *Journal of Chemical Education*, *97*(9), 3436–3440. https://doi.org/10.1021/acs.jchemed.0c00797

Rivaz, M., Momennasab, M., & Shokrollahi, P. (2015, October 1). *Effect of collaborative testing on learning and retention of course content in nursing students*. Journal of Advances in Medical Education & Professionalism. Retrieved September 20, 2021, from http://jamp.sums.ac.ir/index.php/JAMP/article/view/354.

Rudawska, A. (2017). Students' Team Project Experiences and Their Attitudes Towards Teamwork. *Journal of Management & Business Administration. Central Europe*, *25*(1), 78–97. https://doi.org/10.7206/jmba.ce.2450-7814.190

Rupnow, R. L., LaDue, N. D., James, N. M., & Bergan-Roller, H. E. (2020). A perturbed system: How tenured faculty responded to the COVID-19 shift to remote instruction. *Journal of Chemical Education*, *97*(9), 2397–2407. https://doi.org/10.1021/acs.jchemed.0c00802

Sansom, R. L. (2020). Pressure from the pandemic: Pedagogical dissatisfaction reveals faculty beliefs. *Journal of Chemical Education, 97*(9), 2378-2382. https://doi:10.1021/acs.jchemed.0c00657

Siegel, M., Roberts, T., Freyermuth, S., Witzig, S., & Izci, K. (2015). Research and teaching: Aligning assessment to instruction: Collaborative group testing in large-enrollment science classes. *Journal of College Science Teaching*, *044*(06). https://doi.org/10.2505/4/jcst15_044_06_74

Smith, K. C., & Sepulveda, A. (2018). Students' perceptions of common practices, including some academically dishonest practices, in the undergraduate general chemistry classroom laboratory. *Chemistry Education Research and Practice*, *19*(4), 1142–1150. https://doi.org/10.1039/C8RP00058A

Sturges, J. E., & Hanrahan, K. J. (2004). Comparing telephone and face-to-face qualitative interviewing: a research note. *Qualitative research*, *4*(1), 107-118. https://doi.org/10.1177/1468794104041110

Thaanyane, M., & Jita, T. (2024). Enhancing 21st Century Skills Through Digital Collaborative Assessment Practices in Lesotho Schools. *Educational Administration: Theory and Practice*, *30*(11), 495-502.

Theobald, E. J., Eddy, S. L., Grunspan, D. Z., Wiggins, B. L., & Crowe, A. J. (2017). Student perception of group dynamics predicts individual performance: Comfort and equity matter. *PLOS ONE*, *12*(7), e0181336. https://doi.org/10.1371/journal.pone.0181336

Wilson, K. (2020). Balancing the disruptions to the teaching and Learning Equilibrium—responsive Pedagogic approaches to teaching online during the Covid-19 pandemic in general chemistry classes at an Arabian Gulf University. *Journal of Chemical Education*, *97*(9), 2895–2898. https://doi.org/10.1021/acs.jchemed.0c00702

Zgheib, N. K., Dimassi, Z., Bou Akl, I., Badr, K. F., & Sabra, R. (2016). The long-term impact of team-based learning on medical students' team performance scores and on their peer evaluation scores. *Medical Teacher*, *38*(10), 1017–1024. https://doi.org/10.3109/0142159X.2016.1147537

Zipp, J. F. (2007). Learning by exams: The impact of two-stage cooperative tests. *Teaching Sociology*, *35*(1), 62-76. https://doi.org/10.1177/0092055X070350

Corresponding Author Contact Information:

Author name: Allan Jay Dinglasan

Department: Teacher and Special Education

University, Country: Central Michigan University, USA

Email: dingl1a@cmich.edu

Please Cite: Dinglasan, A. J., & Weible, J. L. (2025). Higher Education STEM Faculty Views on Collaborative Assessment and Group Testing. *Journal of Research in Science, Mathematics and Technology Education, 8*(SI), 315-334. DOI: https://doi.org/10.31756/jrsmte.4113SI

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential 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: Data available upon request.

Ethics Statement: This study obtained institutional review board approval prior to conducting any research related to this study.

Author Contributions: All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Received: March 16, 2025 ▪ Accepted: May 27, 2024