

LEVELED ASSESSMENTS IN HIGH SCHOOL BIOLOGY AND THEIR EFFECTS ON
STUDENT WELL BEING

by

Beth Leigh Lakin

A professional paper submitted in partial fulfillment
of the requirements for the degree

of

Master of Science

in

Science Education

MONTANA STATE UNIVERSITY
Bozeman, Montana

July 2026

©COPYRIGHT

by

Beth Leigh Lakin

2026

All Rights Reserved

TABLE OF CONTENTS

1. INTRODUCTION AND BACKGROUND.....	1
Context of the Study	1
Focus Questions	2
2. CONCEPTUAL FRAMEWORK.....	3
Rural Schools and Differentiation.....	3
Summative Assessment and Differentiation	5
Summative Assessment, Student Self-Esteem and Self-Efficacy	6
Conclusion	8
3. METHODOLOGY	9
Demographics	9
Research Design.....	10
Data Collection and Analysis Strategies	11
Assessment Scores	12
Student Survey	13
Interviews.....	13
Teacher Journal	14
Analysis Strategies.....	15
Assessment Data	15
Qualitative Data	15
Conclusion	15
4. DATA ANALYSIS: CLAIMS AND EVIDENCE.....	17
Results.....	17
Test Performance.....	17
Student Confidence	20
Perceived Fairness	23
5. VALUE: REASONING AND REFLECTION	25
Claims From the Study	25
Student Performance	25
Student Confidence	26
Perceptions of Fairness	26
Value of the Study and Consideration for Future Research	27
Impact of Action Research on the Author	28
REFERENCES CITED.....	30

APPENDICES	35
APPENDIX A: IRB APPROVAL	36
APPENDIX B: STUDENT SURVEY INSTRUMENT	38
APPENDIX C: GROUP INTERVIEW GUIDE	40

LIST OF TABLES

Table	Page
1. Table 1. Data Triangulation Matrix	11
2. Table 2: Example of CDE Performance Expectations (CDE, 2022)	12

LIST OF FIGURES

Figure	Page
3. Figure 1. Comparison of student scores on the traditional test vs treatment, (N=12).....	18
4. Figure 2. Distribution of scores on traditional test vs treatment, (N=12).	18
5. Figure 3. Performance change by student, (N=12).	19
6. Figure 4. Score on the treatment vs traditional test, (N=12).....	20
7. Figure 5. Pre- and post-treatment survey responses to questions related to test worry and anxiety. (n=8).	21
8. Figure 6. Student survey data for the statement “I have done well in science classes in the past,” (n=8).	22
9. Figure 7. Student survey data regarding general beliefs about success in science, (n=8).	22
10. Figure 8. Student Survey data about perceived fairness of the leveled test model, (n=8).....	24

ABSTRACT

Traditional summative assessments often fail to accurately measure student understanding, limiting meaningful feedback and negatively impacting student confidence, performance, and perceptions of learning. To determine how leveled assessments impact student performance, confidence and perceptions of fairness, students in high school biology were administered a unit test in three leveled versions. Students selected their test version in consultation with the teacher and could opt into a lower-level assessment if they found their selection to be too difficult. Student performance on the leveled assessment led to modest, statistically insignificant student performance compared to a prior traditional assessment. Student confidence increased slightly although students indicated a strong preference for leveled assessments citing agency and appropriate difficulty as positives. This initial study indicates that leveled assessments are a possible way to more accurately measure student performance and give better feedback.

CHAPTER ONE

INTRODUCTION AND BACKGROUND

Context of the Study

In my career, I have only taught science in small rural schools. Usually, I am the only high school science teacher, and most classes contain students at all ability levels. This often poses challenges for summative assessments as high achieving students can achieve perfect scores that might not fully probe the frontiers of their understanding and other students return blank tests or tests that guess at answers and come back covered in red ink. For struggling learners, this situation can make unit assessments feel like inevitable failures. It can also contribute to a negative teacher–student dynamic regarding success in future science courses I may teach and weaken students’ beliefs in their ability to succeed in science overall.

As a result of this challenge, I have considered giving leveled assessments. The goal of a leveled assessment is to have students complete a task that is appropriately leveled for their skills. With a leveled assessment, students would be able to complete all questions on the test and fully demonstrate their knowledge on a standard. A more fully completed assessment would facilitate giving better feedback to students. With these improvements, students are more likely to see assessments as part of the learning process rather than punitive.

I teach at Ouray High School in Ouray, Colorado. Housed in a pre-kindergarten through 12th grade building, Ouray High School serves 51 students. Science offerings include earth science, biology, chemistry, physics, AP Biology, and AP Environmental Science. Class sizes range from 4-15 students. As a small school, Ouray’s mission is to provide an “exceptional

education” in a “safe, supportive, and inclusive environment” (Ouray School, n.d.). Overall, Ouray High School students outperform the state on the Colorado Measures of Academic Success (CMAS) Science Assessment with 50% of Ouray students meeting or exceeding standards compared to 25% of Colorado students, with similar results over the prior five years (Colorado Department of Education, 2025).

I completed this action research (AR) project with my second semester biology students by offering them choice in their assessments. Students selected assessments leveled as “approaching expectations,” “meets expectations,” and “exceeds expectations” based on their own self-assessment and teacher guidance. The purpose of this study was to understand if leveled assessments would allow students to feel more successful, receive more useful feedback on their assessments, and, in general, increase positive feelings around science and learning.

Focus Questions

My focus question was, How do leveled assessments affect student confidence and performance in high school biology?

My sub-questions include the following:

1. How does using leveled assessments affect student confidence in biology class?
2. How do students perceive the fairness and helpfulness of leveled assessments in biology?

CHAPTER TWO

CONCEPTUAL FRAMEWORK

Although Ouray High School demonstrates strong overall performance on the state science assessment, many students still fail to meet standards, and very few exceed them. The largest gains in test scores can be made by offering appropriate challenge and feedback to both low- and high-performing students. Like many rural schools, Ouray High School faces unique challenges in serving diverse populations in fewer class sections. Teachers may instruct many courses with little peer support while also serving a variety of abilities within a single classroom. Utilizing differentiated summative assessments is a time effective approach that can also support student well-being in the classroom.

Rural Schools and Differentiation

Rural school districts often have uniquely complex and varied learner demographics within each classroom. Rural schools often cannot provide the number of courses offered by larger schools particularly in Advanced Placement or college concurrent levels (Hart et al., 2023). These districts often have higher than average proportions of English Language Learners (ELL) (Johnson et al., 2018) and disproportionately high numbers of students affected by adverse childhood experiences (Zolkoski et al., 2021). Similarly, these districts must also meet the needs of special education students and gifted students, often within the same classroom (Stambaugh & Wood, 2015). With each of these groups of students facing unique challenges, rural teachers must find ways to serve their learners within a single section of a course while managing

individualized education plans (IEPs), 504 plans, advanced learning plans (ALPs), culturally and linguistically diverse students, and other individual student needs.

To meet the needs of a class where each student is treated as an individual, teachers are tasked with differentiating instruction. To adequately differentiate, instruction must be tailored with respect to content, process, products, and learning environment (Tomlinson, 2017). Differentiation helps teachers “better ensure that all learners are receiving respectful work, while they address the concepts required by the state content standards and make meaningful academic progress” (Richards & Omdal, 2007, p. 426). Differentiation is an accepted way to make sure the needs of all learners are being met by ensuring “all versions of tasks that give them the opportunity to grow” (ASCD, n.d.). Meaningful progress for all students requires teachers differentiate.

Rural teachers often work independently in their department which creates challenges with implementing differentiation. Van Geel et al. (2018) identified a variety of factors that increase complexity in differentiation implementation. Among those factors, *group composition* and *school support* have a particularly strong impact on rural teachers. Due to their small size, rural schools often can provide little to no grouping of students into courses and teachers often have no subject area peers in their district with which they can collaborate (Inouye et al., 2023). Furthermore, teachers often feel stress in not being able to provide individualized curricula for each student because of time pressures (Pozas et al., 2023). This is largely because differentiated instruction “cannot be done ‘on the fly,’ but rather in a planned and purposefully reflected process” (Pozas et al., 2023). Although differentiation is considered the gold standard in instructional practice, rural teachers may need to focus their attention on the most time effective

practices to teach and evaluate each individual student. One way that teachers can choose to do this is through differentiating summative assessments.

Summative Assessment and Differentiation

A key component of differentiation is that all students work toward the same standards but assessing them in the same way does not allow every student to demonstrate what they have learned. Given that students have varying abilities and will reach learning goals at varying levels of success, assessments must be designed such that all students are able to demonstrate knowledge on the assessment (Kaur et al., 2018). When teachers give high quality assessments, it encourages further learning and gives students direction on how to continue their progress in meeting or exceeding standards (Stiggins, 2007). Assessments that are left largely blank or in which students guess at random cannot be used to accurately assess learning and give appropriate feedback. In the same way that teachers differentiate instruction for student success, assessments need to be constructed to accurately measure that success.

Summative assessments are conducted after instruction to measure a students' attainment of learning goals. Summative assessments can take a variety of formats including (a) written tests, (b) portfolios, and (c) performance assessments (Moon, 2005; Stefanakis, 2010). In a survey of the literature, Brownlie et al. (2024) identified six facets of effective summative assessment: (1) validity, (2) reliability, (3) fairness, (4) authenticity, (5) flexibility, and (6) other. Of particular importance in differentiation of assessment is the idea of flexibility. Brownlie et al.'s survey showed if "each student has an equal chance to demonstrate their knowledge or understanding, it can be possible to compare achievement and draw accurate conclusions" (p. 36). Often, summative assessments do not reflect student knowledge because of flaws in task

design including achieving a level of task difficulty in which students can show their skills.

Black et al. (2011) found “some tasks were beyond some pupils so that they produced nothing of merit, whilst other tasks had too low a ceiling so that the strongest were not stretched” (p. 455).

If such assessments are administered, only a select group of students will have the opportunity to adequately show their achievement on the standards. Creating assessments such that students can show their knowledge is pivotal to respecting student efforts on the assessment and building a respectful culture of learning.

Summative Assessment, Student Self-Esteem and Self-Efficacy

Differentiated instruction has ties to improved student outcomes in self-esteem and other measures of student well-being. Roy et al. (2015) found that differentiated practices reduced feelings of competitiveness and increased self-assessment ability. When instruction is highly individualized, special education students experience an increased academic self -concept (Alnahdi et al., 2022). Assessments that focus on individual growth rather than comparing between students are more likely to have positive impacts on student socio-emotional outcomes (Pozas et al., 2021). Much less research has been completed on the impact of assessments and assessment feedback on student psychology.

Student beliefs about their science self-efficacy have broad implications for their success and continued pursuit of science. When students believe that they can succeed in science, they tend to “select such tasks and activities, work hard to complete them successfully, persevere in the face of difficulty, and be guided by physiological indexes that promote confidence as they meet obstacles” whereas students who do not believe they will succeed tend to avoid such work (Britner & Pajares, 2006, p. 486). Science-self efficacy can be a stronger predictor of

performance than demographic information such as gender, ethnicity, and parental background suggesting that promoting self-efficacy can have an equity impact in the classroom (Britner & Pajares, 2006). Of the four proposed components of self-efficacy, mastery experience is the strongest predictor of self-efficacy (Talsma et al., 2018). Giving students work that allows them to showcase their knowledge should increase mastery beliefs and improve emotional attitudes towards science and their science self-efficacy.

Students are emotionally invested in their educational outcomes whether because of intrinsic motivation or because of messaging about school outcomes and future success. Connections have been found between “students’ emotions and their motivation, learning strategies, cognitive resources, self-regulation and academic achievement” (Christie & Morris, 2021, p. 150). Self-regulation theory suggests that students who see the benefits of assessment may earn higher grades as opposed to seeking “to blame schools or teachers for poor assessment results” (Lee et al., 2013, p. 248). Students who see assessments as beneficial to learning are more likely to seek challenge, persevere, and take educational risks while students who see assessment as punitive are more likely to avoid new skills, seek what is easy, and avoid initiative (Stiggins, 2007). Student emotional responses must be considered for optimal educational and personal growth of students.

Students’ emotional responses to their coursework, including summative feedback, have broad implications. Emotional responses in the learning environment affect a student’s ability to learn information (To et al., 2023). A student’s emotions can affect “cognitive processes, attention, memory, reasoning and information retrieval” which can have substantial impacts on their ability to learn and to demonstrate their learning (Hopwood et al., 2024, p. 194). Feedback

that is perceived by students to contribute to skill development is most likely to generate positive emotions (Chen & Brown, 2018). It is imperative that the feedback given by teachers in summative assessments does not activate emotions that will negatively impact learning. The literature regarding assessment and emotions has been largely limited to the higher education context. The research that has been completed shows that students who receive comments that are critical or upsetting tend to become demotivated and do not use the feedback to improve (Ryan & Henderson, 2018). In some situations that feedback can hinder future learning and cause students to change career aspirations (Crossman, 2007). Zero grades are perhaps the most extreme version of negative feedback and are often awarded when students leave questions blank demonstrating no level of understanding (Paul et al., 2018). When teachers receive no feedback on a student's knowledge, this results in feedback that can be damaging to a student's desire to learn.

Conclusion

Applying differentiation tactics to summative assessments will allow students to best demonstrate their knowledge of the target concepts by focusing on what they do know rather than on what they do not. Differentiated assessments will also increase measures of student well-being. Biology is often a decision point for students about whether they will continue a rigorous science education or transition to simply fulfilling graduation requirements. Facilitating their learning with assessments that honor the work that they have done, rather than the concepts they haven't learned will increase morale in the classroom and improve attitudes towards science.

CHAPTER THREE

METHODOLOGY

Demographics

The purpose of this study is to examine the impact of leveled assessments on student performance and confidence in high school biology class. I wanted to understand how giving students an opportunity to give higher quality answers that actually reflected their learning could improve test scores and boost student confidence. An additional sub-question seeks to quantify student perceptions of fairness in the exams. While the assessments explicitly address the same standards, it is important that students are able to understand the use of leveled assessments so as not to undermine class morale.

Ouray High School is comprised of 54 students in grades 9-12 and emphasizes having high standards in a supportive learning environment. As the primary science teacher in a small school, all students take at least two years of science in my classroom and many take three or four years of science with me. It is important that I find ways to balance high expectations with making student efforts feel valued and that they know their learning is supported in my classroom since they're likely to be back in my room the following year.

This study was conducted with twelve students in high school biology, with the majority being girls ($N=12$). One student is on an individualized learning plan and three are on advanced learning plans. This group is characterized by excellent discussion between students of similar abilities and growing levels of discussion within table groups of mixed abilities. Ouray School only offers one section of each course which necessitates use of a nonexperimental research

design as students are not assigned randomly between sections. The research methodology for this project received an exemption by Montana State University's Institutional Review Board and compliance for working with human subjects was maintained (Appendix A).

Research Design

This study utilized a mixed-methods research design to capitalize on the strengths of both qualitative and quantitative methods (Mertler, 2017). Qualitative data is good at answering complex questions around ideas like confidence and fairness (Gaudet & Robert, 2018). For questions of how performance is improved, quantitative data is important. Nonexperimental research design is that in which “the groups already exist and the experimenter cannot or does not attempt to manipulate an independent variable” (Lobmeier, 2010, p. 911). A one-group pretest-posttest design allows for one group to be “its own control” (Spector, 1981, p. 29). This design assumes that any changes are attributable to the treatment.

Instead of providing all students with the same end of unit assessment, students were offered a choice of three leveled assessments for two units in the second semester. The leveled assessments for the treatment units were written targeting different performance expectations issued by the Colorado Department of Education (2022) (Table 2). These leveled assessments were used for a unit on homeostasis early in the second semester. The leveled assessments are designed to help students demonstrate their learning as few students have no knowledge at all about a standard.

There are three variations of the assessment for each unit: (1) the A assessment that differentiated between meeting and exceeding standards, (2) the B assessment to differentiate between approaching and meeting standards, and (3) the C assessment that differentiated

between partially meeting expectations and approaching expectations. An example of the Colorado Performance Standards is included in Table 2. Questions on the A level assessment generally required more analysis and independent thought from students to demonstrate exceeding the state performance expectations. The C level exam provided some additional scaffolding for students to demonstrate their knowledge. These scaffolds included sentence stems, word banks, and multiple-choice questions paired with free response explanations. Selection of the student tests were made based on a combination of student choice and teacher feedback. Providing students with appropriate supports to avoid blank answers and other non-informative responses such as “IDK” (common student shorthand for “I don’t know”) allow for teachers to give useful feedback to students.

Table 1. Data Triangulation Matrix.

Data Collection Instruments	Focus Questions		
	How do leveled assessments affect student confidence and performance in high school biology?	How does using leveled assessments affect student confidence in biology class?	How do students perceive the fairness and helpfulness of leveled assessments in biology?
Pre and Post Tests	X		
Semi-structured interviews	X	X	X
Student Perception Survey	X	X	X
Teacher Notes and Observations	X	X	X

Data Collection and Analysis Strategies

This study examines whether leveled assessments enhance students’ ability to demonstrate their learning, improve exam performance, and influence their self-confidence. This

study used a mixed-methods design that incorporated both qualitative and quantitative data to answer the research questions.

Table 2: Example of CDE Performance Expectations for evidence outcomes (EOs) (CDE, 2022)

Standard	Partially Meets Expectations	Approaches Expectations	Meets Expectations	Exceeds Expectations
GLE 2.12	Identify natural selection as a cause of adaptation in populations. Identify extinction as a possible effect of the endangerment of species. [2.12.a, 2.12.b]	Relate a simple explanation of natural selection as a cause of adaptation in populations. Relate familiar evidence of how environmental conditions cause population increase or decrease and can affect extinction and emergence of species. [2.12.a, 2.12.b]	Explain or evaluate the evidence supporting claims about examples referenced in the EOs that changes in environmental conditions can cause population increase or decrease, extinction of existing species, adaptations of populations, or emergence of new species through the process of natural selection. [2.12.a, 2.12.b]	Explain or evaluate scientific information about changes in environmental conditions and predict their likely effects on population increase or decrease, extinction of existing species, adaptations of populations, or emergence of new species through the process of natural selection, including examples not referenced in the EOs. [2.12.a, 2.12.b]

Assessment Scores

Performance was evaluated using assessment scores. Test scores were collected for a control unit before the treatment, the control unit test was the same for all students. The instruction methods for the control unit and the assessment units were held constant, so these test scores were used in data analysis to compare student performance on a standard assessment to

the treatment. The assessments were written to target CDE performance standards. Students selected their leveled assessment in consultation with their teacher based upon self-evaluation and performance throughout the unit on formative assessments. Scores were reported for each standard as *partially meets expectations*, *approaches expectations*, *meets expectations*, or *exceeds expectations*.

Student Survey

A student survey was administered to understand how student perceptions and attitudes change after the treatment. The instrument is modeled after Britner and Pajares (2006) who examined the factors determining self-efficacy with answers recorded as a Likert scale with four options. The questions focused on Mastery Experiences and Emotional States (Jones et al., 2023; Figure 1). Two open-ended questions directly addressing the sub questions have been added to garner additional qualitative responses (Appendix B). This survey was conducted after the pre-treatment exam and after the second treatment exam.

Interviews

Interviews were used to gather information about student confidence and perceptions of fairness. Group interviews in combination with individual interviews and open-ended survey questions were used. Group interviews offer students the opportunity to respond to each other and likely go deeper than they would in a one-on-one interview (Holstein & Gubrium, 2003). Homogenous groups of students were interviewed to increase the likelihood of discussion of learning challenges with each other; this grouping is consistent with focus group methodologies (Smithson, 2019). All students were assigned to groups for an interview although group sizes

were not consistent. A semi-structured group interview (Appendix C) guide was developed to probe student experience.

Individual interviews, in addition to the group interviews, were conducted to give space to students who are uncomfortable speaking in a group setting or for whom discussions of educational confidence are tied to deeply personal life experiences or histories (Smithson, 2019). Time limitations required limiting individual interviews to participants who participated less in the group interview setting or that had unique perspectives to contribute.

Teacher Journal

Finally, a teacher journal was kept during the treatment period. Teacher reflections and observations were recorded for each assignment and assessment with feedback given on the performance assessment scale. Observations included completeness of assessments to observe whether students were able to show their knowledge and of student responses to feedback. Reflections were naturalistic observations (Teachers College, n.d.). Unstructured observations allow for teachers “to shift focus from one event to another as new, and perhaps more interesting, events arise” (Mertler, 2019, p. 131). The teacher journal filled in gaps in the qualitative data generated from student surveys and interviews. The additional perspective from my own notes helped to round out the qualitative data from the relatively small class. Additionally, with a small class size, subjective teacher data was used to help to evaluate the main question around student performance.

Analysis Strategies

Assessment Data

Analysis of the pre- and post-test score data was completed with the Wilcoxon ranked sign test. The Wilcoxon ranked sign test considers both the magnitude and direction of changes for non-parametric data (Pett, 2016). This test is frequently used to describe changes and evaluate the effectiveness of interventions (Pett, 2016). The ranked sign test can only examine changes in one group over time but for this nonexperimental design, that is acceptable. Unfortunately, no key subgroups (IEP, ALP, ELL, etc.) had enough members for separate analysis of performance data.

Qualitative Data

The qualitative data including group and individual interviews, survey data, and the teacher journal was analyzed using thematic analysis. Thematic analysis of interview data, particularly transcripts, allows the researcher to distill large amounts of qualitative data into usable themes and patterns (Mertler, 2017). The transcripts of the student responses were reviewed to identify themes that are relevant to the research questions. After reviewing the transcripts, key words can be identified and then coded. The codes can be further grouped into themes (Naeem et al., 2023). These themes can then be further interpreted to reach conclusions about the impacts of the treatment on the sample.

Conclusion

Multiple sources of data were utilized to address the research questions around leveled assessments. Assessment data, student surveys, interviews, and a teacher journal were collected.

These qualitative and quantitative sources were used to analyze the impacts of leveled assessments on performance and student confidence.

CHAPTER FOUR

DATA ANALYSIS: CLAIMS AND EVIDENCE

Results

Student performance was higher on the leveled assessment than on the traditional assessment although the improvements were not statistically significant. The improvement in performance was stronger for lower performing students. Student confidence seemed to slightly improve and students overwhelmingly perceived the leveled assessments as fair and desirable.

Test Performance

Statistical analyses using both descriptive and inferential statistics indicated that leveled assessments modestly increased student performance relative to a traditional non-leveled assessment ($N=12$). Using the Wilcoxon signed rank test, the leveled assessment treatment showed a large but non-statistically significant positive effect on student test performance ($Z=1.69$, $p=0.09$, $r=0.53$). The mean student score increased modestly from 78.3% to 80.5% and median score remained static across the assessment types at 80% (Figure 1).

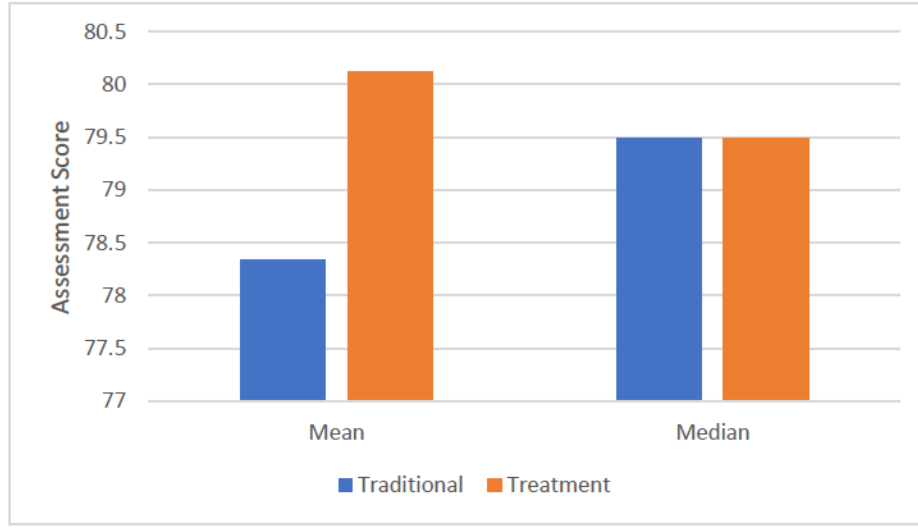


Figure 1. Comparison of student scores on the traditional test vs treatment, ($N=12$)

The treatment decreased the spread of student scores relative to a traditional test. The spread of the scores decreased on the leveled assessments compared to the traditional assessment ($\sigma_{trad} = 7.57$, $\sigma_{treat} = 5.78$) (Figure 2).

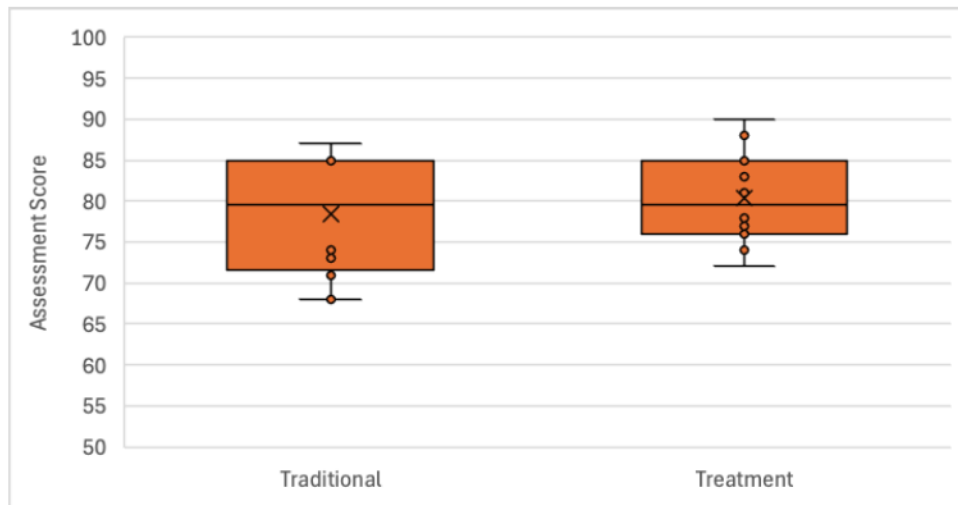


Figure 2. Distribution of scores on traditional test vs treatment, ($N=12$).

The ratio of the standard deviations was less than 4 indicating that the spreads were not statistically different ($\sigma_{treat}:\sigma_{trad} = 1.31$). Most students' scores increased with the leveled assessment compared to the traditional assessment (Figure 3).

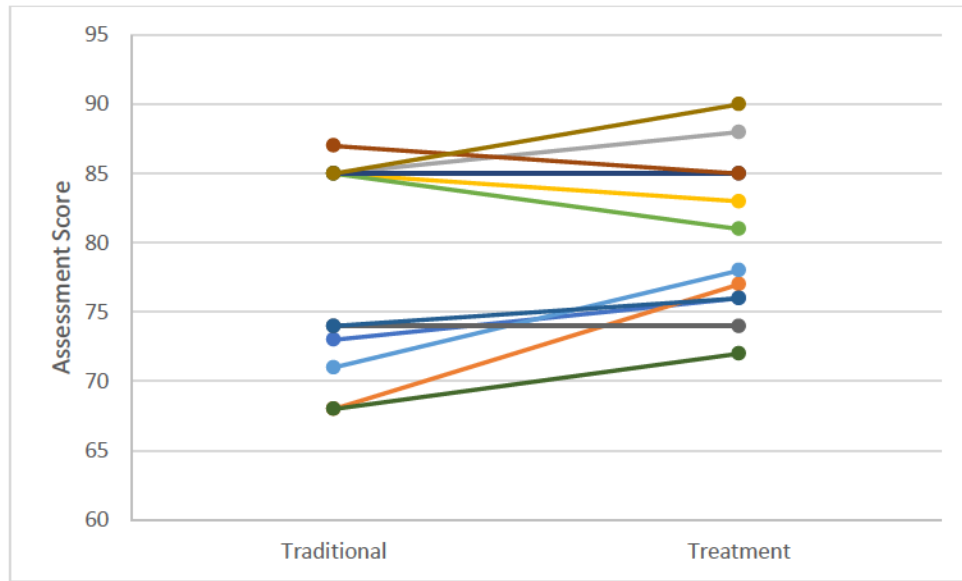


Figure 3. Performance change by student, ($N=12$).

Students that scored in the 1st and 2nd quartiles of the traditional assessment saw the largest score increases (Figure 4). Students in the 3rd quartile had more mixed results and the student in the 4th quartile had a decrease in score.

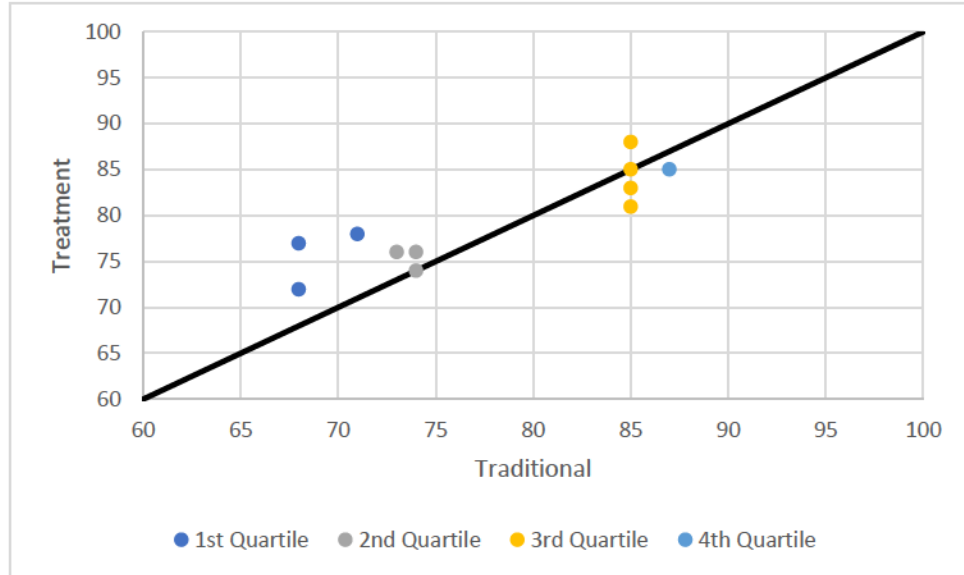


Figure 4. Score on the treatment vs traditional test, ($N=12$).

Student tests were substantially more complete with only three blank answers among all students on the leveled assessment as noted in the teacher journal.

Student Confidence

Student confidence showed small positive shifts in the Student Survey data. Across nearly all items on the Likert-type survey, responses shifted towards higher confidence after the leveled assessment. The sample size was small ($n=8$) due to absences and school activity interruptions during the data collection period which complicates data interpretation.

Additionally, the magnitude of the shifts was small, often a net of one student moving in a positive direction. There was a slight increase in students generally feeling less nervous around test taking (Figure 5). In each of three survey questions related to “worry,” being stuck on a problem, or “feeling nervous” there was a shift towards feeling more confident. While on both the pre- and post-treatment surveys 50% of students said they either *disagreed* or *strongly*

disagreed with the statement “Before a science test, I can calm myself down if I start to worry,” after the treatment no students *strongly disagreed* with the statement and an additional student *strongly agreed* with the statement as compared to the pre-treatment survey. The biggest gain was seen in the question “During a science test, I can stay focused even when I feel nervous” where in the pre-treatment survey 38% of students *disagreed* with the statement. Although no students *strongly agreed* with the statement in the post-treatment survey, only one student did not *agree* with it.

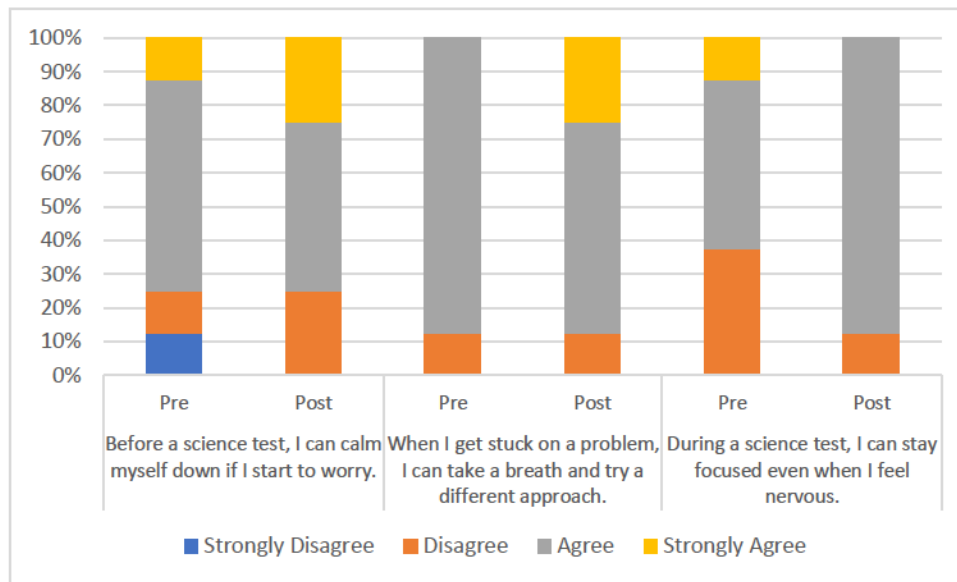


Figure 5. Pre- and post-treatment survey responses to questions related to test worry and anxiety. ($n=8$).

A similar slight positive effect was also observed in more general questions about how students have done in science in the past. Before the treatment, 50% of students *disagreed* with the statement that they had done well in science classes in the past but after the treatment only 25% *disagreed* (Figure 6).

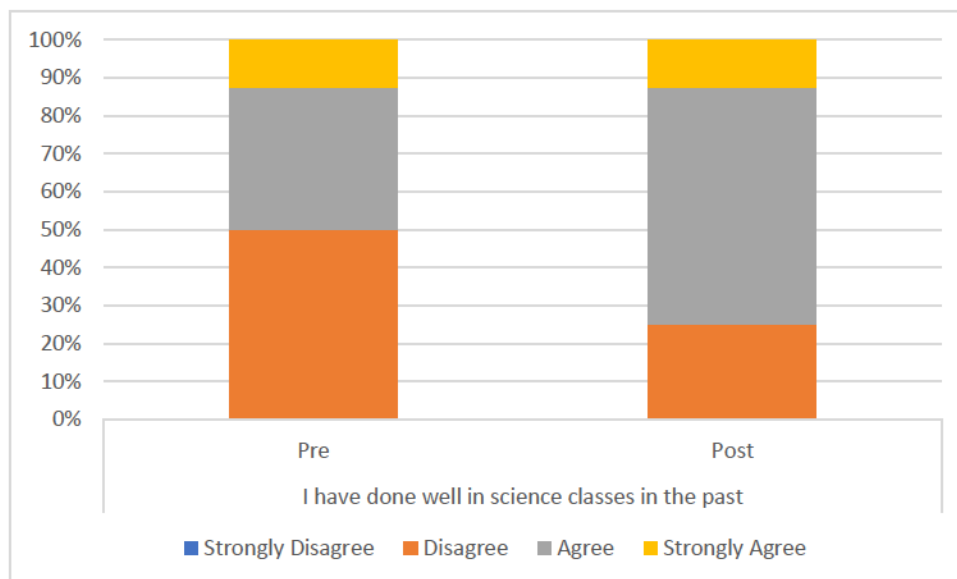


Figure 6. Student survey data for the statement “I have done well in science classes in the past,” (n=8).

More general questions about the students’ conceptions of themselves as science students also showed a decrease in students *strongly agreeing* with statements regarding their success in science. However, there were also decreases in the number of students who *disagreed* or *strongly disagreed* with the statements (Figure 7).

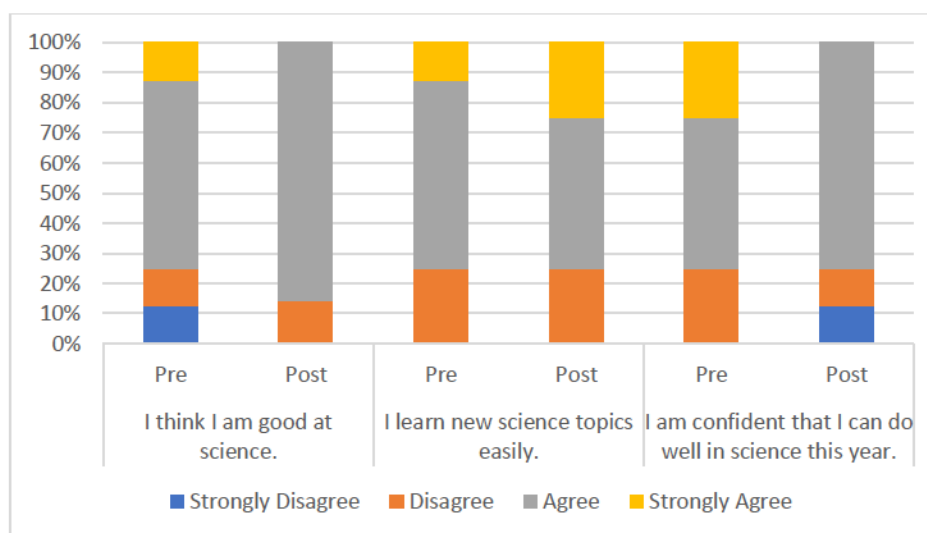


Figure 7. Student survey data regarding general beliefs about success in science, (n=8).

In student interviews, students ($n=11$), were positive about the leveled assessments. While no students directly delved into the idea of confidence, even when directly prompted, they spoke about the idea of agency and how it was nice to have a choice. One student said, "...it was nice to, like, have an easier level so that if you didn't understand it, you could understand the next one better" which was a sentiment echoed by several other students. Only one student mentioned feeling some embarrassment about asking to go down to a lower level. In general, all students said they would be okay with leveled assessments being used in the future. Many students also expressed a preference for the leveled assessment model

Perceived Fairness

Students expressed strong agreement with the fairness of leveled assessments. This was indicated both on the Student Survey and in student interviews. The student survey indicated a strong belief in the fairness of leveled assessment before the treatment with 87.5% *agreeing* and 12.5% *strongly agreeing* with both the statement "The leveled test model is fair for me" and the statement "The leveled test model is fair for all students" (Figure 8). One additional student *strongly agreed* with each statement in the post-treatment survey.

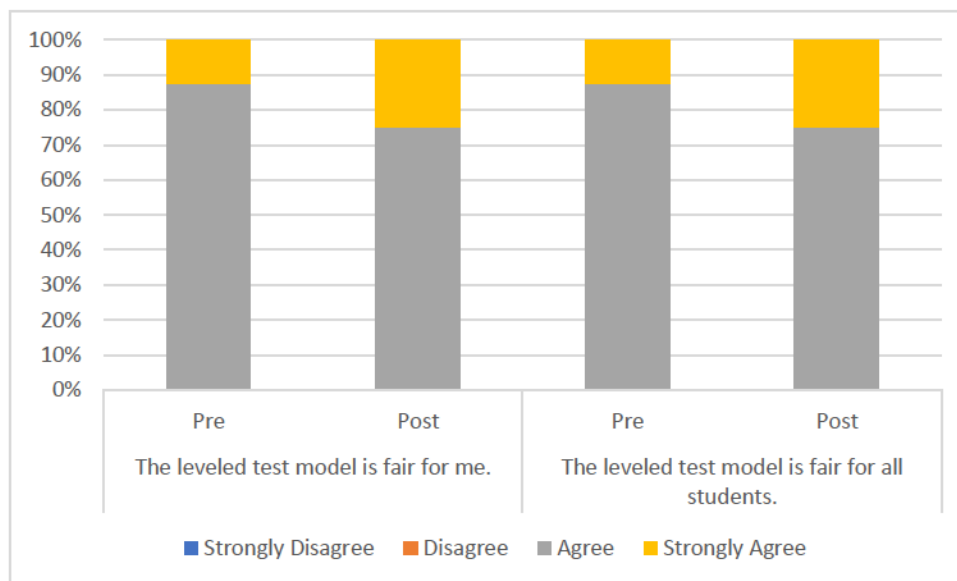


Figure 8. Student Survey data about perceived fairness of the leveled test model, ($n=8$).

The student interviews showed even stronger beliefs in the fairness of leveled assessments. One student seemed almost surprised that I would ask if the leveled assessments are fair. They quickly responded, “It’s literally based on you.” Another student said, “Biology is easier for some people, so then you would give them a harder test, you know? And then, it’s more of a struggle for other people. And then you would give them a test like that matches their skill level.” There was a strong belief that everyone had the opportunity to succeed and that the leveled assessment was “‘more fair’ than the normal one.”

CHAPTER FIVE

VALUE: REASONING AND REFLECTION

Claims From the Study

The goal of this action research was to determine if leveled assessments impact student performance and student confidence in biology class. The leveled assessment resulted in a modest increase in student performance although the results were not statistically significant. Results on student confidence were mixed and students strongly professed belief in the fairness of leveled assessments.

Student Performance

Leveled assessments may improve student performance, especially for lower-level learners. The mean score of the assessment increased and though the score increase was not significant, the effect size was large. Students in the first and second quartiles showed the most improvement and the number of blank responses decreased. Matching the tests to student learning, as demonstrated on formative assessments and in student opinion, allowed for more learners to adequately demonstrate their learning. Struggling learners were supported by leveled assessments the most. These leveled tests represent a meaningful step towards ensuring that assessments are “respectful work” (Richards & Omdal, 2007, p. 426) where each student has an equal opportunity to demonstrate their mastery of standards. Despite the lack of statistical significance, the large effect size and descriptive trends support continued exploration of leveled assessments, particularly regarding their impact on student confidence.

Student Confidence

Leveled assessments resulted in small positive impacts to student confidence especially with respect to students choosing their test level. The Likert survey indicated mainly small positive shifts in confidence while in interviews students emphasized choice and control as positives in the leveled assessment. Confidence appeared to be positively impacted by students being able to access the test questions which decreased students becoming stuck on a question. This positive effect also seemed to impact lower learners the most; these students benefited from being able to engage with appropriately challenging questions, reducing the likelihood of becoming stuck or leaving questions unanswered. In contrast, higher-performing students showed more mixed changes in confidence, indicating that the structure of the assessment may not have supported all learners equally. Overall, these findings indicate that leveled assessments support confidence primarily by increasing students' ability to engage with assessment tasks, rather than by fundamentally changing students' perceptions of their academic ability. This finding is consistent with Usher and Pajares (2008) finding that mastery experiences are challenging but achievable.

Perceptions of Fairness

Students overwhelmingly perceived leveled assessments to be as fair or more fair than traditional assessments. This belief remained consistent throughout the treatment period. Students emphasized that they were able to choose the level of their test which made the leveled assessment "more fair than the normal one." Students did not appear to mind variation in grades and seemed to understand that the different levels of the test reflected different levels in understanding. These findings are consistent with the psychometric perspective of fairness in

assessment. This perspective finds that fairness is “ensuring equity in the assessment process so that everyone receives the result they deserve according to the extent they have demonstrated their learning in the assessment process” (Rasooli, 2024).

In addition to the importance of assessments reaching students at an appropriate level, students seem to think about their ability to choose their test as an important factor in fairness. This emphasis on student choice is consistent with Self-Determination Theory which posits that providing meaningful choice supports students' need for autonomy, which is associated with greater internalization of learning and more positive educational experiences (Guay, 2022). For my students, that positive educational experience seems to have extended to seeing their assessment as very fair.

Value of the Study and Consideration for Future Research

Summative assessments are an important piece of information in the science classroom that inform instruction, provide feedback to students, and impact student motivations. Initial results from this study indicate that leveled assessments can have a positive impact on student achievement, particularly for generally lower performing students. I believe that by focusing student attention on demonstrating the knowledge they do have, leveled assessments can shift emphasis away from deficits and toward growth.

While this study was small and only included one treatment test, initial results indicate that leveled assessments have positive impacts on student performance. Students, particularly students who generally struggle, were more able to demonstrate the knowledge they do have. Additionally, I felt I was able to give students better feedback to point them in a positive learning direction.

I plan to implement this model in additional tests in biology this year and possibly extend it to additional courses next year. I would also like to explore digital leveling options to support students selecting a different level mid-test without worrying about the perception of other students. Additionally, it became clear that students need more information about what the performance standards are and what it means to earn a particular grade on a standard and on a test overall. Without this key step, students who are used to getting near perfect scores on assessments may lose buy-in on scores that indicate they have met but not fully exceeded state performance standards.

Impact of Action Research on the Author

This action research project has really changed my view of differentiation. In the past, I was guilty of considering differentiation to be a burden on teachers rather than a key to helping students succeed. The student emphasis on feeling the challenge at whatever level they were working really struck a chord with me. While I don't know how much these findings will impact inquiry work in my classroom, this project has made me more open to general differentiation within the classroom and a strong proponent of differentiation on assessments.

Students' beliefs about their ability to succeed in science can influence their engagement and willingness to continue studying STEM subjects. Leveled assessments allow me to maintain high academic expectations while providing students with opportunities to demonstrate learning at an appropriate level of challenge. Although this study did not find substantial increases in student confidence, interview responses suggested that some students felt the assessments were more accessible and better reflected their understanding. A future challenge for me will be to

support this growth mindset with reflection activities. Building this growth mindset can help encourage students to consider continued study in STEM.

I also think student concerns about how to get a perfect score or hit a desired benchmark deserve significant investment by individual teachers and by schools and districts. In my Advanced Placement courses, I have put a lot of emphasis on understanding the standards set by the College Board. Through creating, grading, and discussing these leveled assessments with students, I have realized that I need to communicate more clearly what different levels of progress toward the standards look like in my grade-level courses. While this work will benefit all students, it will particularly help mitigate the concerns of the higher performing students who had assessment scores that were more neutrally or negatively impacted by the leveled assessments. At Ouray School, most of these students will rise to high expectations when educators clearly communicate what excellence looks like. This will benefit the students' growth process and preparation for advanced courses, support positive communication with families, and promote a growth mindset.

Holding all students to high standards has been a hallmark of my philosophy as a teacher. This action research has opened my eyes to the power of differentiation but also reaffirmed my belief in strong consistent standards. I am excited for the possibilities for student reflection, building a growth focused classroom, and making my high standards more transparent that have arisen from this project.

REFERENCES CITED

- Alnahdi, G. H., Lindner, K.-T., & Schwab, S. (2022). Teachers' implementation of inclusive teaching practices as a potential predictor for students' perception of academic, social, and emotional inclusion. *Frontiers in Psychology*, 13, Article 917676. <https://doi.org/10.3389/fpsyg.2022.917676>
- ASCD. (n.d.). *Differentiated instruction: Teaching with student differences in mind – Module 1: Reading: Respectful differentiation*. Retrieved from https://pdo.ascd.org/LMSCourses/PD11OC138M/media/DI-Instruction_M1_Reading_Respectful_DI.pdf
- Black, P., Harrison, C., Hodgen, J., Marshall, B., & Serret, N. (2011). Can teachers' summative assessments produce dependable results and also enhance classroom learning? *Assessment in Education: Principles, Policy & Practice*, 18(4), 451–469. <https://doi-org.proxybz.lib.montana.edu/10.1080/0969594X.2011.557020>
- Britner, S. L., & Pajares, F. (2006). Sources of science self-efficacy beliefs of middle school students. *Journal of Research in Science Teaching*, 43(5), 485–499. <https://doi.org/10.1002/tea.20131>
- Brownlie, N., Burke, K., & van der Laan, L. (2024). Quality indicators of effective teacher-created summative assessment. *Quality Assurance in Education*, 32(1), 30–45. <https://doi.org/10.1108/QAE-04-2023-0062>
- Chen, J., & Brown, G. T. L. (2018). Chinese secondary school students' conceptions of assessment and achievement emotions: Endorsed purposes lead to positive and negative feelings. *Asia Pacific Journal of Education*, 38(1), 91–109. <https://doi.org/10.1080/02188791.2018.1423951>
- Christie, H., & Morris, N. (2021). Assessment and emotion in higher education: the allure of blogging. *Research in Post-Compulsory Education*, 26(2), 148–163. <https://doi.org/10.1080/13596748.2021.1909922>
- Colorado Department of Education. (2022). CMAS Science (Grade 11) 2023 performance level descriptors. https://www.cde.state.co.us/assessment/cmas_hs_sci_plds
- Colorado Department of Education. (2025). *Spring 2025 CMAS Data and Results*. <https://ed.cde.state.co.us/assessment/cmas-dataandresults-2025>
- Crossman, J. (2007). The role of relationships and emotions in student perceptions of learning and assessment. *Higher Education Research & Development*, 26(3), 313–327. <https://doi.org/10.1080/07294360701494328>
- Gaudet, S., & Robert, D. (2018). *A journey through qualitative research: From design to reporting*. SAGE Publications.

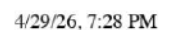
- Guay, F. (2022). Applying self-determination theory to education: Regulations types, psychological needs, and autonomy supporting behaviors. *Canadian Journal of School Psychology, 37*(1), 75–92. <https://doi.org/10.1177/08295735211055355>
- Hart, C. M. D., Linden, R., Jacob, B., & Loeb, S. (2023). Online course-taking and expansion of curricular options in high schools. *Peabody Journal of Education, 98*(1), 26-48. <https://doi.org/10.1080/0161956X.2023.2160108>
- Holstein, J. A., & Gubrium, J. F. (Eds.) (2003). Inside interviewing. SAGE Publications, Inc., <https://doi.org/10.4135/9781412984492>
- Hopwood, N., Palmer, T. A., Koh, G. A., Lai, M. Y., Dong, Y., Loch, S., & Yu, K. (2024). Understanding student emotions when completing assessment: technological, teacher and student perspectives. *International Journal of Research & Method in Education, 1*–16. <https://doi.org/10.1080/1743727X.2024.2358792>
- Inouye, M., Houseal, A. K., Gunshenan, C., McReynolds, A., & Perkins, M. (2023). Exploring collaborative professionalism as a means of virtually supporting rural teachers. *The Rural Educator, 44*(1), 14–27. <https://doi.org/10.55533/2643-9662.1313>
- Johnson, J., Ohlson, M. A., & Shope, S. (2018). Demographic changes in rural America and the implications for special education programming: A descriptive and comparative analysis. *Rural Special Education Quarterly, 37*(3), 140-149. <https://doi.org/10.1177/8756870518771381>
- Jones, E. A., Piontek, J., Walden, L. C., & Harrell-Williams, L. M. (2024). Development and validation of the sources of research self-efficacy scale. *Journal of Psychoeducational Assessment, 42*(1), 29–45. <https://doi.org/10.1177/07342829231204507>
- Kaur, A., Noman, M., & Awang-Hashim, R. (2018). Exploring and evaluating differentiated assessment practices of in-service teachers for components of differentiation. *Teaching Education, 30*(2), 160–176. <https://doi.org/10.1080/10476210.2018.1455084>
- Lee, M., Lin, T., & Tsai, C. (2013). Proving or improving science learning? Understanding high school students' conceptions of science assessment in Taiwan. *Science Education, 97*(2), 244-272.
- Lobmeier, J. H. (2010). Nonexperimental designs. In N. J. Salkind (Ed.), *Encyclopedia of research design* (Chapter 271). SAGE Publications. <https://doi.org/10.4135/9781412961288.n271>
- Moon, T. R. (2005). The role of assessment in differentiation. *Theory Into Practice, 44*(3), 226–233. https://doi.org/10.1207/s15430421tip4403_7

- Naeem, M., Ozuem, W., Howell, K., & Ranfagni, S. (2023). A step-by-step process of thematic analysis to develop a conceptual model in qualitative research. *International Journal of Qualitative Methods*, 22, 1–18. <https://doi.org/10.1177/16094069231205789>
- Noman, M., & Kaur, A. (2020). Differentiated assessment: A new paradigm in assessment practices for diverse learners. *International Journal of Education and Cognitive Sciences*, 1(3), 1–7.
- Ouray School District R-1. (n.d.). About us. <https://www.ourayschool.org/apps/pages/about>
- Paul, C., Webb, D. J., Chessey, M. K., & Lucas, J. (2018). Pondering zeros: Analysis of a decade of blanks and missed quizzes. *arXiv*. <https://doi.org/10.48550/arXiv.1807.09711>
- Pett, M. (2016). *Nonparametric statistics for health care research*. SAGE Publications, Inc, <https://doi.org/10.4135/9781071878705>
- Pozas, M., Letzel-Alt, V., & Schwab, S. (2023). The effects of differentiated instruction on teachers' stress and job satisfaction. *Teaching and Teacher Education*, 122, 103962. <https://doi.org/10.1016/j.tate.2022.103962>
- Pozas, M., Letzel, V., Lindner, K.-T., & Schwab, S. (2021). DI (Differentiated Instruction) does matter! The effects of DI on secondary school students' well-being, social inclusion and academic self-concept. *Frontiers in Education*, 6, Article 729027. <https://doi.org/10.3389/feduc.2021.729027>
- Richards, M. R. E., & Omdal, S. N. (2007). Effects of tiered instruction on academic performance in a secondary science course. *Journal of Advanced Academics*, 18(3), 424–453. <https://doi.org/10.4219/jaa-2007-499>
- Roy, A., Guay, F., & Valois, P. (2015). The big-fish–little-pond effect on academic self-concept: The moderating role of differentiated instruction and individual achievement. *Learning and Individual Differences*, 42, 110–116. <https://doi.org/10.1016/j.lindif.2015.07.009>
- Ryan, T., & Henderson, M. (2018). Feeling feedback: Students' emotional responses to educator feedback. *Assessment & Evaluation in Higher Education*, 43(6), 880–892. <https://doi.org/10.1080/02602938.2017.1416456>
- Smithson, J., (2019). Group Interviews, In P. Atkinson, S. Delamont, A. Cernat, J.W. Sakshaug, & R.A. Williams (Eds.), SAGE Research Methods Foundations. <https://doi.org/10.4135/9781526421036750847>

- Spector, P. E. (1981). One-group designs. In P. E. Spector, *Research methods in industrial and organizational psychology* (pp. 29–39). SAGE Publications.
<https://doi.org/10.4135/9781412985673.n4>
- Stambaugh, T., & Wood, S. M. (2015). *Serving gifted students in rural settings*. Taylor & Francis.
- Stefanakis, E. H. (2010). *Differentiated assessment: How to assess the learning potential of every student*. Jossey-Bass.
- Stiggins, R. (2007). Assessment through the student's eyes. *Educational Leadership*, 64(8), 22–26.
- Talsma, K., Schütz, B., Schwarzer, R., & Norris, K. (2018). I believe, therefore I achieve (and vice versa): A meta-analytic cross-lagged panel analysis of self-efficacy and academic performance. *Learning and Individual Differences*, 61, 136–150.
<https://doi.org/10.1016/j.lindif.2017.11.015>
- Teachers College (n.d.). Columbia University, Office of the Institutional Review Board Research instrument examples [PDF]. Retrieved from
https://www.tc.columbia.edu/media/administration/institutional-review-board-/guide-and-resources---documents/Published_Study-Material-Examples.pdf
- To, J., Gutterman, J., & Lipnevich, A. A. (2023). Students' emotions in feedback engagement. In A. A. Lipnevich, J. To, & K. H. K. Tan (Eds.), *Unpacking students' engagement with feedback: Pedagogy and partnership in practice* (pp. 26-40). Routledge.
- Tomlinson, C. A. (2017). *How to differentiate instruction in academically diverse classrooms* (3rd ed.). ASCD.
- Usher, E. L., & Pajares, F. (2008). Sources of Self-Efficacy in School: Critical Review of the Literature and Future Directions. *Review of Educational Research*, 78(4), 751-796.
<https://doi.org/10.3102/0034654308321456>
- van Geel, M., Keuning, T., Frèrejean, J., Dolmans, D., van Merriënboer, J., & Visscher, A. J. (2018). Capturing the complexity of differentiated instruction. *School Effectiveness and School Improvement*, 30(1), 51–67. <https://doi.org/10.1080/09243453.2018.1539013>
- Zolkoski, S. M., Aguilera, S. E., West, E. M., Miller, G. J., Holm, J. M., Sass, S., & Stokes, E. L. (2021). Teacher perceptions of skills, knowledge, and resources needed to promote social and emotional learning in rural classrooms. *The Rural Educator*, 41(3), 1-11.

APPENDICES

APPENDIX A: IRB APPROVAL



APPENDIX B: STUDENT SURVEY INSTRUMENT

Instructions:

Students responded to each statement using a 4-point Likert scale administered via Google forms.

1 = Strongly Disagree 2 = Disagree 4 = Agree 5 = Strongly Agree

1. I have done well in science classes in the past.
2. I usually do well on science tests.
3. When I perform well on a science task, I feel confident about my abilities.
4. When I find out a science test is coming, I feel tense. *(R)*
5. Before a science test, I can calm myself down if I start to worry.
6. During a science test, I can stay focused even when I feel nervous.
7. When I get stuck on a problem, I can take a breath and try a different approach.
8. After a test, I read my teacher's feedback.
9. Reviewing a graded test helps my confidence for the next one.
10. I think I am good at science.
11. I learn new science topics easily.
12. I am confident that I can do well in science this year.
13. The leveled test model is fair for me.
14. The leveled test model is fair for all students.

Open-ended questions:

15. Tell me more about the fairness of the leveled assessment:
16. Did you leave a question blank or answer "IDK" on your most recent test? If so, why?

(R) = reverse-coded item

APPENDIX C: GROUP INTERVIEW GUIDE

Group Interview Guide – Leveled Assessments in Biology

Introduction Script (Read Aloud)

Thank you for being here today. I'm interested in learning about your experiences with leveled assessments in biology. There are no right or wrong answers—I'm just trying to understand your honest perspectives. What you share will remain confidential and will not affect your grade in any way.

Group Norms

- One person speaks at a time
- Feel free to disagree—just be respectful
- Everyone's experience is valid—speak from your own perspective
- You may pass on any question

Interview Questions

1. How would you describe a leveled assessment to someone not in our class?
2. Do you feel leveled assessments help you show what you know in biology? Why or why not?
 - Can you think of a specific example from a recent biology exam?
 - How did you feel about choosing your level?
3. Have leveled assessments affected your grades in biology? In what way?
 - Do your scores reflect your understanding of biology concepts?
 - Have your scores improved, declined, or stayed the same?
4. Did leveled assessments change how you viewed or used feedback?
 - Was the feedback more or less useful?
 - Did the feedback feel any different?
5. Do leveled assessments help you understand biology better—not just get a grade?
 - Do they help you learn topics like natural selection, cellular processes, or photosynthesis more deeply?
 - What would make them more helpful?
6. Do you think leveled assessments are fair for all biology students? Why or why not?
 - Do students at different levels get equal chances to succeed?

- Is the grading system fair?
7. How would you feel about having leveled assessments in your science class next year?