

Northwestern College, Iowa

**NWCommons**

---

Master's Theses & Capstone Projects

Education

---

Spring 2023

## **Standards-Referenced Grading Assessments: Frequency and Student Preference**

Courtney Waterbury

Follow this and additional works at: [https://nwcommons.nwciowa.edu/education\\_masters](https://nwcommons.nwciowa.edu/education_masters)



Part of the [Education Commons](#)

---

**Standards-Referenced Grading Assessments: Frequency and Student Preference**

Courtney Waterbury

Capstone Project: An Action Research Project

Northwestern College, Orange City, Iowa

Dr. Theresa Pedersen

April 30, 2023

### **Abstract**

The purpose of this action research was to investigate whether a noticeable difference in student performance could be identified when using end of unit assessments versus smaller frequent assessments after a lesson. The researcher conducted this study in a high school general education classroom with five sections of an American History course. The action research project took the scores from three standards and compared the results of all students who took the assessment after a lesson for one unit as well as at the end of a unit for the other unit. The findings showed a trend of students performing better in frequent assessments however the results were insignificant. The research was conducted in order to guide the researcher in future assessment practices as they work to improve student learning.

*Keywords: assessment, frequency, standards-referenced, achievement,*

**Table of Contents**

Abstract.....	2
Introduction.....	4
Literature Review.....	7
Methodology.....	16
Participants and Research Site.....	16
Research Questions.....	17
Intervention & Timeline.....	17
Variables.....	18
Measurement Tool(s).....	19
Anticipated Statistical Analysis.....	19
IRB.....	20
Data Collection.....	16
Data Analysis.....	21
Discussion	
Findings.....	27
Alignment to Research.....	27
Limitations.....	27
Impact on Teaching and/or Learning.....	28
Future Research.....	28
Conclusion.....	29
References.....	31

### **Standards-Referenced Grading Assessments: Frequency and Student Preference**

Educators assess student work in order to identify progress and there are different ways in which progress can be measured (Blevins-Knabe, 2019). Assessment should not merely be done to students; rather, it should also be done for students, to guide and enhance their learning (National Council of Teachers of Mathematics, 2000). The most familiar assessments are formative and summative assessments. Benefits of standards-based summative assessment are that it allows for comparing learner performances across diverse populations on clearly defined educational objectives and standards; it provides reliable data that can be used for accountability purposes for various stakeholders, and it can inform educational policy. (Shute, 2017). Further, Shute and Rahimi stated, “Formative assessment is intended to support teaching and learning. It is incorporated directly into the classroom curriculum and uses results from learners’ activities as the basis on which to adjust instruction to promote learning in a timely manner” (p. 2).

Assessment is a major part of education to provide feedback on student learning although the methods of assessing vary based on subject, grade level, and teacher preference among other variables. Assessment tends to focus on assessment of and for learning rather than having a stronger emphasis on assessment for learning (Shute, 2017). Assessments are used to communicate with students their learning which can become challenging without consistency and common language (Blevins-Knabe, 2019). Blevins-Knabe goes on to explain, “Faculty who conducts assessment do not need to assess identical sets of student assignments or other evaluative materials. In actuality, assessing general education using a variety of methods may yield more consistent findings” (p. 32). These sources are highlighting the importance of assessment while also pointing out the challenges educators face while having to sort through

and identify the best form of assessment for their classroom. Assessments provide educators an opportunity to evaluate student learning and it is important for educators to seek the best form of assessment to provide feedback to their students. The problem is the variety of options available when choosing how to assess students in a summative form. By focusing on the best method to assess students, educators can provide quality feedback on student learning connected to the standards. Educators are committed to helping students find success in the classroom and assessing students in a meaningful way. The demand for more information to guide educators in making a choice regarding forms of assessment is a necessary step forward.

The two main categories of assessment are formative and summative, however within these two categories are a variety of methods to distribute an assessment. Wehlburg highlights challenges stating, “Unfortunately, assessment as a whole has been used as a tool for accountability and accreditation. This has been especially true for general education assessment” (p. 90). Students take assessments as a necessary step in their education progress. Educators use these assessments to understand student learning and for students to showcase what they have learned. By creating assessments that are intended to support student learning as best as possible, assessments can be tailored for specific situations to meet the needs of the majority of students. The purpose of assessments is for educators to provide feedback to students, for students to showcase their learning, and for progress to be checked. The purpose of this action research is to find an effective method for assessing students based on standards in a secondary social studies classroom.

The research gathered for this action research project were found using the DeWitt Library at Northwestern College. A variety of sources were gathered covering topics on

assessment. These topics related to assessment and student achievement looked into demographics, student mentality, variation in assessment, classroom strategy, standards alignment, issues related to assessment, and innovation for assessment. These studies show the variation that exists within assessment as it is such a large topic. Focusing specifically on the social studies classroom as there is a gap in research in that area specifically but other subject areas can be used to provide a reference point.

In order to move forward with the action research an analysis of the literature will be done to gain a deeper understanding of what the research indicates for assessments in the standards-referenced classroom. These sources will look at assessment in a variety of ways and review what has already been learned throughout other studies. Starting with a breakdown of how demographics influence student achievement on assessments. Then sources will show how student mindset influences their perceived success. The next portion will dive into how variations in testing can impact student achievement on assessments. Finally, analyzing how classroom structures and strategies can influence students' assessments. The variety of sources which will be presented are intended to show importance of assessment and the variety of areas that can be looked into further in regards to student achievement. The sources also show the areas for further research regarding social studies assessments and the ways in which standards can be assessed to better student learning.

## Review of the Literature

### Demographics Influence on Assessment

Changes in assessment practices are often met with resistance. As schools shift towards standards-referenced grading there are misconceptions that make the transition challenging. According to Peters et al. (2017), schools transitioning must deal with prior beliefs, perceptions, and practices with regard to grading, which can result in a less successful implementation and lack of compliance. Peters showed concerns regarding the fairness of the grading system stating, “fairness should be a transcendent value, the basis on which all educational decisions are made. He questioned the fairness of any system that subjects students to mysterious, changing expectations” (p. 20). These challenges with assessing were also echoed by Vogler et al. (2020), stating the flexibility granted to educators generates inconsistencies and potential challenges to different learners. According to the article, *The Effect of Learning Environment, Inquiry and Student Learning Interest on Student Social Studies Learning Assessment*, students’ assessments were impacted by their learning environment in a social studies classroom (p. 4). In the academic journal *European Alliance for Innovation (EAI)*, Arthur et al. (2022) studied performance in mathematics which showed students interests and classroom environment had an impact on student performance.

When researching the impact of accountability tests results and variables that influence the results Vogler et al. (2020) found, “In addition to poverty, the impact of gender and race/ethnicity on student achievement has been well documented in the literature” (p. 207). This statement was supported by the data collected by Vogler et al. (2020), as the Levene’s test was used to account for assumptions and variables while looking to understand the relationship between instructional time and gender, ethnicity, and poverty. Peters et al. (2017) highlighted



variations in grading and how it influences students reaching their potential. Peters et al. (2017) goes on to emphasize the “Is the purpose of assessment and grading to select talent or develop it? If it is indeed to develop it, then it entails a different approach to our work” (p. 21). Peters shows assessments can have a negative impact on students’ mindset in regards to reaching their potential so educators can show care when using assessments in order to allow all students to see how they are capable. In contrast, Rosfaini et al. (2018) found that the learning environment carries more weight in regards to student achievement. Rosfaini et al. (2018) stated, “A positive learning environment significantly correlated with student academic progress, therefore, recommendations were made for the improvement of the learning environment” (p. 1). Thinking more outside the box, Arthur et al. (2020) identifies factors that positively impact student achievement including understanding the history of a subject. Arthur states, “When the history of mathematics is adopted in teaching and learning it will enhance student’s mathematics thinking ability which will result in better performance of the students in the classroom” (p. 190). Many factors combine to influence student achievement which work together to impact the classroom setting. Rosfaini and Peters both echo the importance of a positive learning environment to help students succeed. Arthur highlights the need for students to understand why they are learning about a specific topic and the history of that topic. Vogler’s research shows the correlation to factors outside of students control that influence their learning. Ultimately, there are a variety of factors that contribute to a learning environment for students; educators aim to support all students that enter their classroom, so recognizing the various elements that play into that success is helpful as educators build an environment for improvement.

### **Students Mentality Impacting Assessment**

Secondary schools across the United States are shifting towards standards-referenced grading. This transition focuses on student mastery of content skills on their own timeline. Lewis (2022) explained the goal of standards-referenced grading stating, “Eventual mastery matters; students are provided multiple opportunities to demonstrate mastery of each standard and are not penalized for failing to master a standard on a given attempt” (p. 68). Lewis (2022) highlights the negative impacts test anxiety has on college aged students. According to their findings nearly 40% of students have test anxiety and this is connected to how students perceive the level of difficulty on an exam. The integration of growth-mindset alongside standards-referenced grading has been shown to reduce anxiety among students leading to positive changes in students. Lewis stated, “Studies have shown that a growth mindset is associated with improved achievement; moreover, even short interventions can foster an increased growth-mindset” (p. 69). This information compiled by Lewis demonstrates the potential that standards-referenced grading and using a growth-mindset can have on student outcomes.

The research conducted by Rupp et al. (2006) echoes the benefits of standards-referenced grading on students’ mental health. Rupp discusses the goal of schools from all over of trying to achieve improved learning for all students. Standards-referenced grading has emerged as a way to accomplish this, “Standards-referenced grading is not without flaws, challenges arise in the consequences that can arise from grading as schools are held to specific standards” (p. 330). The benefits of standards-referenced grading come from the improved mental health of students but Rupp finds in his research concerns regarding the evaluation of learning that comes from standards-referenced grading. Rupp et al. (2006) stated, “Overall, this study shows that the proficiency classifications of the standards-based assessment reflect the diagnostic profiles on a select set of component skills of reading rather poorly” (p. 330). Rupp continues to discuss the

findings further placing an emphasis on needing to strive to find assessment practices which are beneficial for all students. Standards-referenced grading is a complex system to integrate into schools the researchers highlight the importance of taking care to ensure the system is benefiting students in more ways than one.

Analyzing the impacts of student motivation and success in the classroom, researcher Kuo, et al. (2021), was able to research if there was a connection between psychosocial skills and academic success at a young age to success as a secondary student. Through the study Kuo, et al. (2021) was able to determine, “Our findings show that students who have sufficient core academic skills and better psychosocial skills in middle school are more likely to succeed in high school” (p. 138). The psychosocial skills looked at throughout the study include social control, self-regulation, and motivation contributed to students finding success at the secondary level. The research noted middle school is a key transition period for students and the changes they experience in their education connect to future success. The findings made by Kuo, et al. (2021) can help educators support students, “It is hoped that the findings will provide educators a better understanding about the significance of PSFs in middle school to help students succeed in high school” (p. 151). The results of this study agree with previous findings which connect students being able to self-regulate to success.

Assessments are used across the education field to determine student learning, Hinett points out the inconsistencies in assessment within institutions and how that can lead to students struggling. Hinett, et al (1996) found, “Assessment for learning may promise better quality higher education. Unfortunately, it could also be seen as a dereliction of standards and standards are widely seen as what quality is about” (p. 8). The findings of Hinett agree with the findings of other researchers. Standards-referenced grading can be a strength to students and educators as

traditional assessment fails to assess high quality learning. The research by Hinett looks more specifically at what students expected to get out of their education, a different perspective than others have taken. The findings did show the importance of assessment and how finding value in feedback can have a larger impact on the student than the educator may have intended.

### **Assessment Variation in Relation to Student Achievement**

Assessments are used across the United States to measure students learning but there is a lack of consistency that Baker points out. Baker, et al. (2014), “Our national data shows essentially no improvement over a number of years, most recently reported in the 12th grade results (NAEP, 2014). Unlike countries with traditional reliance on tests and competitive performance, such as China, South Korea, and France, in the United States our distinct political divisions result in crosscutting tensions” (p. 2). Baker points to the inconsistencies that is holding back the United States on a global stage. Variation studied by Baker at the national level led to negative results when comparing the United States to other nations, this leads to a response to try and pull the United States back to the top in education. “Our nation has responded to this general problem by demanding more of schooling, by increasing its mechanisms of accountability, and by depending more and more on the data from standardized tests of academic achievement” (p. 22).

Kantor, et al. (2011) looked into testing at the elementary level finding different forms of assessment were not creating differences in results for students, “The results did not support the idea that reliability would be better for DA [dynamic assessment] compared to standard task administration” (p. 319). Breaking down data showed the researcher there was not a significant difference in results for different assessment which is in contrast with Baker (2014) which found

differences in assessment reflect the scores of students across the country. The studies have variations in the age of students who were studied.

Assessment variation can take on a variety of forms, Elmahdi, et al. (2018) looked into the impacts of technology on assessment to determine if technology impacts student engagement as well as student assessment results. The findings of this study showed a positive correlation between the use of technology and student results on assessment. Elmahdi, et al. (2018) described the positive impact using a technology resource called Plickers supported student learning and overall outcomes. These findings contradict with Kantor who found variation in assessment did not necessarily improve results. These studies differ with one another in regards to the ages sampled. Elmahdi, et al. (2018) described the positive effects of using Plickers as, “it is found that students’ engagement is improved, when the teacher uses Plickers for formative assessment which leads to creating an effective learning environment that promotes learning” (p. 187). This research shows the positive effects different forms of assessment can have on student learning, specifically the use of technology in the classroom.

Mastery based assessments are giving students credit for the skill once they have shown mastery, this enables them to move on once they show knowledge of a skill. In Harsy, et al. (2020) the results of this study showed, “We found that mastery-based testing led to students feeling the assessments better reflected their content knowledge as well as higher end-of-semester course grades with fewer hours spent studying outside of class” (p. 1). This study was conducted on students in the secondary classroom for calculus II students. The study showed success in students’ overall grades on assessments as well as end of semester grades, the feedback Harsy (2020) received showed students felt more confident in the content as well as and it encouraged them to revisit topics they had struggled with.

Assessment variations exist in a variety of ways. Harsy (2020) showed students assessing using a mastery-based test was positive in their overall learning, and Elmahdi (2018) showed using technology in the classroom correlated to better test results. Kantor (2011) however showed challenges at the lower level with students struggling to identify a difference in learning by using differentiated assessments. Baker (2014) was looking at the United States in comparison to other countries around the world and highlighted the weakness the United States has because of inconsistencies in education with some areas finding more success than others.

### **Classroom Strategy Influenced by Assessment**

Classroom setup impacts student learning and therefore assessment. Dennis (2010) is able to identify learning variation benefits student learning in the elementary classroom. According to the findings of this study, Dennis (2010) found “Certainly, what was learned through this research is that struggling young adolescents demonstrate complex, heterogeneous reading abilities requiring significantly different instructional interventions” (p. 290). Recognizing learners are different, Dennis highlights the importance of varying classroom instruction in order to meet the needs of students. The study by Dennis (2010) specifically looked at students who were identified as struggling readers by a state assessment. Through this research Dennis identified different strategies which were effective for students but most importantly the fact that some strategies work for certain students while other strategies benefit others, it is not a one size fits all approach.

Looking into standardized testing and how that should influence academics, De Lisle (2008) recognized the differences that exist throughout school districts. De Lisle highlighted the need for teachers to be trained on how to use the data the state collects. De Lisle (2008) stated, “Empowering teachers to use data better requires building assessment and data literacy” (p. 107).

Finding that teachers struggled to use the data collected is in contrast with Dennis (2010) who found teachers are using different strategies for students as they should but the larger issue is within classrooms differentiation is constantly needed.

Embracing a conclusion on the balance between assessment and classroom procedure, Shapiro, et al. (1992) described the balance that must be struck between assessment and curriculum. “Assessment, however, must be viewed as only part of a total process that includes implementing interventions derived from that assessment as well as measuring their outcomes” (p. 294). Having classrooms with students of different levels poses challenges to educators who must work to meet the needs of all students. Similar to Dennis (2010), Shapiro emphasizes teachers be constantly checking to ensure learners are understanding the material through formative checks. Shapiro’s research setting included classrooms with students who have special needs to determine assessment strategies which were effective and found using a problem-solving approach was successful.

Standards-referenced grading implementation is not simple as Knight acknowledged; however, it leads to improved teaching practices and a variety of other positive changes according to the research conducted by Knight, et al. (2019). Students Knight, et al. (2019) found standards-based grading positively impacted teachers in a variety of ways extending past assessment. “High school teachers in this study who adopted SBG practices believed it made their planning, instruction, assessment, and environment more purposeful and successful” (p. 89). These findings support the implementation of standards-based grading, Knight does acknowledge challenges teachers may face while transitioning, but overall, the evidence presented showed the new grading method would support student learning as they adjusted to the transition as well.

## **Conclusion**

Research shows the demographics of a classroom has an influence on student performance as adjustments need to be made to meet the needs of the students who are present (Arthur, et al. 2020; Vogler, et al. 2022; Peters et al. 2017; Rosfaini et al. 2018). Students' mindset in the classroom is a major factor in their success, recognizing their perception of an assessment paired with growth-mindset thinking leads to improved scores (Lewis, 2022; Rupp, et al. 2006; Hinett, et al. 1996; Kuo, et al. 2021). Ultimately variation in assessments creates challenges for educators as they adapt to their students' needs, but finding strategies which are effective is essential to measuring students learning (Elmahdi, et al. 2018; Harsy, et al. 2020; Kantor, 2011; Baker, 2014). According to studies conducted on classroom variation, students thrive in a variety of academic settings so teachers being able to utilize assessment in order to modify their teaching is important to help students succeed (Knight, et al. 2019; Shapiro, et al. 1992; Dennis, 2010; De Lisle, 2008). Assessments are a major part of classrooms around the world and striving to create assessments which foster student development, enable students to have a growth-mindset, show improvements in learning, and adapting to the needs of a classroom are all areas where educators continue to learn and grow.



## **Methods**

### **Participants**

This action research project was conducted in a 10<sup>th</sup> grade social studies classroom. This is a public school district is located in a rural eastern Iowa. There are five sections of American History for a total of 98 students. Students range in age from 15 to 16. Of the 98 students there are 57 boys and 41 girls. There are eight students who are identified and have an Individualized Education Plan (IEP). The Iowa Department of Education reported these students are 84.1% proficient according to the 2021-2022 ISASP English Language Arts assessment. The demographics of this district are 90.3% white according to the Iowa Department of Education public school enrollment data for the 2022-2023 school year.

The setting of this action research project is in a general education classroom. Students are one to one with apple laptops. Students have information available to them through the school's use of canvas. This provides students access to the lecture slideshows, recordings of the lecture, and study resources. The sample selection includes all students on the roster for the 2023 spring semester who agreed to participate in the study, 98 students' total. The researcher is the social studies teacher for all students in this study.

### **Data Collection**

This action research project is intended to learn about the impact of various testing methods on student performance as well as preference. The research is looking to ideally find what has a positive impact for both student performance and student preference. This information can be used to guide future units and be shared with fellow teachers to provide additional options for how they may want to set up tests. Students prior to this study have limited

input in how the class is tested. The following questions are the focus of this action research study:

Do students prefer traditional end of unit assessments or more frequent smaller assessments?

Do students perform better in end of unit assessments, more frequent assessments, or is there no change in performance?

The researcher collected data from three standards which include cause and effect, economics, and primary source analysis. These standards were assessed two times over the course of the study and are based on the Iowa Core Social Studies Standards from 2017. The first assessment of each standard was in unit seven where students assessed after the corresponding lesson, so assessments were one page long. The second assessment was given in an end of unit format so the three standards were assessed on the same test. The formats were similar, the only adjustments were made based on the content. These assessments were given over an eight-week period, the first three assessments were spaced out roughly one every three class periods, the researcher follows a day one and day two block schedule. The second form of assessment at the end of the unit was given at the end of a five-week unit, spring break was included during this time. Timeliness of feedback is recognized as important and all assessments were returned to students within two class periods. The research was collected over an eight-week period total.

The researcher collected data from the standards by administering two assessments per standard. For each assessment the researcher provides participants with access to a study guide, a recording of the lectures, and the slide show. This information is all distributed through canvas, a reminder announcement is also sent to students through their email two days before an assessment. On the day of an assessment students have the first ten minutes of the 85-minute block to ask questions and discuss material. Students then turn in their phones to the “parking

lot” on the teacher’s desk. During the test students will be silent until all students have completed the assessment. The researcher then grades the assessments using the rubric and standard scale, grades are given back to students within two class periods. Students have an opportunity to reassess if they choose within two weeks of receiving grades. Data is then organized by the researcher by class period and the grade received, there are six potential grading groups a test could be placed in. This is intended to show the researcher the overall trend of scores as well as a more specific class by class breakdown.

After students had received grades for their assessments, students were then prompted to review their grades and then complete the survey. This survey given during class time took approximately five minutes for students to complete. Data for the survey was collected through canvas, an online communication tool used by the researcher’s district. The researcher explained to students prior to its completion the information provided was to be used for an action research project and the results of what students shared would be then used to determine how to assess in the future potentially. The data from the survey was collected using a Likert scale to show how students felt about each type of test, they could potentially like or dislike both testing styles.

The independent variable is the format and timing of the assessment. This is true because the research is able to determine if the assessment will be given following a lesson that is connected or if the assessment will be administered to students at the completion of the unit. The format is chosen based on the standards aligned in the gradebook which the research has predetermined according to the content and after breaking down standards with their PLC group, instructional coach, and district curriculum director. The research has options based on the standard for what formats will best fit the content, so the format is selected by the researcher.

The dependent variables are the students' scores and opinions on the tests and survey administered. Following the completion of an assessment, scores will be given to students based on the information they are able to demonstrate. Grades will be based on a rubric for consistency between students as well as between units. The survey information will be collected from canvas and students will be giving feedback on a Likert scale as well as having an opportunity to provide open ended feedback for the researcher.

In order to maintain reliability in the grading the standards, the researcher uses rubrics to have maintain similar expectations. The researcher also uses the same scale in order to give students a grade, so the same amount of information was needed to give students a meeting. The scale used by the researcher's district is standards-referenced. In order for students to receive an exceeding, 100%. they need to answer an additional question which is optional. If they do not answer the exceeding question, the best possible grade they can receive is a meeting, 90%. Students have an option to reassess within two weeks of the grade being entered on infinite campus, for the sake of consistency the data reflects students having the opportunity to reassess rather than all students first attempts. No student is required to reassess, they go through the process if they so choose. The Likert scale was used in the survey to obtain student feedback in addition to an open-ended response box where students could provide more specific feedback to the researcher if they choose.

In order to understand the data, the researcher is comparing the scores of students on each standard as well as an overall grade. The research is using this data to identify if there is a difference in student performance based on the timing of tests. The data collected from the survey is being used to gauge how students feel about the different formats of testing. If there is no significant difference in overall performance, the researcher plans to respect student

preference based on what the survey results show. If no significant difference is shown between test results and students are split on assessment preference, the researcher will continue to learn and collect data and take that as a sign to use both types of assessment in addition to continuing to look for more potential.

The researcher recognizes the importance of maintaining the privacy of participants in this study. The data was kept on the researcher's computer which is locked with a unique password and only those with permission were given access. Names were not associated with any data unless it was necessary. Paper versions of the test were stored in the researcher's locked cabinet with the key remaining in the researcher's possession so it could not be accessed by anyone who did not have permission.

An application for this action research project was submitted to the Northwestern College Institutional Review Board was approved and permission was granted to conduct this study. This research involved normal educational practices by comparing assessment strategies in addition to student feedback in regards to assessment preferences. The researcher recognizes how important it is to protect the confidentiality of student data, to obtain informed consent from students' guardians, and presenting the complete data to be interpreted accurately and presented.

## Findings

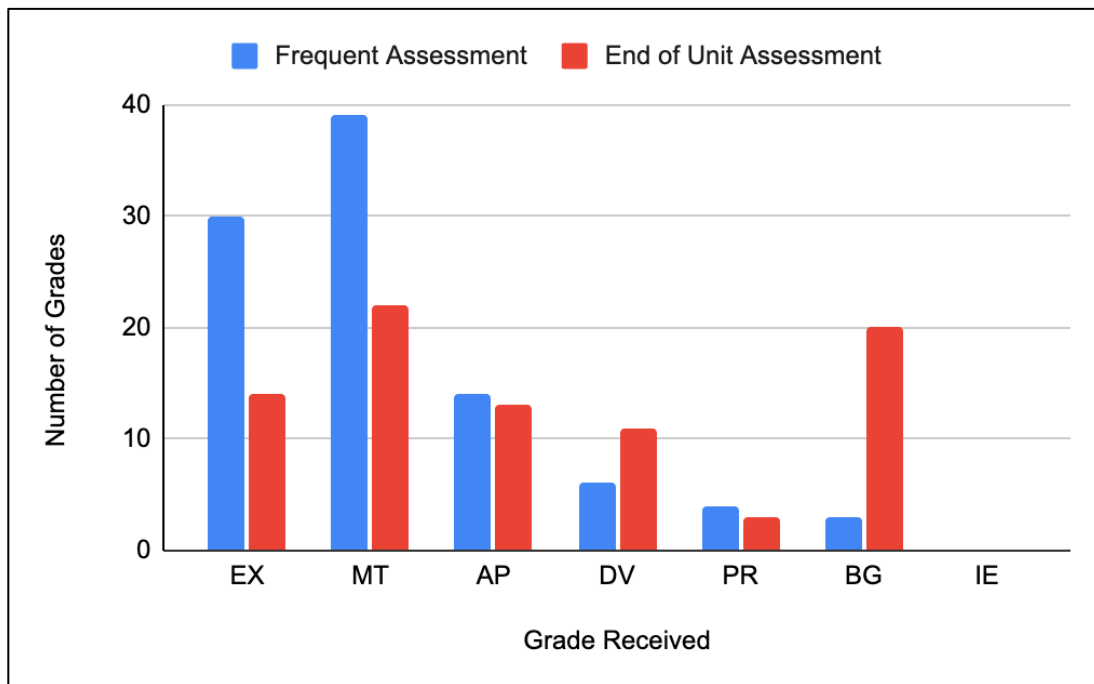
### Data Analysis

This action research project focused on student achievement when comparing assessments given at the end of the unit versus after a lesson. The quantitative data record the grade students received for three types of standards. Additional quantitative data was recorded based on the feedback and opinions of students which they shared personal opinions on the testing strategies. The grading scale used is standards referenced which means grades are rounded to fit into a grade based on the rubric. The grading scale is shown in Figure 1 for reference. Figure 2 shows the grades students received on the economics standard for each type of test. The researcher recorded the number of students who received each grade for both the frequent assessment and the end of unit assessment.

**Figure 1**

*Central-DeWitt Standards-Reference Grading Scale*

EX (100%)	Exceeding	The student has demonstrated an in-depth understanding of the learning goal beyond what was taught.
SE (95%)	Securing	
MT (90%)	Meeting	The student has met the learning goal.
AP (85%)	Approaching	
DV (75%)	Developing	The student has demonstrated the foundational skills and knowledge for the learning goal. (almost there)
PR (65%)	Progressing	
BG (60%)	Beginning	The student has <u>some</u> understanding of the learning goal.
IE (50%)	Insufficient Evidence	The student has attempted but has not demonstrated enough understanding to earn a beginning mark.

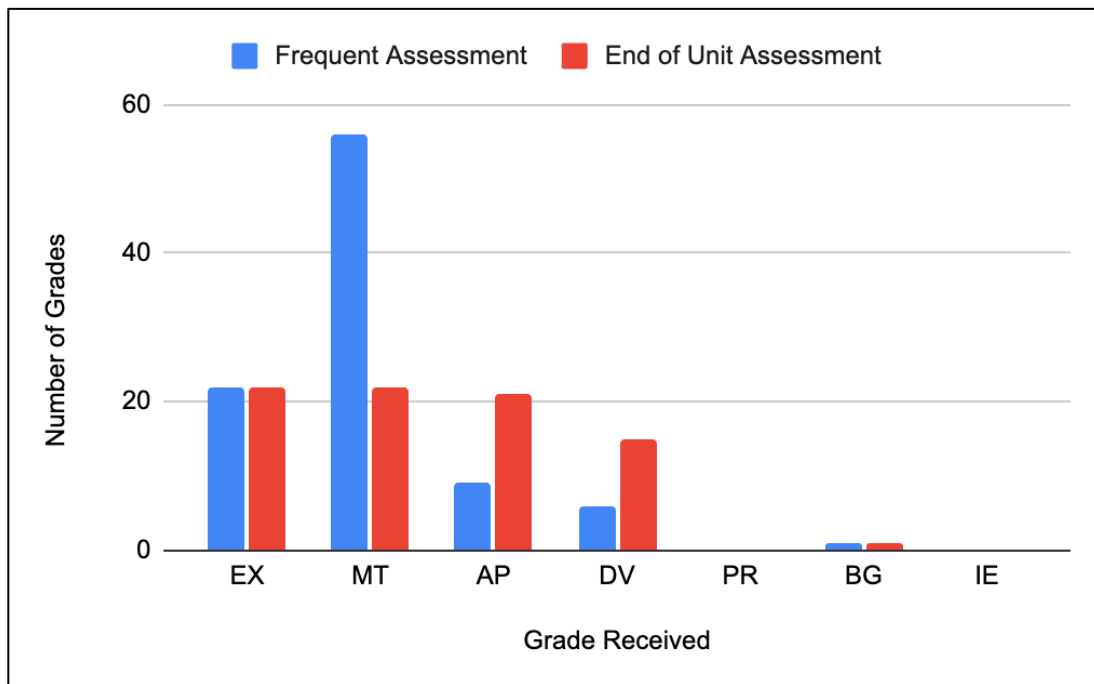
**Figure 2***Comparison of Economic Assessment Results*

A dependent sample t-test was conducted in order to identify if there was a significant difference between the assessment type and student results. In the frequent assessment the (M=89.47, SD=10.24) is compared to the end of unit assessment (M = 80, SD= 15.58) a two-tailed t-test revealed an insignificant difference  $t(96) = 2.21, p > .05$ .

Quantitative data was collected to compare the assessment results of frequent testing compared against end of unit assessment results for primary and secondary source standards. Figure 3 compares the results of the frequent assessment and end of unit assessment for the primary and secondary sources assessment. The research recorded the test scores for students in both types of tests. These scores are assigned based on the standards referenced grading chart in Figure 1. The researcher used a rubric to consistently assign grades for students work, each student took both forms of assessment.

**Figure 3**

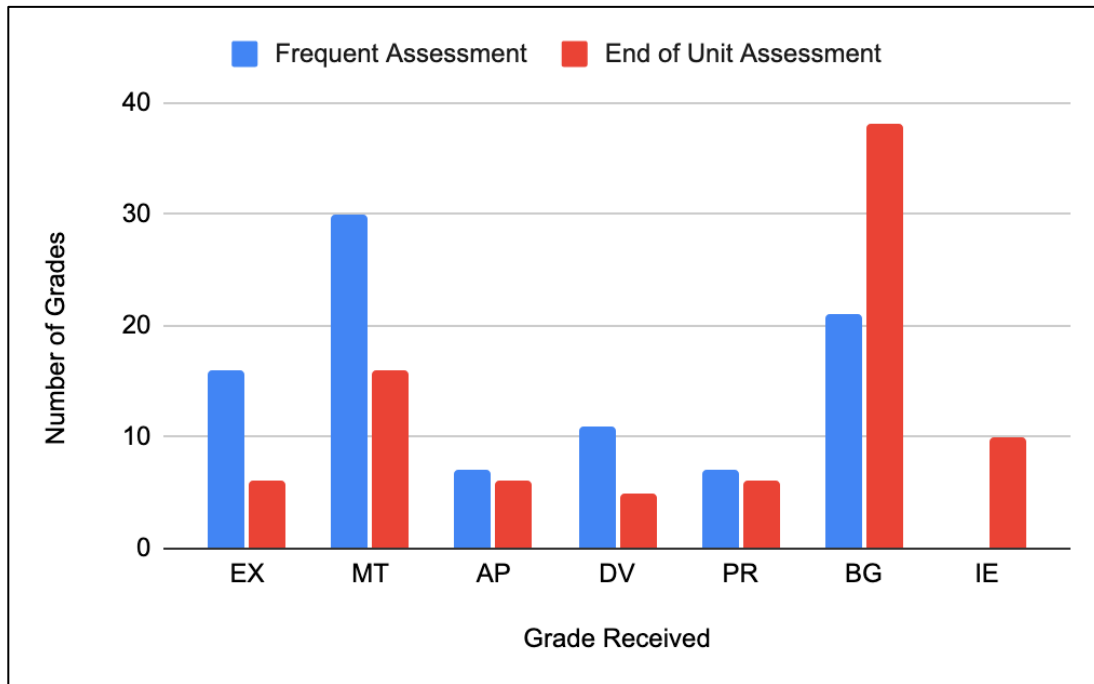
*Comparison of Primary and Secondary Sources Assessment Results*



The researcher used a dependent sample t-test in order to identify if there was a significant difference. For the primary and secondary sources standard the frequent assessment ( $M=90.58$ ,  $SD=7.06$ ) compared with the end of unit assessment ( $M=88.53$ ,  $SD=9.50$ ), showed an insignificant difference  $t(95) = 0.12$ ,  $p > .05$ .

Figure 4 compares the assessment results for the cause-and-effect standard between frequent assessment and end of unit assessment. Qualitative data was collected in order to compare the two types of assessment results. The researcher had students compete an assessment in both formats for the same standard. Students were graded based on a common rubric for consistency and then given a grade based on Central Dewitt's standards-referenced grading rubric. Grades were then placed in the corresponding category to be represented in the formulas as well as the graph for a visual perspective.



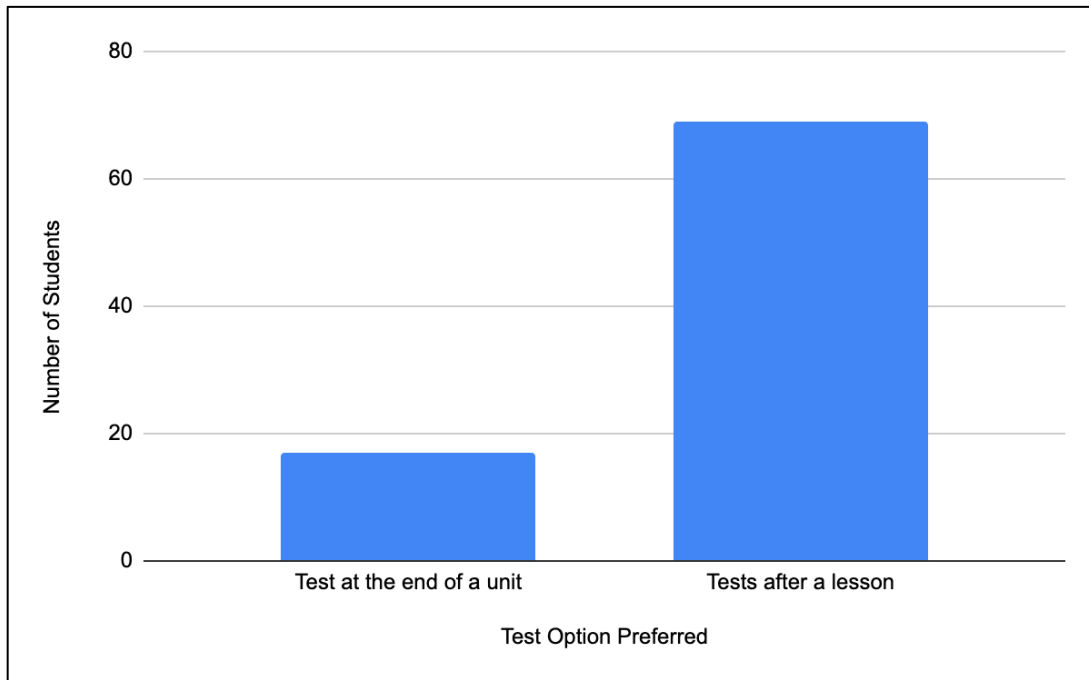
**Figure 4***Comparison of Cause-and-Effect Assessment Results*

The researcher conducted a dependent sample t-test to understand if there was a significant difference. For the cause-and-effect frequent assessment ( $M=80.815$ ,  $SD=14.73$ ) compared with the end of unit assessment ( $M=70.05$ ,  $SD=15.70$ ), showed an insignificant difference  $t(93) = 4.84$ ,  $p > .05$ .

Upon completion of the assessment's students were asked to provide feedback regarding their personal preference for the type of test. Students were informed data was being collected based on scores as well as their opinions. The quantitative data was collected by having students respond to the survey available on canvas. This survey was quick, the average completion time took just over three minutes, and the survey was done during class time. Figure 5 shows how students chose for themselves which test they liked best based on scores and personal preference. Figure 6 shows how students felt about each type of test which was rated using a Likert scale.

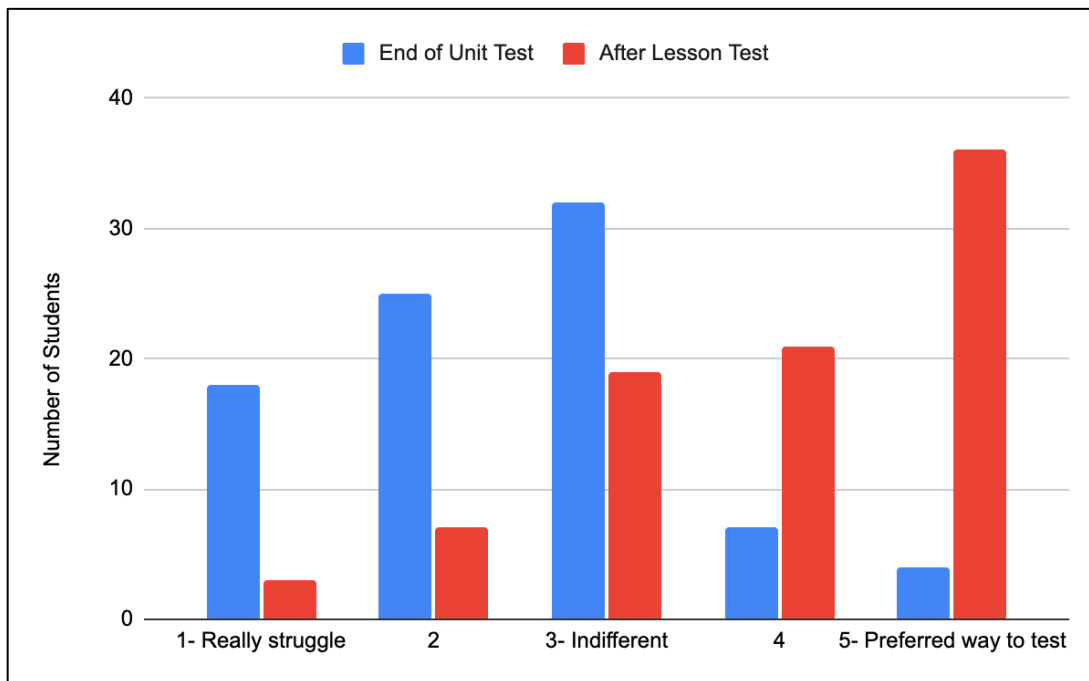
**Figure 5**

*Student Survey Responses to Test Preference*



**Figure 6**

*Student Survey Response to how they feel about each type of test*



The researcher conducted a dependent sample t-test to understand if there was a significant difference between students' feedback for end of unit assessments ( $M=3$ ,  $SD=1.01$ ) versus frequent assessments ( $M=3$ ,  $SD=1.13$ ). For the frequent assessment compared against the end of unit assessment there showed an insignificant difference  $t(85) = 6.49$ ,  $p > .05$ .

Throughout this action research project, the insignificant differences include the difference between cause-and-effect assessments, economic assessments, and primary and secondary sources assessments. The information collected from the survey was also insignificant as there was not a significant difference in feedback from students about which assessment they favored.

## Discussion

### Summary of Major Findings

In comparing frequent assessments to end of unit assessments the data showed students were more likely to meet the standard in the frequent assessment for the economic standard, primary and secondary sources standard, and the cause-and-effect standard. The survey results also indicated students were much more likely to prefer taking frequent smaller tests compared to the end of unit assessments. This preference is expressed by students and then elaborated on in the comment section of the survey. The insignificance is consistent with the study by Harsy et al. (2020), which found no significant difference in testing methods. Students commented explaining their choice which allows their responses to be better understood. Comments from students will be quoted in the phrasing they were submitted. "I do like tests after a lesson however I think tests after a unit is would more useful in the long run as it takes less time and is able to do least 3 grades at once per test instead of doing 1 grade per test. Unit tests take more studying though." Some students expressed concern with how assessments which are more like quizzes might not prepare them as well for college. Other opinions included, "I prefer testing after a lesson. Although there's a lot of tests, it's easier for me to remember the material. Maybe consider fewer tests?" Student feedback echoed many of the same points highlighting how students preferred the frequent tests because it was less stressful, easier to study for, and fresh in their minds.

### Limitations of the Study

The demographics of this course were not diverse with the students coming from similar socioeconomic backgrounds and similar ethnicities. The makeup of the research participants was also influenced by students who are taking an AP course in place of general history so there are

very few honors students. This study was completed towards the end of the school year as students had been assessing using the frequent format for the majority of the year so they were more accustomed to that format instead of the end of unit assessment. This study was conducted in a high school history course. Assessments can be difficult to replicate because the content changes from unit to unit, it builds differently than other subject areas.

### **Further Study**

This action research project looked into assessments and if students were more successful with frequent assessments or end of unit assessment. The researcher was interested in how changes to assessments can impact student success. Another potential study could include using a body of participants by looking into different subject areas. Replicating the idea of smaller assessments versus a large assessment can be done in different ways. Subject areas can adjust the idea to their content and could learn from one another potentially with how subjects assess students differently. Results could lead to a greater consistency across content areas or learning from one another to find new ways to assess students with greater success.

Another variation to the action research project could be to look into student's study habits. Many students had commented on their studying in the survey so looking at how students' study, when they start studying, how confident they are in their ability to study, or whether they study at all could be useful to understand. Results could shape the way students are prepped for assessments in the future with time set aside to talk about study habits.

### **Conclusion**

This action research project was intended to learn about student performance on assessments, looking into how the frequency of assessments can impact student performance and preference. This research was conducted using both a survey and data collection. Data was collected using a survey to identify student preference on testing, students rated the assessment options using a Likert scale and picked between the two options which they would prefer. Data was also collected on assessment performance for each of the three standards, each standard was assessed two different ways for all students. The two testing options were frequent, which had students take a shorter assessment the next class following a lesson and the second testing option was an end of unit assessment where multiple standards were assessed at one time.

A two tailed t-test was conducted to determine the findings were insignificant between the frequency of assessments and student performance. The information collected allowed the researchers to make adjustments to their classroom assessment. Due to findings being insignificant between the assessments the feedback from students can then influence the assessment choices used. Continuing to learn and make adjustments based on student performance and feedback is part of good practice.

The findings were insignificant for the student survey which asked students to choose the type of test they prefer. Data was collected throughout an 8-week period for multiple assessments, there were three standards total which were used to compare assessment results. After students received their grades, they completed a survey during class time which indicated the findings were insignificant. A two tailed t-test was conducted to come to this conclusion. Although the findings were insignificant, when asked to choose between the two assessment frequencies an nearly 75% chose the frequent assessment format.

Moving forward the results of this action research project indicated insignificant findings overall. The researcher will still use the information gathered when making decisions moving forward regarding how to test students. Continuing to learn about student's study habits and seeking information about other assessment strategies will be a focus moving forward. The outcome of this action research project allowed students to have a say in their learning and how they are assessed in class which created a positive response from students. The conclusion of this study can provide information regarding testing methods in the secondary classroom as educators search to find the best way to show student learning.

### References

- Baker, E. L., & Gordon, E. W. (2014). From the Assessment OF Education to the Assessment for Education: Policy and Futures. *Teachers College Record*, 116(11), 1–24. <https://doi-org.ezproxy.nwciowa.edu/10.1177/016146811411601107>
- Blevins-Knabe, B. (2019). Issues in General Education Assessment at Community Colleges. *New Directions for Community Colleges*, 186(186), 31–36.
- De Lisle, J. (2008). Can Standards-Referenced, Large-Scale Assessment Data Lead to Improvement in the Education System?: Judging the Utility of Student Performance Standards in the Primary School National Assessments of Educational Achievement. *Caribbean Curriculum*, 15 <http://ezproxy.nwciowa.edu/login?url=https://www.proquest.com/scholarly-journals/can-standards-referenced-large-scale-assessment/docview/2564506794/se-2>
- Dennis, D. V. (2010). "I'm Not Stupid": How Assessment Drives (In)Appropriate Reading Instruction. *Journal of Adolescent & Adult Literacy*, 53(4), 283-290. <https://doi.org/10.1598/JAAL.53.4.2>
- Elmahdi, I., Al-Hattami, A., & Fawzi, H. (2018). Using Technology for Formative Assessment to Improve Students' Learning. *TOJET : The Turkish Online Journal of Educational Technology*, 17(2) <http://ezproxy.nwciowa.edu/login?url=https://www.proquest.com/scholarly-journals/using-technology-formative-assessment-improve/docview/2025352702/se-2>
- Harsy, A., & Hoofnagle, A. (2020). Comparing Mastery-Based Testing with Traditional Testing in Calculus II. *International Journal for the Scholarship of Teaching and Learning*, 14(2)



<http://ezproxy.nwciowa.edu/login?url=https://www.proquest.com/scholarly-journals/comparing-mastery-based-testing-with-traditional/docview/2535072996/se-2>

Hinett, K., & Knight, P. (1996). Quality and assessment. *Quality Assurance in Education*, 4(3), 3-3+. <https://doi.org/10.1108/09684889610125832>

Kantor, P. T., Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (2011). Comparing two forms of dynamic assessment and traditional assessment of preschool phonological awareness. *Journal of Learning Disabilities*, 44(4), 313-321. <https://doi.org/10.1177/0022219411407861>

Knight, M., & Cooper, R. (2019). Taking On a New Grading System: The Interconnected Effects of Standards-Based Grading on Teaching, Learning, Assessment, and Student Behavior. *National Association of Secondary School Principals. NASSP Bulletin*, 103(1), 65–92. <https://doi.org/10.1177/0192636519826709>

Kuo, Y., Casillas, A., Allen, J., & Robbins, S. (2021). The moderating effects of psychosocial factors on achievement gains: A longitudinal study. *Journal of Educational Psychology*, 113(1), 138-156. <https://doi.org/10.1037/edu0000471>

Lewis, D. (2022). Impacts of Standards-Based Grading on Students' Mindset and Test Anxiety. *Journal of the Scholarship of Teaching and Learning*, 22(2). <https://doi.org/10.14434/josotl.v22i2.31308>

National Council of Teachers of Mathematics (NCTM) (2000). *Principles and Standards for School Mathematics*. Reston VA: National Council of Teachers of Mathematics.

Peters, R., E.D.D., Kruse, J., P.H.D., Buckmiller, T., P.H.D., & Townsley, M. (2017). "It's Just Not Fair!" Making Sense of Secondary Students' Resistance to a Standards-Based

Grading. *American Secondary Education*, 45(3), 9-28.

<http://ezproxy.nwciowa.edu/login?url=https://www.proquest.com/scholarly-journals/just-not-fair-making-sense-secondary-students/docview/1938071902/se-2>

Rosfiani, O., Ma'ruf Akbar, & Neolaka, A. (2018). *The Effect of Learning Environment, Inquiry and Student Learning Interest on Student Social Studies Learning Assessment*. European Alliance for Innovation (EAI). <https://doi.org/10.4108/eai.21-11-2018.2282251>

Rupp, A. A., & Lesaux, N. K. (2006). Meeting expectations? An empirical investigation of a standards-based assessment of reading comprehension. *Educational Evaluation and Policy Analysis*, 28(4), 315-333. <https://doi.org/10.3102/01623737028004315>

Shapiro, E. S., & Ager, C. (1992). Assessment of Special Education Students in Regular Education Programs: Linking Assessment to Instruction. *The Elementary School Journal*, 92(3), 283–296. <http://www.jstor.org/stable/1001981>

Shute, V. J., and Rahimi, S. (2017) Review of Computer-Based Assessment for Learning in Elementary and Secondary Education. *Journal of Computer Assisted Learning*, 33: 1– 19. doi: 10.1111/jcal.12172.

*Studies on Education, Science, and Technology 2022* (2022). In Noroozi O., Ed., Sahin I., Ed. (Eds.). International Society for Technology, Education, and Science. <https://eric.ed.gov/?id=ED623088>

Vogler, K. E., Schramm, S., Allan, A., & Parler, J. D. (2020). Using Instructional Time and Race/Ethnicity to Predict Social Studies Accountability Test Results. *Journal of Educational Research and Practice*, 10(1), 199-220.

<http://ezproxy.nwciowa.edu/login?url=https://www.proquest.com/scholarly-journals/using-instructional-time-race-ethnicity-predict/docview/2488222766/se-2>

Wehlburg, C. M. (2010). Meaningful General Education Assessment that is Integrated and Transformative. *New Directions for Teaching and Learning*, 121(121), 89–97.