

Saint Mary's College of California

## Saint Mary's Digital Commons

---

MATL Action Research Projects

---

Fall 2023

### Oral-Based Exams: Implications for Special Education and English Learning Students

Stephen Apel

Follow this and additional works at: <https://digitalcommons.stmarys-ca.edu/matl-action-research>



Part of the [Educational Leadership Commons](#), and the [Teacher Education and Professional Development Commons](#)



This work is licensed under a [Creative Commons Attribution-Noncommercial-Share Alike 4.0 License](#).

Oral-Based Exams: Implications for Special Education and English Learning Students

An Action Research Project

Presented to

The Faculty of the Kalmanovitz School of Education

Saint Mary's College of California

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts in Teaching Leadership

By

Stephen Apel

Fall 2023

Copyright © 2023 by Stephen Apel

All Rights Reserved

This action research project, written under the direction of the candidate's master's project advisory committee and approved by members of the committee, has been presented to and accepted by the faculty of the Kalmanovitz School of Education, in partial fulfillment of the requirements for the Master of Arts in Teaching Leadership degree.

---

Candidate: Stephen Apel

---

Date

Master's Project Advisory Committee:

---

Research Advisor: Monique Lane, Ph.D.

---

Date

---

Faculty Advisor: Heidimarie Rambo, Ph.D.

---

Date

---

Program Director: Monique Lane, Ph.D.

---

Date

---

Dean: Carol Ann Gittens, Ph.D.

---

Date

## **Abstract**

Oral-Based Exams: Implications for Special Education and English Learning Students

By

Stephen Apel

Master of Arts in Teaching Leadership

Saint Mary's College of California, 2023

Monique Lane, Ph.D., Research Advisor

Special education (SPED) and English learning (EL) students have historically struggled with standardized testing. These tests often use a multiple-choice question-style assessment to measure knowledge and skills. Similarly, teachers, schools, and districts use these tests regularly to gauge their own student populations' performance. The literature reviewed depicted multiple-choice tests as having many flaws with accuracy and accessibility for students with different learning and expression styles. Whereas in other literature another style, the oral exam, offered the opportunity for test takers to reveal a greater depth of knowledge. This action research project attempted to gauge the impact on SPED and EL students when a switch from a multiple-choice test to an oral discussion-based examination occurred. Findings displayed that both groups increased their mean test scores when completing the oral exam. Interestingly, general education students also increased their mean scores as well.

## **Dedication**

This project is dedicated to all my students, parents, and my wife, who have formed me into the teacher and leader I am today. Throughout my life and career, I have also been blessed with many experiences that, looking back on them, have shaped how I interact with everyone. I would like to first thank the students whom I have already taught; you have shown me that teaching is worth every moment. I have educated students from a diverse array of backgrounds, learning abilities, and ages. These students have and continue to positively challenge me every day; they have taught me a great amount about resilience and perseverance. These learners continue to inspire me and constantly make me want to improve my practice as an educator.

My parents' examples of leadership, nurturing, and care have also contributed to my blossoming into a teacher leader and, ultimately, this work. My mother was an elementary school teacher and principal who always emphasized learning and education for her students and myself. She still recalls students she taught, and we have even shared some of these same students. Her example of care and dedication has rubbed off on me, to say the least. My father was a hard-working quality control inspector for a housing development company and one of the leaders for my local Boy Scout of America troop. He took an active role, which he volunteered for, and led by example. He was a coach to some but the mentor many of the boys needed at the time. This has influenced me personally and professionally by developing compassion and empathy for students and people overall.

Lastly, to my wife. Thank you for supporting me throughout my personal and professional life. I love that we share a commitment to better the community we grew up in and want to contribute to the betterment of the world. I would not have been able to accomplish this research without your support at home.

## **Acknowledgments**

I would like to acknowledge my site colleagues who have completed the Saint Mary's College of California Master of Teaching Leadership (MATL) program. You inspired me to challenge myself and complete the program. The information provided prior to and support throughout the program has been unmatched.

I would also like to acknowledge my MATL cohort, though a smaller group this year, everyone's willingness to contribute, volunteer, and work together was amazing. The support offered was always positive! Though the task of this large project was daunting at first, everyone was full of encouragement which we used to then chip away and make this project manageable.

Finally, I would like to acknowledge the professors, Faculty Advisor, and Research Advisor, of the Master of Arts in Teaching Leadership program at Saint Mary's College of California. You provided a priceless experience for me and my peers. Your guidance through this program and project will remain unmatched.

## Table of Contents

	<b>Page</b>
List of Figures .....	ix
Chapter	
I. Introduction .....	1
Statement of the Problem.....	4
Purpose of the Research.....	6
Action Research Question.....	7
Limitations .....	8
Positionality of the Researcher .....	9
Definitions of Terms .....	10
Implications.....	12
II. Literature Review .....	14
Overview of Literature Review .....	15
Theoretical Rationale .....	16
Review of Related Research .....	20
Summary .....	36
III. Method .....	37
Setting .....	38
Demographics of the Classroom.....	41
Data Collection Strategies.....	43
Procedures .....	48
Plan for Data Analysis .....	51
Summary .....	53
IV. Results.....	55
Overview of Methods and Data Collection .....	56
Demographics of the Participants .....	57
Analysis of Multiple-Choice Testing and Oral Examinations.....	58
Analysis of Conversational Tally Sheet.....	63
Analysis of Oral Examination Questionnaire Exit-Ticket .....	65
Summary .....	69



V. Conclusions .....	71
Summary of Findings.....	73
Interpretation of Findings .....	79
<i>Closing Achievement Gaps</i> .....	79
<i>Demonstration of Greater Depth of Knowledge and Higher Scores</i> .....	80
<i>Increased Opportunity for Student Success</i> .....	82
Reflection on Limitations .....	83
Summary .....	85
Plan for Future Action .....	86
References.....	91
Appendices.....	94
A. 1920s, Great Depression, New Deal: Test .....	94
B. Cold War Assessment Questions .....	99
C. Cold War Discussion Exam Grading Rubric .....	100
D. United States History Cold War Discussion Exam Conversational Tally Sheet ..	101
E. Discussion Exam: Before, During, and After Questionnaire .....	102

## List of Figures

Figure

1. Multiple-Choice Test: Mean Scores of Different Student Groupings .....59
2. Oral Examination Test: Mean Scores of Different Student Groupings .....61
3. Multiple-Choice Test vs. Oral Assessment: Mean Scores.....62
4. Student Use of Evidence vs Teacher Prompt for Depth: Mean Scores .....65
5. Before, During, and After Questionnaire: Student Grouping Mode.....68

## **Chapter I**

### **Introduction**

As children enter American classrooms, more students are identified as students with special needs. Additionally, the number of English learning identified students is increasing country-wide (CDE, 2022). These students will not all learn and obtain information the same. Some will be kinesthetic learners, whereas others may learn best through linguistics or interpersonal education. (Gardner, 2011). Knowing that all students learn differently, some educators have hypothesized that some students may show learned knowledge, skills, and information differently. Although some teachers use various assessment methods, the multiple-choice standardized test is still one of the most common classroom exams. Student assessments suited to individual learning styles instead of one standardized test may benefit their long-term educational goals and shrink achievement gaps. Ultimately, assessing students with various testing styles might assist the student's long-term success as course assessments could reflect more accurate outcomes. The oral examination is one type of assessment that seems to have fallen into disuse yet shows continued promise for student achievement.

According to the California Department of Education (2022), between 2010 and 2018, the number of students enrolled in special education rose from 680,164 to 795,047, an approximate 16.8% increase. Additionally, between 2011 and 2022, the enrollment of English learners (ELs) in California schools held more than one million enrolled students (CDE, 2022). Both student populations consistently underperform on the California Assessment of Student Performance and Progress (CAASPP) test, a standardized assessment. The 2021-2022 English Language Assessment (ELA) CAASPP results show that 65.26% of students with disabilities did not meet their grade-level standards. In the Mathematics section, 73.72% of students with

disabilities did not meet the standard. These numbers closely resemble the EL population's performance on the same exam. The ELA CAASPP scores of English learners displayed that 63.50% did not meet the standard, and in the mathematics section, 70.33% of ELs did not meet the standard. Tests like the CAASPP use multiple choice and typed responses to gather the data from these students.

Similarly, standardized multiple-choice testing and written assignments are often used in general education classrooms to measure student learning, knowledge, and skills. This information is critical to understanding because these two populations of students historically struggle in reading and writing. If they cannot read the question, of course, they will underperform on these exams.

In 2001, the Bush administration passed the No Child Left Behind Act (NCLB) in the United States. The goal of NCLB was to improve school and student outcomes. The legislation built in a layer of accountability for the academic institutions as well to ensure students were performing up to standards. The United States implemented standardized testing to measure student achievement. The NCLB produced positive and negative consequences for the educational system.

A few positives from the passage of NCLB are that More money was spent per student to improve skills. Students performed better on math assessments at lower grade levels and populations that tested historically low improved scores (Dee et al., 2010). However, there were some negative impacts too. School performance became tied to these tests, whether they wanted them to be or not, and students and teachers alike focused on these assessment performances (Dee et al., 2010). Fixation of increasing test scores can be seen by the narrow test preparation activities teachers used to yield results (Dee et al., 2010, p. 190). These exams appeared in

written format, including the standardized multiple-choice testing (SMCT) format. To meet the demands of NCLB, teachers focused on content and skills that would appear on the exams (Pedulla et al., 2003).

Today, remnants of this system play out in the American classroom. Teachers, governmental organizations, and private industries often still use a written assessment, or a SMCT, to gauge and assess student learning. For example, the National Assessment for Educational Progress (NAEP) test, also referred to as “the nation’s report card,” uses multiple choice questions in addition to constructed written responses to measure student achievement (McFarland et al., 2019). Though these exams give scores and student data, they do not always truly reflect the learning of all students. Non-white students such as Black, Indigenous, and People of Color (BIPOC) students historically struggle to score the same or better than their White or specific Asian American counterparts on these exams. A better way to assess these students is first to understand if they have learned the information but cannot express it through a standardized test.

According to Howard (2001), one of his teacher-participants reported that her African-American students scored consistently higher on assessments judging reading comprehension if given in an oral-based format (Howard, 2001). Many research studies have backed the use of oral examinations as beneficial to accurately measuring student understanding (Hazen, 2020; Huxham et al., 2012; Sayre, 2014). Across different fields, students have shown that they are successful when oral assessments are used instead of SMCT and written exams. This is because these exams allow the expression of more profound knowledge and the use of circumlocution by students to prove knowledge and skill. The examiner can ask follow-up or clarifying questions to probe deeper and make the tester elaborate on answers if needed (Hazen, 2020; Huxham et al.,

2012; Sayre, 2014). Though these studies were conducted with college-level students and smaller class sizes, the oral exam shows promise for increasing test scores. The oral exam can potentially reveal information or skills thought to not exist in the student's brain when using a standardized testing model.

### **Statement of Problem**

Teachers and schools around the United States remain focused on testing students for achievement data, which may remain the same. However, a shift in how these data are collected is needed to display better what all students genuinely know. Multiple-choice exams have many issues in accurately reporting information on student knowledge. Students may need help understanding questions or vocabulary on the exam. Students can also eliminate answers and potentially guess the correct answer without knowing it (Turhan, 2020). However, the oral-based assessment may solve the inaccuracies and inequities of the SMCT and constructed response assessment formats.

There are several problems with using standardized testing models as the only means of assessment. First, on multiple-choice testing, students can give narrow answers and guess the correct answer without knowing the exact answer (Reich, 2009). Another disadvantage of multiple-choice testing is using academic vocabulary or complex syntactic structure within the multiple-choice questions/answers, which may be above the learners' academic literacy level. Thus, their responses may not indicate their understanding of the content but whether they understood the question's vocabulary or syntactic structure. An equally important point to discuss is that on written exams, students can merely list knowledge about a subject without answering a prompting question due to misreading or misunderstanding the prompts (Boedigheimer et al., 2015). Oral examinations, however, offer solutions to these issues. The oral examination method

allows the examiner to probe student knowledge deeper on topics to assess whether a student understands the topics accurately. Additionally, this type of assessment can offer the examiner opportunities to redirect students toward answering the assessment's questions if they meander off-topic or engage in a cycle of information dumping (Boedigheimer et al., 2015). Aside from SMCT, there are several ways to prove knowledge that are more relevant to dealings outside of the academic world.

Outside of schooling, multiple-choice testing is not used daily for many people residing in the United States. Discussions, conversations, and presentations, however, are very popular in the business world. By not practicing these presentations and discussions of information, the schooling system is doing a great disservice to our American students by not developing the skills necessary to be competitive in the work world (Hazen, 2020). No matter what line of work one enters, people must have practical communicative skills to relay their ideas clearly and effectively.

Throughout my time as an educator, I have had the opportunity to work with countless numbers of BIPOC, special education, and English learning students who struggled with these standardized tests. Many of the students who took these tests wanted to avoid viewing their scores on exams as they were predicting failure based on their previous testing experiences. Due to this, I began exploring new ways to assess student learning without a traditional multiple-choice test. Most of the assessments I have used to date have had some element of writing included. Whether that be an essay, short answer, creation of a slideshow, poem, or something to that effect. It has only recently occurred to me that more oral-based assessments would likely produce higher assessment scores and outcomes. Since spoken language develops before reading and writing, the students would have more opportunities to express what they do know using

their current language abilities (Krashen & Terrell, 1998). Since this skill emerges first, assessing using the students' most robust and developed means of expressing knowledge makes sense. Additionally, students who struggle with proficient literacy could use circumlocution to prove knowledge without particular academic vocabulary.

I am currently an 11th-grade United States History teacher in a middle-class suburban school. I have three sections of United States History totaling 90 students. I have found that, for the most part, my special education population and English learners have lower test scores on SMCTs. I have tried standardized testing models using multiple-choice questions and written responses, but overall, these methods have yet to be successful. These populations continue to lack the ability to perform well on these depth-of-knowledge tests. I am interested in trying a different approach to testing using oral assessment, which has shown success in BIPOC groups. I would like to know if incorporating oral assessment might improve my special education and English learners' scores and better represent the information they know about United States History.

### **Purpose of Research**

The purpose of this research was designed to analyze the effect of shifting assessment formats from SMCT or written exams to oral-based assessments on Special Education and English learning students to measure whether they can demonstrate a more accurate representation of their learned historical content in a mainstream United States History classroom. I hypothesized that multiple choice testing and writing assessments did not show the correct depth of students' content knowledge and were instead providing misinformation to teachers. Special education and EL students historically perform poorly on these assessments,



and therefore offering them this opportunity would increase their chances of passing the assessment and, ultimately, the course.

My school district and school site have provided various assessment tools for all students. The two most common tools they have asked teachers to explore for assessments have been multiple-choice tests and writing assignments. Though these tests and writing assignments begin to address the problem of extracting learned historical information, these types of assignments still stifle how learners can communicate as many have difficulties with writing or specific academic language. The tests especially create difficulty for special education students and English language learning students. As our school and district look to improve test scores, they continue to employ testing methods that may not show accurate student data. By shifting the assessment format away from a formal paper-and-pencil SMCT or writing test towards an oral-based format, the student will have the opportunity to express information while speaking even if they lack academic vocabulary, have issues with spelling, or have issues with reading.

I wondered if my SPED and EL students could display more knowledge of unit-specific historical information and skills if they were assessed using an oral discussion assessment instead of a SMCT or a written task. Discussion-based formats such as a paired communication protocol, a Socratic seminar, and a one-on-one oral exam were implemented as an assessment to gather this data. It is possible that if the oral examination was brought into the general education classroom these students would display higher levels of understanding concerning the whole unit of study rather than taking points off for individual mistakes.

### **Action Research Question**

This study was intended to examine the effectiveness of oral-based assessments as a measure to gauge learning for high school SPED and EL students. The paired communication

protocol, Socratic seminar, and one-to-one oral examination formats were used alongside a rubric to gauge student understanding of the historical material studied and analyzed on a previous occasion. The question that drove this study was: *How will shifting assessments from a multiple-choice test to an oral-based assessment better represent High School special education and English learning students' knowledge of United States History?* My hypothesis focused on the idea that some students perform poorly on exams because these students learn and express knowledge differently, especially students who are deemed English learners or special education. These students need to express their knowledge in a way less rigid than a standardized multiple-choice test. I hoped to see increased United States History test scores for Special Education and English learning 11th-grade students if given a test in an oral-based format.

### **Limitations**

Several limitations may have affected the outcome of this research project. This study was conducted over eight weeks with 90 students in three sections of United States History classes. One of the limitations of this study was time. As a full-time teacher of two different history courses, United States History and Advanced Placement World History, along with conducting research, gathering data, and analyzing it, it was challenging to maintain the study in such a short period. Additionally, since I conducted this study during the fourth quarter of the year, I could not gather data over one year with the same students. By not being able to conduct the research for a full year and all my United States units of study, students could potentially have scored higher simply because they understood this unit of the Cold War the best. This might have hindered the study as well since only one formal one-on-one oral examination with the teacher was given instead of multiple throughout the school year. Another limitation of the study was working with a smaller sample size. Out of the 84 students in the study, 15 of these

students were designated as special education students, and 16 were designated as English learning students. This limited the amount of data I could gather and analyze specifically on EL and special education students as opposed to general education students who were non-EL and non-special education. Lastly, some students were naturally hesitant to participate in oral-based exams as a discussion test between teacher and student was not the normal exam structure that they were used to. This could have potentially skewed the results of their exam.

### **Positionality of the Researcher**

I am a White cis-gender man born in the United States. At this point, I have taught history at a high school level for nine years and have been working with EL and special education students for that duration as well. I have no diagnosed learning disability or designation of being an English learner, nor have I ever had these designations. I grew up in a home that I would identify as socio-economically middle class. My mother was a primary school teacher and eventually an elementary school principal. Therefore, she knew how to better my education and get the support I needed when I was not up to parental or state standards. My parents encouraged me and my siblings to pursue higher education, where I further strengthened my skills in the English language. I have also always been an outgoing and talkative individual who did well on presentations but did not always perform well on standardized tests or writing samples. The idea that not all students will learn or perform the same is founded in the theory that humans have multiple intelligences and will display them differently (Gardner, 2011). When teaching, I found myself naturally interested in measuring student success and outcomes when different supports or assessment models were used particularly those that involved discussing the course material.

I researched how the implementation of oral examinations, as opposed to standardized multiple-choice testing, on special education and English learning students would affect their achievement of these assessments. Having seen the standardized testing data for Special Education and English learning students, I saw the need to implement a better testing model to assess these groups of students. I believed that offering a different model for testing would produce higher outcomes for all learners, but particularly the EL and special education children. To minimize bias in the research, I used a rubric to grade the oral examination. Additionally, I used empirical research to help guide how the oral examination took place, how it was graded, and how to measure the success of the new assessment itself. This research also guided a few of the structures I introduced to students throughout the research project such as providing the students with the set of exam questions at the beginning of the unit and providing class time to practice peer discussions leading up to the assessment held by me.

### **Definitions of Terms**

The following definitions have been provided to bring clarity to this action research project. They are here with the intent to provide the reader with a further understanding of the definition in the same way as the researcher. These definitions are sourced from a variety of different organizations and publications.

### ***BIPOC***

“Black, Indigenous, and People of Color is a term used especially in the United States to mean Black people, Indigenous American people, and other people who do not consider themselves to be White.” (Cambridge, 2023).

### ***English Learning (EL) Students***

“English-language learners are students who are unable to communicate fluently or learn effectively in English, who often come from non-English-speaking homes and backgrounds, and who typically require specialized or modified instruction in both the English language and in their academic courses.” (Great Schools Partnership, 2013).

### ***Multiple Intelligences***

“As a species, humans have a set of autonomous intelligences and are not just limited to one type of intelligence. Some intelligences lay in logical and linguistic abilities while others are rooted in kinesthetic, musical, interpersonal, intrapersonal, and more” (Gardner, 201, p xii).

### ***Oral Assessment/Examination***

“...Oral exams (orals) provide an assessment approach ideally suited to encouraging students to reflect on and contextualize their experiences in field courses, as well as offering a face-to-face opportunity for the instructor to assist students in achieving deeper levels of understanding and more effective communication skills. The oral [exam] can also provide a culminating summative experience that allows students to build from formative assessments administered throughout the course, particularly journaling and discussion” (Hazen and Hamann, 2020, p. 130)

### ***Special Education (SPED)***

“Specially designed instruction, at no cost to parents, to meet the unique needs of a child with a disability.” (US Department of Education, 2023). “Children who typically qualify for special education services include those with the following disabilities: Intellectual disability, Hearing impairments (including deafness), Speech or language impairments, Visual impairments (including blindness), Serious emotional disturbance, Orthopedic impairments, Autism spectrum

disorder, Traumatic brain injury, other health impairments, Specific learning disabilities, Developmental delay (Purdue University, 2023).

### ***Standardized Testing***

“A standardized test is any form of test that requires all test takers to answer the same questions, or a selection of questions from common bank of questions, in the same way, and that is scored in a “standard” or consistent manner, which makes it possible to compare the relative performance of individual students or groups of students. While different types of tests and assessments may be “standardized” in this way, the term is primarily associated with large-scale tests administered to large populations of students, such as a multiple-choice test given to all the eighth-grade public-school students in a particular state...” (Great Schools Partnership, 2015).

### ***Multiple Choice Test***

A multiple-choice test or question is one in which you are given a list of answers, and you have to choose the correct one (Cambridge, 2023).

### **Implications**

The purpose of this study was to test if students could perform better or demonstrate the knowledge, they had learned in a more accurate way than the limitations set within a traditional exam structure. This study can potentially reveal shortcomings of the multiple-choice test formats. If this study successfully reveals accurate student knowledge about a specific content area, in this case, history, the oral exam may apply to additional content areas. This may shrink equity and achievement gaps in the classroom as some students lack the specific vocabulary to answer questions on paper-based exam formats. In the oral discussion-based format, the students also have the opportunity to ask clarifying questions or potentially get questions rephrased so they understand how to answer a particular question. Similarly, suppose only a basic answer is

provided by the student at first. In that case, the teacher has the opportunity to probe and nudge the student into revealing more information they know about the given subject. Thus, the student could score better on the entirety of the exam and do better in the course.

The oral exam may also have additional implications for special education or English-learning students. Students with physical impairment might have a better chance at expressing the entirety of their knowledge in a better way suited to their bodies. For example, if a student had dyslexia, they might not have as difficult of a time answering a question if the question is read to them. Similarly, if a student has dysgraphia, expressing their knowledge through conversation may be easier than writing out their answers. EL students may be able to use the vocabulary they know to answer the question without the use of academic vocabulary. Therefore, both groups' needs are being met and potentially in a better way.

## **Chapter II**

### **Literature Review**

The purpose of this research is designed to explore the effects of shifting assessment formats from a standardized multiple-choice test (SMCT) or written exams to oral-based assessments on the academic achievement of special education (SPED) and English learning (EL) students. I am conducting this study with the hope that SPED and EL students will demonstrate a greater depth of their learned historical content in a mainstream United States history classroom.

As children enter American classrooms, more students are being identified as special education and English learners (CDE, 2022). All students entering our schools, regardless of their learning abilities, will not all learn and obtain information in the same way. Some will be kinesthetic learners, whereas others may learn best through linguistic or interpersonal education methods (Gardner, 2011). Knowing that all students learn differently, some educators have embraced the proposition that students may show learned knowledge, skills, and information in different ways too. Though some teachers use various assessment methods, the standardized multiple-choice test (SMCT) is still one of the most common classroom exams. However, the SMCT does not always accurately reflect the information students truly know and understand due to issues with literacy and content. (Reich, 2009). One of the issues with SMCT exams is students' inability to access academic vocabulary or simply understand what a particular question is asking. If students can be assessed using a method better suited to their learning style, such as oral assessment, they may benefit in terms of their educational outcomes. These benefits might include success in their courses, attainment of long-term educational goals, increased confidence, and the reduction of known achievement gaps. In addition, oral examinations offer



better skill preparation for the workplace environment (Hazen, 2020). This is due to the continuing practice of public speaking, presentation, and communication of intellectual ideas.

Moreover, research has suggested that a unique benefit of the oral examination is that it can be used by the proctor to reveal a greater depth of knowledge over traditional exam formats. The examiner's utilization of rephrasing prompts, asking follow-up questions, or nudging testers to clarify vague answers all contribute to better exam scores (Boedigheimer et al., 2015; Hazen, 2020; Hazen & Hamman, 2020, Sayre, 2014).

Hence, the research question that guided my project was, *in what ways will shifting assessments from a multiple-choice test to an oral-based assessment affect high school special education and English learning students' representation of knowledge learned in a United States history class?*

## **Overview of the Literature Review**

The action research project examined the impact that shifting exam formats from SMCT to oral examination would have on 11th-grade United States history students who were both special education and English language students. This chapter begins with the theoretical rationale. The work of Howard Gardner's multiple intelligence theory, Benjamin Bloom's Taxonomy of Higher-Order Thinking, and Krashen and Terrell's theory on second language development all contribute to this study as they provide the theoretical rationale. Next, the Review of Related Research discusses the empirical research covering general education classrooms and students with differing abilities, academic English language skills: SPED & EL development, standardized and multiple-choice testing issues, and oral exams and discussion. The research was completed to reveal the potential pitfalls of the traditional exams in classrooms

across the country and to better understand how individuals learn and express knowledge in diverse ways through language.

The articles gathered and reviewed were collected to gain an understanding of the benefits and shortcomings of traditional testing practices and how they serve students with special needs and English learners, along with whether the oral examination had benefits for any learners or test takers who were not labeled as the groupings above. These articles were assembled from databases and sources such as EBSCO, Education Source, ERIC, and Google Scholar. The key search terms used were: *oral exams, oral examination, oral communication, oral performance exam, alternative forms of assessment, testing, multiple-choice, test design, communication, discussion, English language learner, language, language acquisition, special education, and students with disabilities.*

## **Theoretical Rationale**

### ***Multiple Intelligences Theory***

Howard Gardner's multiple intelligences theory supports the research question of this thesis. Gardner's multiple intelligences theory provides a theoretical framework for explaining how people learn and express knowledge. Gardner's original theory proposed six different forms of intelligence in 1983. However, that number has increased to nine different forms of intelligence (Gardner, 2011). The following is a list of the nine intelligences: naturalistic, musical, logical–mathematical, existential, interpersonal, linguistic, bodily–kinesthetic, intra–personal, and spatial intelligence. For instance, bodily-kinesthetic intelligence expression is where an individual can express their learning through movement and actions. Depending on the cultural context, certain intelligences may be deemed more important than others. In literate

societies, a highly intelligent person may be seen as such because of their level of literacy, knowledge, or expression of language (p. 354)

Intelligence can also be multilayered. Each individual can possess one or more of these intelligences to make up their unique learning abilities (p. 382). Within individual intelligence, people can learn and express knowledge differently (p. 350). For example, within linguistic intelligence, there are many domains, including semantics, syntax, phonology, and oral. An individual could be more proficient in one of the domains and not others but still be considered proficient or have competency in that intelligence. This research portion is integral to understanding because it reveals the differences inside the intelligence umbrellas.

Gardner's theory of multiple intelligences and the domains within these intelligences presents educators with information to better support classroom learners by individually tailoring their education. By doing so, these students will be able to acquire information more easily and in a way that they can process and comprehend the knowledge being taught. These multiple intelligences can be further supported by adapting lessons or information into Bloom's taxonomy of higher-order thinking. By doing so, students can be challenged to manipulate information to build a better understanding and comprehension of the material or skill that is being taught.

### ***Higher-Order Thinking Skills Theory***

An additional theorist supporting this research's foundation is educational psychologist Benjamin Bloom. His work on higher-order thinking skills (HOTS) is titled *Taxonomy of Educational Objectives: The Classification of Educational Goals*, initially published in 1956. This theory outlined several levels of possible learning objectives. The learning objectives set by Bloom assisted in revealing how much learning occurred for a person and how well that person could use the information or skill. According to Bloom, these learning objectives can be

categorized into six levels of learning objectives: knowledge, comprehension, application, analysis, synthesis, and evaluation (1956). The most superficial part of learning occurs at the knowledge level, as this can simply be memorizing facts or information without understanding the material. The level that reveals the greatest depth of learning would be the evaluation level. At this level, learners evaluate information based on prior information or skills learned.

Bloom's research established a tool to gauge the depth of knowledge and applicable skills taught. When specifically applied to the classroom environment, a teacher can implement the six learning objectives to construct questions or simulations to accurately measure if students truly understand and can use the taught information constructively.

When trying to develop and assess information for students with special needs and English learning students, a teacher must keep in mind the multiple intelligences and higher-order thinking theories. They must also be aware of where their student learner(s) are with their language skill and development and how to best teach and work with these individuals. Educators who are working with this student clientele should be mindful of *The Natural Approach to Language Acquisition* (1998) as the theory suggests how language can be developed. Though this theory specifically focuses on developing a second language and the best practice for this, some of it may be applicable to learners who struggle with language in general.

### ***Second Language Acquisition and Second Language Learning Theory***

A third theory was presented in the updated 1998 book *The Natural Approach to Language Acquisition* by Stephen Krashen and Tracy Terrell. This book builds on Krashen's original works from his 1981 book, *Second Language Acquisition and Second Language Learning*. Together, Krashen and Terrell provide a theoretical framework and give hypotheses for explaining how people learn a second language. I will outline two of the major hypotheses

relevant to my work. They theorize that language develops in two ways. The first way language can develop is through acquiring language. Language acquisition occurs when the student learns through communication in natural settings and situations (Krashen & Terrell, 1998, p. 18).

Usually, language students who are acquiring a language are not worried about the structures of language, grammar, or other rules of language. The second way the theorists present how a second language can be learned is through “learned language.” In this method, the learner is consciously or formally learning the language and using the available language tools, such as grammar or rules. They further their theory by saying language acquisition is how an individual understands the language, and language learning is to assist the language student in monitoring, editing, and adjusting their language.

Krashen and Terrell present another relevant hypothesis they call *the input hypothesis*. This is where the language learner already has some knowledge and structures of the language but is presented with a portion of new information and given context around the material by someone with a higher understanding of the language. It assists the learner because it provides a scaffold to the new information and continues to help build understanding. Ultimately, if done enough, the language acquirer can move to a new stage of language development (Krashen & Terrell, 1998, p. 33). Overall, the language learner will only learn and advance if they understand the message of the language. In relating this theory to my research, if oral examinations are used with this theory, it will provide more opportunities for the language learner to understand what is being asked and respond appropriately (Boedigheimer, 2015; Hazen, 2020; Huxham et al., 2012). It also provides an opportunity for the examiner to give feedback, assess the tester's skills, and provide similar or new support if the student needs it (Hazen, 2020).

When all three theories are used in conjunction, each serves to help students develop knowledge and skills that could be applied to the oral examination. Combined with Bloom's taxonomy, Gardner's multiple intelligence theory is beneficial as it offers a model for measuring and assessing learning objectives without one rigid assessment such as a SMCT. Bloom's taxonomy can be applied to specifically design assessments for students who learn or express learning differently. In application to this research, Bloom's taxonomy was used to create variable difficulty questions based on the levels to measure student outcomes. Rephrased questions also created an opportunity for great depth responses demonstrating higher-level thinking. Additionally, when developing the methods of this research, Krashen's theory was necessary as it offered insight into how language naturally develops for language-learning students on their own and in conversation with more advanced speakers. By using the oral examination, students with developing language skills did not have to worry about language rules or grammatical structures. They could focus on presenting their genuine understanding of the material.

### **Review of Related Research**

This review of related research includes four major areas of Interest. First, a discussion of general education classrooms that have students with differing abilities that display achievement gaps. Second, an examination of how basic and academic language develops with special attention given to English learners. Third, a review of findings on how standardized and multiple-choice testing has issues for test takers. Then, a critical look at the impacts of implementing oral examinations. The findings of these studies indicate that oral examinations and the use of discussion in classrooms can benefit learners of all abilities and serve as a further learning/teaching tool. However, some stress may be involved for the examinee.

### *Students With Differing Abilities and Educational Settings*

In the United States, learners from diverse backgrounds are mandated by law to be placed in the least restrictive learning environment, which is commonly a general education classroom setting (US Department of Education, 2017). In these classrooms they are expected to thrive, learning required content materials and hopefully meeting state performance measures. The subsequent articles have been gathered and analyzed to determine how special education and English learners perform in American schools; and if they are accessing the curriculum similarly to general education students whose primary language is English and who do not have special learning circumstances.

Special education students have shown difficulty keeping the same or similar achievements to students who do not have special needs. Compared to students without disabilities, SPED students showed lower achievement in reading (Gilmour et al., 2019). A meta-analysis of 23 studies analyzed the reading achievement of students with disabilities versus those without. The meta-analysis gathered research on three published reports, eight dissertations, and 12 journal articles between 1997 and 2017. The information gathered was on students of varying grades from elementary, middle, and high school levels. It also gathered data on students with differing disabilities from the listed grade levels. It should be noted that to gather enough significant data for middle and high schoolers, they were categorized into one data set and compared to the reading achievement of the elementary tester. The meta-analysis took particular care in removing duplicated information, non-English studies, and only selected studies that met the inclusion criteria of studies conducted from 1997-2017. Answers were gathered from multiple choice question testing.

Conclusions from this study displayed that there was an average reading growth gap of

roughly three years between students with disabilities versus students without disabilities (Gilmour et al., 2019). There was no statistically significant data difference between the performance of elementary students with disabilities and middle/high school students with disabilities. Therefore, the three-year reading growth gap remains throughout grade levels. If these students cannot access the curriculum at the same speed as their peers without disabilities, they will fall further behind. Special education students' lag in reading skills poses a legitimate concern for their learning in general education, passing their assessments, and ultimately passing their courses.

If information and skills were taught differently, appealing to the multiple intelligences theory, these students could potentially shrink that gap. This research related to my study because it highlighted the reading achievement gap of students with special needs compared to their non-special needs peers. Lastly, since this information was gathered using multiple-choice testing, there may have been fewer opportunities for testers to demonstrate a full understanding of the reading. Similarly, students learning English as a new language have a clear disadvantage in the academic setting.

Between 2014 and 2015, Miley and Farmer set out to measure if there was a significant difference in English Language Arts and Math proficiency levels for students deemed English learning who achieved proficient scores on WIDA-Access versus non-English learners. Data was collected from eight elementary and middle schools in rural Tennessee, grades three through eight. The data was collected from 200 English learners and 102 non-English learning participants. In addition to this, they used available data from 302 elementary and middle schools that participated in the WIDA-Access, Tennessee's English language proficiency test, and Tennessee Comprehensive Assessment Program (TCAP) exams.



Miley and Farmer's analysis claims that EL students, though deemed proficient, performed lower than their non-English learning counterparts. They state that English proficiency does not equal academic English proficiency and that these ELs were being assessed using the same standards since their proficiency was labeled. (Miley and Farmer, 2017). Since these students are being labeled as proficient, they are expected to perform just as well in content areas (Miley and Farmer, 2017). In the discussion of their findings, they claimed that students of language, to be measured equally, need between five to ten years to acquire cognitive academic language proficiency (CALP), based on the Cummins 1999 research. This was an important point for the researchers as they also claimed that these EL-proficient students were expected to learn content similarly to their non-English learning peers even though they had not mastered the language. This research study illuminates that both non-proficient ELs and proficient ELs are not able to learn the content or curriculum like their non-English learning peers.

These studies depict that achievement gaps between general education native English speakers, English learners, and special education students exist. They also illustrate that these students are not accessing the curriculum similarly to their counterparts in some cases due to the lack of proper support. If academic scaffolding was better provided for these student groups, they might have a better chance of accessing the curriculum and skills.

### ***Academic Language and English Language Development***

The development of language and academic language takes place over an extended period. Some individuals need explicit instruction and assistance to develop the language skills necessary to thrive in the academic setting. English language learning students could benefit from specific methods implemented by teaching professionals to develop these language skills and academic traits.

Zwiers (2007) explored academic language growth in general education classrooms. The focus was on how teaching practices in mainstream classrooms assisted the development of academic language for EL students. The study aimed to determine what methods teachers implemented to build and strengthen academic language within their rooms. Zwiers acquired data by observing three seventh-grade classrooms, teachers, and their students of varying subjects: Language Arts, History, and Science. The researcher selected four focal students deemed intermediate English learners based on their California English language development test (CELDT) scores. These classes contained additional EL students ranging from eight to twelve in each. He noted and tracked the interaction, involvement, use of visual aids, gestures, and other means of creating comprehensible academic language using a marking system. Aside from observations, Zwiers collected written data through writing samples, tests, and other products.

Zwiers concluded that if teachers understand who their student population is and what difficulties they have with obtaining academic language, they can engage their students and facilitate the development of academic language skills, which will improve their academic language abilities. His findings suggested using five cognitive skills in the classroom could develop these attributes. The five cognitive skills are to be implemented to create intrinsic interest in students and engage them to express opinions and information instead of filling out passive worksheets. The cognitive skills are as follows: 1) cause and effect, 2) comparing, 3) persuading, 4) interpreting, and 5) taking other perspectives (Zwiers, 2007, p. 21). This is similar to the theory of higher order thinking skills as well. As teachers use these theories and strategies, they can assist students in building language and academic skills they can take into the future.

Zwiers's work is relevant to my research because it displays how academic language

does not just emerge but rather must be developed. Teachers can develop these skills through specific teaching methods and engaging their students meaningfully. A few of the most relevant pieces of this research were the use and observation of student discussions, teacher reflections on the discussion, and teacher interventions in the discussion. For example, during a small group discussion, the educator asks questions or for elaboration. In the case of my research, the oral examination aims to do this, allowing the teacher to evaluate and continue developing academic language skills.

A peer-reviewed journal article by Margarita Calderón and her colleagues sheds light on the difficulties English language learning students face in the United States school system and how to address some of the language achievement gaps. Calderón and her team gathered and analyzed a collection of studies from the elementary through high school levels. Their analysis is found in *The Most Effective Instruction for English Learners*, published in 2011. At a secondary level, many ELs were placed in specialized English language development classrooms; whereas in elementary schools, EL students were placed in a general education classroom and pulled for English language development. When these students were not in the specialized setting, they remained in their general education classes in a “sink-or-swim instructional situation” (Calderón, 2011, p. 106). Without offering these students proper support, these achievement gaps will not close in language or specific content areas. (Calderón, 2011 p. 106). Additionally, if these students are placed in a general education classroom, they are not offered proper scaffolding to access the content and continue their learning. The researchers offer some insight on how to support EL students better so that they can effectively learn the language.

The researchers concluded that there were significant achievement gaps between the EL and non-EL populace. (Calderón, 2011). To minimize this gap, they suggested three ideas. 1)

Teachers and schools must dedicate themselves to instructing EL students in a whole-school intervention model. 2) Language development needs to be focused on in the early grades, preschool to third grade. 3) Teachers need to be educated on language acquisition through professional development to understand how to develop and assist these learners. If schools and teachers can achieve these, combined with the multiple intelligences theory, they could make language development unique to the learner and potentially further build their language toolkit. This research is essential to my study because it not only displays the disparity of achievement between EL and non-EL students based on their language proficiency, but it also displays the power of academic interventions to support EL learning and the impacts these interventions may have.

Together, these studies display that by building scaffolding and support for both English learners and academic English skills, students have a better opportunity to develop their English language and hopefully shrink the achievement gap. Additionally, they will have the opportunity to excel in other academic areas like specialized content, i.e., history, mathematics, and science. Once basic language skill evolves into academic language, students can apply it to the state standards and performance exams, though unfortunately, most remain in the form of multiple-choice tests (MCT), which have their unique issues (Reich, 2009).

### ***Standardized and Multiple-Choice Testing***

Research about standardized testing reveals many shortcomings of the testing type that impact both instructional practices and student learning. Teachers must adjust their teaching curriculum to match the testing materials and students are expected to meet government standards based on these tests. The MCT also has issues with testing equity, accessibility for language learning styles, and problems with accurately depicting student knowledge (Dee et al.,

2010; Pedulla et al., 2003; Turhan, 2020). Though students are learning standardized material, which may be positive for the growth of the student, teachers' autonomy to educate other content, skills, and topics is waning. Teachers adjusted their curriculum and placed greater attention on tested materials to produce better results for students and schools on standardized tests. Due to this, non-tested material falls subject to less focus for teaching (Dee et al., 2010; Pedulla et al., 2003). Student learning and outcomes are also affected by standardized testing...For example, many standardized exams use multiple choice as the primary exam method; this allows students to learn how to create better test results even though they may not know the correct answers. For example, testers can eliminate or guess correct answers (Turhan, 2020).

Pedulla et al. (2003) investigated the effects of high-stakes standardized testing on teaching and learning. The research team collected surveys on teachers' attitudes toward state testing programs. A Likert scale was used for most of the 80-question survey data. It also contained a written answer section where teachers could place statements. The survey targeted nine different topics, but for this study, we will discuss the findings of two. "1) Relationship of the mandated test to the state curriculum frameworks and standards. 2) Perceived effects of the state-mandated test" (Pedulla et al, 2003). The researchers collected 4,195 kindergarten-12th grade teacher surveys from states across the USA. The data collected from high schools was twice that of elementary and middle schools.

The findings for the two relevant focus areas suggested that 1) teaching of tested material increased. In contrast, non-tested material and skills decreased (p. 61), and 2) Teachers did not believe that the state standardized testing accurately represented student education, especially English language learners. (p. 116). The research does not explicitly mention why these teachers believe this, but it may have to do with English learners having equal literacy access to questions

and answers. This is relevant to this study because it reveals that educational professionals do not believe that these standardized tests accurately measure all students' achievement (p. 129).

These findings are also problematic because the study reveals that tests are used as an accountability measure for teachers and students, and they may determine the futures of both parties (p. 100-102). The study also displays that teachers are spending more time in exam preparation of tested content than in other areas (p. 126). If educators had data showing that these tests measure academic and content competency correctly, they might be more likely to have more faith in these tests. Until then, educators will search for and implement methods they can use to gauge classroom learning.

Multiple-choice exams, used in many American schools and classrooms, have glaring assessment equity and accuracy issues (Reich, 2009). Reich conducted a study in New York State within one history class. 13 tenth-grade students were a part of the study, and they learned about World War 2 and the beginning of the Cold War. Reich gathered data using several methods. He gathered multiple-choice tests, think-aloud questions, observations, and interviews about question-and-answer selections from the study students.

Research demonstrated that scores on the multiple-choice tests varied from actual knowledge in some instances. Some of the students would guess the correct answers without knowing the factual information. Reich believed that literacy and test-taking abilities, which Reich referred to as test-wisness, were one of the main culprits to the variations and answer selections (Reich, 2009, p. 346). Reich's interviews with students after their exams revealed this information. For example, during the study, a student was able to narrow down multiple answers to a single answer based on context clues within the writing of the question and because of a map that was provided. However, during this student's conversation with Reich, it was clear he did

not know the material, even though his multiple-choice answer was correct. This depicts an equity issue in multiple-choice questions because not all students have the same testing-taking abilities or academic literacy. Unfortunately, the same equity issue emerges for the accuracy of the entirety of the exam style itself due to its makeup of these MCT questions.

Reich's work is relevant to this study because it assists in displaying that multiple-choice testing data is not 100% accurate at measuring student knowledge. It also reveals how critical academic literacy is to taking an assessment. Individuals can use context clues in the language of the question to narrow down potential answers. Furthermore, this highlights an equity issue embedded in the exam structure, as not all test takers have these vital skills. Additionally, it displays a need for a more accurate way to convey students' proper knowledge of topics.

Another study by Turhan (2020) further backs up these findings.

Turhan (2020) conducted a mixed-method research study in Istanbul, Turkey, revealing issues similar to Reich's study (2009). Though Turhan's study does not speak ill of the multiple-choice test, it does depict similar happenings in the exam method and questions. Interestingly, Turhan's study aimed to gather students' opinions on different question types. Question examples included: open-ended questions, matching, multiple choice, true and false. The study included quantitative data from 355 students and qualitative data from 12 students in a government-run school. The study participants were in 10th, 11th, and 12th grade. The information was collected between 2018-2019. Methods used to gather data included: The use of an Academic Motivation Scale, developed by another researcher Bozanoglu (2004), written short answers, true and false questions, and multiple-choice questions from students. Lastly, interviews were conducted with teachers.

The findings of this study revealed that students primarily preferred two types of

question types. The first was multiple choice questions because students could figure out the answers to questions even if they did not know the correct answer. (Turhan, 2020, p. 136). Students expressed that it might also be easier to prepare for this exam type. The second preferred question type by students was the open-ended question. Students claimed that these questions allowed them to score some points based on the knowledge they learned instead of not receiving a whole point on a specific question (Turhan, 2020, p. 136). Lastly, something worth researching more was that students who preferred the multiple-choice test said that if they were teachers, they would employ the use of the open-ended question type (Turhan, 2020, p. 139).

Though not conducted within the United States or a similar European nation, the study adds weight to this topic as the subjects of the study are in similar age ranges, grade levels, and were historically using similar testing methods that the United States implements (Turhan, 2020, p. 133). The study is also important because it reveals students' preferred testing methods and similar problems in the multiple-choice question/exam. Additionally, it shows that students prefer open-ended questions to show their knowledge of a subject or skill in a non-limiting way. This is relevant to this study because the oral examination aims to use open-ended questions to allow students to reveal their learned information on their own without the limitations of set answers.

These studies work together to tell a story of how multiple-choice questions and exams are not a full picture of student understanding of the curriculum learned. Test results can be skewed since students can narrow answers or potentially guess correct answers without knowing the material. These skewed testing results could show mastery of the curriculum when in reality, students do not have this mastery. Similarly, students who know they will be tested using multiple-choice questions instead of an in-depth testing method can adjust their study habits to



what they might see fit. These studies also showed that MCTs have equity issues for students who lack literacy and test-wiseness skills. EL Students, when the test is given in English, may not be able to understand what the question was asking, what answers mean, or read/understand the specific academic vocabulary used. Some language learning students may also lack test-wiseness tools based on their language understanding. These testers might be unable to pick up on content context clues to help answer questions or eliminate wrong answers. These together display an unfair disadvantage for testers not familiar with the testing language. One way for teacher examiners to get around literacy issues is to use an alternative testing method called the oral examination.

### ***Examiners, Examinees, and The Oral Examination***

Research indicates oral exams give testers a better chance at expressing their knowledge about content or skills over other testing types (Boedigheimer, 2015). This testing method also lent itself to examiners to nudge testers into giving more information, get students to clarify vague statements, and answer questions more thoroughly (Boedigheimer, 2015; Hazen, H. 2020). Additionally, from the teachers' perspective, this testing method also revealed the depth of knowledge and grasp of complex concepts testers had on the curriculum (Boedigheimer, 2015; Hazen, H. 2020). Finally, the oral exam may benefit students based on specific learning styles and educational needs (Huxham et al., 2012).

Oral examinations provide a means to better understand what testers truly know about curricula. Boedigheimer et al. (2015) populated scoring of oral exams, students' written work, student interviews, and informal observations into a study to explore if oral exams provided a richer measure of student performance and enhanced student skill. Research and data were collected over ten years on approximately 1,000 American Air Force cadets enrolled in upper-

division mathematics courses. Twenty-five instructors proctored the oral exams twice in each course. Students were provided with the test questions about a week before the exams. The instructors used two different rubrics to grade and provide scores. The first rubric was a general rubric for holistic grading. The second was to assign point values to specific exam problems/questions. Researchers analyzed the exam scores to collect quantitative data and used interviews and observations for the qualitative portion of the study.

The research suggested that students improved their scores on the oral exam over time (p. 102). Scores were also generally higher than written work and tests (p. 110). Unique to this exam style, testing scores presented a method to assess concepts and a deeper understanding of the curriculum (p. 113). Another unique aspect is that students get instant feedback and scores for their work (p. 107). Additionally, examiners could redirect students who needed clarification on questions or used methods to avoid answering the prompt questions (p. 108). Lastly, students who came to the first exam unprepared amended their study habits to prepare themselves for the next test better (p. 108).

This research is relevant to this study because it displayed some of the benefits of using oral examinations and how oral exams could be implemented in the testing environment effectively. It provided a large sample size showing that students could perform well on oral examinations. It also displayed how the oral exam can be continuously refined for both the instructor's and the student's benefit. For example, this type of exam can be individualized and repeated if necessary. This article also depicted how the oral exam is more equitable for students as prompts from the instructor could lend a helping hand and assist students in expanding or clarifying their explanations. Furthermore, it addresses more equity issues for literacy, as students have time to analyze the questions and ask for clarification for misinterpretations or

unfamiliar vocabulary. Oral examinations are not limited to the United States Air Force or mathematics courses; they can be applied in civilian student settings with differing content.

A second study focused on the student perceptions of oral examination was completed in 2020 by Hazen. This study analyzed the student's view of the testing process and the exam type's positives and negatives. The research was conducted over three years and involved 61 students in an upper-division geography class. Hazen used the mixed method approach to collecting data by gathering student rubric scores and questionnaires. The researcher provided 25 questions on class materials in advance to students. Hazen asked three of the 25 questions from the supplied exam list during the testing period. Following the exam, an anonymous questionnaire was given. This questionnaire included the use of the Likert scale.

The researcher concluded that students' general perception of the oral exam was that the test type depicted an accurate understanding of class materials better than a written format. Only five of 61 students disagreed with this. Two-thirds of the students also expressed that they could develop their ideas into more complex thoughts (p. 598-599). Similar to Boedigheimer et al. (2015), students expressed that one benefit was getting immediate feedback, corrections for answering the questions correctly and being able to answer questions in greater depth. (p. 603) Students also commented on the negative aspects of the oral exam.

Not all test-takers enjoyed the exam type. Testers who did not like the style indicated that the oral exam had some drawbacks. The negative aspects included: the test structure affecting their physical/emotional state, a challenging exam pace, and the structuring of their answers could have been better. These students' physical/emotional state was challenged because they believed the oral exam caused additional stress as it was more intimidating than other testing types (p. 601). Some said that the pace of the exam was difficult since they had to think quickly

and did not have sufficient time to analyze their answers and correct mistakes. Finally, some testers commented that the written exam, if given instead, would have offered a better method to structure answers more cohesively and comprehensively (p. 601). Hazen noted this but also commented that these opinions are not grounds to dismiss the use of the oral exam structure altogether (p. 601). The study, overall, communicated that there were far more positives than negatives.

Hazen's (2020) research is highly applicable to this study because it provided the exam structure, pretesting, and testing procedures and had similar data collection methods. For example, the implementation of preset questions provided to examinees ahead of time, the use of a rubric, and the subsequent utilization of a student questionnaire for qualitative data purposes were all used in my research. Besides these, the study also compared scores from the oral exam to a different exam type. One aspect Hazen acknowledges but does not entirely delve into is the ability to use oral-based exams for students with different academic abilities.

Huxham et al. (2012) contribute to research on oral exams as it compounds evidence supporting how this testing type reveals a greater depth of course knowledge. The study also gives merit to the claim that oral exams can assist differently abled students in reaching superior assessment scores. The researchers aimed to collect data to answer three research questions concerning the stress students feel towards the exam whether scores would differ between question/testing types, and student perceptions of the strengths and weaknesses of the oral assessment. Sixty-one students provided data for the research through questionnaires and rubric-graded test scores. The students were split into two groups and randomly selected for an oral or written exam. Upon request, testers could switch exam types. During the oral exam, ten volunteers were used as interviewers who administered the assessment. Afterward, all students

were provided with a questionnaire at the end of the exam.

The researchers concluded several findings like that of the other two studies. Students did better on oral exams than on written tests and they displayed a greater depth of knowledge. An additional similarity was that interviewers could clarify and rephrase questions when needed. The researchers commented that rephrasing questions made the exam more accessible to confused testers, creating a more authentic individual score. One new study finding suggested that students with reading disabilities could potentially reach a superior score with the oral exam as it offered greater accessibility to understanding and answering questions according to personalized needs. During the study, a dyslexic student who was originally selected for the written exam requested to change to the oral examination. This is an important finding because it might also apply to language learners. This student was not alone in switching to their desired testing preference.

A different tester switched to the oral assessment as well, while two switched to the written exam. This shows that some examinees prefer to take tests according to their unique learning or expression preferences. These change requests were significant because they provided a concrete example of the desire to express knowledge differently for test-takers to yield their best results. This speaks to Gardner's theory of multiple intelligences (2011).

Overall, these studies are essential because they demonstrate the ability of the oral exam to reveal more profound knowledge. They demonstrate the malleability of oral exams for examiners through the utilization of clarifying and rephrasing questions. (Boedigheimer et al., 2015; Hazen, H. 2020; Huxham et al., 2012) The research suggested that oral exams may be beneficial to students who have different academic abilities or learning preferences (Huxham et al., 2012). Lastly, not all test-takers enjoy the oral examination; some students expressed

increased anxiety levels because of the type of assessment.

## **Summary**

My action research study intended to measure the impact on SPED and EL students when a switch from a multiple-choice test to an oral discussion-based examination took place. Previous research supported the concept that the multiple-choice test had issues of accessibility for some students and problems in measuring knowledge or skills accurately. Additional research suggested that the oral assessment could be a viable alternative as it can indicate a greater depth of knowledge a test-taker possesses over the MCT. The oral discussion-based assessment also displayed a potential to solve a few of the issues of equity the MCT struggled with.

The theoretical framework used to support this research was primarily Gardner's theory of multiple intelligences (2011). This theory puts forth the idea that individuals learn and express knowledge or skills in different ways. Therefore, to assess an individual accurately, different testing methods need to be used according to the unique intelligence of the test-taker. The oral exam can serve as one more added testing method educators could employ in their classrooms to assess their students who are struggling with traditional multiple-choice tests. Furthermore, Bloom's taxonomy theory (1956) provides my research with different levels or categories of understanding for individuals. The last theory by Krashen and Terrell (1983) outlined the concept of different stages of language acquisition and use. Additionally, their work suggested that language learning students will acquire language best when an emphasis on communication occurs first rather than grammar. This was of particular importance to my research project as it bolstered the idea of discussions having power in the classroom as a learning and assessment tool.

## **Chapter III**

### **Methods**

Special education (SPED) and English language (EL) learning students do not meet state assessment standards (CAASPP, 2021-2022). Overall, this seems to be an issue of accessibility to the curriculum and assessments themselves (Gilmour et al., 2019; Miley & Farmer, 2017; Calderón, 2011). All public school students are expected to take these standard-based multiple-choice tests (SMCT) and perform proficiently. Many teachers use these multiple-choice tests in their classrooms to measure student knowledge. However, these multiple-choice assessments are riddled with inequities due to problems with literacy and test-wiseness and show inaccuracies in examinee knowledge (Pedulla et al., 2003; Reich, 2009; Turhan, 2020). These test issues create problems for the SPED and EL students making it more difficult to perform up to these state standards, showing major achievement gaps between these student groups and general education students. (CAASPP, 2021-2022) This has been the case for several years and is true for this school site (CAASPP, 2021-2022). According to the School Accountability Report Card (2022), 8.5% of students at this study's school site were English language learners, and 14.5% were identified as SPED. That means 23% of the school's population was disadvantaged when taking standardized multiple-choice tests. Considering these statistics and the testing data, the school site was challenged with teaching these learners old and new curricula, skills, and test-wiseness.

After reviewing extant literature, SPED and English language students could perform better on tests if allowed to display their knowledge using a different assessment type. Studies on oral examinations demonstrated testers had better access to the test and an understanding of what was asked of them. They also showed a greater depth of content knowledge (Boedigheimer et al., 2015; Hazen, 2020; Huxham et al., 2012). One scholar noted that a college student with dyslexia

opted to switch to an oral exam format when originally selected for a written exam. Thus, highlighting a preference towards the oral test based on the unique student needs (Huxham et al., 2012). The research on oral assessments was primarily conducted by university professors on the undergraduate and graduate levels, without focusing on particular groups of students. I plan to expand this research into a high school setting and see if oral assessments impact SPED and English learners' test scores. Hence my research question was, *How will shifting assessments from a multiple-choice test to an oral-based assessment better represent High School special education and English learning students' knowledge of United States History?* This chapter describes the setting and classroom makeup, the data collection strategies and procedures used to gauge exam performance, and the plan for data analysis.

### **Setting**

The high school in which this study took place was in the East Bay region of Northern California. It resides in a suburban neighborhood with only a few businesses in the immediate area, one mile or less, but had many within 10 miles. The town was once a rural agriculturally focused community, but more recently has been built up. More housing and businesses have been brought to the area allowing more people and families to move from urban areas to this growing town. The approximate average price for a home in the area was \$500,000 in 2022, and the approximate average household income was \$110,000 (US Census Bureau, 2022). Additional census data displayed that only about 7% of the city's residents were in poverty, lower than the 12.8% national average (US Census Bureau, 2022). These figures are significant when comparing the city's poverty rate with the number of socioeconomically disadvantaged students at the school. About 37% of students were in this category during the 2022-2023 school year. In the previous two years, these figures fluctuated between approximately 38% from 2020-2021



and 33% from the 2021-2022 school years (CDE, 2022). Therefore, in the 2020 school year, the site qualified for Title 1, Part A. Receiving the Title 1, Part A status means that the school qualified for some federal funding to assist socio-economically impoverished students (CDE, 2022). The funding from Title 1 Part A served roughly 950 enrolled students in 2021. By allocating these funds properly, the school has reached an impressive 98% graduation rate, well above the state median in the 2021-2022 school year (CDE, 2022).

School enrollment during the time of the study was approximately 2,500 ninth-12th grade students. The racial and ethnic makeup was approximately 1,200 Hispanic or Latino students (48%), 600 White students (24%), 250 Black or African American students (10%), 130 Filipino students (5%), 100 Asian (3%), 10 American Indian or Alaskan Native students (less than 1%), 10 Hawaiian or Pacific Islander students (less than 1%), 170 students who identified as two or more races (7%), and 30 students (1%) who did not provide racial or ethnic data. Based on these numbers, the school population is primarily Hispanic or Latino but is diverse in race and ethnicity.

School data also displayed that just shy of 200 students (8%) were English learners during this study. Using the data published most recently, the 2021-2022 English Language Proficiency Assessment for California (ELPAC) scores indicated the following proficiency levels for EL learners that attended the school: 10% were Level 1, minimally developed, 27% were somewhat developed, Level 2, 44% were moderately developed, Level 3, and 19% had a well-developed English Language proficiency, Level 4 (ELPAC, 2022). This is important to the setting as it displays that over 80% of English learners at the school site needed assistance in improving their proficiency. Another focal group of the school was special education students.

Almost 400 students were identified as needing special services for disabilities, making up 15% of the student body on campus.

In the 2021-2022 school year, the site employed over 100 teachers, of which 95 were fully credentialed teaching in their field, four were fully credentialed teaching outside their field, two were teachers without credentials, two were interns, and eight were unknown (CDE, 2022). The most recent demographic data on teachers revealed that in the 2018-2019 school year, there were approximately 70 teachers who identified as women and 50 teachers who identified as men. Approximate racial and ethnic percentages concerning teacher data were gathered from this same school year. The racial and ethnic makeup of the faculty was as follows: 80% White or Caucasian, 9% Hispanic or Latino, 3% Black or African American, and 1% were two or more races. The data displays that the teaching staff does not racially or ethnically mirror the students that they are serving. Though there were differences in demographics, teachers, alongside administrators, counselors, and staff worked diligently to prepare all students for college attendance after high school. In 2023, over 250 of the roughly 500 senior students met the A-G requirements in order for them to attend a four-year college, either a University of California (UC) or California State University (CSU), directly upon graduation. This impressive school site is only one of three sites in the district.

The school is one of three comprehensive high schools in the district serving two towns and three unincorporated areas for a total population of roughly 130,000 in 2020 (US Census Bureau, 2021). Two other schools make up the entire district; one was an alternative school, and the other managed students on independent study. In 2022, the comprehensive schools collectively had an enrollment of nearly 8,000 students and a graduation average of 97% (CDE, 2022). Due to the global COVID pandemic, college readiness for students in California was not

measured between 2020-2022; however, in 2019, the district average for students' readiness for college was 57% (CDE, 2022). Thus, most of the student population from all three sites were ready for college after attending one of the three high schools.

### **Demographics of the Classroom**

I conducted my research in three different United States (US) history classes. This course was a high school graduation requirement, meaning that all students who attend school must pass the class if they intend to pass high school. This also meant that students took this class more seriously than one of their elective courses. Overall, there were 91 students in the three sections combined. Out of the 91 students, 16 were designated as special education students, and 16 were designated as English learning students. Two students were identified as being both SPED and EL students. Most students were in 11th grade, although seven were in 12th grade. The US history classes were similar in content but different in terms of who they served. I will refer to these different periods with pseudonyms. Period A was a general education US history class, Period B was a co-taught US history class, and Period C was a US history class with English language support. Period A was the largest class, with 35 students, Period B had 29, and Period C had 27 students. Each period is discussed below individually.

Period A, the general education United States history course, had the largest grouping of students. The student-to-teacher ratio was 35:1. This class had 31 eleventh-grade students and four twelfth-grade students repeating the history course. It had a mixture of native-English-speaking students, English-learning students, and special education. School data was used to collect most of the information on gender. Modifications were made if a student requested to be identified differently than what was initially given to the school site. There were 19 students who identified as men and 16 who identified as women. There were also a variety of races and

ethnicities present in the classroom. The racial breakdown of this class was: 16 White students, seven Black or African American students, five Asian or Pacific Island students, two Alaskan Native or Native American students, and five declined to state. Additionally, our school has a separate marker for indicating individuals who are ethnically Hispanic. Eighteen students identified themselves as ethnically Hispanic.

Period B was a co-taught US history course with me and an additional teacher. The second teacher held a special education credential for mild to moderate special education students. This co-taught class normally had a cap of 28 students set by the district, but a special expectation was made for an additional student bringing the total up to 29 students. This made for a unique class as it was still a smaller class size, down from 35 students. This allowed the student-to-teacher ratio to drop from 35:1 to 14.5:1 since there were two teachers. This co-taught class was designed to meet special education students' needs within a general education history setting. Period B also had more special education students than a general education classroom. More support, such as scaffolds or modifications, were made for students who struggled with class content and skills needed to engage in lessons fully. One student in this class was identified as both special education and an English language learner. There were 18 students who identified as men, 9 who identified as women, and two who used the pronouns, they/them. The racial breakdown of this class was: 13 White students, seven Black or African American students, six Asian or Pacific Island students, one Alaskan Native or Native American student, and two that declined to state. Additionally, ten students identified themselves as ethnically Hispanic.

Period C was the US history class that supported English language learners. All the EL students in the course were either Level 3 or Level 4 per the ELPAC score, except one student who was a Level 1. A few students in the class had also been redesignated to be fluent in

English. Though the course is designed to assist ELLs especially, the school also places general education students proficient in English in the class. It is important to note that these general education proficient students were never identified as English learners. Though these redesignated and general education students were placed in the course without a choice, they assisted the EL learning students quite a bit by offering them support throughout the year. I.e., helped with vocabulary, read text, when necessary, worked on projects together, and ultimately conversed about materials, content, and skills pertinent to the class. One student in this class was identified as both an English language learner and special education. There were 18 students who identified as men and nine who identified as women. The racial breakdown of this class was: 12 White students, three Black or African American students, four Asian or Pacific Island students, and three Alaskan Native or Native American students. Additionally, 15 students identified themselves as ethnically Hispanic.

All students in the three US history classes were chosen for the study as I needed to gather as much data on general education students, special education (SPED), and English learning (EL) students as possible. Though my focus was on SPED and ELs, general education students were included for comparison. A letter was read in class and sent home to inform students and families of the study. The letter specifically included language that no names or unique identifiers were to be used, and therefore I could include every student in the study. I also allowed time for questions from students and made myself available to answer emails concerning the study.

### **Data Collection Strategies**

Several instruments were used to collect data for this study to gauge if shifting student assessments away from multiple-choice and towards oral examinations impacted the depth of

knowledge and scores. These tools were used by students, teachers, and sometimes both. I implemented a quantitative data collection method but gathered evidence using a variety of instruments. A multiple-choice exam (see Appendix A), an oral exam question set (see Appendix B), and a grading rubric (see Appendix C) were used together, and a before/during/after questionnaire (see Appendix E) that used a Likert-type scale offered quantitative data. As for the qualitative data, an observational instrument called the conversational tally sheet (see Appendix D) was used. The first to be discussed will be the multiple-choice exam.

### ***Multiple-Choice Exam***

To collect baseline data, a multiple-choice exam (see Appendix A) was created by myself and my Period B co-teacher. However, the California State Standards for 11th grade US History were used to guide the creation of the questions. The multiple-choice exam was on the Great Depression. A study guide was also created and provided to the students before the examination by me and the Period B co-teacher. Before the multiple-choice test was given and graded, a working day was given to the students to work on the study guide and prepare for the assessment. This provided the students with an opportunity for assistance in answering some of the study guide questions that would appear on the exam.

The assessment questions were placed in a packet, and an individualized blank scantron was created for each student. The exam also utilized closed-ended questions where one question and multiple answers were provided. However, only one correct answer would appear out of these numerous answers. There were 40 questions given on the exam. This test included 23 multiple-choice questions, a small section of six true/false questions, and a section of 11 matching questions. These were given because it is a common practice on other standardized multiple-choice tests.

Ultimately, the multiple-choice exam was useful for collecting baseline data from my classes on student performance for the multiple-choice exam. These data were needed to compare against the oral examination rubric scores. Together, these two assessments were used to determine if there was an impact on the expression of student knowledge and scores when testing methods were modified.

### ***Oral Examination Question Set and Rubric***

I developed the oral examination question set (see Appendix B), but the Period B co-teacher made adaptations and corrections. This was a question set that included 15 questions used on the oral exam. Though the question set contained 15 questions, ultimately, only 14 were selected from the set to be used due to time restrictions for the unit and research. Of the remaining 14 questions, question 1 was used for all students as a control question, the student selected one question, and the teacher selected one question for a total of three questions used for scoring. The purpose of the question set was to provide the students with a tool to prepare for the exam. Data concerning the question set was collected using the exam questionnaire exit ticket.

Once the oral exam question set was developed, a rubric was needed to measure testing results. The oral exam rubric (see Appendix C) was the third instrument to be taken and developed. This rubric was originally taken from Hazen and Hamann's (2020) study and then adapted to suit the needs of high school students and this research. The amendments were made because the initial rubric was too vague for high school classrooms. The edits to the rubric can be seen in each column, where I added information about the student's use of evidence and how their provided answer connects to the question and other topics previously discussed in class. Again, the rubrics' primary purpose was to give a numerical score to compare to baseline data.

### ***Conversational Tally Sheet***

The conversational tally sheet (see Appendix D) was created next. I developed this tool independently but it was commented on and amended by Period B's co-teacher. This was manufactured to track the number of times students demonstrated a particular trait during class conversations. Traits that were tracked included: preparation and materials brought to the discussion, the number of times specific evidence was used like unit vocabulary, the number of times examinees connected their answer to the question or an additional unit topic, and the number of times the examinee needed a teacher nudge or assistance to move the conversation forward. This tally sheet instrument was used both by the teacher and the student for practice leading up to and on the exam. It was presented and taught as a positive instrument for the students to feel more comfortable and confident in their answers rather than feel they were receiving degrading remarks on their discussion-based oral examination. The students used the tool with a peer leading up to the exam to practice answering questions and receive meaningful feedback to improve their conversational skills before heading into the high-stakes exam with the teacher. The reason behind the implementation of the tool was to get the students used to a person documenting their conversations on paper and ease anxiety about the oral examination.

### ***Oral Examination Questionnaire Exit-Ticket***

The last instrument to be developed was the oral examination questionnaire (see Appendix E). This was created again independently by me and given as an exit ticket. The questionnaire consisted of 10 questions broken up into three sections, *Before the Exam*, *During the Exam*, and *After the Exam*. Each question utilized a Likert-style scale to gather a response. This meant that each question was a statement, and students were asked to select a single answer that best represented their attitude toward the statement. All statement responses had five choices



ranging from “Strongly Disagree to “Strongly Agree,” with a neutral choice, too. The options and values are as follows: “Strongly Disagree” (1 point), “Disagree” (2 points), “Neither Agree nor Disagree” (3 points), “Agree” (4 points), and “Strongly Disagree” (5 points). Each of these point values was used to calculate relevant statistical data afterward.

The first section, *Before the Exam*, included three questions to determine student attitude and preparedness for the oral assessment. This section included statements such as, “I felt confident going into the discussion.” The following section, *During the Exam*, had two questions and measured the examinee's nervousness and comfortability as the assessment was conducted. These statements were “I became more nervous with the type of testing as the exam was going on.” and “I became more comfortable with the type of testing as the exam was going on.” The statements aimed to gather information on the possible confidence levels as the exam proceeded. The final section, *After the Exam*, included five questions. This set of questions aimed to collect information post-test on student perspective of the oral examination method, whether students thought the oral exam matched their learning styles, and their preferences on assessment types, multiple-choice and written tests versus the oral examination. An example statement from this questionnaire portion was, “I think the discussion exam fits my learning/testing style more than a multiple-choice or written exam.”

Through this tool, I aimed to obtain student perspectives and insight into their attitudes concerning the oral examination before, during, and after the final assessment. A secondary purpose of the questionnaire was to gather student data on the preference for an oral examination versus the traditional testing methods like the multiple-choice test or written assessment. This tool also gathered student thoughts on the fairness of the oral exam gauging their understanding of material and how well it matched their learning style or not.

## **Procedures**

This study took place over eight weeks. Most of the time was spent developing the instruments used in this study and teaching the unit on the Cold War. The study occurred in the third and fourth quarters of the school year. This was from early March to mid-May. The study consisted of three phases. The first phase was the pre-intervention phase, where I collected baseline data using the multiple-choice exam (one week). The second phase, or the intervention phase, was when the oral examination was introduced and discussed with each class, the teaching of the Cold War content took place, peer discussion practice, and the oral examination (six weeks) happened. The last or post-intervention phase included using the oral examination questionnaire exit ticket (one week). The post-intervention phase was critical to gather student perspectives on several aspects of the research. The three intervention phases will be discussed in detail below.

### ***Pre-Intervention***

Before beginning the intervention, all three US history classes were conducted normally. No major changes to the curriculum, skills, or testing methods were made. This was done to ensure that the data gathered at the end of the unit would be similar to any previous units or tests. Two units before the intervention phase was on the topic of the Great Depression. The Great Depression unit was taught using lessons developed in prior years but modified slightly according to the student needs in the current year and classroom. At the end of the unit, a study guide was given to students before their examination. They had a working day to complete the study guide individually, with a partner, or a larger team. Opportunities to ask questions or clarify information were given at the same time. Additionally, students were told they could use their study guide on the exam if they completed their study guide.

The following week, the multiple-choice test was administered. The students were given a testing packet with questions and asked to enter their answers on a scantron. All students took the examination during the class except absent students and students with special accommodations as outlined in their Individualized Education Plans (IEPs). These students, upon request, took the test in the school's Testing Lab. This was a supervised area separate from the general education classroom. The test was securely administered in this room by a paraprofessional. The tests were brought to this room before the students took the exam. The exams were then delivered to me after the completion of the test.

Password-protected software called *Illuminate* was used to grade the exams. This software scanned students' responses and compared them to a test key I had created before the scan. Once scanned, *Illuminate* provided which answers were correct and incorrect and the overall score. I used this software to mark the scantrons accordingly and show each student their total score in person or via the grade book software, Canvas. Finally, all data from this examination were downloaded into Excel files onto a password-protected computer to retain the information for future comparison.

### ***Intervention***

After the school returned from Spring Break, the classes began a new unit of study, the Cold War. The Cold War unit was taught to students in all three periods based on prior year lessons. Only small modifications were made based on the necessity of the student needed. At the start of the unit, the exam question set was also given to the students. Students were encouraged to use the question set throughout the unit as we discussed and learned each topic and subtopic of the unit. Toward the end of the unit, students were given a workday without new content to answer questions thoroughly. They were allowed to complete the assignment

independently, with a partner, or larger group. During class, students were provided opportunities to ask clarifying questions or clear up misunderstandings. On a different day and after direct teaching, students were given the conversation tally sheet and instructed on how to use it. They were then asked to find a partner and practice answering the oral exam questions from the given set. The students became familiar with the tally sheet and understood how to provide meaningful feedback to their peers for conversations. An additional opportunity to ask clarifying questions or comment on the tally sheet was provided but not taken advantage of.

The following week, students were asked to join the teacher for their independent discussion-based oral examination on the Cold War. They were told they could bring notes, their question sets, and anything else they had prepared for the exam. Test-takers sat away from peers but still within the same room. Students were then tasked with answering three of 14 questions. They were encouraged to use as much evidence as possible and connect their answer to any relevant unit topic. The examiner used the conversational tally sheet during the exam to keep track of evidence and unit connections the students made and the number of times they required a nudge or teacher assistance to keep the conversation moving in the correct direction to answer the question in an in-depth way accurately. The conversational tally sheet was used during all three questions. After the student left the testing area, a score was decided for each question using the rubric. The student was then asked if they wished to see the rubric score for each question or if they preferred a total score to be posted into their online grade.

As for the exam, all students were asked question 1; “The Cold War was a war of ideas. Can you explain a conflicting political idea between the USSR and the USA?” Question 1 was used as a control question as it was very general and could be answered using many pieces of evidence. Additionally, it was a lower-end question on Bloom’s taxonomy of higher-order

thinking and thus did not require an enormous amount of critical thinking. Students were given a choice of the remaining questions for their second question/answer. The idea behind this was to see if they would naturally navigate to higher-order thinking questions and if their confidence in answering questions affected their ability to provide evidence and impacted their score on the self-selected question. Lastly, a third final question was selected by me. The chosen questions required a demonstration of deeper understanding and sat higher on Bloom's taxonomy.

### ***Post-Intervention***

The questionnaire was explained and provided to students during the class period following the oral examination. If they needed assistance in reading the questions, I offered individual assistance. Students were instructed to answer honestly and to the best of their ability. They were also asked to put their names on the questionnaire for informational and disaggregation purposes. Clarification was provided that their names would not be passed along to the researcher or used in the study in any way. After the questionnaire was completed, the questionnaire was collected by me when students indicated completion with a raised hand. Questionnaires were then placed in a locked cabinet. Each questionnaire answer was documented in a Google Spreadsheet to be disaggregated based on the needs of this study and further analyzed.

### **Plan for Data Analysis**

Each instrument was used to address the research question, *How will shifting assessments from a multiple-choice test to an oral-based assessment better represent High School special education and English learning students' knowledge of United States History?* Study participants took a multiple-choice question exam to provide baseline information (see Appendix A), then ultimately took a discussion-based oral examination based on the question set provided to them

(see Appendix B) and scored through a rubric (see Appendix C). Scores for all students were gathered from both test types. Focal group data were then disidentified and disaggregated. Following this, Google Sheets was used to find the mean, median, and mode using the data. This data was then compared to display if a significant difference in testing results emerged. The post-exam questionnaire was also used to gather student perspectives on the oral examination. The questionnaire answers were transcribed into Google Sheets, and data were deidentified and further disaggregated. This questionnaire utilized a Likert-style scale for responses. The answers provided were translated into a numerical value of 1-5, thus allowing for statistical analysis. Students who were identified as being both SPED and EL students were included in the disaggregated data for both independent groups. For example, Student X was SPED and EL. Student X was included in the disaggregated SPED data. Student X was also included in the separate disaggregated EL data. I chose to do it this way as I did not believe two students would provide enough data to draw a significant conclusion. These methods were employed to gather the quantitative data element of the research.

Conversational tally sheets were retained to keep track of students' use of evidence and teacher intervention occurrences. This data also assisted in the research by revealing whether the students prepared for the oral assessment. Furthermore, grouping the data by SPED and EL learning students independently should display whether these students were ready for the examination or relied on materials brought to the exam for assistance during the discussion. Likewise, the tally sheets would show how often students used evidence, connected answers to other relevant topics, and required teacher assistance to redirect answers or prompt greater depth of knowledge. I looked for individual themes or trends that emerged in SPED, then English learning students, and then general education students. Once completed, I compared these

themes and trends to each other. Patterns that were identified are outlined in Chapter IV. These instruments and data collection devices were used to triangulate my data to make for a stronger research study.

## **Summary**

This research investigated whether changing student assessment away from traditional practices such as the multiple-choice and written exam methods and moving towards an oral examination style would impact student achievement for SPED and English learning students. Before the research, I observed that SPED and ELs needed help with traditional assessment methods, such as the multiple-choice test (MCT). Due to only having two students who were both SPED and EL students, I chose to include them in both data sets, SPED and EL independently instead of their own disaggregated grouping. I hoped this study's intervention would close the achievement gap for these learners and offer them a more equitable chance at test-taking. This study took place over eight weeks through three phases: pre-intervention, intervention, and post-intervention. The first week was dedicated to taking a multiple-choice assessment (see Appendix A) to gather baseline data. Most of the time, six weeks were spent in the intervention phase, where the oral examination question set (see Appendix B) was introduced, taught, practiced, and finally used for a student-graded test. The rubric (see Appendix C) and conversational tally chart (see Appendix D) were also introduced during this time. Finally, in the last week, participants completed a questionnaire (see Appendix E) regarding their attitude toward the oral exam.

This chapter discussed the large school and community setting and classroom demographics. It further introduced the research instruments used in the study and data collection/analysis methods. These were explained in detail with the replication concept in mind.

The methods used to gather data were specifically chosen to triangulate data for creating well-grounded research. The next chapter will explore the data obtained and provide a detailed analysis.



## Chapter IV

### Results

The purpose of this action research project was to study how the implementation of an oral-based exam over a traditional exam type would impact student achievement for various subgroups. Therefore, the question that guided my research was: *How will shifting assessments from a multiple-choice test to an oral-based assessment better represent High School special education and English learning students' knowledge of United States History?* The leading reason I chose to explore this topic was that I noticed my English language learners (EL) and special education (SPED) students consistently scored lower on multiple choice question tests in class. Additionally, these groups historically have scored lower on standardized state tests, which often use multiple-choice as a means to assess student achievement (CAASPP, 2021-2022). The reason many scholars and researchers hypothesize that this happens is a difference in testing skills known as test-wiseness and issues with student literacy (Pedulla et al., 2003; Reich, 2009; Turhan, 2020).

After reviewing the extant literature, SPED and EL students could perform better on tests if allowed to display their knowledge using a different assessment type. Studies on oral examinations demonstrated test-takers had better access to the test and an understanding of what was asked of them. They also showed a greater depth of content knowledge (Boedigheimer et al., 2015; Hazen, 2020; Huxham et al., 2012). These ideas are backed by major educational and psychological theorists like Bloom and Gardner. Howard Gardner (2011) claims that people have multiple intelligences and, therefore, learn and express knowledge differently based on their unique and personal characteristics. This bolsters the thought that not all students' achievement should be measured by the use of one standardized test. Likewise, these tests do not always

display the true depth of knowledge a person has (Reich, 2009; Turhan, 2020). By the application of Benjamin Bloom's taxonomy of higher-order thinking, a greater depth of knowledge can be revealed.

Bloom's taxonomy of higher-order thinking skills (HOTS) (1956) was also relevant to this research. In order to reveal how much a student understood about a topic, different levels of thinking were examined. The knowledge or simple memorization of information is the lowest level whereas the evaluation of that same information is the highest. In this theory, students who can evaluate data have a greater understanding of the material. Bloom's taxonomy of HOTS in combination with Gardener's hypothesis would mean that each individual would express their depth of knowledge differently based on their unique learning characteristics. For example, student A may need to evaluate information differently than student B, and though the information would be the same, the expression would be or could be different. Therefore, in this research study, a different means of assessment, the oral assessment, was used after the collection of data from a multiple-choice test to measure student achievement differences.

### **Overview of Methods and Data Collection**

This study took place over eight weeks. The majority was spent teaching the unit on the Cold War and assessment tools so the students would understand what to expect during the assessment. The study occurred in the third and fourth quarters of the school year. This was from early March to mid-May. The study consisted of three phases. The first phase was the pre-intervention phase, where I collected baseline data using the multiple-choice exam (one week). During the Intervention phase, or phase two, I taught the Cold War, introduced the idea of the oral examination, the students participated in peer discussion practice, and I conducted the oral examination. The intervention phase took approximately six weeks to complete. In the last phase

or post-intervention phase, I gave an oral examination questionnaire exit ticket. This took one week to complete with all class periods. This post-intervention phase was critical because it gathered student perspectives on several aspects of the research.

Several instruments were used to collect data for this study to gauge if shifting student assessments away from MCTs and towards oral examinations impacted the depth of knowledge and scores. I used the quantitative data methods to obtain the necessary data for this study. These tools were used by students, teachers, and sometimes both students and teachers. A multiple-choice test (see Appendix A), an oral exam question set (see Appendix B), and a grading rubric (see Appendix C) were used together. I also gave a before/during/after questionnaire (see Appendix E) using a Likert-type scale and an instrument called the conversational tally sheet (see Appendix D).

### **Demographics of Participants**

Participants in this action research study originally included all of my 91 students enrolled in my three United States history courses. However, because of scheduling changes, i.e. students moving schools or class transfers, and other conflicts, my participant pool dwindled to 84 participants. Fifty-three were general education students (63%) who participated in the study, 15 were special education students (18%) who participated, and 16 were English language learning students (19%) of the 84 participants. Their ages ranged from 16 to 18 years old. Although there were students who left the study, the participant population of the study remained quite diverse. The racial and ethnic makeup of participants were as follows: 7% Alaska Native, 17% Pacific Islander, 20% Black or African American, 46% White, and 10% Declined to state. Additionally, when our school collects demographic data there is an option to check that students are of Hispanic descent, separate from or in addition to the other groups above. Students who

were identified as having Hispanic descent in this study made up 45% of the participants. Lastly, 65% identified themselves as being a man and 35% identified themselves as being a woman. Adjustments were made from school-provided data upon student request for the use of different pronouns. This was done for two students.

### **Analysis of Multiple-Choice Testing and Oral Examinations**

Two tests were given to the participants in this study. The first was a 40-question multiple-choice test and the second was a three-question rubric scored oral exam. Student test scores for both assessments were calculated first in their unique styles and then awarded an alphabetic letter grade A-F. Once complete, each letter was given a simple numerical value A=4, B=3, C=2, D=1, F=0. This was done so the test score data could be analyzed and compared directly between the two exam types. The findings most relevant to this part of the study were mean scores for my students. Below, the mean score findings from each assessment and three subgroups are discussed individually first and then compared against each other afterward. The subgroups were general education students, special education (SPED), and English language learning (EL) students.

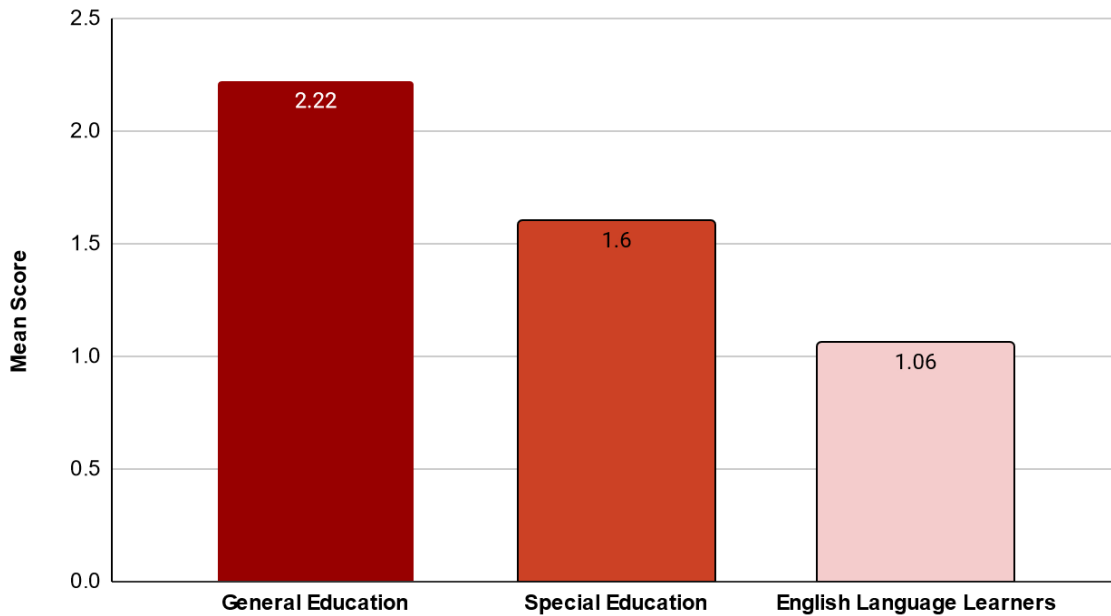
#### ***Multiple-Choice Testing***

A Multiple-choice test (see Appendix A) was given to students after completing a unit on the Great Depression. The multiple-choice test (MCT) included 23 standard multiple-choice questions with several options of answers to choose from, six true or false questions, and a matching section with 11 questions. I analyzed grade results by finding the mean scores of different groups relevant to the study. Figure 1 displays the mean score found for each of the subgroups, which was disaggregated from the full study group ( $N=84$ ). The data reveal that general education students scored better (2.22) than both SPED (1.6) and EL learning students

(1.06). Special education scored the second best in our three-group disaggregation of data whereas the EL students scored the lowest. The difference between mean scores for the highest and lowest scoring groups was 1.16. Furthermore, English learning students scored more than one letter grade equivalent (10%) lower than the general education students. Thus, illuminating the large achievement disparity between these two groups. These findings are consistent with standardized test results given by schools, districts, and the state of California (CAASPP, 2023).

**Figure 1**

*Multiple-Choice Test: Mean Scores of Different Student Groupings*



*Note.* Alphabetic grades based on a standard grading scale (A=100%-90.0%, B=89.9%-80.0%, C=79.9%-70.0%, D=69.9%-60.0%, F=59.9%-0%) were simplified to a numerical value to better display performance (A=4, B=3, C=2, D=1, F=0).

\*The number of students who participated in the study and received a grade for this multiple-choice test ( $N= 84$ ), general education students ( $n=53$ ), special education students ( $n=15$ ), and English learning students ( $n=16$ ).

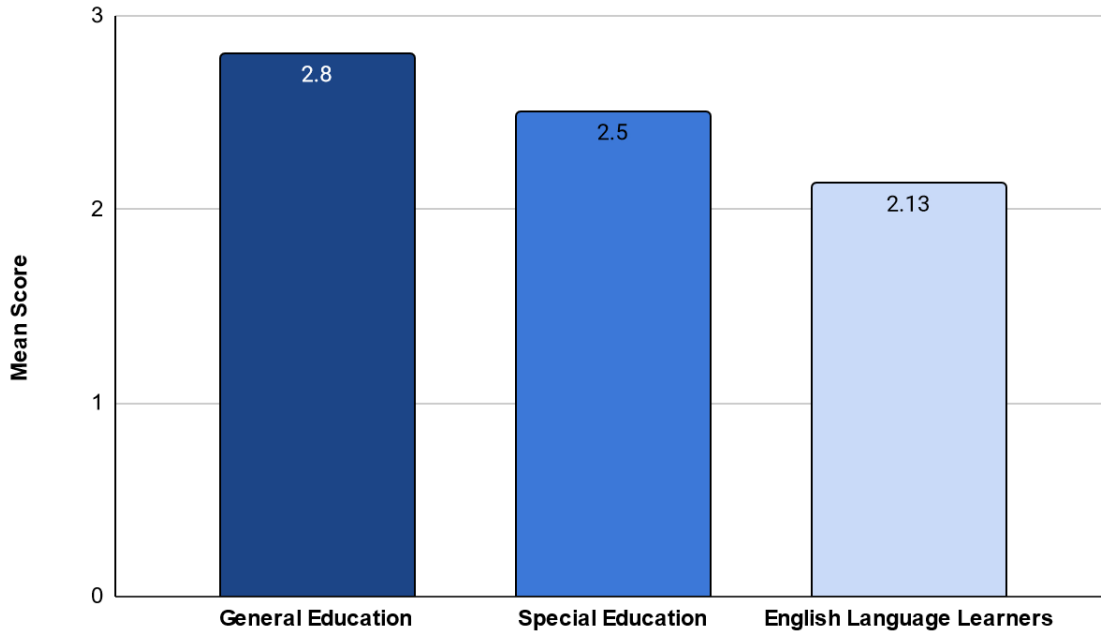
### *Oral Examination*

Figure 2 displays the disaggregated mean results of the oral examination given on the Cold War unit (see Appendix B). The 15 test questions on the exam were given to students at the beginning of the unit. Question number 15 was eliminated due to time restrictions, making the oral examination now out of 14 possible questions, 1-14. Three questions were chosen out of the remaining 14 at the time of the test and discussed with the student during the oral assessment. Question 1 was always given and the other two were randomly chosen for each student. Students were able to use notes or bring up their oral examination questions to the discussion table to assist them during the assessment. Students were encouraged to do this. Bringing notes or the question set did not negatively impact their grade. Grades were calculated for all three questions using a rubric. Rubric scores were then awarded an alphabetic letter grade and a corresponding simplified value (A=4, B=3, C=2, D=1, F=0). These were then used to find the mean scores for participant subgroups.

General education students, like the multiple-choice test, still scored the highest (2.8). They were followed by special education students (2.5) next and finally EL students (2.13) last. These mean scores display that each subgroup had a “C” average on the oral assessment, though general education students were closer to achieving the “B-” average than the other two subgroups. Additionally, though the EL group was in the “C” range, it was closer to a “C-”. Important to note as well is the difference between the highest and lowest mean scores for the oral exam. The difference was a mean score of .67, far lower than the multiple-choice test mean score differences between the same two groups. This will be discussed further in Figure 3. To gain a better understanding of how the two exams’ results faired against each other, Figure 1 and Figure 2 have been compounded into one larger figure, Figure 3, which is discussed below.

**Figure 2**

*Oral Examination Test: Mean Scores of Different Student Groupings*



*Note.* Alphabetic grades based on a standard grading scale from a rubric were simplified to a numerical value to better display performance (A=4, B=3, C=2, D=1, F=0).

\*The number of students who participated in the study and received a grade for this multiple-choice test ( $N=84$ ), general education students ( $n=53$ ), special education students ( $n=15$ ), and English learning students ( $n=16$ ).

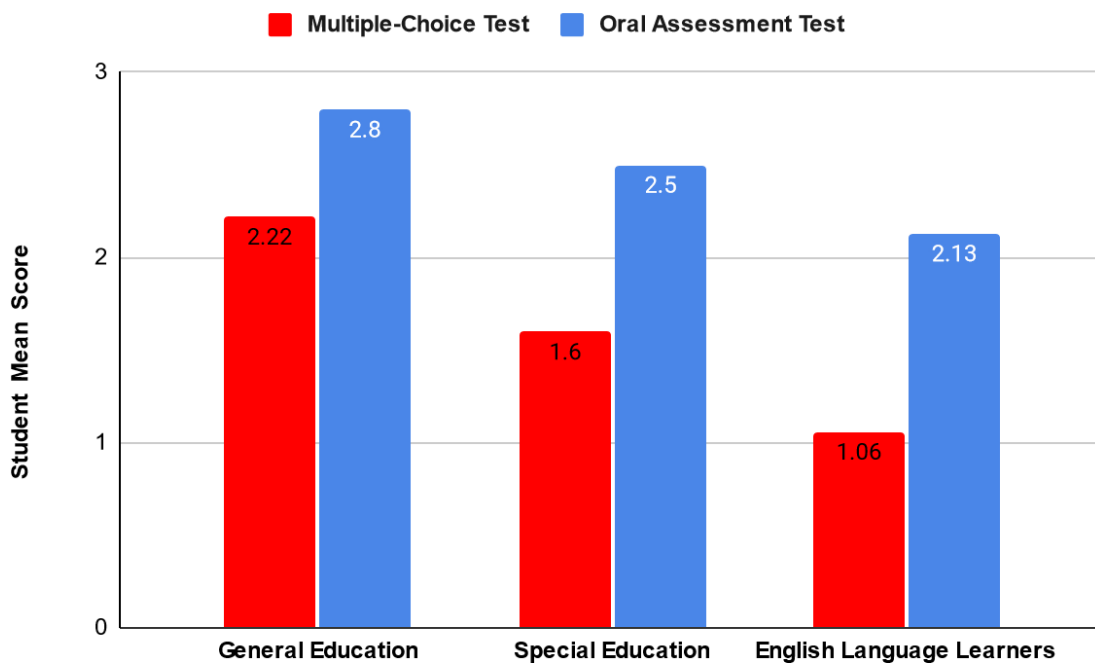
***Comparison Between the Multiple-Choice Test and Oral Examination***

Figure 3 displays the difference in mean scores between the two assessments given. Overall, the mean scores of every subgroup were higher on the oral examination when compared to the MCT. Mean scores increased from the MCT to the oral assessment for each subgroup in the following way: general education students by +.58, special education increased their mean score by +.90, and English language learning students increased their mean score by +1.07. The greatest increase in mean scores came from the English language learning students with SPED shortly behind them. This means the difference for EL students equated to one complete letter

grade increase; moving the groups' mean-based letter grade from just above a "D" to slightly above a "C" on a standard grading scale. Interestingly, the subgroup with the least mean score increase was the general education students. Though they scored the best on both exam types, the improvement from the MCT to the oral test was the lowest of all subgroups. If they had the same improvement rates as the others, they would have increased their mean score to a "B-" range.

**Figure 3**

*Multiple-Choice Test vs. Oral Assessment: Mean Scores*



*Note.* Due to the different scoring methods used on the two exams, raw score totals were converted into an alphabetic grade based on a standard grading scale and awarded a simplified numerical value (A=4, B=3, C=2, D=1, F=0).

\* The total number of students who participated in the study and received grades for both tests ( $N= 84$ ), general education students ( $n=53$ ), special education students ( $n=15$ ), and English learning students ( $n=16$ ).

The difference between the highest and lowest subgroup mean score for the MCT was 1.16 and the same subgroup difference for the oral test was .67, much lower. This means that scores between subgroups for the MCT were greater than one letter grade equivalent and less



than a letter grade for the Oral test. Additionally, all student subgroup means for the oral examination were within the same letter grade band of a “C” and equivalent to approximately 7% or less from each other. Together, these data highlight that the oral exam subgroup mean scores and letter grades were closer together displaying the shrinking of an achievement gap between subgroups over the MCT. For reference, I use a 10% band difference between each letter grade “A” through “D” (A=100-90%, B=89.9%-80%, C=79.9%-70%, D=69.9%-60%, F=59.9%-0%). My analysis, interpretation, and implications of these results will be discussed in greater detail in the concluding chapter, Chapter V.

### **Analysis of Conversational Tally Sheet**

Another piece of evidence and data that I collected was from the Conversational Tally Sheet (see Appendix D). This instrument was used to collect data while the oral assessment and discussion were being conducted. As students discussed the answers to their chosen questions, I added a tally mark under a certain category depending on what they were talking about. Tallies were collected for use of evidence, connections to other unit topics, and the number of times they required teacher assistance in the form of prompting for elaboration, expansion, or the rephrasing of a question to get them to answer on a deeper level. These data were collected separately for all three oral exam questions asked and totaled for each student. Data were then populated for all classes, disaggregated, and charted in Figure 4.

Figure 4 displays that almost all students gave approximately 13 pieces of evidence during their discussions. General education students' mean score of evidence revealed 13.94 pieces of evidence used, so they were close to breaking the 13-piece threshold set by all subgroups. Special education was the group with the second highest use of evidence (12.62), followed by ELs (12.31). English learners used the least amount of evidence of any group.

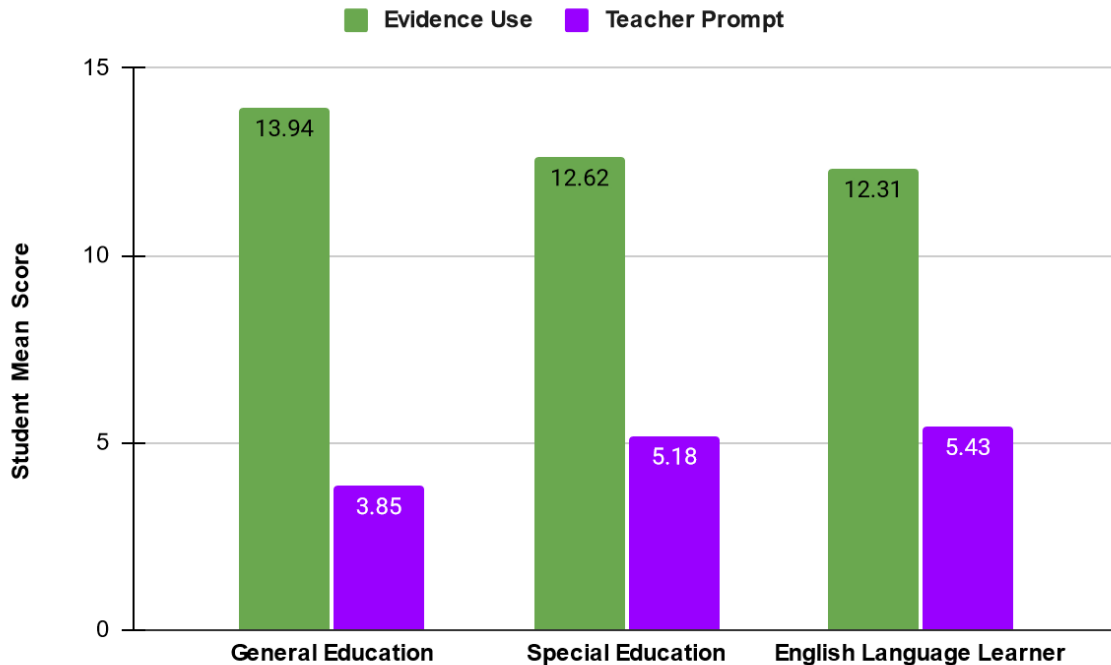
Students were allowed to bring self-created notes to the exam so the use of evidence for all students, I felt, was high. Also, I discounted any evidence misunderstood by the students.

Figure 4 also depicts how much teacher prompting, elaboration requests, or rephrasing of questions the different subgroups required. On occasion, some of the prompting asked for additional uses of evidence from students to support their claims. EL students required the most teacher intervention of any subgroup. Mean scores revealed they needed 5.43 separate instances of teacher assistance. Special education was the group that required the second most teacher involvement occurrences with a mean score of 5.18 times. Lastly, the general education subgroup relied on the teacher prompting the least with only a mean score of 3.85 times for teacher interventions. Overall, the SPED and EL students required roughly the same amount of teacher involvement occurrences.

The data highlighted in Figures 3 and 4 demonstrate that although EL students required the most amount of teacher prompting and redirection, they were also the subgroup that increased their mean score on the oral exam over the MCT test the most. This trend was also followed by special education and lastly general education students. Without this teacher prompting element for elaboration, redirection, and rephrasing, many of the SPED and EL students would have left their answers too broad and not revealed important knowledge or connections they knew about on the subject. These were critical to answering the questions properly and thoroughly. This was true for a few of the general education students too, however, most elaborated on their own.

**Figure 4**

*Student Use of Evidence vs. Teacher Prompt for Depth: Mean Scores*



\*The number of students who participated in this part of the study ( $N= 84$ ), general education students ( $n=53$ ), special education students ( $n=15$ ), and English language learning students ( $n=16$ ).

### **Analysis of Oral Examination Questionnaire Exit-Ticket**

The Oral Examination Questionnaire Exit-Ticket (see Appendix E) was given post-intervention but included questions for each of the phases of the study. Students reported responses using a Likert-type 5-point scale on 10 statements related to their attitude toward the multiple-choice test, the oral exam, their learning style, and the general impact of the intervention. Reporting options ranged from *Strongly Disagree* (1 point) to *Strongly Agree* (5 points). Also included was a *Neither Agree nor Disagree* (3 points). It is important to note that the students had earned grades for the assessments before taking the questionnaire, which may have influenced their responses. I analyzed the results by collecting responses from each period

and transcribing the data responses into the appropriate numerical value. Data were then combined and disaggregated into general education, SPED, and EL learning students. Once these groupings were completed, the modes were found for each Likert-type question. Due to papers with no names, the number of participants was reduced for this instrument ( $N=68$ , general education students  $n=31$ , students with disabilities  $n=11$ , and English language learning students  $n=14$ ).

For this piece of the research, I analyzed the mode for all questions given, however, I will discuss the five most relevant questions. Question 1 asked students, “I prepared more for the discussion exam than I would have for a multiple-choice test.” The general education students (53.1%) and English language learning students (57.1%) both “Agreed”. For SPED students (63.6%), their most common answer was “Neither agree nor disagree”. This was interesting because it suggests an indifference to the testing type. However, the findings are especially intriguing when comparing subgroup mode answers to Question 3, “I felt nervous because the exam was different from what I have experienced before.” All subgroups “Agreed” with this prompt. Even though the SPED subgroup was indifferent to studying for this new type of exam they “Agreed” that they were nervous before the exam. They were no more anxious though than the other two subgroups.

The mode findings for Question 1 when used in conjunction with the mode findings of Question 9, “I believe if another discussion test was given I would improve my score” offered similar interesting results. General education students and SPED both commonly responded that they “Strongly Agreed” with the statement. Though SPED responded neutrally to Question 1, Question 9 may suggest that both of these groups, especially special education, would know how

to better prepare for the next oral exam and maybe influence the future study habits of students identified as SPED.

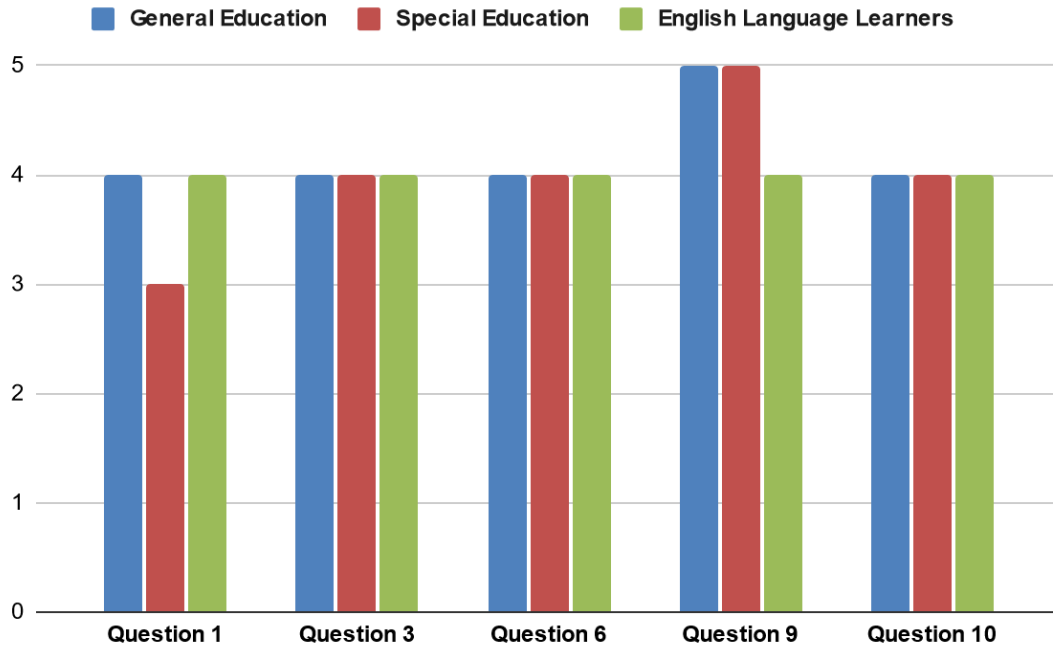
The mode response from all subgroups to Question 6, “I think the discussion exam fits my learning/testing style more than a multiple-choice test or written exam,” was that they *Agreed*. Although not every single participant responded that they “Agreed” or “Strongly Agreed” to this question, the vast majority did (85.7%). Only 14.3% of respondents “Disagreed” or “Strongly Disagreed” that the oral exam fit their learning style more than the MCT. This is important as it highlights the ideas of Gardner’s multiple intelligences (2011) and that not all students will learn or express knowledge in the same way.

Finally, mode responses from Question 10, “Although I would have preferred a different testing format (i.e. multiple choice, written test, etc.), I think the discussion test accurately measured my knowledge on the subject.” indicated that all subgroups “Agreed”. Eighty-seven percent of General education students “Agreed” with the statement. Whereas 81.8% of SPED “Agreed” and 71.4% of English language learning students “Agreed”. This question was of particular importance as I wanted to see if students believed the exam structure allowed them to express their answers completely, accurately, and to the best of their ability. These findings suggest that students thought they did.

Figure 5 below displays the modes for the disaggregated participant responses for questions 1, 3, 6, 9, and 10 on the Oral Examination Questionnaire Exit-Ticket. A score of five on the figure below indicates that a group’s mode was “Strongly Agree”. Four indicates “Agree”. Three indicates “Neither Agree nor Disagree”. Two is “Disagree” and one indicates “Strongly Disagree”. The majority of the responses to these questions were either “Agree” or “Strongly Agree”.

**Figure 5**

*Before, During, and After Questionnaire: Student Grouping Mode*



\*The number of students who participated in this part of the study ( $N=68$ ), general education students ( $n=31$ ), special education students ( $n=11$ ), and English language learning students ( $n=14$ ).

\* See Appendix E for a full list of the questions.

This quantitative data supports the idea that using a different method of assessment can benefit learners over traditional methods like the multiple-choice test. None of the mean scores for the subgroups were lower than the multiple-choice test, meaning that all groups scored better on the oral assessment than the MCT. These data support the notion that special education and English language learning students who take multiple-choice tests are likely to have a lower outcome or achievement than students in the general education population. It also bolsters the argument that oral assessments give more opportunity to display a greater depth of knowledge than other testing methods (Boedigheimer et al, 2015; Hazen, 2020).

## Summary

The purpose of this study was to determine whether changing exam styles would impact the achievement of SPED and EL learners. The intervention that I implemented was to change assessment forms from a multiple-choice test to an oral discussion-based examination. The data collection portion of the project took place over eight weeks and was collected through a variety of quantitative methods. Data instruments that were used included the following: a multiple-choice test, an oral exam and rubric, a conversational tally chart, and a before, during, and after questionnaire.

Quantitative data were collected in three phases a pre-intervention, intervention, and post-intervention phase. During the pre-intervention phase, baseline data on the multiple-choice test were gathered and analyzed. Mean scores for my three focus subgroups, general education, special education, and English language learning students, were found and compared to similar data collected during the intervention phase. In this phase, I collected rubric scores from oral examinations. I determined that the switch the oral discussion-based assessments increased the mean scores and achievement of all subgroups. The greatest increase in mean scores came from the EL student group, followed by the SPED group, and lastly the general education students.

Additional quantitative data from the conversational tally chart suggested that EL students required the most teacher intervention occurrences for redirection and prompting for elaboration than any other group. Though they needed the most occurrences, I determined they had the most growth because of this opportunity during the oral discussion-based exam. The tally chart data also found that SPED students required the second most amount of teacher intervention occurrences followed by the general education students. General education students

required the least amount of assistance during the exam but also showed the least amount of mean score increases of any group.

Finally, data from the post-intervention phase questionnaire revealed that most SPED individuals did not prepare differently for the oral discussion-based exam than a multiple-choice assessment but general education and EL students prepared more. Though most SPED students did not prepare differently, they felt as though they were nervous because of the testing method because they had not experienced it before. The same “Agree” response was popular for the two other groups too. Additionally, if given an opportunity to take another oral test, all subgroups felt as though they would increase their scores positively. Lastly, the questionnaire data also found that almost all students “Agreed” that the oral exam measured their knowledge accurately and fit almost all participants' learning styles except for 14.3% of questionnaire respondents.

In the next chapter, I will discuss the results and interpretations of this data. Results and findings from this study are compared to previous studies discussed in the literature review. Chapter V is the final chapter of this action research project and includes how this study might impact my future classroom and others. Finally, it includes a discussion of how this work has impacted me, and my plans as a transformative teacher leader.



## **Chapter V**

### **Conclusions**

As students enter American classrooms, more are identified as special education (SPED) students. Alongside these students, English language (EL) learners increasingly populate U.S. schools (CDE, 2022). These students face additional barriers to their education that general education and native English-speaking students will not encounter, such as having difficulty with reading and literacy, as well as accessing the curriculum and assessments similarly to students now identified as SPED or EL students (Calderón, 2011; Gilmour et al., 2019). On top of these challenges, every student will learn and express knowledge and skills differently according to Gardner (2011). His multiple intelligences theory, which has been revised several times, breaks down student learning and expression into different intelligence categories. Each individual has a unique makeup of intelligences which will determine how they learn and express information the best. Consequently, that means if an educator wants to teach and measure knowledge in the most meaningful way, each student should be taught and assessed using their unique set of intelligences.

This also means if classroom structures, lessons, and assessments remain stagnant, teachers will only serve a small portion of their student clientele. Furthermore, educators who can adapt to the unique needs of their students will have a better and more accurate understanding of student knowledge and skills. In order to gauge understanding of taught material and skills teachers rely on assessments regularly. Although some teachers use varying strategies for assessing their students, the multiple-choice test (MCT) remains a very popular option. From other research conducted, ultimately these MCTs were found to have glaring inaccuracies for scoring. They were inaccessible to certain students with literacy issues, and they

have loopholes that can be exploited by particular test takers who have prior knowledge or test-wiseness skills (Pedulla et al., 2003; Reich, 2009; Turhan, 2020).

Knowing that the MCTs have clear issues, I set out on a path to explore different options for more equitable assessment styles. I landed on the oral discussion-based assessment format due to the research I found conducted by other scholars. Research for the oral discussion-based exam has been conducted by numerous educators, but I focused on three in particular: Hazen (2020), Boedigheimer et al., (2015), and Huxham et al. (2012). These researchers concluded that oral discussion-based exams are effective at revealing great depth of knowledge from test takers, they provided an opportunity to students for redirection and teacher prompting if need be, as well as a noticeable increase in test preparation on the student's behalf. Additionally, students reported that they believed this test to be an accurate representation of their knowledge and they enjoyed the immediate feedback and grade from the testing style.

For my research, I honed in on students who may be at a disadvantage by the MCT test and could benefit the most from a switch in the assessment style. Therefore, I focused on special education and English language learners. When collecting data, I also gathered information on general education students to see how the intervention affected their achievement scores. A second reason why these students were included in the study was to compare the findings for each subgroup against each other.

To guide my action research project, I created a purpose statement and a research question. The purpose of this action research project was to study how the implementation of an oral-based exam over a traditional exam type would impact student achievement for various subgroups. Hence, the question that guided my research was: *How will shifting assessments from*

*a multiple-choice test to an oral-based assessment better represent high school special education and English learner students' knowledge of United States History?*

Chapter IV provided the findings of the action research project, the instruments used, and an outline of the phases from beginning to end. The findings indicated that all student subgroups, general education, special education (SPED), and English language (EL) learning students, performed better on the oral discussion-based examination over the multiple-choice test. The findings also revealed that the ELs showed the greatest mean score increase of any subgroup followed by the SPED group, and lastly, the general education students. This chapter is arranged into five sections: summary of findings, interpretations of findings, reflection on limitations, plan for future action, and the summary. The summary of findings section will include a discussion of the data discoveries from the instruments in all three phases: the pre-intervention, intervention, and post-intervention. In the next sections, I will give my analysis and interpretations of the findings. The third section will be a breakdown of the specific limitations of this study. This will be followed by a summary of the chapter. The final section will be a plan for future action in my practice because of the research I have carried out.

### **Summary of Findings**

For this study, I implemented a quantitative data collection method through three phases and used four instruments. The three phases included the following: pre-intervention, intervention, and post-intervention. The data-gathering instruments were a multiple-choice test, a rubric-graded oral discussion-based exam, a conversational tally sheet, and a before, during, and after questionnaire. These four evidence-collecting instruments were used to triangulate data to increase accuracy and validity and minimize biases in the findings of the study. I began the study with 91 students, but due to scheduling issues and student class transfers my participant count

was reduced to 84 students. Participants were from diverse backgrounds, genders, ages, grade levels, and learning abilities. For a complete analysis of participants see the Demographics of Participants section in Chapter IV. Once data were collected, I disaggregated them into three target subgroups. The subgroups I analyzed were general education, SPED, and EL students. The methods and timing of the tools will be discussed next.

The first phase was the pre-intervention where I gathered baseline information from student scores on a 40-question multiple-choice U.S. history test (see Appendix A). This tool was employed to gather baseline data before anything new was introduced. Next was the intervention, where I taught the Cold War, introduced the concept of the oral exam (see Appendix B), grading rubric instrument (see Appendix C), and any skills needed to complete this new assessment type. Students were also given the opportunity to practice discussions with partners during this phase, and they took the Cold War oral discussion-based examination. While the oral exam took place, I used a conversational tally sheet (see Appendix D) to gather the number of times the students used evidence, connected multiple unit concepts, and required teacher prompting or redirection. The last phase of the research was the post-intervention phase. During this portion of the study, I gave the before, during, and after questionnaires (see Appendix E). This measured students' attitudes toward the different testing methods, preparation for the oral exam, their learning styles, and accuracy of the oral assessment for gauging their content knowledge.

### ***Multiple-Choice Test Versus Oral Examination***

Analysis of the pre-intervention phase multiple-choice test (MCT) (see Appendix A) and intervention oral examination (see Appendix B & C) yielded several findings. Overall, the findings demonstrated that all student subgroups' mean test scores were higher on the oral

assessment than the MCT. The general education students scored the highest out of all subgroups on both assessments but had the smallest mean score growth increase when exam types were switched. These students increased their mean scores from 2.22 to 2.8, a positive average score increase of +.58. This equates to approximately half a letter grade better on the oral examination for my class. The mean score letter grade equivalent is as follows: A=4, B=3, C=2, D=1, F=0. The SPED subgroup's findings were also positive. This group had a mean score of 1.6 on the MCT and increased this to a mean of 2.5 on the oral exam. This was an increase of +.9 in the mean score. This is just shy of one complete letter grade increase on the exam for these students in my course. English language learners increased their scores dramatically. On the MCT, their mean score was 1.06; but on the oral assessment it was 2.13. This was a mean score increase of +1.07. This is the equivalent of just above one letter grade higher on the oral exam than the MCT.

Though the EL subgroup did not score the highest out of the groups tested, they displayed the most growth and the greatest improvement. This was followed by the SPED group and then the general education students. When using the mean score to find the average letter grades the EL and SPED groups moved from a "D" to a "C". The general education students remained in the "C" range for both exams but were just short of breaking the "B" threshold. They were .2 mean score points away from the letter increase. This assists in partially answering my research question as it displays that my focus groups can increase their test scores when given an oral discussion-based exam instead of an MCT.

### ***Conversational Tally Sheet***

The conversational tally sheet (see Appendix D) was used to collect data during the exam. The instrument gathered information on the number of times students used evidence,

connected information to different unit topics, and the number of times they required some type of teacher intervention such as prompting for elaboration, redirection of answers, or the rephrasing of questions. Students had the opportunity to practice with a partner and get used to how this instrument worked. They did this before the exam took place and had the chance to ask clarifying questions about it. One of the reasons this tool was implemented was to display to students that marks on a paper during the exam could indicate a positive attribute, therefore reducing the amount of anxiety or nervousness students felt if they saw me making such marks or tallies.

The results from the conversational tally chart depicted two major occurrences. The first is that almost all groups used approximately 13 pieces of evidence during the oral assessment. Mean scores for the subgroups are as follows: general education students 13.94, SPED 12.62, and EL students 12.31. The difference between the highest use of evidence and the lowest was 1.63 or just over one piece of evidence. The second major finding was the number of times each group required teacher prompting or assistance. Mean scores for the subgroups are as follows: general education students, 3.85, SPED 5.18, and EL students 5.43. The difference between the highest mean number of teacher interventions and the lowest was 1.58 or just over one and a half prompts, redirections, or rephrases. Analyzing the prompting data alongside the mean score results data from the switch of MCT testing to the oral assessment displays that EL students required the most teacher intervention occurrences but had the greatest score increase. This trend was followed by the SPED students, and lastly the general education population. This further assists in answering my research question as it shows that simply redirecting, rephrasing, or prompting for elaboration for SPED and EL students can reveal a greater depth of understanding

than the MCT. It also assists SPED and EL students with explaining their answers in unique ways that the MCT would not allow.

### ***Before, During, and After Questionnaire***

The before, during, and after questionnaire (see Appendix E) was used to collect information regarding the students' attitudes towards the different testing methods, their study habits, learning styles, and more. The tool utilized a Likert-like scale to gather participant responses. The potential answers ranged from "Strongly Agree" to "Strongly Disagree" with a neutral response of "Neither Agree nor Disagree". Due to no-name papers, there were fewer counted responses for this instrument ( $N=68$ ). General education student participants fell to ( $n=31$ ), SPED was reduced to ( $n=11$ ), and EL students down to ( $n=14$ ).

Results from this instrument yielded a few pieces of interesting data. When presented with question 1, both general education (53.1%) and EL students (57.1%) said that they prepared more for the discussion exam than they would have for an MCT. This may explain their increased mean scores in the oral exam over the MCT. This additional preparation may be the result of students realizing that they would have to prove their understanding of content material at a deeper level than other testing types. This assumption is supported by Boedigheimer et al. (2015). However, SPED students responded in a neutral manner to this same statement. Thus, SPED students displayed an indifference to the testing type. Additionally, participants were asked if the switch from the MCT to the oral examination made them feel nervous, question 3, and the most common response from all subgroups was "Agree". This finding was congruent with the works I found during my literature review.

The most telling finding from this instrument came from question 9, "I think the discussion exam fits my learning/testing style more than a multiple-choice test or written exam."

The two most common responses from all subgroups indicated that they either “Strongly Agreed” or “Agreed” (85.7%) with only 14.3% of respondents answering “Disagree” or “Strongly Disagree”. If this response was flipped, this result would also suggest that most of the participants who took this questionnaire felt as though the multiple-choice test would not have fit their learning/testing style best. This finding corresponds with Gardner’s theory of multiple intelligences where individuals should learn and express knowledge according to their unique abilities to best display their understanding (Gardner, 2011).

Question 10 asked if students felt as though the oral exam accurately measured their knowledge of the subject. Eighty-seven percent of general education students “Agreed” with the statement, whereas 81.8% of SPED “Agreed” and 71.4% of EL students “Agreed”. When question 9 and 10 responses are analyzed collectively, it displays those students preferred the oral examination and felt as though it was a fair assessment of their knowledge. These findings supported previous scholarly research of Boedigheimer et al. (2015), Hazen (2020), and Huxham et al. (2012) who purported that the oral examination could reveal a greater or equal depth of knowledge than other examination types. Additionally, the oral examination supports students who express knowledge in different ways or have disabilities that may impede test-taking when completing a traditional exam type like the MCT (Huxham et al., 2012). This instrument assists in partially answering my research question as it displays that more students will *prepare* for the oral discussion-based assessment than the MCT, which could be a factor in why certain subgroups increased their mean test scores dramatically. Additionally, it seems as though students' attitudes towards the oral exam are generally positive as they believe it measured their knowledge accurately.



## **Interpretation of Findings**

Based on the analysis of quantitative data gathered from four instruments over my eight-week action research period, I was able to draw the following conclusions regarding the use of oral discussion-based exams: it closed achievement gaps for SPED and EL students, students demonstrated a greater depth of knowledge than the MCT, mean scores for all subgroups were higher for the oral exam when compared to the MCT, and the oral exam offered more opportunities for student clarification, elaboration, and redirection. I was not able to conclude that oral examinations would eliminate all inequities for all students or that it fits all students' learning styles. Additionally, I was unable to conclude that oral exams contributed to more or less anxiety for participants.

### ***Closing Achievement Gaps***

Scores for special education and English language learners on standardized multiple-choice testing have found that time and time again these students fall short of predetermined benchmarks and perform lower than their general education peers (CDE, 2022; CAASPP, 2023). Multiple-choice testing is riddled with issues of inequities when it comes to literacy and test-wiseness skills (Reich, 2009; Turhan, 2020). Certain students are able to perform better on this type of test because of outside knowledge and context skills learned that others do not possess (Turhan, 2020). Meaningful change in the education or testing systems must occur to assist students who struggle with the multiple-choice test. The oral assessment may be one solution to shrinking or completely closing one achievement gap seen in traditional testing methods.

The overall scores for SPED and EL students increased dramatically on the oral examination over the multiple-choice test. Before the implementation of the oral exam, the achievement gap in my class was greater than one full letter grade between general education

students and those identified as SPED or EL. After the intervention, this achievement gap between these groups shrunk to approximately 7%, and all subgroups scored within the “C” letter grade range. A letter grade higher improvement for SPED and EL subgroups. If students can better express their knowledge for these tests, why not offer them in a variety of modalities? Gardner’s multiple intelligence theory (2011) discusses the importance of recognizing that individuals have one or more domains of intelligence. This means that people learn, and express knowledge based on their unique intelligence domains.

Therefore, students should be able to express their knowledge and skills in a way that fits them particularly if that means achievement gaps will be shrunk or closed completely. The oral examination has shown great promise as an accurate and fair testing method for the benefit of the teacher and student (Boedigheimer et al., 2015; Hazen, 2020; Huxham et al., 2012). It supports the development of better or more frequent study sessions (Boedigheimer et al., 2015), and supports multiple learning styles (Huxham et al., 2012). This markedly answers a major portion of the action research project's question of what implications switching to an oral assessment from a multiple-choice test would have on special education and English language learners.

### ***Demonstration of Greater Depth of Knowledge and Higher Scores***

Through the implementation of the oral discussion-based assessment, I was able to reveal a greater depth of knowledge from all subgroups and participants who took the exam. Every student required at least one prompt for redirection of answers, a prompt for elaboration, or they expressed the need for a rephrasing of a question. Without the oral exam testing structure, students are left to their own devices to read and answer a question properly even if they do not understand it completely. Thereby, in some cases, triggering a best-guess scenario if they were

taking an MCT and potentially answering correctly, creating an inaccuracy of true knowledge in their test score (Reich, 2009).

My research on the revealing of in-depth answers by use of oral exams for social science classes is congruent with Hazen (2020) and Hazen and Hamann (2020). For SPED and EL students the oral examination gave more accessibility to the test and allowed them to understand the level of performance needed to answer questions properly. Therefore, with prompts, they were able to explain their answers in greater detail and in ways that they could not express through a multiple-choice test. For example, if they could not remember a specific content or academic vocabulary word or phrase, they could describe what they were talking about in their own vernacular until I understood what they were attempting to say. Similarly, one of my students who was tested had dysgraphia, a disability that impacts the person's ability to handwrite legibly. This student performed very well on the oral exam and finished the test much faster than they would have had it been written. This is similar to Huxham et al.'s (2012) experience with a student who had dyslexia. I believe that the oral exam could continue to lend itself as a far more accessible testing method for other students who have physical or mental disabilities in the future.

Since the oral examination has shown that it can reveal more true and accurate test scores and greater depth of knowledge from test-takers it is natural to conclude that assessment scores will be higher. However, results may be equal to open-ended questions for a written exam format (Huxham et al., 2012). My findings suggest that, at least in my study, oral exams show promise over the MCT test.

### *Increased Opportunity for Student Success*

Due to the nature and structure of the oral assessment, there is an increased opportunity for student success (Hazen 2020; Huxham et al., 2012). Since this exam allows for direct contact between the student and teacher during the testing period, clarification of questions, redirection of answers, and prompting for elaboration are easily accomplished. As a result, students get real-time immediate feedback and can adjust their answers accordingly therefore getting a greater opportunity for success on the test (Boedigheimer et al., 2015). Boedigheimer et al. (2015) conducted a longitudinal study over 10 years on oral assessments for more than 1000 U.S. Air Force cadets. This study concluded that this type of exam gave “instructors considerable insight into students’ understanding, and, in general, students value the opportunity to demonstrate their abilities via this kind of assessment” (p. 99). The researchers also shared that cadet test-takers appreciated the immediate feedback and the opportunity, by design, to retake the assessment if necessary. Unlike the MCT, the oral assessment allows more than one correct way to answer a question by a test-taker and makes retakes much easier to manage for teachers and students alike.

Considering the structure of the exam, it gives support to those who struggle with literacy and language too. Both Hazen (2020) and I gave the test question sets early to participants. This allowed test-takers to study them, ask clarifying questions, and prepare notes or answers going into the assessment. This allowed these particular students, and all students for that matter, the time needed to process the language of the test leading up to the exam. This processing time is necessary to understand the message of what needs to be accomplished for the testing task. This is an important concept borrowed from Kashen and Terrell’s (1998) work regarding language acquisition. Additionally, students who struggle with presentations or anxiety could practice their

responses leading up to the exam day with others building confidence in their answers and themselves. However, this opportunity will not negate anxiety or nervousness for all test-takers.

Overall, the oral examination offers greater flexibility and opportunity for students to perform to the best of their ability. Since most research has been conducted with college level students or their equivalent, more research for oral examinations needs to be completed with high school students. Additionally, more research should be conducted on the oral examination with specialty subgroups to further the findings of this action research project and prior research studies.

### **Reflection on Limitations**

Five limitations to this study were apparent to me. The first limitation I recognized was the inability to test my student participants on the same unit topic using the two different exam styles. Due to the timing constraints of the data collection window and the rapid pace of the United States history class moves on our high school campus, I was unable to do the oral examination and multiple-choice test on the Great Depression and New Deal unit. I believe if this change took place, the study would be even stronger. There are several solutions to this limitation. One solution would be to split the class into two halves and have one section complete a multiple-choice test and the other half the oral examination. Another solution would be to have all students take both tests concurrently, during the time of one-on-one discussions with me other students could be completing the multiple-choice test in another space monitored by additional staff.

A second limitation of this study was that the oral examinations became very time-consuming. I had planned to complete all oral discussion-based exams over one 88-minute class period. However, the oral assessment took almost two complete 88-minute class periods to

complete, and four by-appointment makeups. Due to this, I delayed the beginning of my next unit of study for all of my U.S. history classes. This limitation and problem should be carefully considered for those who wish to replicate the study with their students if faced with time constraints by outside sources or plagued by large class sizes. A solution to this problem could be to select a smaller group of participants from multiple classes or approach the unit of study knowing that the oral assessment will take place over several exam days.

A third limiting factor I dealt with leading up to the oral examination was that some students were nervous about the change in assessment methods. Since some students had never taken an oral discussion-based exam they expressed anxious feelings to me. These students agreed to take the exam, but not be near their peers or during their specific class period as they felt vulnerable. A small but important accommodation for these students was made. I allowed these students to come in during a different class period if they wished, by teacher approval, or make an appointment for another time. I believe this instance will be the case for any teacher-researcher electing to replicate their study with students. One solution would be to have a written exam and an opt-in oral examination option for those who feel more comfortable with the exam discussion structure. Both the written exam and oral exam can be graded using the same rubric.

Another limitation of my data analysis is that I had an extremely small pool of students who were both special education and English learners. Two students fit this criterion. Since this grouping was so limited, I chose to analyze them as a part of both groups separately instead of as a part of their own overlapping group. I also chose to do this because I did not believe the data from two students would result in an expression of significant data comparable to the larger groups.

Lastly, I was both the teacher and researcher during this study which could have altered the responses given during the oral discussion-based exam or potentially the questionnaire given in the post-intervention phase. Unfortunately, due to the oral assessment structure and link to a student's grade, it is difficult to create anonymity when functioning as the teacher. One solution would be to have a colleague give the oral assessment and questionnaire, and then hand over grades and responses.

## **Summary**

Special education and English language learning students often underperform on standardized tests given by teachers, schools, districts, and government entities (CDE, 2022). These tests often employ multiple-choice questions as the means to collect information on knowledge and skills from students. This type of assessment, however, has issues with accessibility for some learners and often records inaccurate scores due to test-wisness skills by test-takers (Reich, 2009, Turhan, 2020). Moreover, individuals learn and express knowledge and skills differently through unique intelligence domains (Gardner, 2011). Educators and exam developers should keep this in mind when developing tests as one type will not fit all learners. Through the articles, books, theories, and research analyzed, I found that oral discussion-based examinations displayed promise to address the inequities of traditional testing methods while also offering a solution to inaccurate student test scores. Therefore, I decided to conduct research in my classroom on oral discussion-based exams. The following research question kept me focused during the project: *How will shifting assessments from a multiple-choice test to an oral-based assessment better represent high school special education and English learning students' knowledge of United States History?*

I enacted the plan by designing a three-phase action research project. In phase I, the pre-intervention phase, I assigned a multiple-choice test to my students. Phase II, Intervention, I taught the Cold War, introduced the concept of the oral exam, explained the tools and instruments we would use during the test, and had the students take the oral assessment. In Phase III, post-intervention, students completed a questionnaire regarding their attitudes about the individual tests, their test preparation habits, and learning styles, amongst other items.

The findings and conclusions of this study offer educators insight into the power of implementing oral examinations in their classroom over traditional testing methods such as the multiple-choice test. Results depicted that all subgroups of students performed better on the oral exam than on the MCT. Special education and English language learning students increased their scores dramatically and shrunk the achievement gap considerably between themselves and general education students. An additional conclusion I was able to draw was that the SPED and EL students utilized the supports and opportunities that the oral exam offered more than the general education group which is what likely contributed to their increased scores and overall success in minimizing the achievement gap.

Limitations of the study included the inability to test students on the same unit material for both the MCT and oral assessment, the oral exam became time-consuming, some students expressed feelings of nervousness about the change in testing methods, and that I had to assume the role of both the teacher and the researcher for this project.

### **Plan for Future Action**

The comprehensive findings of my research suggest that oral examination reveals a greater depth of knowledge for student subgroups when compared to the multiple-choice testing method. I plan on presenting my findings for the oral examination to my history department, the



professional learning community for English language teachers, and my site administrative team. My hope for these presentations is to educate and get other teachers and leaders on campus excited about the use of the oral discussion-based assessment. A major section at the beginning of the presentation will be dedicated to the theory of Gardner's Multiple Intelligences (2011). The idea of starting with this theory is to remind educators that each student learns differently and will express this information uniquely. Therefore, teachers need to seek out more ways to teach students according to their needs and find other assessment methods to extract taught information that will address inequities in testing. An additional section would include a few of the major pieces of literature I used for this research to depict the power and accuracy of oral discussion-based assessments.

A major movement at our school currently is offering differentiation for lessons and assessments to assist all students; however, this plan originated to assist special education students. Therefore, I believe that this research presentation and plan would align perfectly with our school goals. I will also reach out to our English language learning (EL) PLC teachers to see if they would be interested in learning more about the implementation of the oral exam and the use of discussions in their classrooms to strengthen the skills of these learners. Since the discussion-based exam requires practice and discussion between students leading up to the assessment, students can practice the use of basic language and academic language. Additionally, this would support the use of cognitive skills for EL students which would in turn benefit their overall language proficiency and growth as a student.

Secondly, because of my research, I would like to explore, implement, and add additional assessment methods into my teaching arsenal. I would like to do this to continue to address different learning and expression styles inside my classroom. To do this, I want to continue

researching Gardner's theory of multiple intelligences and find relevant articles on how to reveal intelligence domains from individuals. I am hoping to find credible software I can give my students at the beginning of the year. Doing so would allow me to tailor my teaching throughout the year to meet the needs of all or most of my learners as best as possible. I believe this will pick up where I left off by addressing more inequities and hopefully give a voice to a greater group of students in the classroom. An instructional method I am interested in pursuing for my classes is project-based learning (PBL).

A small goal of mine is to learn more about project-based learning (PBL) and how to effectively bring it into a classroom that has set pacing standards. In a PBL course, students learn content material and skills through the creation of a project over a set time. Usually, these projects try to tackle real-world issues or attempt to answer complex scenarios or questions. Projects can be unique, and the presentation of the final product can vary from individual to individual or team to team. I believe that this could be a natural progression for my courses and assessments as each student or group can present information in the way they see fit best. Furthermore, I suspect that PBL and history content can go together easily by attempting to link past to present movements, actions, events, or more together. This type of classroom would also lend itself to integrating technology into the classroom and allowing students to use that modality as they see fit. With PBL instruction, I foresee that pacing at my site could become an issue unless these projects have specific checkpoints and deadlines that are strictly followed by students and teachers. Likewise, projects would also have to have integrated benchmarks and required material for students to meet state standards and common assessment goals set by the district office.

Lastly, I would like to join my school site's Multi-tiered System of Support Team (MTSS). This team is responsible for identifying students in need of intervention and supporting them through anything that they may be dealing with. This team primarily focuses on the achievement of the student, however they do get other school teams involved, such as our school's Care team who can recommend or give professional assistance for social or emotional issues. The end goal of the MTSS team is to assist the students in passing their classes with good grades. One idea behind this team and the intervention is to get these students to their high school graduation and give them a shot at college should they choose that life path. Our site MTSS team is currently comprised of administrators, counselors, teacher-leaders, and other support staff.

By joining the team, I would be able to continue to mature as a transformative teacher leader on campus and actively work with colleagues and students. I believe that I could assist the team by introducing or reintroducing the idea of Gardner's multiple-intelligences theory (2011) and finding out how particular students learn best. From that point, I could work with the MTSS team, targeted students, and the student's specific teachers to adapt lessons or assessments for their unique expression. Of course, this would take a lot of work and teacher cooperation for this intervention to work; however, if it improves even one grade for the students, I will find it a success.

The student participants in this study were introduced to and completed an oral discussion-based examination, a valuable skill in the world both inside and outside the classroom. Due to the oral assessment intervention and movement away from the multiple-choice test, I determined that the students can be successful at revealing a greater depth of knowledge when allowed to express their understanding in different ways. If students continue to

practice these discussion skills, I believe they will be able to express a greater understanding and communicate their ideas better than those strictly studying for multiple-choice testing. I surmise that this study, the theoretical rationale, and the literature review have had wide-reaching impacts on how I approach teaching and they will continue to influence what I do in my classroom and beyond. In closing, the MTSS experience along with my other plans would extend my learning experience as a transformative teacher leader who is always open to growing, changing, and adapting in the name of better teaching and solving inequities seen in our schools.

## References

- Bloom, B. S. (1956). *Taxonomy of educational objectives: The classification of educational goals*, David McKay Company, Inc.
- Boedigheimer et al. (2015). Individual oral exams in mathematics courses: 10 years of experience at the air force academy. *Problems, Resources, and Issues in Mathematics Undergraduate Studies*, 25(2), 99–120. <https://doi.org/10.1080/10511970.2014.906008>
- Calderón et al. (2011). Effective instruction for English learners. *The Future of Children*, 21(1), 103–127. <https://doi.org/10.1353/foc.2011.0007>
- California Assessment of Student Performance and Progress. (2023) *English language arts/Literacy and mathematics: Smarter balanced summative assessments*. <https://caaspp-elpac.ets.org/caaspp/DashViewReportSB?ps=true&lstTestYear=2022&lstTestType=B&lstGroup=4&lstSubGroup=160&lstGrade=13&lstSchoolType=A&lstCounty=00&lstDistrict=00000&lstSchool=0000000&lstFocus=a>
- California Department of Education School Accountability Report Card. (2022). <https://www.caschooldashboard.org/reports/07617210730457/2022>
- Cambridge Dictionary. (n.d.). BIPOC abbreviation. In *Cambridge Advanced Learner's Dictionary & Thesaurus*. <https://dictionary.cambridge.org/us/dictionary/english/bipoc?q=BIPOC>
- Cambridge Dictionary. (n.d.). Multiple Choice Test. In *Cambridge Advanced Learner's Dictionary & Thesaurus*. <https://dictionary.cambridge.org/us/example/english/multiple-choice-test>
- Cummins (1999). Research news and comment: Alternative paradigms in bilingual education research: Does theory have a place? *Educational Researcher*, 28(7), 26-41. <https://doi.org/10.3102/0013189X028007026>
- Dee et al. (2010). The impact of no child left behind on students, teachers, and schools. *Brookings Papers on Economic Activity*, 2010(2), 149-207. <https://doi.org/10.1353/eca.2010.0014>
- English Language Proficiency Assessments for California. (2022). *Summative ELPAC*. <https://caaspp-elpac.ets.org/elpac/DashViewReportSA?ps=true&lstTestYear=2022&lstTestType=SA&lstGroup=1&lstSubGroup=001&lstGrade=13&lstSchoolType=A&lstCounty=00&lstDistrict=00000&lstSchool=0000000&lstFocus=c>
- Gardner, H. (2011). *Frames of mind: The theory of multiple intelligences*. Basic Books.

- Gilmour et al. (2019). Are students with disabilities accessing the curriculum? A meta-analysis of the reading achievement gap between students with and without disabilities. *Exceptional Children*, 85(3), 329–346. <https://doi.org/10.1177/0014402918795830>
- Great Schools Partnership. (2013). English Language Learner. In *Glossary of Education Reform*. <https://www.edglossary.org/standardized-test/>
- Great Schools Partnership. (2015). Standardized Test. In *Glossary of Education Reform*. <https://www.edglossary.org/standardized-test/>
- Hazen, H. (2020). Use of oral examinations to assess student learning in the social sciences. *Journal of Geography in Higher Education*, 44(4), 592-607. <https://doi.org/10.1080/03098265.2020.1773418>
- Hazen and Hamman (2020). Assessing student learning in field courses using oral exams. *The Geography Teacher*, 17(4). 130-135. <https://www.tandfonline.com/doi/full/10.1080/19338341.2020.1828137>
- Howard, T. (2001). *Why race and culture matter in schools: Closing the achievement gap in america's classrooms*. Teachers College Press.
- Huxham et al., (2012). Oral versus written assessments: A test of student performance and attitudes. *Assessment & Evaluation in Higher Education*, 37(1), 125-136. <https://doi.org/10.1080/02602938.2010.515012>
- Krashen, S. & Terrell, T. (1998). *The natural approach to language acquisition*, Prentice Hall Europe.
- McFarland et.al. (2019). *The condition of education 2019* (NCES 2019-144). U.S. Department of Education. National Center for Education Statistics. <https://nces.ed.gov/pubsearch/pubinfo.asp?pubid=2019144>
- Miley and Farmer, (2017). English language proficiency and content assessment performance: A comparison of English learners and native English speakers achievement. *Canadian Center of Science and Education*, 10(9), 198-207. <http://doi.org/10.5539/elt.v10n9p198>
- Pedulla et al. (2003). *Perceived effects of state-mandated testing programs on teaching and learning: Findings from a national survey of teachers* (ED481836). ERIC. <https://www.bc.edu/research/nbetpp/statements/nbr2.pdf>
- Purdue University. (2023). Special Education. *What is special education?* <https://online.purdue.edu/blog/education/what-is-special-ed>
- Reich, G.A. (2009). Testing historical knowledge: Standards, multiple-choice questions and student reasoning. *Theory & Research in Social Education*, 37(3), 325-360 <https://doi.org/10.1080/00933104.2009.10473401>

- Sayre, E.C. (2014). Oral exams as a tool for teaching and assessment. *The Journal of the Australian Science Teachers Association*, 60(2), 29-33.
- Turhan, (2020). Why do students prefer different question types? *International Journal of Progressive Education*, 16(3), 132-142. <https://doi.org/10.29329/ijpe.2020.248.10>
- United States Census Bureau, 2022. *Quick facts*.  
<https://www.census.gov/quickfacts/oakleycitycalifornia>
- United States Department of Education, National Center for Education Statistics. (2023, May). *Students with disabilities: Condition of education.*, Institute of Education Sciences.  
<https://nces.ed.gov/programs/coe/indicator/cgg>.
- United States Department of Education. (2017, May 3). *Sec. 300.114 lre requirements*.  
<https://sites.ed.gov/idea/regs/b/b/300.114>
- Zwiers, (2007). Teacher practices and perspectives for developing academic language. *International Journal of Applied Linguistics*, 17(1). <https://doi.org/10.1111/j.1473-4192.2007.00135.x>

## Appendices

### Appendix A

#### 1920s, Great Depression, New Deal: Test United States History

- How did the 1920s spending habits of Americans contribute to the beginning of the Great Depression?
  - American people only bought essential items to survive which slowed the economy and businesses down.
  - American people racked up huge bills and debt on credit that they couldn't pay back easily.
  - American people went out to party often; drinking, smoking, and dancing. This created too many businesses and there was not enough consumer money to support them
  - None are correct.
- Which definition is best fitting for the term: **Stock Market**
  - A place where accountants from banks work on an everyday basis.
  - A place where businesses can invest in new product ideas that are just being invented.
  - A place where people can buy/sell investments in publicly traded companies.
  - A place where people can use credit or get loans for items or large purchases such as houses.
- Louis Armstrong, Langston Hughes, and Jazz are all examples of people and things that were popular. They were also all associated with which larger event from the 1920s?
  - The Three Rs of the New Deal
  - Speakeasies
  - Talkies
  - Harlem Renaissance
- All of the following are new "freedoms" women experienced during the 1920s except for what?
  - They gained voting rights and could participate in elections.
  - Going to college more frequently to pursue a higher education and career instead of a husband.
  - They experienced new social freedoms (i.e. smoking, drinking, dance clubs, etc.)
  - They gained new careers like stock brokers, CEOs, company presidents, chief business managers, etc.
- What was the main trigger to the start of the Great Depression?
  - The banks collapsed
  - The stock market prices increased to an all-time high and normal people wanted to sell their stocks.
  - The stock market crashed
  - The bank foreclosed on many farms because farmers couldn't pay their bills.
- The American Great Depression causes many families to suffer tragedies, what was one of the common things that happened to families?
  - Men abandoned their families to "ride the rails" to look for work.
  - More children worked in factories to make up for the lost income in their families.
  - Women stayed home so the men could work.
  - Many families were supported by the government early on.
- The following are all causes of the Great Depression except?
  - People rushed to the banks and took their money out which collapsed the banking system.
  - The stock market crashed on Black Tuesday.
  - The President at the time told the country they did not have enough gold to support the US dollar.
  - Farmers had too much surplus, as a result of WW1, which dropped the prices of what they produced and they couldn't pay their bills because of it. The new deal would pay them to get rid of these crops later.



8. Dorothea Lange was a famous Photographer: she took this photo below to convey a message to the United States people. What was the message?



- a. To tell the nation that everyone was going through the same struggle
  - b. To display that the nation could survive this problem
  - c. To Show the strength of the nation and that rich people suffered too
  - d. To tell them to get over their hardships and get a job
9. What was the Dust Bowl?
- a. A series of giant dust storms across the Midwestern United States
  - b. A failed attempt to invade Germany during WWI
  - c. A failed attempt by President Hoover to save American People during the Great Depression
  - d. A series of small dust storms across the Eastern United States
  - e. None of the above
10. Who was responsible for the Dust Bowl?
- a. Several generations of farmers overworking the soil and a drought
  - b. President Herbert Hoover
  - c. President Woodrow Wilson
  - d. The new technology allowed one generation of farmers to extract nutrients from the soil.
  - e. Both B & C
11. How did banks work during the 1920s?
- a. People's deposits only sat in the bank vaults.
  - b. People's deposits were used for loans and stock investments by banks.
  - c. People's deposits are sent to the Federal Reserve for safekeeping.
  - d. People's deposits are always split between a checking and savings account.
  - e. Both C & D are correct.

12. All of the following are reasons why banks crashed during the Great Depression, **except** for what?
- The banks were invested in the stock market
  - People rushed the banks and pulled their money out
  - Banks poorly invested people's money and the investments failed
  - Money became hyper-inflated and banks shut down because their money was worthless.
13. The Hoover Dam was the largest project started before FDR took office. The dam project provided many things for the nation **except for one of the following**. Which does not belong?
- Construction jobs were offered to unemployed men.
  - Electricity for thousands of residents.
  - A model for working conditions, set up by the US government, for future businesses.
  - A symbol of the United States' power even in a time of depression.
14. What was the New Deal?
- A deal between the United States and France to provide food for starving US citizens
  - Herbert Hoover's plan to get people back into houses and fix the US banking system
  - FDR's failed attempt to "pack" the courthouse
  - FDR's radio announcements to calm the nation during the Great Depression
  - A series of Government Programs designed to "save" the US economy and people during the Great Depression
15. Why would FDR and the US government pay American farmers to burn crops during the great depression?
- To restore regular prices for crop sales
  - To force Americans into buying canned food because it stays good longer.
  - To reduce surplus/overproduction
  - Both A & B
  - Both A & C
16. The original Social Security Administration offered assistance to all of the following groups **except** which?
- Single men
  - Retirees
  - Families with dependent children (such as foster youth)
  - People with disabilities (mental or physical)
17. The WPA mostly focused on building what?
- Highways, bridges, and dams
  - Airports, trails, parks
  - Harbors
  - Hiking Trails
  - They didn't build things, they were artists
18. Hoover's Initial Response to the Great Depression was what?
- Immediate government action
  - Do nothing and let the economy run its natural course; eventually started the Hoover Dam project.
  - Provided emergency housing and loans for bank bailouts
  - He started the New Deal
  - None of the above

19. What was the Public's response to Herbert Hoover's Presidency at the end?
- Supported the President
  - Elected Herbert Hoover to another term
  - Didn't support the president and elected someone else
  - Loved President Hoover because of the New Deal and allowed him to serve six terms
20. What was the General Public's initial response to Franklin D. Roosevelt's (FDR) Presidency?
- Thought of him as a "Lazy Do Nothing"
  - Supported him and "loved" his taking action response to the Great Depression
  - Hated him because he put the United States in a massive amount of debt
  - They neither loved nor hated him, they rather saw him as an average president.
21. What piece of technology was important to rural Americans during the Great Depression and Dust Bowl?
- Television
  - Refrigerator
  - Radio
  - Microwave
22. The overall impact of the Great Depression on government was \_\_\_\_\_
- The federal government expanded its power by starting the Social Welfare State.
  - The federal government's power was limited.
  - Businesses would be less regulated by the Federal government.
  - Banks would be less regulated and checked on by the federal government.
23. Many Historians argue that the Great Depression was not solved by FDR and his programs rather it was solved by...?
- Herbert Hoover and his programs
  - The GI bill allowed former military men to go back to school and get better jobs
  - Wealthy business owners donated money to the working man and his families
  - The US entered WW2 and kick-started the economy
  - None of the above

### **True (A) / False (B)**

24. The Civilian Conservation Corps (CCC) offered jobs to young men during the Great Depression and focused on environmental conservation.
25. The Federal Deposit Insurance Corporation (FDIC) insures private bank accounts from losing all their money to this day.
26. Over 3000 banks closed their doors permanently during the Great Depression, never to open again.
27. Even with having a job during the Great Depression it was difficult for families to get by and get all that they needed to survive comfortably.
28. Hoovervilles were the nicknames given to the communities built out of materials found and collected by homeless families.
29. The Great Depression is primarily known as an economic disaster; however, it was a mental and emotional one as well.

## Matching:

**Directions:** Bubble the letter or letters that correspond with the answer you would like. For Example, if you wanted Agricultural Adjustment Act (AAA) as your answer you would bubble both letters A & B.

A. Security and Exchange Commission	B. Bonus Army	C. Relief	D. Glass-Steagall Act	E. Franklin D. Roosevelt (FDR)
AB. Agricultural Adjustment Act (AAA)	AC. Black Tuesday	AD. Recovery	AE. Herbert Hoover	BC. Reform
BD. World War 2				

30. President \_\_\_\_\_ was raised poor and believed that the economy and people could fix the depression themselves without government intervention
31. \_\_\_\_\_ was the name given to the worst day of the Stock Market crash
32. \_\_\_\_\_: This was a part of the New Deal targeted at solving unemployment issues, the housing crisis, and foreclosures.
33. \_\_\_\_\_ was a government program to assist farmers during the Depression. It stabilized food, crop, and livestock prices by destroying surplus goods.
34. \_\_\_\_\_: This was a part of the New Deal targeted at solving issues related to businesses, the stock market, banking, and more. The concept for this one was to make sure issues did not happen again.
35. President \_\_\_\_\_ won by a landslide victory, winning every state in his election except for 5 of them. He also had Polio, a disease that paralyzed him from the waist down.
36. \_\_\_\_\_ is a government program and group that controls investment laws surrounding the stock market and other types of investments.
37. \_\_\_\_\_ were World War 1 veterans that protested in Washington D.C. because they were not paid their "benefits" from their service time.
38. \_\_\_\_\_: This was a part of the New Deal targeted at businesses and companies in general. The concept was to try and get these sectors of the economy back to pre-depression conditions.
39. \_\_\_\_\_ was the law that made it illegal for banks to use clients' money for stock market investments.
40. When \_\_\_\_\_ began it provided more "jobs" than any FDR / New Deal program and it actually brought the United States out of the Great Depression as it assisted in the opening of factories as well.

## Appendix B

### Cold War Assessment Questions:

**Directions:** At the end of the Cold War Unit, in late April/early May, you will be assessed using a discussion instead of a multiple-choice test or written essay/short answer. The discussion will be one-on-one between the teacher and student with no others around. Under these directions, you will find 15 questions to study and work on throughout the unit. On the day of the test, I will select three of these questions to ask you.

- You may use notes, maps, and other assignments we've completed in class.
- You will not have access to the internet during the exam discussion.
- I expect that you prepare for our conversation and do not try to do this conversation on the fly.

#### 15 Exam Questions you could potentially be asked for our one-on-one discussion.

1. The Cold War was a war of ideas, can you explain a conflicting political idea between the USSR and the USA?
2. Why were alliances so important during the Cold War and what did they mean for the USA and/or USSR?
3. What was NATO and the Warsaw Pact? Why were they created by the Superpowers?
4. What is the difference between capitalism and communism? Why would these two government and economic systems dislike each other?
5. Why did the USA and USSR create economic support plans for Europe following WW2? How would these economic support plans be seen as threatening by each Superpower side?
6. What was McCarthyism / Red Scare and why were people fearful of communism?
7. How were nuclear weapons a part of the Cold War? What was their importance? How did it affect the societies of the USA or the USSR?
8. What was the Cuban Missile Crisis and why do historians see it as the most dangerous part of the Cold War?
9. How was espionage/spying used during the Cold War? What Information were the Superpowers seeking from the other?
10. How was Latin America involved in the Cold War? Why was the USA concerned with their increasing political change?
11. Explain how the Cold War changed American society, either in a positive way, negative way, or both. Provide examples.
12. A major strategy the USA used during the Cold War was containment. Please explain what containment was and give at least one example of how the USA used it.
13. Why were proxy wars used during the Cold War? Provide an example of one and discuss the Importance of the war.
14. Why was the Korean War fought? How is it related to the Cold War? What was the result? How has it impacted today's world?
15. Why did the Cold War involve the Vietnam War? What was the result of this war? How did the length and objective of the war change how many Americans viewed our government and military?

## Appendix C

Name:  
Period:

### Cold War Discussion Exam Grading Rubric

	0 Points	1-2 Points	3-4 Points	5 Points
	Answer illustrated <b>No or little knowledge/</b> understanding of the topic. No examples used or applied in answers. No analysis offered.	Answer displayed <b>some knowledge</b> of the topic, but <b>little understanding</b> of how the topic connected to the larger unit of study. At least one example used or applied in answer. Little analysis offered in answer.	Answer displayed a <b>clear understanding</b> of the topic and a decent connection to the larger unit of study. Some examples (1-2) used or applied in answers. Decent analysis offered to create an argument about the topic, connections, evidence, etc.	Answer showed a <b>sophisticated understanding</b> of the topic and a clear connection/importance to the larger unit of study. A multitude of examples (3+) used or applied in answers. Analysis offered an in-depth argument about topic, connections, evidence, etc.
<b>Question 1</b>				
<b>Question 2</b>				
<b>Question 3</b>				
	0 Points	1-2 Points	3-4 Points	5 Points
<input type="checkbox"/>	Answers <b>confused</b> or <b>incoherent</b> .	Answers showed <b>some structure</b> , but ultimately were hard to follow.	Answers were <b>generally well thought out</b> but delivered very tentatively.	Student provided <b>confident, coherent, compelling</b> answers.

## Appendix D

Student Name \_\_\_\_\_  
 Period \_\_\_\_\_

### United States History Cold War Discussion Exam

#### Conversational Tally Sheet

<b>Preparation &amp; materials brought to discussion</b>	None	A few notes or preparation material brought	A substantial amount of notes or preparation materials brought
<b>A tally will be added for each time one of the criteria is met.</b>			
	1	2	3
<b># of times <u>specific evidence</u> used</b>  Unit vocabulary, events, places, ideas, concepts, etc.			
<b>Connections/Links to diverse unit topics</b>  Connected question/answer to different unit topics.  Example: Development of nuclear bomb to espionage and cuban missile crisis			
<b># of times teacher prompt or nudged</b> <input type="text"/>  If the student seems stuck, in a place where they could expand, or connect the information to different ideas/concepts, the examiner took the opportunity to prompt.]			

## Appendix E

### Discussion Exam: Before, During, and After Questionnaire

**Directions:** Read the following statements carefully. Then, circle the answer you feel best represents your feelings about the discussion exam.

#### **BEFORE THE EXAM**

1. I prepared more for the discussion exam than I would have for a multiple choice test.

Strongly Disagree      Disagree      Neither agree nor disagree      Agree      Strongly Agree

2. I felt confident going into the discussion.

Strongly Disagree      Disagree      Neither agree nor disagree      Agree      Strongly Agree

3. I felt nervous because the exam was different from what I have experienced before.

Strongly Disagree      Disagree      Neither agree nor disagree      Agree      Strongly Agree

#### **DURING THE EXAM**

4. I became more nervous with the type of testing as the exam was going on.

Strongly Disagree      Disagree      Neither agree nor disagree      Agree      Strongly Agree

5. I became more comfortable with the type of testing as the exam was going on.

Strongly Disagree      Disagree      Neither agree nor disagree      Agree      Strongly Agree

#### **AFTER THE EXAM**

6. I think the discussion exam fits my learning/testing style more than a multiple choice test or written exam.

Strongly Disagree      Disagree      Neither agree nor disagree      Agree      Strongly Agree

7. I was glad that the test was in a discussion format.

Strongly Disagree      Disagree      Neither agree nor disagree      Agree      Strongly Agree

8. If another discussion/oral test was given, I know how to better study for this type of test.

Strongly Disagree      Disagree      Neither agree nor disagree      Agree      Strongly Agree

9. I believe if another discussion test was given I would improve my score.

Strongly Disagree      Disagree      Neither agree nor disagree      Agree      Strongly Agree

10. Although I would have preferred a different testing format (i.e. multiple choice, written test, etc.), I think the discussion test accurately measured my knowledge on the subject.

Strongly Disagree      Disagree      Neither agree nor disagree      Agree      Strongly Agree