A PHENOMENOLOGICAL STUDY OF THE LIVED EXPERIENCE OF SECONDARY

WORLD LANGUAGE TEACHERS WHO USE PROFICIENCY-BASED RUBRICS FOR

ASSESSMENT

By

James Donovan

Bachelor of Music in Music Education from the University of New Hampshire (2002)

Master of Music Education from Gordon College (2007)

A DISSERTATION

Presented to the Affiliated Faculty of
The College of Graduate and Professional Studies
at the University of New England
Submitted in Partial Fulfillment of Requirements
For the Degree of Doctor of Education

It was presented on

4/13/2023

and reviewed by:

Dr. Andrea Disque, Ed.D., Lead Advisor
University of New England
Dr. Jennifer Galipeau, Ed.D., Secondary Advisor
University of New England
Dr. Andrew Stephens, Ed.D. Affiliate Committee Member
Northeastern University

Doctor of Education Final Dissertation Approval Form

This Dissertation was reviewed and approved by:

Lead Advisor Signature: *Andrea Disque, Ed.D.*

Lead Advisor (print name): Andrea Disque, Ed.D.

Secondary Advisor Signature: *Jennifer A. Galipeau*

Secondary Advisor (print name): Jennifer A. Galipeau

Date: April 13, 2023

A PHENOMENOLOGICAL STUDY OF THE LIVED EXPERIENCE OF SECONDARY WORLD LANGUAGE TEACHERS WHO USE PROFICIENCY-BASED RUBRICS FOR ASSESSMENT

ABSTRACT

The purpose of this qualitative transcendental phenomenological study was to explore the lived experience of Massachusetts public secondary (middle school/high school) world language teachers who utilized the ACTFL proficiency-based rubrics to evaluate student performance environments (American Council on the Teaching of Foreign Languages, 2012). This study was guided by the research question, "What are the lived experiences of secondary world language teachers who used Proficiency-Based Rubrics to evaluate student performance?" Husserl's (1964) transcendental phenomenology served as the study's conceptual framework, and Andrade and Brookhart's (2020) classroom assessment theory served as the theoretical framework.

Semi structured interviews were used to gather data from eight participants. The emergent themes which evolved from this data were that performance feedback focuses on students' individual growth, that environment dictates the level of adherence to the ACFTL rubrics, and that the proficiency-based classroom creates a safer and more collaborative learning environment. The findings of the study suggest that there was inconsistent implementation and usage of the ACTFL proficiency-based rubrics by participants and that only some of the expected changes to instruction occurred as a result of the adoption of these rubrics.

ACKNOWLEDGEMENTS

To my parents, Michael and Mary Ellen Donovan, and my in-laws, Paul, and Charleen Sastri, your influence on my life and in the lives of the people I care most about in this world has been greater than you will ever know. Thank you for being such great role models for all of us. To Mom, Dad, and Charleen, thank you for all the times you helped Katie and me get the boys places while I did my schoolwork or came over to watch the kids so that we could go out during this period, I am eternally grateful.

Finally, to the incredibly talented cohort of classmates that I had the privilege of completing this program with, especially the "collective of over-thinkers," as we affectionally named ourselves. Michele, Machelle, Laura, Maggie, B.C., and Eric, thank you for your friendship and kindness.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

Since the early 20th century, teachers, including secondary teachers, across the United States have adopted and utilized the 100-point/A-F grading scale as a means of reporting student achievement on assignments and in reporting end-of-quarter grades (Brookhart et al., 2016). Despite widespread usage, the shortcomings of the 100-point/A-F scale for grading student work has been known for over a century (Brookhart et al., 2016; Carey & Carifio, 2012; Guskey & Brookhart, 2019; Starch & Elliott, 1912). In her seminal study, Brookhart (1993) suggested that grades function more as a "token economy" (p. 139) within the classroom, which Lipnevich et al. (2020) described as an "amalgamation" (p. 483) of competing value systems held by the individual teacher or school as opposed to a true reflection of the skills or knowledge developed in the classroom. Olsen and Buchanan (2019) and Vatterott (2015) further noted that while teachers may attempt to use grades to inspire or incentivize good behavior and performance in the classroom, grades are also used at times to punish low-performing students. These, and other internal threats like the implicit bias of teachers, call into question what final grades represent when appearing on a student's transcript, especially when considering the implications that a student's transcript has on their application process for post-secondary educational institutions and programs (Brookhart et al., 2016, Finn et al., 2019; Fornaciari & Lund Dean, 2013; Olsen & Buchanan, 2019; Quinn, 2020; Urhahne & Wijnia, 2021; Weed, 2018).

Many researchers, including Guskey and Brookhart (2019) and Coussens-Martin (2019), argued that the way to address the lack of validity in traditional grading is to adopt standards-based grading (SBG) in place of the 100-point/A-F scale; Brookhart and Guskey, notably, have argued for this position since the 1990s. In SBG, students are provided feedback using a rubric on both the cognitive aspects of their tasks and the non-cognitive aspects. Coussens-Martin

(2019) noted that this bifurcated approach to providing feedback supports the students in building self-efficacy rather than using grades to build "student compliance, reward, or (as a form of) punishment" (p. 59). In SBG, it is common for students to be categorized based on level descriptors like proficient, advanced, needs improvement, etc. (Guskey & Brookhart, 2019); teacher work may also be categorized using similar descriptors by teachers' SBG. Percell (2017) argues that the most effective feedback teachers can provide their students through SBG is specific process-oriented feedback about the student's own learning as opposed to specific product outcomes.

Another alternative to traditional grading practices is proficiency-based learning. Advocates of proficiency-based learning argue that its emphasis on continuous improvement (learning) and not grading distinguishes it as an alternative to traditional grading practices (Twadell et al., 2019). While proficiency-based learning is synonymous with mastery-based grading and competency-based grading practices within the literature, it is considered distinctly different from SBG (Vermont Agency of Education, 2018). Understanding this is important as it can be easy for educators to confuse SBG and proficiency-based learning. For example, Welsh (2019) noted that standards-based grading is fundamentally a reporting system of individual student achievement compared to state or local standards. Conversely, Twadell et al. (2019) posited that "proficiency-based instruction is not for the teacher to find out where students are in their learning; rather, it is for students to find and understand where they are in their own learning" (p. 131), and it is this information which is ultimately reported to students.

Unlike SBG, there is disagreement among proponents of proficiency-based learning as to how traditional grades could be used in reporting student performance over time (Twadell et al., 2019; Vermont Agency of Education, 2018). Teachers utilizing proficiency-based learning in

their classroom focus on helping their students assess their own level of proficiency, establish goals, and help them achieve these goals through a process rooted in self-reflection. As a result, as Twadell et al. (2019) suggested, traditional grading practices (A/F-100-point scale) are not necessary to assess student growth and performance on either report cards or when assessing individual assignments. Some proponents posited that the adoption of proficiency-based learning should provide more consistent scoring across classes and a clear separation between scoring for content knowledge and behaviors or applications of that knowledge when using rubrics (Vermont Agency of Education, 2018).

In 2012, the American Council on the Teaching of Foreign Languages (ACTFL) moved to integrate proficiency-based learning into the teaching of Foreign Languages across the United States by publishing their ACTFL Proficiency Guidelines Document (2012). This was a continuation of work that ACTFL has been engaged in since the 1970s when they first experimented with the development of a proficiency-based model of instruction (Warford, 2000). Specifically, Warford (2000) notes that ACTFL partnered with ETS (Educational Testing Service) to design a credible proficiency-based instructional model for schools based on the model the State Department of the United States used. The ACTFL Proficiency Guidelines Document (2012) established precise levels of proficiency (low, mid, intermediate, high, etc.). It included rubrics to be used by teachers when assessing students listening, speaking, writing, and reading skills (Custable et al., 2019).

The ACTFL Proficiency Rubrics use a 0-4-point scale and clear performance descriptors (0-4) for each category and sub-category (American Council on the Teaching of Foreign Languages, 2012). ACTFL (2012) advocates for teachers to utilize these rubrics while grading student work throughout the school year. Final classifications of students into their proficiency

category does not occur until the end of the grading period (American Council on the Teaching of Foreign Languages, 2022). Twadell et al. (2019) outlined this process, noting that teachers should

> create opportunities for students to produce evidence that they can directly review and judge against a desired state of proficiency (typically level 3 of proficiency scale.) In this segment of instruction, teachers often determine the current status, and perhaps the final status of a student's proficiency. (p. 31)

While the classroom experiences and learning opportunities at the beginning of the grading period are equally important when assessing for proficiency, determinations must be based on current performance and not earlier levels of performance.

This research study sought to understand the lived experience of secondary world-language teachers who utilized the ACTFL proficiency-based rubrics. In particular, this study sought to understand how these teachers used these rubrics when evaluating student performance. As proficiency-based learning is adopted more widely, specifically the ACTFL standards, which will be required of all world language teachers in Massachusetts in 2023 (Massachusetts Department of Elementary and Secondary Education, 2021a), understanding this phenomenon is increasingly important for world language teachers.

Definition of Key Terms

American Council on the Teaching of Foreign Languages. American Council on the Teaching of Foreign Language (ACTFL) is a professional organization for foreign language teachers in the United States. In 2012, ACTFL released a policy document that advocated for proficiency-based grading practices to be adopted by teachers and school districts nationally (American Council on the Teaching of Foreign Languages, 2012).

A-F Grading Scale. The A-F reporting scale was developed to report summative achievement or progress for students in the K-12 environment. The A-F scale is most commonly broken into bands by letter (A, B, C, D, and F). Each band has a range of between nine and ten points. Each letter band can be subdivided into three smaller bands in some scale iterations, each with its own numeric range. For example, a B- often represents final grades between 80-82, a B between 83-86, and a B+ for grades between 87-89 (Brookhart et al., 2016).

Cognitive Assessment. Students are asked to synthesize their knowledge and skillsets to solve a task or problem set (Olsen & Buchanan, 2019).

Grade Point Average (GPA). A student's cumulative grade point average for a specified period in core courses during an academic year (Weed, 2018).

Non-Cognitive Assessment. Students are assessed based on completion, participation, effort, or engagement in a task or problem set (Olsen & Buchanan, 2019).

Proficiency-Based Grading Practices. Proficiency-based grading practices assess students along a continuum of learning using a four-point scale corresponding to designations of beginning, developing, proficient, and expanding, in place of a standard 100-point/A-F score. This practice is synonymous with mastery-based grading and competency-based grading practices (Vermont Agency of Education, 2018).

Proficiency Scale. Proficiency scales are "a tool that displays a collection of related learning targets and scores for determining a student's current level of performance" (Hoegh et al., 2019, p. 4).

Proficiency Rubric. Proficiency rubrics are used by teachers when assessing elements of a proficiency scale and can be used during formative classroom experiences (Marzano et al., 2017)

Standards-Based Grading (SBG). Teachers within the same content area use the same rubric for grading. Grades are taken as each standard or curriculum framework is taught, and the feedback comes to the students through multiple streams, both cognitive and non-cognitive (Coussens-Martin, 2019).

100-Point scale. The 100-point reporting scale is used for reporting summative achievement or progress for students in the K-12 environment. The 100-point scale is a numerical average of all work assigned and graded during an academic term. This scale is often used synonymously with the A-F Scale (Brookhart et al., 2016).

<div align="center">

Statement of the Problem

</div>

Since the publication of the ACTFL Proficiency Rubrics in 2012 and the introduction of the Seal of Biliteracy, world language teachers across the United States have been adopting and incorporating proficiency-based learning instructional practices and scales into their classrooms (Commission on Language Learning, 2017). But it is not just world language teachers who have been adopting proficiency-based learning practices. In 2012, the state of Maine required that all graduation diplomas be based not on grades but on the demonstration of skills and knowledge across eight subject areas, including world languages (Cover, 2018). This mirrored the adoption of proficiency-based educational practices by 16 states between 2012 and 2018 across the United States (Cover, 2018; Massachusetts Department of Elementary and Secondary Education, 2021; Stump et al., 2017; Vermont Agency of Education, 2018). When adopting proficiency scales, teachers must consider how the proficiency scales will impact their assessment of students (Hoegh et al., 2019). It is important to notice the difference between a proficiency scale and a proficiency rubric. Proficiency-based rubrics are used by teachers to assess formative experiences during a grading period and are derived from the elements of defined categories

within proficiency scales. Proficiency scales are "a tool that displays a collection of related learning targets and scores for determining a student's current level of performance" (Hoegh et al., 2019, p. 4).

The use of proficiency scales as a part of determining student grades often appears in the literature (Hoegh et al., 2019; Twadell et al., 2019; Vermont Agency of Education, 2018). The Vermont Agency of Education (2018) posited that the adoption of proficiency-based learning, and more specifically, the assessment rubrics used by teachers, must provide an opportunity for teachers and school administrators to determine and define the purpose of grading and what exactly will be reported to students and families within these scales. Early adopters of proficiency-based learning, and therefore proficiency scales, often converted their proficiency-based ratings from scales of students into a numerical average, consistent with the traditional 100-Point/A-F Scale, to align with existing grading policies within their school or district (Guskey, 2020; Vatterott, 2015). While this was likely done for expediency, taking this step was at odds with the published research on the successful adoption of proficiency-based learning (Guskey, 2020; Guskey & Brookhart, 2019; Twadell et al., 2019; Vermont Agency of Education, 2018). For example, Twadell et al. (2019) argued that rubrics provide teachers with the ability to define the intended learning targets and that these targets are not "summative experiences" (p. 65) in and of themselves. Instead, these targets, encapsulated in the rubric, provide the framework for understanding students' level of proficiency during a unit of study or exploration and something which they can reflect upon in their learning. Twadell et al. (2019) further argued that while students' performance may be recorded in a grade book, only the final grade should be understood as representative of their work over the grading period.

Twadell et al. (2019) posited that proficiency-based learning is intended to break the heavily scaffolded and often passive learning environments, which are very common in schools. Instead, they argued that the proficiency-based learning "model of lesson design and delivery is more effective in developing efficacy and proficiency because students are the primary generators of learning" (Twadell et al., 2019, p. 27). Twadell et al. (2019) made the case that to achieve the aims of proficiency, the classroom experience must evolve into a differentiated environment for each student and not be dominated by teacher or class-wide needs. Twadell et al. (2019) aligned with Guskey (2020) in stating that the purpose of the rubrics in proficiency-based learning is to not just to be able to rate the finished products in the class but instead to set goals at the beginning of the work and to measure progress along the way.

Since 2012, public school world language teachers in Massachusetts have adopted the ACTFL proficiency-based rubrics as a primary assessment and communication tool (Massachusetts Department of Elementary and Secondary Education, 2021a). There was a gap in the literature pertaining to how using proficiency-based rubrics changes the way a teacher grades their students. Therefore, this study sought to explore the lived experiences of secondary world-language teachers currently utilizing the ACTFL proficiency-based rubrics and, specifically, how they used proficiency-based rubrics to evaluate student performance.

Statement of the Purpose of the Study

The purpose of this qualitative transcendental phenomenological study was to explore the lived experience of Massachusetts public secondary (middle school/high school) world language teachers who utilized proficiency-based rubrics to evaluate student performance. Proficiency-based rubrics are an assessment tool utilized by teachers who have adopted the proficiency-based learning model of instruction (Custable et al., 2019). Proficiency-based rubrics utilize either a

five or a four-point scale corresponding to designations of beginning, developing, proficient, and

expanding (Vermont Agency of Education, 2018). Included in proficiency-based rubrics are

descriptions of the targeted element, skill, behavior, or product being assessed for each rating:

beginning, developing, proficient, and expanding (Twadell et al., 2019).

Research Question

This research study sought to answer the following research question:

What are the lived experiences of secondary world language teachers who used

Proficiency-Based Rubrics to evaluate student performance?

Conceptual and Theoretical Framework

A conceptual framework is described by Bloomberg and Volpe (2019) as the

"scaffolding" which underpins the study within the body of research and theory (p. 163). In this

study, the researcher used Andrade and Brookhart's (2020) classroom assessment theory (CA) as

the theoretical framework for this study and Husserl's (1964) transcendental phenomenology as a

central component of the conceptual framework. Each of these theories examines a lens of the

transformation teachers experience after adopting proficiency-based learning practices,

specifically the use of rubrics, and is important to understanding the phenomena.

When conducting any phenomenological study, a central conceptual framework of the

study is the work of Edmund Husserl, who pioneered this qualitative methodology (Patton,

2015). Patton (2015) described phenomenological research as an approach to research where the

"exploration of how human beings make sense of experience and transform experience into

conscious" (p. 115). Bloomberg and Volpe (2019) described phenomenological research as

studying a small sample of subjects through an "engagement to develop patterns and relationships of meaning" (p. 54).

Phenomenological research often follows one of two branches within the broader phenomenology theological framework. Transcendental phenomenology was articulated by the German philosopher Edmund Husserl (1964), and hermeneutic phenomenology was later articulated by Martin Heidegger (1962). For the purposes of this study, the researcher utilized Husserl's transcendental theoretical framework, which seeks to understand the lived experiences of individuals by gathering descriptions, explications, and interpretations of their lived phenomena (Patton, 2015).

Peoples (2021) noted that Husserl believed that the act of thinking was always connected to an event or an endpoint in some form and that the more an individual can reflect upon their experience, the clearer the essence of that experience could be understood. Patton (2015) noted that phenomenological research seeks to understand the "essence or essences" of a shared experience (p.116). Peoples (2021) described the aim of phenomenological research as being able to describe the "phenomena exactly as they appear in an individual's conscious" (p. 5). Further, Peoples (2021) noted that the goal of the researcher is to understand the phenomena as it exists and not to attempt to compare it with the experience of others.

Classroom assessment theory is defined by Andrade and Brookhart (2020) as being the "process through which teachers and students gather, interpret, and use evidence of student learning" (p. 351). Andrade and Brookhart (2020) posit that a variety of data support the work of teachers in CA, which includes but is not limited to "diagnosing student strengths and weaknesses, monitoring student progress toward meeting desired levels of proficiency, assigning grades, and providing feedback to parents" (p. 351). Within CA, Andrade and Brookhart (2020)

suggest that a broader variety of data sources be included when determining final grades. For example, Andrade and Brookhart (2020) suggest using computer diagnostic scores, peer and self-assessments, quizzes, tests, conversations with teachers, and observations when making decisions about how students learn. This approach is often associated with the constructivist instructional design, where new knowledge or skills are built upon previously constructed learning (Windschitl, 2002). Brookhart and Andrade (2020) envisioned CA as a merger and extension of two previously articulated constructivist learning theories, self-regulated and co-regulated learning.

Assumptions, Limitations, and Scope

In this study, the researcher made several assumptions. Bloomberg and Volpe (2019) noted that assumptions often include specific references to the potential bias of the researcher and any other idea that the researcher believes to be true but lacks evidence to support the claim. In this phenomenological study, the researcher assumed that the study participants were honest in their responses while participating in the semi structured interviews and that they accurately represented their experience with the central phenomenon. The researcher recruited participants who used the 2012 ACTFL proficiency rubrics in their classroom for instructional and assessment purposes; no distinction was made for those who did so on their own initiative, as part of a curricular change within their school or department, or those who began teaching in a school where these practices were already in place. Throughout the study, it was assumed that the researcher would be able to monitor and keep in check any personal biases by journaling before and after each interview (Peoples, 2021).

While phenomenological research has many noted strengths, Bloomberg and Volpe (2019) posit that it has well-established limitations. A potential limitation of this study was the

overall sample size of eight participants. While a goal in phenomenological research is to understand the shared essence of the experience had by one or many individuals, the ability to generalize the findings of this study for world language teachers broadly is limited. Further, this study targeted secondary teachers who were utilizing proficiency-based rubrics, thus narrowing the field of potential participants statewide; it is important to keep in mind that most world language instruction occurs during a student's secondary experience (Commission on Language Learning, 2017). Finally, the scope of a study includes the external conditions which act as boundaries for the researcher while the study is ongoing (Bloomberg & Volpe, 2019). Importantly, while teachers in all subject areas can anchor their classroom experience in proficiency-based education, this study focused on world language teachers who were currently using proficiency-based rubrics in their classroom environment, which are based on the 2012 ACTFL Guidelines (American Council on the Teaching of Foreign Languages, 2012).

Rationale and Significance

For nearly as long as teachers have used the 100-point A/F Scale to grade students, the limitations of this method have been understood (Brookhart et al., 2016). In their review of a hundred years of research on grading practices, Brookhart et al. (2016) described the transition away from the proficiency-based grading scales of the 19th century to the now traditional, 100-point/A-F scale, chiefly because of its efficiency. Brookhart et al. (2016) describe the traditional 100-point/A-F Scale as a "hodgepodge" (p. 826) of competing roles and value systems. Lipnevich et al. (2020) described the traditional 100-point grading scale "amalgamation" (p. 484) of cognitive product outcomes (assessments) and non-cognitive process outcomes and not a reflection of what students can do with their learning. This is particularly important given the

long-term implications final grades can have on a student's post-secondary opportunities (Coussens-Martin, 2019; Hopfenbeck; 2019; Weed, 2018).

Proponents of proficiency or standards-based grading methodologies at the secondary level argued that without the need to consistently rate and grade every piece of work, students would be able to have higher-level conversations with teachers about their own learning and performance in the classrooms (Guskey, 2020; Vermont Agency of Education, 2018). Conversely, Vatterott (2015) has advocated for a hybrid approach where the rubric grade is converted into a numerical average to anchor the usage of rubrics within the traditional grading paradigm, thus making the transition more palatable and achievable for all stakeholders. As noted previously by Guskey and Brookhart (2019), this type of rubric rating to grade conversion can introduce the same types of limitations as the traditional 100-point/A/F grading scales.

This study may benefit teachers and administrators in secondary schools who currently utilize proficiency-based rubrics for evaluating student performance or those who may be adopting these practices. For teachers, the findings of the study could provide them with an understanding of how the adoption of proficiency-based rubrics in their classrooms may change how they evaluate their students. This study may also assist building principals, curriculum directors, and district administrators to help frame their internal discussions in planning for the adoption of proficiency-based learning environments and grading their students. This is particularly important as the underlying philosophies and practices of proficiency-based learning, specifically in grading, will likely be new for many teachers, and, as Link (2019) notes, teachers have "little training on grading practices. Many teachers apply the same practices they experience as students or mechanically follow school or district grading norms or policies" (p. 189).

Summary

The purpose of this transcendental phenomenological study was to explore the lived experiences of secondary world-language teachers who used ACTFL proficiency-based rubrics for assessing student work. Since the ACTFL Proficiency Guidelines were published in 2012, world language teachers across the United States have been adopting ACTFL's proficiency rubrics and scales into their classroom instruction, and states like Massachusetts are integrating the rubrics into the World Language Curriculum Frameworks (Massachusetts Department of Elementary and Secondary Education, 2021a). This study used a conceptual framework based on Husserl's (1962) transcendental phenomenology and Andrade and Brookhart's (2020) classroom assessment theory as the theoretical framework to understand better the individual experiences of the research participants from data collected through semi structured interviews. This study may also benefit stakeholder groups in public schools and districts considering, or in the process of, adopting and implementing proficiency-based learning in world language classrooms.

CHAPTER 2: LITERATURE REVIEW

The grading practices of secondary school teachers in the 20th century were often directed toward creating efficiencies for the teacher based on the number of students they have in class, and to simplify how they communicate a student's level of achievement with parents and the community (Brookhart et al., 2016). During this time period, student achievement was most often conveyed as a numeric average using a 100-point scale or as a letter grade, with each letter representing a specific band of the 100-point scale (Brookhart et al., 2016). During the 1960s to the 1970s, high schools began to adopt GPA scales as a means of identifying their highest-performing students (Weed, 2018). As a measure of student achievement, GPA is important for students because it is a key metric used by colleges and universities when making admissions decisions and financial awards (Coussens-Martin, 2019).

Lipnevich et al. (2020) pointed out that final grades often represent a student's performance in a particular course relative to an amalgamation of values, biases, beliefs about learning, and learners held by the teacher and not the individual student's abilities. Beginning with Brookhart's (1993) study, many researchers have demonstrated the countless internal and external threats to the validity of student grades which include teacher bias based on personality type, physical appearance, pronunciation of last name and gender, the weight given to cognitive and non-cognitive tasks, and the potential impact of governmental policy (Bygren, 2020; Fornaciari & Lund Dean, 2013; Guskey & Link, 2019; Homberg et al., 2019; Olsen & Buchanan, 2019; Quinn, 2020; Urhahne & Wijnia 2021). Bonner and Chen (2021) disagreed with this position in citing data demonstrating a 65-75% correlation between achievement and grades. However, as they note, this leaves a variance of between 25-35%, which they attribute to the factors cited above (Bonner & Chen, 2021).

Beginning in the 1990s, researchers began to publish findings and recommendations for educators to shift away from traditional 100-point A/F grading practices and toward standards-based grading practices (Brookhart, 1993; Brookhart et al., 2016; Guskey & Brookhart, 2019). Brookhart et al. (2016) and Guskey (2020) have advocated strongly for adopting standards-based report cards in place of the current system to improve the quality of feedback and communication between students, teachers, and families. In standards-based grading, students are rated against or compared to an established grade-level 'standard' which is developed by a school, district, or state (Guskey, 2020). Their research, combined with the work of others, has prompted teachers, schools, and states to adopt this model of reporting and communication with parents over the last 30 years (Bonner et al., 2018; Cover, 2018; Kenna & Russell III, 2018; Stump et al., 2017).

However, standards-based grading was not the only potential solution to address the inequities of traditional grading practices identified at the end of the 20th century; proficiency-based grading was also offered as a possible solution (Brookhart et al., 2016; Carey & Carifio, 2012; Guskey & Brookhart, 2019; Marzano et al., 2017; Olsen & Buchanan, 2019; Twadell et al., 2019). In proficiency-based grading, students and teachers work to develop a student's proficiency with an established skill or competency, and the student is assessed using a rubric that articulates designations of beginning, developing, proficient, and expanding, in place of a standard 100-point/A-F score (Custable et al., 2019). This practice is synonymous with mastery-based grading and competency-based grading practices (Vermont Agency of Education, 2018). Some organizations, including the American Council on the Teaching of Foreign Languages in 2012, adopted a proficiency-based approach to the teaching and grading of world languages (American Council on the Teaching of Foreign Languages, 2022). Unfortunately, as Guskey and

Brookhart (2019) noted, research is beginning to show that teachers utilizing standards-based or proficiency-based approaches to grading demonstrate some of the same biases as teachers using the 100-point/A-F scale, especially when utilizing conversion scales that turn rubric scores into traditional grades; a practice for which there is significant disagreement within the literature, but proponents see as a way to make adoption easier for teachers and schools (Guskey, 2020; Language Testing International, 2021; Moskal, 2002).

The purpose of this qualitative transcendental phenomenological study was to explore the lived experience of Massachusetts public secondary (middle school/high school) world language teachers who utilized proficiency-based to evaluate student performance. Proficiency-based rubrics are the assessment tool utilized by teachers who have adopted the proficiency-based learning model of instruction (Custable et al., 2019). This literature review examined the evolution of grading practices at the secondary level since the beginning of the 20th century in the United States, culminating in proficiency-based learning and the ACTFL guidelines for proficiency.

Conceptual and Theoretical Framework

The role of a conceptual framework is to explain and anchor the different relationships within the ecosystem of a research study: topic, theories, research questions, methodology, potential outcomes, research goals, and the significance of the study (Ravitch & Carl, 2021). Ravitch and Carl (2021) make the analogy that a conceptual framework is like a "compass" that "situates the study in its theoretical, conceptual, axiological, and practical contexts" (p. 34). At the center of this research study is an examination of secondary teachers who have adopted proficiency-based grading or those who use proficiency-based rubrics in their classrooms. Proficiency-based learning was developed and implemented in response to the identified failings

of traditional grading practices, which developed over the first half of the 20th century (Brookhart et al., 2016).

The impact of teachers' practices and school-based grading policies for secondary high school students extends well beyond the classrooms and teachers where they originate (Weed, 2018). Grades are often an amalgamation of competing or dissonant values that may or may not reflect what a student has learned in a grading period and is now able to do (Brookhart, 1993; Brookhart et al., 2016; Guskey & Link, 2019; Lipnevich et al., 2020). Teachers and schools can work to develop a coherent and transparent grading system that articulates the construction of and frames decision-making when assigning grades (Lipnevich et al., 2020).

Central to the conceptual framework of this study was the philosopher Husserl's (1964) transcendental phenomenology and Andrade and Brookhart's classroom assessment theory. Edmund Husserl's (1964) transcendental phenomenology is a qualitative methodology for conducting research (Patton, 2015). Moran (2005) suggested that Husserl saw phenomenology as the "evolution of modern philosophy" (p. 174) and that it "correctly articulates the truest sense of the necessary correlation between subjectivity and objectivity" (p. 175-176). This description aligns with Patton's (2015) description of phenomenological research as an approach to research where the "exploration of how human beings make sense of experience and transform experience into conscious" (p. 115) and Bloomberg and Volpe (2019) description of phenomenological research as studying a small sample of subjects through an "engagement to develop patterns and relationships of meaning" (p. 54). Transcendental phenomenology is commonly used in qualitative studies and also served as the methodology of this study.

Classroom assessment theory (CA) is a newly developed theoretical framework by prominent researchers Heidi Andrade and Susan Brookhart. Andrade and Brookhart (2020)

posited that the optimal learning environment has shifted from the "transmissionist" (p. 350) philosophies that underpin much of the 20th and early 21st-century learning environments to one where students and teachers work collaboratively to help students identify and achieve their own proficiency in the curriculum or skill being developed. Finally, Andrade and Brookhart (2020) note that developing the student's ability to self-regulate during learning is key to the learner's success.

Classroom Assessment Theory

In this study, the researcher utilized the theoretical framework of classroom assessment (CA) as articulated by Andrade and Brookhart (2020). In developing this framework, Andrade and Brookhart (2020) examined the evolution from transmissionist philosophies, which focus on the transmission of knowledge from teachers to students and assessed via tests, to more behaviorist and constructivist theories, which increased in prominence in the early 2000s (Windschitl, 2002). They noted that the more modern iterations of constructivist theory "emphasize the importance of student agency as a matter of the self-regulation of learning" (Andrade & Brookhart, 2020, p. 350). In developing CA, Andrade and Brookhart (2020) expanded the phases and areas of self-regulation of learning developed by Pintrich and Zucho, to include the co-regulation of learning, which "includes teachers and materials" (p. 351). Andrade and Brookhart (2020) ultimately defined CA as

> a process through which teachers and students gather, interpret, and use evidence of student learning for a variety of purposes, including diagnosing student strengths and weaknesses, monitoring student progress toward meeting desired levels of proficiency, assigning grades, and providing feedback to parents. (p. 351)

Importantly, Andrade and Brookhart (2020) distinguished between classroom assessments and other normative, evaluative, or high-stakes assessments and do not consider these when articulating CA. Instead, CA focuses entirely on utilizing formative and summative assessments within the classroom experience to develop student learning. Within the framework of CA, Andrade, and Brookhart (2020) suggested that a wider variety of data sources, including computer diagnostic scores, peer and self-assessments, quizzes, tests, conversations with teachers, and observations, should be included when making decisions about what students know and are able to do.

The key to understanding CA and its significance is recognizing the philosophical underpinnings that make up the theory. In particular, Andrade and Brookhart (2020) explored the role of assessments within the learning process and the intersection between self-regulation and co-regulated learning in CA. Constructivists argue that learning is the outcome of experiences where students build upon previous knowledge to construct new knowledge or skills (Andrade & Brookhart, 2020). Windschitl (2002) noted that it was often more challenging to adopt and incorporate constructivist practices within the classroom than many early adopters first recognized. Chen and Bonner (2017) noted the lack of "coherence" (p. 20) between teacher training and practice as a root cause of this phenomenon.

Based on the work of Pestalozzi, Dewey, and Rousseau, constructivism became associated with the child or student-centered movements of the 20th century (Windschitl, 2002). Later, Windschitl (2002) noted that constructivism evolved again by introducing co-constructive classroom experiences where students and teachers learn together. Andrade and Brookhart (2020) described classrooms built around self-regulated learning theory as classrooms

where students construct their own meaning in a shared learning environment. In co-regulated learning theory, Allal (2020) stated that learning results

> from the joint influence of student self-regulation and of sources of regulation in the learning environment: namely, the structure of the teaching/learning situation, the teacher's interventions and interactions with students, the interactions between students, and the tools used for instruction and assessment. (p. 332)

It is important to note that CA is a relatively new theory of learning, and it is appropriate to consider this a potential weakness. However, the long-standing credibility and contributions of Andrade and Brookhart to the fields of assessment, standards-based grading, and measurement are substantial over the course of their careers. Therefore, the researcher was confident that CA was appropriate for this study.

Transcendental Phenomenology

While Husserl had begun to develop phenomenology as early as 1905, he didn't begin publishing and speaking about phenomenology until 1913 with the publication of Ideen I (Moran, 2005). Moran (2005) described Husserl as believing phenomenology to be the final iteration of transcendental philosophy, which had been first articulated by Plato and is connected to both philosophy and science (Moran, 2005). Moran (2005) noted that Husserl believed that a researcher could not consider an object without taking into consideration its context. Peoples (2021) further described Husserl's thinking by noting that meaning is always connected to an event or an endpoint in some form and that the more an individual can reflect upon their experience, the clearer the essence of that experience could be understood. Peoples (2021) cautioned that when conducting a phenomenological study, the researcher must not attempt to compare the phenomena they are studying with the experience of others. This aligns with

Husserl's assertion that what exists only exists within the consciousness of those who experience it (Moran, 2005). Therefore, it is only the individual phenomena that can be studied, and the findings cannot be generalized beyond those who had the conscious experience.

A clear conceptual framework serves an important role in a research study as it is the binding force that connects the problem of practice to the research questions being examined, the theoretical frameworks which underpin the research, the desired outcomes for the study, and the methodology employed by the researcher(s) (Ravitch & Carl, 2021). The purpose of this qualitative transcendental phenomenological study was to explore the lived experience of Massachusetts public secondary (middle school/high school) world language teachers who utilized proficiency-based rubrics to evaluate student performance. Assessing student work and progress through the act of grading has been a core responsibility of teachers for more than a century (Brookhart et al., 2016). In this study, the researcher utilized classroom assessment theory (CA) as articulated by Andrade and Brookhart (2020) and Husserl's transcendental phenomenology as theoretical frameworks (Moran, 2005). Specifically, this study seeks to understand the phenomena experienced by teachers who used proficiency-based rubrics to evaluate student performance.

Grading Practices

At the turn of the 20th century, proficiency-based grading was the norm in public schools across the United States (Coussens-Martin, 2019; Olsen & Buchanan, 2019; Weed, 2018). With the rapid expansion of public secondary education in the early part of the 20th century, teachers and schools quickly transitioned away from a proficiency approach of grading students and toward the 100-point/A-F scales to be more efficient in their grading practices with the increase of student population (Brookhart et al., 2016). As the 100-point/A-F scale became the most

common grading scale for public secondary students, there was a concerted effort to adopt the grading practices of colleges and universities, like GPA, to help identify the highest performing students and make prospective students more attractive during the application process (Coussens-Martin, 2019; Weed, 2018).

Beginning in the 1990s, researchers identified multiple threats to the internal validity of the 100-point/A-F grading scales (Bonner & Chen, 2021; Brookhart, 1993; Brookhart et al.,2016; Bygren, 2020; Finn et al., 2019; Lipnevich et al., 2020; Malouff & Thorsteinsson, 2016; Quinn, 2020; Urhahne & Wijnia, 2021; Weed, 2018)**.** Brookhart (1993) noted the inherent conflict of interest that teachers find themselves in when grading their own student's work and found that teachers often considered the short and long-term implications of the grade when assigning them to students. Further, Brookhart (1993) found that teacher grade books often included both cognitive and noncognitive tasks. Bonner and Chen (2021) also found inconsistencies within summative course grades caused by the inclusion of nonachievement and non-cognitive factors within a student's grade.

While cognitive tasks can often be easy to identify (summative assessments), non-cognitive tasks can include effort, ability, growth, means and time of submissions, and behavior which obscure what the grade represents (Guskey & Link, 2019; Olsen & Buchanan, 2019). Olsen and Buchanan (2019) posited that grades are often thought of as a representation of what a student has mastered or achieved when instead, they are a "tool to inspire. It is a tool to control" (p. 2005). This is similar to Vatterott's (2015) suggestion that grades are used to shape the behaviors desired of students and do not reflect what they can do. Chen and Bonner (2017) refer to this type of grading as "rooting or pulling for students" (pg. 20). These arguments are also similar to Brookhart et al. (2016) "hodgepodge" (pg. 826) description of grading practices where

they noted that it is impossible that grades could represent both achievement and control simultaneously. This research led to a push for standards-based grading practices and the introduction and use of rubrics to assess student work as opposed to the 100-point/A-F grading scale (Brookhart et al., 2016; Guskey, 2020; Guskey & Link, 2019).

At the same time, proficiency-based grading was offered as another potential solution to the shortcomings of traditional 100-point/A-F grading practices referenced above (Vermont Agency of Education, 2018). Custable et al. (2019) posited that what distinguishes proficiency-based grading practices from traditional and standards-based, was the collaborative nature of the relationship between students, their teachers, and the desired outcomes. Custable et al. (2019) further suggested that adopting proficiency-based grading practices fostered the self-efficacy of students.

Grading practices, like all aspects of education and our public and private models of education, have evolved greatly since their introduction to society in the early colonies (Brookhart et al., 2016). When grading, teachers most often utilize a combination of non-cognitive and cognitive factors when grading student work, which includes effort, ability, achievement, and behavior (Brookhart et al., 2016; Olsen & Buchanan, 2019; Randall & Engelhard, 2010). Olsen and Buchanan (2019) noted that there is often only a "moderate correlation" (p. 2006) between a student's grade and their performance on common measures of achievement. Further, Olsen and Buchanan (2019) asserted that grades are often thought of as a representation of what a student has mastered or achieved when instead, they are often used to incentivize or compel behaviors in students through compliance.

18th and 19th Centuries

In the 18th and 19th centuries, teachers provided feedback on students' progress to families in oral and or narrative forms, in some cases going door to door (Brookhart et al., 2016). At this time, individual teachers and schools would set their reporting intervals (monthly, quarterly, semi-annually, or annually) and evaluate using their proficiency criteria. In their historical review of grading practices, Brookhart et al., (2016) presented that narrative (written) reports of student progress on skills such as "penmanship, reading, or arithmetic" (p. 805) became increasingly common at both the primary and secondary levels in the 19th century.

Vatterott (2015) notes that early colonial schools were understood to be an outgrowth of religion, and the desire to create a strong "moral education" (p. 6), was understood to be the central purpose of schooling. Good discipline, Vatterott (2015) notes, was also understood to be an essential virtue of students at this time and, as such, it was important that schools promote and recognize "hard work, neatness, effort, and penalize late work, tardiness, cheating, etc." (p. 7). Lipnevich et al. (2020) refer to this type of grading practice as an "amalgamation" (p. 484) and, like Brookhart (1993), Lipnevich et al. (2020) note the inherent dissonance when trying to synthesize behaviorism and achievement into a numeric average. Additional research has shown that teachers grading practices are often based on their personal experience (Olsen & Buchanan, 2019), personality type (Fornaciari & Lund Dean, 2013), biases (Bygren, 2020; Finn et al., 2019; Malouff & Thorsteinsson, 2016; Quinn, 2020; Urhahne & Wijnia, 2021) and professional development or lack thereof (Olsen & Buchanan, 2019).

The 20th and 21st Centuries

Between 1870-1910, the number of secondary schools in the United States increased from 500 to 10,000 (Guskey, 1996; Olsen & Buchanan, 2019). This expansion was due to the waves of European immigrants, the introduction of child-labor laws, changes in agricultural

practices, and increased state requirements for compulsory educational experiences for secondary students (Lleras-Muney, 2002). During this period, American secondary schools came to adopt the ordinal (100-point) percentage grading system to assess proficiency (Coussens-Martin, 2019; Olsen & Buchanan, 2019; Weed, 2018). It is important to note that the most common grade for a 19th or early 20th-century student being assessed using the 100-point proficiency scale would have been 50 (Carey & Carifio, 2012).

In the early 20th century, the 100-point proficiency scale evolved into a 100-point percentage scale (Brookhart et al., 2016). Brookhart et al., (2016) noted that schools moved away from narrative descriptions and to the 100-point percentage scale because the previous model was considered "time-consuming" for teachers, especially with the overall increase in students (p. 805). An unintended consequence of this migration was that the reported average of submitted assignments transformed the conversation between teachers and their students away from proficiency to what students can actually do, and to a discussion about amalgamated grading formulas (Guskey & Brookhart, 2019).

During this time, it became increasingly important to educators at both the college/university and secondary levels to be able to identify and rank students in an effort to find the highest-performing students (Coussens-Martin, 2019; Weed, 2018). Coussens-Martin (2019) noted that educators believed that ranking would increase students' intrinsic motivation to perform better in school. However, to do so, schools and educators needed to move away from the previous proficiency-based model of reporting on student achievement toward a more empirical model that would provide the averages needed to rank students easily (Coussens-Martin, 2019). Using previously developed scales like the 100-point scale and, at the college level, the 4-point interval scale, which would later be known as the GPA scale, secondary

schools, colleges, and universities began to offer scholarship opportunities or admission to post-secondary institutions and programs based on class rank (Coussens-Martin, 2019, p. 5).

The A-F scale, which sits on top of the 100-point average scale, was first introduced in the 1920s and used by 80% of all high schools across the United States by the 1940s (Brookhart, 1993; Guskey, 1996; Lee, 2020; Weed, 2018). In its most common form, students earn an A when their final grade falls between the range of 90-100, B in the range of 80-89, C in the range of 70-79, D in the range of 60-69, and 59 and below are commonly an F (Weed, 2018). The simplicity that came with having a letter grade instead of the numerical average that underpins it on the transcript contributes to why this system is still widely used today in both high schools and at the university level (Weed, 2018). It is important to note, however, that an institution can adjust the A-F scale and that many colleges and high schools have now moved to a weighted GPA (Wetzler, 2019). Lipnevich et al. (2020) noted that there are three distinct 100-point/A-F grading formulas. In the first formula, the point values for all assignments for a grading period total 100-points (Lipnevich et al., 2020). A student's grade in a course would be the total number of points earned out of 100. In the second formula, every assignment has a particular point value. At the end of the quarter, the total number of points earned is divided by the total possible points available, and this percentage becomes the course grade. The third formula is a percentage system (Lipnevich et al., 2020). In this formula, assignments are all scored on a 100% basis, and each assignment or category of assignments (homework, assessments, etc.) makes up a percentage of the overall final grade (Lipnevich et al., 2020).

In the 1930s, grading on the curve became an increasingly popular way to rank students' work (Guskey, 1996). One method highlighted by Guskey (1996) was the 6-22-44-22-6 system, where only 6% of the students in a class would receive an A, 22% would receive a B, 44%

would receive a C, etc. Vatterott (2015) suggested that teachers who used the bell curve came to believe that being successful in the classroom was a rare event and that grades became the "yardstick" (p. 13) by which teachers and students measured themselves. Vatterott (2015) also suggests that the usage of the bell curve in schools made the failure of students acceptable within society.

By the 1960s and 1970s, the percentage of high school graduates seeking a college education increased significantly as high school graduates sought to defer being drafted for the war in Vietnam (Weed, 2018.) To aid their graduates applying to college, high schools began to adopt and report a Grade Point Average (GPA) based on the standard 0-4 interval scale that had been used to assign credits at the college/university level (Weed, 2018). While this system is most effective at distinguishing the highest performing students (i.e., rank), Wetzler (2019) points out that GPA disproportionately impacts lower-achieving students. After conducting a review of the grading scale at the United States Military Academy at West Point, Wetzler (2019) found that the ranges in GPA at the lower end of the scale, C and below, are significantly greater and more significant than those at the higher end of the same scale. For this reason, many college/university systems have moved to a plus/minus GPA to better align with the 100-point A-F scale (Barnes & Buring, 2012). In a plus/minus scale, a 4.0 GPA would be awarded to a student with an A, while an A+ average would have a 4.33 GPA weight, and an A- would be worth 3.67 in the GPA scale (Wetzler, 2019).

A distinguishing characteristic of high school GPA models from those at the college and university level is that high schools often provide bonus weight to college and honors-level courses using a practice called weighted GPA (Merritt, 2021). For example, a tenth grader may earn a 4.5 GPA for an A in their AP Government class because the Advanced Placement course

is considered a college or university-level course. If that same student were to earn an A in an honors-level English course, the weighted GPA would be weighted at 4.25 because it is considered a more rigorous experience than their base, college-prep course. Accordingly, this same student would only earn a 4.0 weighted GPA for an A in a College-Prep level course.

Because of this, discussions about GPA usually begin as soon as a student enters high school (Weed, 2018). These conversations naturally lead to questions about teachers' grading practices, specifically, the fairness or lack thereof in how a teacher constructs and assigns grades (Weed, 2018). These are not idle concerns, as the research calling for structural reforms in grading has existed within the literature since at least the 1990s (Brookhart et al., 2016; Olsen & Buchanan, 2019; Randall & Engelhard, 2010).

The Need for Grading Reform

Since the early 1990s, researchers have determined that student grades are more often representative of an amalgamation of sometimes competing value systems held by the individual teacher or school rather than a reflection of the skill and ability to synthesize these skills in the community (Brookhart, 1993; Brookhart et al., 2016; Lipnevich et al., 2020). The transition to the 100-point/A-F grading scale was done, in part, to make grading more streamlined for secondary teachers as the number of students increased dramatically at the beginning of the 20th century, not because it was the optimal system for reporting student achievement. (Brookhart, 1993; Brookhart et al., 2016; Guskey, 1996; Lee, 2020; Weed, 2018). While the 100-point/A-F scale is the most common format for reporting student achievement at the secondary level, as Brookhart et al. (2016) noted, there have been many recent attempts to reform or replace traditional grading.

Importantly, failings of the 100-point/A-F grading scale have been present since shortly after its adoption and resulted in the first attempts at reforming the practices with, among other things, the introduction of the bell curve in the 1930s (Brookhart et al., 2016; Guskey, 1996; Vatterott, 2015). Olsen and Buchanan's (2019) study highlighted the failure and potential damage to students when teachers use grades on cognitive and non-cognitive tasks as a form of punishment to incentivize positive behaviors or compliance from students. Similarly, researchers have repeatedly found that teachers are susceptible to having their own biases about their students impact the teacher's assessment practices and the grades that students earn (Angelo & Reis, 2021; Brookhart, 1993; Brookhart et al., 2016; Bygren, 2020; Finn et al., 2019; Guskey & Link; 2019; Quinn, 2020; Urhahne & Wijnia, 2021).

Non-Cognitive Assignments

Brookhart et al. (2016) observed that teacher gradebooks are often comprised of both cognitive and non-cognitive tasks. While cognitive tasks can often be easy to identify (summative assessments), non-cognitive tasks can include effort, ability, growth, means and time of submissions, and behavior which obscure what the grade represents (Olsen & Buchanan, 2019; Guskey & Link, 2019). The potential implications for including non-cognitive assignments in grading are significant. For example, Angelo and Reis (2021) noted that when non-cognitive tasks are considered by teachers, female students often score higher as girls are believed to be more "amenable" (p. 105) to the traditional learning environment than boys.

Guskey (2020) classified cognitive and non-cognitive grading tasks into three categories: process criteria, progress criteria, and product criteria. Product criteria include high-point value summative assignments, which Guskey (2020) noted are often used as evidence of what a student can and is able to do. Progress criteria in grading focuses on the cognitive and

noncognitive steps students take while either demonstrating or creating a product (Guskey, 2020). Guskey (2020) described process criteria as being mainly constructed of effort, transactional (submission of) classwork, and homework. When teachers value non-cognitive assessments (process criteria) the same way they value cognitive assessments (product criteria), they inflate formative tasks' value and devalue summative tasks inside of the final course grade (Guskey, 2020).

While it is true that there may be benefits to rewarding or encouraging students with high grades, Olsen and Buchanan (2019) noted that this type of behavior by teachers actually has the most significant impact on high-achieving students. Guskey (1996) argued that the inverse behavior by teachers, punishing a student with a low grade, should have no role in education. Olsen and Buchanan (2019) ridiculed this type of behavior as teachers using grades as both the "sword and the shield" (p. 2005). Similarly, a student that does very little homework on a nightly basis could be impacted significantly if the non-cognitive task of homework weighs heavily within the grading formula of a particular teacher and only marginally in the classroom of a teacher that does not value homework within their grading formula.

The potential danger of overvaluing non-cognitive tasks was noted by Griffin and Townsley (2021) in their study, which analyzed 795 high school students' final grades in math. In particular, Griffin and Townsley (2021) examined "to what extent students' grades were inflated or deflated due to including homework and employability scores in the grade" (p. 1). For the purposes of Griffin and Townsley's (2021) study, employability scores were defined as "points given to students as part of their grades that reflect 21st -century skills demonstrated within the classroom environment" (Griffin & Townsley, 2021, p. 2). These skills included class participation, citizenship, and the timeliness of work submission. In their analysis of the data,

where they were able to separate out the homework averages and employability scores for each student, the researchers found that "336 (43.2%) students had their grades inflated or deflated by 5% or more and 97 (12.6%) students had their grades inflated or deflated by 10% or more" (Griffin & Townsley, 2021, p. 2).

The Presence of Bias

In their comprehensive "*Review On the Accuracy of Teacher Judgements,*" Urhahne and Wijnia (2021) found evidence of teacher bias and inconsistencies of expectations by gender, academic rank, and the rigor of the class.   Specifically, they found that teachers tend to overestimate well-adjusted children and those with high social-emotional functioning. When students exhibit negative behaviors, for example, being messy or hyperactive, they are more prone to negative judgments. Similarly, Urhahne and Wijnia (2021) found that male students who were well organized and wrote neatly scored higher than those who were messy when their work was normed for academic influences. Urhahne and Wijnia (2021) also note that teachers also favor more physically attractive students and those who have better social skills.

Finn et al. (2019) produced a similar finding when they found that obese students may often be considered lazy, less intelligent, and less socially desirable by teachers than their peers. In their study, Finn et al. (2019) also found that although the grading on mechanics and structure were consistent with their non-obese peers, the grades for obese students were slightly lower than their peers. However, it is important to note that Finn et al.'s (2019) study sample size was relatively small, with only N=133 teachers in the UK.

Quinn (2020) noted that evidence exists that teachers may "give a racially biased evaluation of student work" (p. 375). Quinn (2020) further posited that the impact of this can be substantial "given that future teachers base their expectations in part on the biased evaluations of

previous teachers" (p. 375), impacting both short and long-term performance. In Quinn's (2020) study, 1,549 teachers in the United States were provided with a writing sample of a second grader with either stereotypical Black (Dashawn) or Caucasian (Connor) sounding names. Quinn (2020) found that teachers were 4.7% less likely to rate the Dashawn version as being at grade level. Furthermore, Quinn (2020) found that "female teachers may be more likely than male teachers to show bias against Black males" (p. 384) as they were 7% less likely to rate the Dashawn writing example as being at grade level. Guskey and Link (2019) found that teachers' internal biases do not improve independently over time but can be addressed through professional development. Unlike Brookhart et al. (2016) previous findings, Guskey and Link (2019) found that effort was not a significant factor weighed by teachers when grading students.

<div align="center">**Reform Movements in Grading Practices**</div>

To address the inherent dissonance found in traditional grading practices, many researchers have advocated that teachers, districts, and policymakers move to either a standards-based model of assessment or a proficiency-based model of assessment (Brookhart, 1993; Custable et al., 2019; Guskey & Brookhart, 2019; Twadell et al., 2019). In both standards-based grading and proficiency-based grading, teachers utilized rubrics that are intended to define where in the continuum of learning a student may currently exist and, hopefully, eliminate the biases noted previously (Brookhart et al., 2016; Guskey & Brookhart, 2019; Quinn, 2020).

Standards-Based Grading

It is important to note that not all schools and secondary classrooms provide feedback to students using the 100-point or GPA scales alone. Standards-based grading (also referred to as criterion-based grading) has become more common in American schools, although its effectiveness is unclear (Coussens-Martin, 2019; Guskey & Brookhart, 2019). Bonner et al.

(2018) observed that SBG reform efforts have "dominated" education reform in the United States since the 1990s and that reform leaders borrowed heavily from organizational learning models that originated in industry (p. 72).

In standards-based grading, there are clear and objective frameworks by grade level of what students should know and be able to do (Coussens-Martin, 2019). Vatterott (2015) argued that the most important thing in a standards-based grading system is the quality of the feedback and not the grade. Guskey and Brookhart (2019) noted the developing body of research suggesting that SBG is a better communicative tool for progress with students and their families and that there is some correlation between state (federal) testing scores and well-designed SBG rubrics.

Bonner et al. (2018) reported that the adoption of standards-based teaching practices and assessments nationally, as required under the Every Student Succeeds Act (ESSA) of 2016, (including previous iterations of Federal Education Reform going back to the No Child Left Behind Act (NCLB) of 2002), has not resulted in the expected higher levels of alignment in teaching practices. In some cases, Science and Mathematics saw a regression in the desired alignment between practices and assessment data (Bonner et al., 2018). Kenna and Russell (2018) noted that between the adoption of NCLB in 2002 and ESSA in 2016, President Obama signed into law Race To The Top (RTTT) in 2011, which facilitated the adoption of the Common-Core Standards that had been under development by the National Governors Association; the Common-Core was intended to be a universal standards-based framework of curricula which aligned with the federal requirements (Conference of State Legislatures, 2014). Importantly, the Common-Core Standards did not propose changes to world language instruction (Conference of State Legislatures, 2014).

Coussens-Martin (2019) noted that the value of standards-based grading (SBG) is that the feedback comes to the students through multiple streams (cognitive and non-cognitive) and that it is focused on (student) "self-efficacy" rather "than with student compliance, reward, or punishment" (p. 59); this occurs because teachers are not assigning percentage or completion grades to non-cognitive tasks or for desired behaviors. Coussens-Martin (2019) also suggests that high schools can effectively implement standards-based grading and have very positive outcomes. This finding was amplified by Knight (2017), who found that adopters of SBG, in particular those in schools where adoption was widespread, saw an improvement in their planning, instruction, classroom environment, the role of assessment within the classroom, and the behaviors and characteristics of students in the classroom. Knight (2017) noted that teachers observe their students operating with a growth mindset more often in classrooms that use SBG. Quinn (2020) also points out that well-designed rubrics can correct teacher bias when assessing students.

While there is agreement among proponents on the rationale for adopting SBG, there are some clear disagreements. Coussens-Martin (2019) made the case that schools cannot merely plug standards-based grading into the typical high school grading environment, as the validity of the ratings in an SBG system are created differently. Vatterott (2015) suggested that despite the philosophical objections of many, teachers and schools can create conversation charts that convert rubric scores into more traditional grades, which can be used effectively. Knight (2017) suggested that administrators should include all stakeholders in the creation of SBG rubrics as a way to decrease potential pushback from individual stakeholder groups. This argument by Vatterott (2015), as noted earlier, is contradicted by Guskey (2020).

Guskey and Brookhart (2019) noted that while there is a developing body of research on SBG, particularly as a communication tool, effective implementation or failure, and the correlation between state (federal) testing scores and SBG rubrics. Further, Guskey and Brookhart (2019) noted that there has not been any research into the validity of the grades submitted and that of the pitfalls of traditional grading practices SBG was intended to mitigate, like a "hodgepodge grading" (p. 150), remain present in SBG. For example, Guskey (2020) highlighted the dissonance created when teachers convert rubric scores into an A-F letter grade.

Proficiency-Based Grading

Proficiency-based learning practices, which include proficiency-based grading, emerged at the same time as SBG as another potential solution for teachers and districts looking to move away from traditional grading practices; note, proficiency-based learning is synonymous with competency or mastery-based learning (Vermont Agency of Education, 2018). Custable et al. (2019) highlighted that "while there is great value in cooperative learning strategies, proficiency-based grading calls for a deeper and more transformative approach to learning: students need to co-construct their learning with their teacher or, importantly, with their peers" (p. 370). Further, Custable et al. (2019) posited that individual growth, the ability to reperform a task, and building the skills and awareness necessary to be reflective about your own learning are key aspects of proficiency-based learning. To track this growth, Custable et al. (2019) suggested that schools provide a context for the learning using rubrics and communicate it effectively with students, parents, and community members. These rubrics must be highly functional and be able to capture the evidence of what students are able to do with the knowledge and skills they have acquired (Custable et al., 2019).

In proficiency-based grading, students are assessed by using a rubric that articulates the expected outcomes (Custable et al., 2019). In this system, student work is graded quickly by the teacher and continually returned to students for resubmission until they have developed the skill or application (Twadell et al., 2019). Throughout the learning cycle, proficiency-based learning teachers focus on helping their students assess their own level of proficiency, establish goals, and help them achieve these goals through a process rooted in self-reflection through the rubrics (Twadell et al., 2019).

Custable et al. (2019) posited that proficiency exists within a continuum, and rubrics must be built around "student-friendly learning targets" (p. 31), such as proficiency or needs improvement for a single skill. Each level of proficiency must then have a descriptor of the behaviors or expected outcomes of a student at that level (Marzano et al., 2017). For example, when assessing students' vocabulary skills, a proficient student should be able to provide definitions that are complex and original and include supporting examples, while a student that needs improvement may only be able to define those same words using simple language or previously given definitions (Custable et al., 2019).

One of the major critiques of proficiency-based grading is that it incentivizes the "blind" retaking of work just to see if a student's work gets any better and that the actual grading of student work is therefore pushed back toward the end of the grading period (Vatterott, 2015). Custable et al. (2019) noted that it is a common temptation for students to "ease up on effort" to perform at a higher level later (p. 36). Vatterott (2015) argued that students should never be allowed to retest on any summative task unless they have already shown that they have mastered the concepts either in person with the teacher or through other work they have resubmitted. Ultimately, Custable et al. (2019) argued that students come to understand that their

ability to perform and reperform the same skill in multiple contexts is most important if the adoption process is done with fidelity.

Standards-based grading and proficiency-based grading are two heavily researched reform movements that were developed in response to the identified shortcomings of the 100-point/A-F grading scale (Brookhart, 1993; Custable et al., 2019; Guskey & Brookhart, 2019; Twadell et al., 2019). In standards-based grading, students are assessed by comparing their work to established criteria set by the district, county, or state where the student resides. For example, the widespread adoption of Common-Core Standards between 2011-2016 was an effort to standardize across states consistent standards-based curricula (Conference of State Legislatures, 2014; Custable et al., 2019; Kenna & Russell III, 2018). Teachers who utilized standards-based grading practices have been found to provide better feedback to students and their families about the student's learning progress, can build the self-efficacy skills of their students through the student's usage of the rubrics, and improve the instructional environments within their classroom (Coussens-Martin, 2019; Guskey & Brookhart, 2019; Knight, 2017; Vatterott, 2015). Teachers who adopt proficiency-based grading practices teach students how to assess their own progress as compared to clearly defined objectives for their learning and skill development using rubrics (Custable et al., 2019; Twadell et al., 2019; Vatterott, 2015).

Proficiency-Based Learning Environments

While this research study focuses on how secondary world language teachers described their lived experience of using proficiency-based rubrics to evaluate student performance, developing an understanding of the intended instructional environment of a proficiency-based learning environment is important. Marzano et al. (2017) posit that teachers in proficiency-based learning environments use many of the same instructional strategies as their peers in traditional

classrooms; they may simply use these strategies differently. In their work, Marzano et al. (2017) focused on the following strategies: instructional strategies, assessment strategies, and the use of rubrics within the assessment process.

Instructional Strategies

Hoegh et al. (2019) argued that proficiency rubrics "serve as the basis for instruction in a standards-based classroom" (p. 38). In particular, Hoegh et al. (2019) posited that teachers utilizing the rubrics correctly will focus on developing a plan for working with students on the content required by the rubric and that there will be less of an emphasis on any topics or considerations which are outside of the requirements of the rubric. Finally, Hoegh et al. (2019) suggested that by utilizing proficiency-based rubrics, teachers are better positioned to differentiate instruction.

Marzano et al. (2017) found that in traditional secondary classrooms, whole-group instruction, where the teacher teaches all the students in the classroom at the same time, is one of the most used instructional practices, and, therefore, there is a limited ability to differentiate instruction. Marzano et al. (2017) suggest that whole group instruction has a different role in the proficiency-based learning environment, given that students can be working on entirely different assignments simultaneously. In the proficiency-based classroom, whole-class instruction would only be valuable to the teacher when they are when addressing common "cognitive and metacognitive" skills of students (Marzano et al., 2017, p. 71).

In traditional classrooms, developing one-on-one relationships with students is often a goal of teachers using traditional grading practices, but it is not always the reality (Marzano et al., 2017). In proficiency-based classrooms, Marzano et al. (2017) posited that because teachers and students have one-on-one time every day together, their relationship is the primary means of

engagement, not something to be developed outside of the whole-group instruction found in more traditional classrooms. In a proficiency-based classroom, the one-on-one relationship between the teacher and student is the foundation of a student's learning environment (Marzano et al., 2017). Because of this, teachers can set higher and more personalized expectations for students than teachers in more traditional learning environments where there is one standard for every student (Hoegh et al., 2019). Being able to create this type of learning environment is essential for teachers who have adopted proficiency-based grading practices in their classrooms, as it can create a transformational learning environment for students (Twadell et al., 2019).

In traditional classrooms, teachers utilizing small-group instructional strategies may do so to provide differentiated support for students on an as-needed basis (Marzano et al., 2017). Depending upon the instructional plan for the class, teacher-led small-group instructional practices could resemble collaborative peer work, reteaching a concept or conferencing with a student about their work, or previewing a concept that is about to be taught. Marzano et al., (2017) noted that in a proficiency-based learning environment, small-group instruction with identified peer groups should also occur every day in the classroom.

Assessment Strategies in Proficiency-Based Learning Environments

At the center of proficiency-based grading practices is the assessment of student performance (Twadell et al., 2019). Twadell et al. (2019) promoted the idea that there are seven core beliefs that teachers must adhere to when implementing proficiency-based grading practices. Twadell et al.'s (2019) first belief is that teachers must strive to create a learning environment and curriculum around the individual growth of all their students. This growth is achieved by providing students with the ability to reflect on their own learning, identifying the opportunities for growth that they seized, and identifying future areas for growth. In creating

formative assessments, Hoegh et al. (2019) suggested that teachers utilize the learning targets found in the rubric when designing the assessment and teaching the students how to use the rubric to be reflective on their own growth.

In the second belief, Twadell et al. (2019) posited that the student's ability to reperform the task in various settings is how a teacher can assess the growth of their students. Twadell et al. (2019) argued that while there is a role for retaking assessments when a student hasn't performed well, to measure proficiency, students must be able to reperform a task in multiple settings. This type of experience creates the body of work necessary to assess the proficiency level of the student (Twadell et al., 2019, p. 36). Reperforming a task is different than retaking an assessment. The ability to retake assessments is one of the major critiques of the proficiency-based grading as it can appear to incentivize the "blind" retaking of work just to see if a student's work gets any better and that the actual grading of student work is therefore pushed back toward the end of the grading period (Vatterott, 2015).

Twadell et al. (2019) noted that the key to building students' ability to reflect on their own learning requires the ability for students to be able to accept feedback. Hoegh et al. (2019) expanded on this by arguing that when the instructional environment and assessments are designed around the rubric, the teacher's feedback using the rubric is more impactful because the connection between the teacher's feedback and the experience is readily apparent. Twadell et al. (2019) echoed this point when they noted that the key to making sure this is happening is to foster a classroom environment where students

> understand and use the learning targets and standards, understand and use the proficiency
>
> scale, understand and take action to extend or remediate their learning and to share a

common language with their peers, teachers, and interventionists to achieve this goal. (p. 38)

In proficiency-based learning environments, homework can play a very important role. In their next belief, Twadell et al. (2019) argue that homework should be personalized to the needs of the student as opposed to a universal experience for all students. Twadell et al. (2019) noted that homework should always be tied to a learning target to provide evidence of learning or create specific feedback about the learning. Doing so, Custable et al. (2019) suggested, helps to create an environment where homework is not an "act of compliance but as a form of reflection and feedback" (p. 40).

Grades provide teachers and parents or guardians a way to communicate with each other about a student's skill level and how they are progressing (Twadell et al., 2019). Many parents of secondary students experienced more traditional grading practices when they were in middle or high school and are, therefore, very familiar with the 100-point/A-F grading scales. In proficiency-based learning environments, assessments utilizing proficiency rubrics/scales should provide teachers and parents or guardians the ability to speak more concretely about a student's learning progress; this is the fifth of Twadell et al.'s (2019) core beliefs. These conversations can be incredibly powerful when students conduct a meaningful self-assessment and are able to develop the rationale for their final grade collaboratively with their teacher using the same rubric/scale (Twadell et al., 2019).

The final two of Twadell et al.'s (2019) core beliefs involve final summative assessments (exams) and the role of behaviorism in proficiency-based learning. Twadell et al. (2019) posited that while there may be a role for final exams within a student's course, the final exam should be focused on either providing students an opportunity to demonstrate advanced levels of

proficiency or an opportunity to reperform skills and tasks which will establish a level of proficiency for a student. Most importantly, Twadell et al. (2019) highlighted that both final exam designs could be assigned to students in the same class based on their individual needs.

In considering behaviorism, Twadell et al. (2019) outlined scenarios where assessing student behaviors and including them as part of a student's grade are acceptable and, conversely, are not. To many teachers, accepting the premise articulated by Olsen and Buchannan (2019) that behaviors should not be included in a student's grade intuitively makes sense. In Education Week, Ferlazzo (2012) used this scenario to describe the foolishness of taking this approach.

> If someone proposed combining measures of height, weight, diet, and exercise into a single number or mark to represent a person's physical condition, we would consider it laughable. How could the combination of such diverse measures yield anything meaningful? (Ferlazzo, 2012)

 Twadell et al. (2019) similarly posited that if teachers decide that including behavioral expectations within the construct of a student's grade, the behavioral expectations can be included only if they are "explicitly taught by the teacher during the learning process" (p. 45). Hoegh et al. (2019) concurred with this position, stating that by assessing and reporting on student behaviors, albeit separately from the assessment grade, makes the feedback more credible and impactful for students. Further, when the proficiency rubrics include behaviors that "promote academic success" (Hoegh et al., p. 180), teachers are able to speak directly about the behaviors, their impact on the student's work, and how the student can approach their work differently moving forward.

Rubrics in Proficiency-Based Grading

Proficiency-based rubrics, sometimes referred to as proficiency scales, are central to the experience of a proficiency-based learning classroom (Custable et al., 2019; Hoegh et al., 2019; Marzano et al., 2017; Twadell et al., 2019). The use of rubrics, Hawe et al. (2021) noted, is a way to teach students how to self-regulate in pursuit of their own learning and to "acquire evaluative knowledge and skills which will allow them to monitor (their) current understandings and performance, compare these to what is expected or desired and generate internal feedback" (p. 1034). Once these skills are developed, Hawe et al. (2021) stated that students and teachers are finally able to meaningfully assess a student's learning.

Hendry and Anderson (2013) state that rubrics allow students to understand the expectations of the teacher. Hoegh et al. (2019) note that rubrics "show teachers, students, and others what proficiency looks like, what knowledge and skills must be attained for mastery, and how students might go beyond what the standard requires" (p. 26). Further, Hoegh et al. (2019) found that proficiency-based rubrics provide teachers with a framework to design their instructional experiences around.

In their study, Lipnevich et al. (2020) noted that feedback from teachers that is specific to a student's "individual work and unaccompanied by grades" (p. 539) is the most impactful. Custable et al. (2019) suggest that "the rubric should have a proficiency scale, criteria, and a reflective component for both the teacher and student to comment about the evidence" (p. 52). This is important because by having each of these components, the student is better able to be reflective about their own practice, assess their own levels of growth, and gain confidence as a learner (Custable et al., 2019).

Lipnevich et al. (2014) also found that students improved at a higher rate when provided with only a rubric and not a rubric with an attached exemplar. Contrary to this finding, Hendry

and Anderson (2013) posit that the use of rubrics "in context" (p. 1034) with exemplars can help to develop the behaviors of a proficient learner (self-regulation, progress monitoring, etc.). In proficiency-based learning environments, teachers and students work collaboratively to establish individualized goals, which include self-regulation and progress monitoring throughout the process (Marzano et al., 2017) Finally, Custable et al. (2019) suggest that adopting proficiency-based grading practices fosters the self-efficacy of students and learning a new language, particularly with the development of clear communication skills.

Proficiency-Based Grading in World Language

Efficacious learners are mindful that there will be difficulties in acquiring the communication, listening, and cultural awareness required in learning a new language and are intrinsically motivated to achieve this goal (Custable et al., 2019). With the grading of student work replaced with an opportunity to reflect on where they, as individual learners, are based on the criteria set forth in the rubric, students are free to focus entirely on improving their learning and skill development (Twadell et al., 2019). In proficiency-based grading, reflecting on these skill-based and behavioral indicators is as important as the indicators on the product (Marzano et al., 2017).

Custable et al. (2019) argued that the lack of grading periods, often associated with proficiency-based learning environments, aligns well with the reality of learning a language. Specifically, they posited that learning a language is not a linear event and that instead, time, individualized support, and the ability to make mistakes without repercussions are advantageous to the foreign language student. Further, the ability of students and teachers to work collaboratively in the formative assessment process creates the foundation on which students

can, at an extremely high level, assume more ownership for their learning and their performance in the target language (Marzano et al., 2017).

American Council on the Teaching of Foreign Languages

Over the last 30 years, the use of proficiency-based teaching practices has increased nationally, and the usage of proficiency-based rubrics and assessments in secondary-leveled schools has increased at a commensurate level (Custable et al., 2019; Marzano et al., 2017; Twadell et al., 2019). One of the main reasons for this has been the adoption of the American Council of Teachers of Foreign Languages (ACTFL) proficiency standards in 2012 (American Council of Teachers of Foreign Languages, 2012; Johnson, 2008). While it has been common to see proficiency-based or competency-based education referred to when considering vocational settings, it is becoming increasingly common to see proficiency-based learning environments in secondary schools (American Council of Teachers of Foreign Languages, 2021; Johnson, 2008). ACTFL uses three primary rubrics (presentation, interpersonal, and interpretive communication) to assess student performance (Custable et al., 2019). Since their adoption in 2012, the ACTFL proficiency rubrics have become widely used by world language teachers across the country for use in the classroom and in national assessments such as the Standards-Based Measurement of Proficiency (STAMP) exam (Tigchelaar et al., 2017). Teachers use the STAMP exam nationally to determine if students earn the Seal of Biliteracy along with their high school diploma.

ACTFL first experimented with the development of a proficiency-based model of instruction in the 1970s when they partnered with ETS (Educational Testing Service) to design a credible proficiency-based instructional model for schools based on the model the State Department of the United States used (Warford, 2000). The ACTFL Proficiency Guidelines Document (2012) established precise levels of proficiency, which include distinguished,

superior, advanced, intermediate, and novice. This document also included rubrics to be used by teachers when assessing students listening, speaking, writing, and reading skills (Custable et al., 2019). These rubrics utilized a 0-4-point scale and have clear performance descriptors for each category and sub-category (American Council on the Teaching of Foreign Languages, 2022).

As is common in proficiency-based learning environments, ACTFL (2012) recommended that teachers and students utilize these rubrics throughout the year whenever they are assessing student work or performance. While students and teachers assess and provide feedback to the student throughout the various units of study during the school year, final classifications of students into their proficiency category do not occur until the end of the grading period (American Council on the Teaching of Foreign Languages, 2022). In proficiency-based learning environments, only the current performance at the end of the grading period matters when determining the student's level of performance.

Presentation Rubric

The ACTFL presentation rubric was designed to assess one-way communication that is non-interactive (American Council on the Teaching of Foreign Languages, 2012). Students rated in the intermediate proficiency range using the presentation rubric should be able to express their understanding of a topic using multiple sentences (American Council of Teachers of Foreign Language, 2022). Students are expected to be able to be understood by audiences, able to identify and self-correct errors, and be able to maintain the interest of their audience through the use of appropriate vocabulary, expressions, and, where appropriate, gestures (American Council of Teachers of Foreign Language, 2022).

Interpretive Communication Rubric

The ACTFL interpretive rubric was designed to assess the proficiency level of students as they interpret non-participative or overheard conversations (American Council on the Teaching of Foreign Languages, 2012). Teachers utilizing this rubric seek to establish the proficiency level of students based on the student's interpretation of listening examples (American Council of Teachers of Foreign Language, 2022). Using this rubric, students are assessed based on their ability to comprehend the main ideas in both complex texts and simple stories. Teachers utilizing this rubric seek to establish the proficiency level of students who are participating in brief social interactions (American Council of Teachers of Foreign Language, 2022).

Interpersonal Rubric

The ACTFL interpersonal rubric was designed to assess the two-way interactive communication between students (American Council on the Teaching of Foreign Languages, 2012). Teachers utilizing this rubric seek to establish the proficiency level of students who are participating in brief social interactions (American Council of Teachers of Foreign Language, 2022). In these interactions, students are expected to be able to participate in the conversation actively, ask questions of other participants to further the conversation and explore the answers of their classmates, and be able to answer questions of them with multi-sentence responses (American Council of Teachers of Foreign Language, 2022).

<div align="center">

Summary

</div>

This literature review tracked the evolution of grading practices, the identified need to reform grading practices, and the two most significant attempts at reform, standards-based grading and proficiency-based grading practices. While the newest iteration of standards-based and proficiency-based learning environments would be familiar to the teachers and students of the late 19th and early 20th centuries, the use of and centrality of rubrics in proficiency-based

learning environments is noticeably different from the previous iterations (Brookhart et al., 2016; Custable et al., 2019; Twadell et al., 2019).

Proficiency-based rubrics are the assessment tool utilized by teachers who have adopted the proficiency-based learning model of instruction (Custable et al., 2019). Proficiency-based rubrics utilize either a five or a four-point scale which, in world language classrooms, correspond to proficiency levels of the distinguished, superior, advanced, intermediate, and novice (Vermont Agency of Education, 2018). Included in proficiency-based rubrics for assessment are descriptions of the targeted element, skill, behavior, or outcome being assessed for each rating: beginning, developing, proficient, and expanding.

Proficiency-based rubrics are intended to create a more self-efficacious learner and increase the intrinsic motivation of the student and can help teachers design and focus their instruction (Coussens-Martin, 2019; Hoegh et al., 2019; Twadell et al., 2019). A clear weakness identified in the literature is the lack of evidence that adopting proficiency-based learning practices and assessments produces the intended outcomes; especially if it is susceptible to the same failings as the 100-point/A-F grading scales they were intended to replace (Guskey & Brookhart, 2020).

World language teachers across that country have adopted proficiency-based grading practices to be in alignment with the American Council of Teachers of Foreign Language's (ACTFL) 2012 Proficiency Guidelines (American Council of Teachers of Foreign Language, 2012). By adopting these guidelines and practices, an increasing number of world language teachers have begun utilizing interpretive, presentational, and interpersonal rubrics as part of their own classroom experiences. The purpose of this qualitative transcendental phenomenological study was to explore the lived experience of Massachusetts public secondary

(middle school/high school) world language teachers who utilized proficiency-based rubrics based on ACTFL 2012 Guidelines. Custable et al. (2019) note that the use of proficiency-based rubrics in world language classes is advantageous because the acquisition of a foreign language is a non-linear endeavor. Working without grading deadlines helps to foster the optimum learning environment for world language students (Custable et al., 2019).

CHAPTER 3: METHODOLOGY

Proficiency-based educational environments leverage a collaborative learning environment focused on developing a student to their potential through reflection, collaboration, and embracing failure (Twadell et al., 2019). Twadell et al. (2019) noted that teachers in proficiency-based learning environments often use rubrics, which often encapsulate the desired benchmark outcomes. The purpose of this transcendental phenomenological study was to explore the lived experiences of secondary world-language teachers currently utilizing assessment rubrics based on the 2012 American Council of Teachers of Foreign Language (ACTFL) proficiency-based rubrics to evaluate student performance.

Transcendental phenomenology was an appropriate methodology for this study because of the study's focus on a shared phenomenon. Specifically, this study examined how secondary world language teachers described their lived experiences using proficiency-based rubrics to evaluate student performance. Patton (2015) noted that "human beings make sense of (these) experience(s) and transform (them) into conscious" (p. 115). Bloomberg and Volpe (2019) described phenomenological research as studying a small sample of subjects through an "engagement to develop patterns and relationships of meaning" (p. 54).

Phenomenological research often follows one of two models found within the phenomenology theological framework. Transcendental phenomenology was articulated by the German philosopher Edmund Husserl (1964), and hermeneutic phenomenology was later articulated by Martin Heidegger (Heidegger, 1962). For the purposes of this study, the researcher utilized Husserl's (1964) transcendental theoretical framework, which seeks to understand the lived experiences of individuals by gathering descriptions, explications, and interpretations of their lived phenomena (Patton, 2015). Further, this study was grounded in Andrade and

Brookhart's (2020) classroom assessment theory (CA) which focuses on "how teachers and students gather, interpret, and use evidence of student learning" (p. 351).

Research Questions

This research study sought to answer the following research question.

What are the lived experiences of secondary world language teachers who use Proficiency Based Rubrics to evaluate student performance?

This qualitative transcendental phenomenological study sought to explore the lived experience of secondary school world language teachers who use proficiency-based rubrics to evaluate student performance. This study included semi structured participant interviews as the data collection tool. Interviews occurred over Zoom, and participants were public secondary world language teachers in Massachusetts who utilized the ACTFL rubrics in their teaching and possessed, at minimum, their initial teacher license from the Massachusetts Department of Elementary and Secondary Education.

Site Information and Demographics/Setting

To gather data, the researcher recruited Massachusetts world language teachers to participate in the study who are 18 or older, current (full or part-time) teachers in public secondary schools in Massachusetts, including both middle and high schools who utilized rubrics aligned with the 2012 ACTFL proficiency rubrics and possess, at minimum, their initial teacher license from the Massachusetts Department of Elementary and Secondary Education (DESE). As noted previously, no distinction was made between teachers who made the decision themselves to adopt the ACTFL proficiency rubrics, teachers who may have adopted them as part of curriculum change within their specific school or across their department, or those who began

teaching in a district where this practice was already in place. This site was chosen because of the recent adoption of new state frameworks in Massachusetts requiring world language teachers to adopt proficiency-based learning instruction (Massachusetts Department of Elementary and Secondary Education, 2021a). Importantly, these new frameworks required that teachers assess students utilizing rubrics aligned with the 2012 ACTFL proficiency guidelines. For the purposes of this study, a secondary world language teacher was defined as a teacher who teaches a language other than English to students in grades 6-12.

Participant and Sampling Method

Potential participants were asked to self-identify as meeting the study criteria, which included being over 18 years of age, a current (full or part-time) teacher in public secondary schools in Massachusetts, both middle and high schools, who have adopted the 2012 ACTFL Rubrics in their teaching and possess, at minimum, their initial teacher license from the Massachusetts Department of Elementary and Secondary Education; note, minimum licensure requirements include having earned a bachelor's degree. This study sought to explore the lived experience of secondary world language who use proficiency-based rubrics to evaluate student performance. Participants in this study represented a purposeful sample as they "have had a certain experience (or) knowledge of a specific phenomenon" to participate in the study (Ravitch & Carl, 2021, p. 83).

When determining the minimum sample size needed for the study, the researcher noted some ambiguity within the literature in establishing the number. Patton (2015) notes that establishing a set number of participants ahead of the study's beginning is not as important as ensuring that the study samples participants to the point of "redundancy" (p. 300) in participants' descriptions of the phenomena. Merriam and Tisdell (2016) advocate a similar position in

advocating that researchers achieve saturation in their sample. Specifically, Merriam and Tisdell (2016) posited that establishing the exact number of participants is less important than "maximizing" (p. 100) the data through the correct number of participants. Merriam and Tisdell (2016) further noted that the type and number of questions being asked and data being collected often have a significant impact on the overall number of participants needed as well.

Creswell and Creswell (2018) recommend a sample size of between three and 25 participants. In this study, the researcher set the minimum number of participants as eight, consistent with recommendations from Ravitch and Carl (2021). The Massachusetts Department of Elementary and Secondary Education (DESE) (2021b) reported that of the 2020-2021 school year, there was approximately 2,361 full-time equivalent (FTE) world language teachers at the middle and high school levels teaching in Massachusetts public schools.

Participants were required to have, at a minimum, a bachelor's degree to participate in the study. The transition to a proficiency-based model of instruction based on the 2012 ACTFL Guidelines and rubrics at secondary levels (middle and high) was a relatively new phenomenon (Brookhart et al., 2016), and therefore, years of service in teaching were not an important consideration. Further, it was appropriate to include all secondary teachers, not exclusively high school or middle school teachers, so as not to impose an unnecessary constraint on the population.

Specifically, the researcher recruited participants via email using publicly available world language teacher email addresses in districts where the adoption of the ACTFL proficiency standards and rubrics for grading had already occurred. This information was gathered from publicly available information found on school district websites that referred to Massachusetts Foreign Language Association (MAFLA) publications.

The recruitment emails (Appendix A) included the participant criteria and the Participant Information Sheet (Appendix B), which included information about the purpose of the project, why they were being asked to participate, what was involved in the process, potential risks and benefits of participation, privacy, and confidentiality during the study. The recruitment email and Participant Information Sheet asked participants to reply within a two-week window if they wished to participate in the study. Ten days after the initial email was sent, a reminder email was sent to all potential participants again, inviting them to participate in the study. During this period, the researcher began scheduling interviews with those who replied with interest in participating in the study, and all communication occurred using the researcher's University of New England email account. Interviews were scheduled at a mutually agreeable time, and a Zoom meeting link was sent to the participant. After an insufficient number of interested participants responded within the first two weeks, the recruitment email and Participant Information Sheet were sent out again. The researcher utilized the same process outlined above.

Instrumentation and Data Collection

Merriam and Tisdell (2016) posited that "data collection is about asking, watching, and reviewing" (p. 105) the data that needs to be collected in pursuit of the study's goals. Patton (2015) refers to this as seeking "to find out from (the participants) those things we cannot directly see and to understand what we've observed" (p. 426). However, before data collection and analysis could occur, approval for the study had to be obtained from the University of New England's Office of Research Integrity. After securing approval, the researcher recruited and sought the permission of each of the potential participants to participate in the study in accordance with the requirements of the Institutional Review Board.

This study recruited participants through email who were public secondary world language teachers from Massachusetts who evaluated their students' performance using a proficiency-based rubric. These participants were asked to self-identify as meeting the criteria set forth in the study. Interviews were then scheduled and took place via Zoom, where they were conducted, recorded, and then transcribed. Participant names and emails were held during the recruitment, screening, and interview phases of this research on a master list. A master list linked participants with their pseudonyms that were utilized in the study and were secured behind a password-protected Google Drive account, accessed via a password-protected computer, that only the researcher had access to until it was destroyed. The master list was not stored with any data gathered for the study.

Patten and Newhouse (2018) posited that qualitative research "collects data that will be analyzed as words" (p.159). This qualitative phenomenological study utilized semi structured interviews of participants throughout the data collection period to gather data about the study's central research question. Merriam and Tisdell (2016) noted that interviewing is necessary when observing the phenomena is not a possibility in real-time. While structured interviews and unstructured interview protocols were considered for this study, Merriam and Tisdell (2016) noted that structured interviews closely resemble an "oral survey" (p. 108) where there is very little opportunity to ask a participant to expand or clarify their answers. They further suggested that using an unstructured format would require a level of skill that this researcher does not currently possess and further note that the type of interview format is only occasionally found in the literature (Merriam & Tisdell, 2016). Therefore, these two protocols were ultimately deemed unsuitable. Peoples (2021) further posited that semi structured interviews are the preferred method for phenomenological research.

Patton (2015) posited that every interview is a two-way observation and that both participants, the interviewer, and the interviewee, continually assess each other throughout the interview. In particular, Patton (2015) cautioned that the interviewer must develop a level of trust with the participant and be authentic throughout the interview. One of the concerns Patton (2015) noted is "illusory superiority" (p. 429), where the interviewer overestimates their own questioning ability, which can lead to poor data collection and interviewee experience. The solution Patton (2015) proposes to counterbalance this is to develop a thorough and well-designed questioning protocol and to practice the skill of interviewing, listening, and observing non-participatory individuals ahead of the actual period of data collection.

The semi structured interview protocol (Appendix C) included a set list of questions developed by the researcher and was expected to be 60 minutes in duration (Patten & Newhouse, 2018). Importantly, these questions were focused on "experiences and not about thoughts or feelings" (Peoples, 2021, p. 52). Semi structured interviews also provided for flexible wording and ordering of questions during the interview allowing the researcher to adjust questions based on the previous answers or the participant or the natural direction of the conversation (Merriam & Tisdell, 2016). Further, all of the interviews occurred over Zoom as it was desirable to have not only transcription and video recordings available to ensure tone, behaviors, and intonation were accounted for within the transcript and in notetaking (Patten & Newhouse, 2018). For these reasons, the researcher asked that the participants all turn their Zoom cameras on during the interview unless they choose otherwise.

Throughout the study, the researcher continued to make their bias "explicit" (Peoples, 2021, p. 56) by journaling both before and after each interview. All journaling occurred in a password-protected Google Doc, secured in a password-protected Google Drive account, and

accessed via a password-protected computer that only the researcher had access to. Peoples (2021) argued that journaling is not a required step for bracketing analysis as bracketing itself, is a check on bias. However, Peoples (2021) contents that transcendental phenomenology does require the researcher to be reflective about their own thinking and how it may impact their research. Patten and Newhouse (2018) noted that this type of self-disclosure is extremely important in maintaining credibility for both the data collection and analysis of the data parts of this. Roulston (2010), interestingly, argued that researchers conducting phenomenological studies most often develop "romantic" (p. 58) questions that intend to "generate the kind of conversation that is intimate and self-revealing" (p. 56) and lack objectivity. In this phenomenological study, while it was the goal to capture the "personal descriptions of the lived experiences" (p. 434) in as much detail as possible, this researcher was not a world language teacher, nor had they taught using rubrics as the primary means of assessment. It was important that the researcher approach this study and each interview with an open mind and to ensure that their own personal feelings or opinions about either proficiency-based instruction or assessment using rubrics did not introduce bias into the study.

As noted previously, all interview recordings were transcribed using Zoom and reviewed by the researcher for accuracy. The recordings were downloaded from Zoom and secured in a password-protected Google Drive account, accessed via a password-protected computer that only the researcher had access to. The transcripts were then prepared using pseudonyms and by removing all identifying information. The list of pseudonyms and all identifying information was kept in a master list secured in a password-protected Google Drive account, accessed via a password-protected computer that only the researcher had access to. Participants had five

calendar days to review their transcripts for accuracy. If the participant did not reply within five calendar days, the transcript was considered accepted, and the master list and interview recordings were destroyed. From that point forward, the researcher only used the newly deidentified transcripts when conducting the data analysis of this study.

Data Analysis

The goal of this study was not to, as Peoples (2021) described, break the experience of these teachers into segments or parts but to instead understand and "illuminate" (p. 57) their experience as a whole. To achieve this, this study utilized a data analysis procedure articulated by Creswell and Creswell (2018). Specifically, the researcher followed Creswell and Creswell's (2018) five-step approach to data analysis: organized and prepared the data for analysis, read or looked at all the data, coded the data, generated a description and themes, and finally represented the description and themes.

The first and second steps, organizing and preparing the data for analysis, involved transcribing the interviews, deidentifying the participants, a review of the transcripts for their accuracy by the participants, and arranging the data for analysis. To do this, the researcher generated transcriptions of the interviews using the transcription tool in Zoom. The researcher then compared each transcript to the original recording and edited errors in the transcript. Each transcript was also formatted in a manner to better facilitate the analysis of the data gathered; this included line numbering on the left side of the page (Merriam & Tisdell, 2016). Next, the researcher removed any identifying personally identifiable information (PII) provided by the participant in their interview (names, site information, etc.) and replaced each with a pseudonym recorded in the master list in accordance with IRB procedures. The master list and all original

transcripts, and video recordings of the interviews were secured in a password-protected Google Drive, on a password-protected computer that only the researcher had access to. After removing all identifying information, the researcher sent the transcript to the participant so that they could verify for accuracy and asked that it be returned with any corrections within a five calendar day period. If, after five calendar days, no corrections were returned, the researcher considered the transcript as accepted. Once the transcripts were accepted, the master list and the video recordings were immediately destroyed.

In Creswell and Creswell's (2018) third step, the coding of the data begins. Importantly, the coding for this study was done by hand utilizing annotated transcripts of participant interviews. The purpose of the coding process is to break apart and categorize segments of the interviews to help make sense of the experiences of the participants (Creswell & Creswell, 2018). To achieve this, the researcher needed to read and delete any irrelevant information or filler words or sounds to create a clear statement by the participants (Peoples, 2021).

In the fourth step, the researcher established the initial themes that emerged from the interviews. To do so, the researcher attempted to identify the preliminary meaning units in the interviews of the participants. After coding each of the participant's interviews and comparing preliminary meaning units, the researcher established the final meaning units, or themes, that emerged from the data (Peoples, 2021). Creswell and Creswell (2018) posited that themes should display "multiple perspectives from individuals and be supported by diverse quotations and specific evidence" (p. 193). Finally, in the fifth step, the researcher constructed a written narrative of the themes developed, which provided a general description of the phenomena based on the data across each of the interviews. Throughout the study, all the data and documents developed and utilized in steps three-five were secured in a password-protected Google Drive on

a password-protected computer that only the researcher had access to for the duration of the study. Finally, three years from the publication of this study, all of the study data secured in the password-protected Google Drive will be destroyed.

Limitations, Delimitations, and Ethical Issues

Creswell and Creswell (2018) described limitations as "weaknesses in the research that the researcher acknowledges so that future studies will not suffer from the same" (p. 199). Creswell and Creswell (2018) further suggested that by stating the limitations of the study, the researcher can also "provide suggestions for future research, remedies for potential weakness in their study or advance new leads or directions that point to useful applications for knowledge" (p. 199). This transcendental phenomenological study sought to understand the lived experience of public secondary (middle/high) world language teachers in Massachusetts who used proficiency-based rubrics to evaluate student performance. In this study, the researcher set the number of participants as eight, consistent with recommendations from Ravitch and Carl (2021). The Massachusetts Department of Elementary and Secondary Education (DESE) (2021b) reported that of the 2020-2021 school year, there was approximately 2,361 full-time equivalent (FTE) world language teachers at the middle and high school levels across Massachusetts. Further, the number of participants (8) represents a small number of all world language teachers in Massachusetts, and therefore, the data should not be generalized to represent the shared experience of all teachers within a given school or within Massachusetts.

Roberts and Hyatt (2019) described delimitations as the "boundaries of your study" (p. 110). In this study, the researcher confined the eligible participants the researcher recruited to Massachusetts world language teachers who were 18 or older, current (full or part-time) teachers in public secondary schools in Massachusetts, including both middle and high schools who

utilized rubrics aligned with the 2012 ACTFL proficiency rubrics and possess, at minimum, their

initial teacher license from the Massachusetts Department of Elementary and Secondary

Education (DESE). Further, the researcher delimited the potential data gathered in the study by

limiting participants to a semi structured interview and not including a review of documents or

re-interviews as part of the study design. Additionally, the researcher limited the participants to

those who have adopted ACTFL's (2012) proficiency-based rubrics. World language teachers

using earlier editions of the ACTFL proficiency-based rubrics, another proficiency-based

approach, or who had not used ACTFL's 2012 Proficiency Based Rubrics (2012) were not

eligible to be a participant in the study.

To the best of the researcher's knowledge and ability, the research, data collection, and

analysis of data met the ethical standards of the University of New England and the Office of

Research Integrity. In particular, the researcher adhered to a tenant of the Belmont Report,

respect for the participants (Creswell & Guetterman, 2019). To achieve this, the researcher

always maintained the confidentiality of the participants. Throughout the study, the researcher

secured all transcripts, video recordings, and the master list in a password-protected Google

Drive account, accessed via a password-protected computer that only the researcher had access

to. Once each participant had reviewed and verified their transcript for accuracy or if they failed

to respond within five calendar days after a copy of the transcript was provided to them and was,

therefore, accepted by the researcher, the researcher destroyed the original transcript, video

recordings, and the master list for each participant. Finally, all participants participated of their

own volition.

Trustworthiness

During this study, the researcher utilized multiple strategies to increase the trustworthiness of the data (Creswell & Creswell, 2018). In particular, the researcher asked each of the participants to review their interview transcript to ensure accuracy. Throughout the study, the researcher sought to clarify any bias they may have, whether it is by journaling before conducting the interview with the participants or by making sure that any identified biases were clearly stated within the narratives contained within the study (Creswell & Creswell, 2018).

Credibility

Bloomberg and Volpe (2019) described credibility as whether the "participants' perceptions match up with the researcher's portrayal of them" (p. 202). In this study, the researcher took multiple steps to ensure that the credibility of the data was not in question. As noted previously, the researcher made efforts to monitor their own subjective biases and any external factors occurring during the data collection period by journaling before and after each interview (Bloomberg & Volpe, 2019; Peoples, 2021). Each interview was recorded, and transcripts of the interviews were developed, edited for accuracy, and each participant was asked to review the transcript for accuracy. As this is a phenomenological study, the researcher noted any "negative instance or discrepant findings" (Bloomberg & Volpe, 2019, p. 203) that emerged in the data collection period.

Reliability

Merriam and Tisdell (2016) posited that reliability "refers to the extent to which findings can be replicated" (p. 246). In phenomenological research, researchers attempt to understand specific phenomena through the lived experience of an individual, and therefore, the findings of one researcher will not necessarily be the same as another examining the same phenomena (Merriam & Tisdell, 2016.) However, Merriam and Tisdell (2016) noted, when conducting

qualitative research, it's imperative to make sure that the "results are consistent with the data collected" (p. 250). In this study, the researcher utilized a Participant Information Sheet (Appendix B) describing the study's design, and each participant was asked the same questions utilizing a semi structured interview protocol (Appendix C).

Dependability

In this study, the researcher sought to achieve dependability by creating a research process that had clear procedures for data collection, storage, and analysis which could be utilized and reproduced by future researchers (Bloomberg & Volpe, 2019). To do so, the researcher preserved all transcripts and email communications generated while recruiting participants and coding the data, as well as journal entries by the researcher written during the data collection period of this study. These materials will be stored by the researcher for three years after the publication of the study in a password-protected Google Drive account, accessed via a password-protected computer, that only the researcher will have access to until it is destroyed.

Transferability

Bloomberg and Volpe (2019) posited that transferability is the ability of the design of the study to allow readers to determine if similar outcomes could occur in their setting. To achieve this, the researcher must present a complete and thorough description of the research setting, each participant, the data gathered, and the themes generated in the data analysis. This level of detail is what Bloomberg and Volpe (2019) described as a "thick description" and is essential for the reader to develop a "contextualized meaning" (p. 205) of the findings in this study. This is important as it helps the reader to identify and consider what they may or may not experience in

their own setting(s) when adopting proficiency-based rubrics for grading. This study design is transferable to other secondary public-school environments in Massachusetts.

Summary

The purpose of this chapter was to articulate the research methodology of this study. Specifically, this qualitative transcendental phenomenological study explored the lived experience of secondary (middle school/high school) world language teachers who utilize proficiency-based rubrics for grading based on ACTFL 2012 Guidelines. To do so, this study utilized a semi structured interview protocol as the means of data collection from participants, which Peoples (2021) noted is the preferred method for phenomenological research.

Participants were either full or part-time secondary world language teachers who have adopted the 2012 ACTFL Proficiency Guidelines (2012) in their classrooms and were recruited from across Massachusetts. In this study, the researcher set the minimum number of participants as eight, consistent with recommendations from Ravitch and Carl (2021). The Massachusetts Department of Elementary and Secondary Education (DESE) (2021b) reports that of the 2020-2021 school year, there were approximately 2,361 full-time equivalent (FTE) world language teachers at the middle and high school levels across the state and the number of participants equated to approximately .02% secondary world language teachers. Bloomberg and Volpe (2019) note that the "goal of qualitative research is therefore not to produce "truths" that can be generalized to other people or settings but rather to develop descriptive context-relevant findings that can be applicable to broader contexts while still maintaining their content-specific richness" (p. 205).

The researcher then utilized Creswell and Creswell's (2018) five-step approach to data analysis: organize and prepare the data for analysis, read or look at all the data, code the data,

generate a description and themes, and finally represent the description and themes (p. 192-193). The findings of this study should not be generalized to represent the shared experience of all secondary world-language teachers within the state of Massachusetts.

CHAPTER 4: RESULTS

The purpose of this qualitative transcendental phenomenological study was to explore the lived experience of Massachusetts public secondary (middle school/high school) world language teachers who utilized proficiency-based rubrics to evaluate student performance. Proficiency-based rubrics are an assessment tool utilized by teachers who have adopted the proficiency-based learning model of instruction (Custable et al., 2019). Proficiency-based rubrics utilize either a five or a four-point scale corresponding to designations of beginning, developing, proficient, and expanding (Vermont Agency of Education, 2018). Included in proficiency-based rubrics are descriptions of the targeted element, skill, behavior, or product being assessed for each rating: beginning, developing, proficient, and expanding (Twadell et al., 2019).

This research study sought to answer the following research question while utilizing data collected through semi-structured interviews. Eight participants were interviewed using an interview protocol (Appendix C), and names and any identifying information were replaced with pseudonyms. The following question was used to focus on the phenomenon:

What are the lived experiences of secondary world language teachers who used Proficiency-Based Rubrics to evaluate student performance?

The researcher recruited participants via email using publicly available world language teacher email addresses in districts where the adoption of the ACTFL proficiency standards and rubrics for grading had already occurred. This information was gathered from publicly available information found on school district websites that refer to Massachusetts Foreign Language Association (MAFLA) publications. Interviews were scheduled with each participant over Zoom. Subsequent to the interviews, transcripts were prepared and deidentified.

The data was coded after each interview transcript was reviewed by the participant. The purpose of the coding process is to break apart and categorize segments of the interviews to help make sense of the experiences of the participants (Creswell & Creswell, 2018). This chapter will highlight the themes that emerged from the data gathered through the semi-structured interviews. A brief vignette for each of the eight participants, which provides some necessary context for the reader about each participant, will be presented. This section will also include an analysis of the participants' responses during the interviews. Finally, the last section focus' on the emergent themes that were identified in the participants' responses through the coding process. These themes included that performance feedback focuses on students' individual growth, that environment dictates the level of adherence to the ACFTL rubrics, and that the proficiency-based classroom creates a safer and more collaborative learning environment.

Analysis Method

The goal of this study was not to, as Peoples (2021) described, break the experience of these teachers into segments or parts but to instead understand and "illuminate" (p. 57) their experience as a whole. To achieve this, this study utilized a data analysis procedure articulated by Creswell and Creswell (2018). Specifically, the researcher followed Creswell and Creswell's (2018) five-step approach to data analysis: organized and prepared the data for analysis, read or looked at all the data, coded the data, generated a description and themes, and finally represented the description and themes.

The first and second steps, organizing and preparing the data for analysis, involved transcribing the interviews, deidentifying the participant information, asking participants to review their transcript for accuracy, and arranging the data for analysis. To do this, the researcher generated transcriptions of the interviews using the auto transcription feature in Zoom. The

researcher then compared each transcript to the original audio and visual recordings and edited errors in the transcript. Each transcript was also formatted in a manner to better facilitate the analysis of the data gathered; this included line numbering on the left side of the page (Merriam & Tisdell, 2016). Next, the researcher removed any identifying personally identifiable information (PII) provided by the participant in their interview (names, site information, etc.) and replaced each with a pseudonym recorded in the master list in accordance with IRB procedures. The master list and all original transcripts, and video recordings of the interviews were secured in a password-protected Google Drive, on a password-protected computer that only the researcher had access to. After removing all identifying information, the researcher sent the transcript via email to the participant so that they could review it for accuracy asked that it be returned with any corrections within a five-day period. Only one of the eight participants responded with a request that small corrections be made to the transcript. Six of the eight participants accepted their transcript, and two did not reply within five days, and their transcript was considered accepted. Once the transcripts were accepted as being accurate, the master list and the video recording were immediately destroyed.

In Creswell and Creswell's (2018) third step, the coding of the data begins. The coding for this study was done by hand utilizing annotated transcripts of participant interviews. The purpose of the coding process is to break apart and categorize segments of the interviews to help make sense of the experiences of the participants (Creswell & Creswell, 2018). To achieve this, the researcher needed to read and delete any irrelevant information or filler words or sounds to create a clear statement by the participants (Peoples, 2021).

In the fourth step, the researcher established the initial themes that emerged from the interviews. To do so, the researcher attempted to identify the preliminary meaning units in the

interviews of the participants. After coding each of the participant's interviews and comparing preliminary meaning units, the researcher established the final meaning units, or themes, that emerged from the data (Peoples, 2021). Creswell and Creswell (2018) posited that themes should display "multiple perspectives from individuals and be supported by diverse quotations and specific evidence" (p. 193). Finally, in the fifth step, the researcher constructed a written narrative of the themes developed, which provided a general description of the phenomena based on the data across each of the interviews. Throughout the study, all the data and documents developed and utilized in steps three-five were secured in a password-protected Google Drive, on a password-protected computer that only the researcher had access to for the duration of the study. Finally, three years from the publication of this study, all of the study data secured in the password-protected Google Drive will be destroyed.

Presentation of Results and Findings

The first set of questions in the semi structured interviews asked each participant to provide background information, including how long they have been teaching, how many years they had been utilizing proficiency-based rubrics to assess their students, whether it was their choice to adopt the rubric, and the adoption process they went through. The vignettes below were created to provide the reader with important demographic context for each participant in the study. The researcher assigned pseudonyms to protect the confidentiality of all participants.

Victoria

Victoria has been a Spanish teacher for 30 years. She has been using proficiency-based rubrics for the last eight school years. The adoption of the proficiency-based rubrics was driven by her department head, driven by early adopters in her department. Victoria described the

period of adoption as being very "messy and contentious," because of the accelerated pace at which adoption occurred at her school.

Nancy

Nancy has been a French teacher for 26 years. She has been using proficiency-based rubrics for seven years. Nancy had long used rubrics to assess student performance, however, Nancy adopted the ACTFL rubrics seven years ago when she was seeking rubrics that aligned more with the three modes (interpretive, interpersonal, and presentational) that her classroom instruction had naturally evolved to focus on. While Nancy began her adoption process on her own, she was quickly joined by her colleagues at both the middle and high school levels, and they have been working collaboratively throughout the adoption process.

Mary

Mary is a Chinese/Mandarin teacher who has taught in four different districts over the last 10 years at both the middle and high school level. Mary began using proficiency-based rubrics in her first teaching position as a new teacher. That district had a number of schools in the district and had made the transition because of frequent movement of kids between schools. Mary attended two years of training in proficiency-based teaching and assessment and noted that although her classroom experiences are very similar to her Latin language peers. However, Mary shared that her proficiency benchmarks are often set slightly lower than her colleagues.

Madeline

Madeline is a middle school Spanish teacher who has been teaching for 20 years. Madeleine first began using proficiency-based rubrics in a previous district where there had been a structured adoption led by her department head. Madeline shared that in her current district, she

does not have a department head and has had to look outside of her school district for guidance. Specifically, Madeline noted that she stays current with ACTFL publications.

Jennifer

Jennifer has been a world language teacher for 23 years and began using proficiency-based rubrics eight years ago. While she had been aware of the movement toward adopting rubrics and the department had begun to consider implementation, Jennifer shared that in her previous district, she was surprised to learn upon returning to school one fall that all of the world language teachers would be compelled to begin using the rubrics on the first day of school. Jennifer now teaches in a new district and has maintained her usage of the rubrics to assess student performance.

Fred

Fred is in his eighth year of teaching Spanish and his fifth year of utilizing proficiency-based rubrics in the classroom. Fred began using the rubrics after studying them in graduate school and shared that he was given the freedom to experiment with and adopt them in his classroom by his department head. Fred shared that it has taken him about five years to get all of the necessary classroom components in place for him to really transform his instruction to support proficiency-based learning.

Catherine

Catherine has been a high school French teacher for 13 years and has used proficiency-based rubrics for the last 11 years. After encountering the rubrics for the first time in graduate school, Catherine noted that she was given permission to adopt and integrate the rubrics into her classroom. Catherine shared that overt time she has begun to use the rubrics in a more student friendly or student facing way in a unit as opposed to a strictly summative assessment tool.

John

John has been a high school Spanish teacher for 16 years and has used proficiency-based rubrics to evaluate student performance in world language for the last 8 years. John first began to explore proficiency-based learning and assessments when was struggling with grading students in his classes. Proficiency based assessments provided him with a framework and terminology that he could use with his students in that classroom and that aligned with his professional practice. John shared that he first attended the MaFLA Proficiency Academy with colleagues and that they began to personalize rubrics for adoption in their district during the conference. These became the templates used by the district the following year. Recently, John has switched districts.

Understanding the Phenomenon of Using Proficiency-Based Rubrics to Assess Student Performance

During the semi structured interviews, using the interview protocol found in Appendix C, the participants were asked questions about their adoption of rubrics, their usage of the rubrics, how they assess student performance using the rubrics, and how their usage of the rubrics has impacted their instruction. Each section contained between four and seven questions and the questions provided participants with the ability to speak to a different aspect of each of these four areas. During the individual interview sessions, no participant declined to answer a question.

Adoption of the Rubrics

The participants in this study adopted proficiency-based rubrics between five and 11 years ago. All of the participants mentioned that had attended or participated in Massachusetts Foreign Language Association (MaFLA) workshops or conferences on proficiency-based instruction or assessment during their period of adoption. Five of the participants (Catherine,

Fred, Nancy, John, and Madeline) adopted the rubrics at a time of their choosing while three teachers (Mary, Victoria, and Jennifer) were required to adopt the rubrics by their district. Catherine and Fred both chose to adopt the rubrics after being exposed to the ACTFL rubrics in graduate school. Madeline shared that she had begun to implement the rubrics on her own after reading about them in ACTFL publications. Nancy noted that her adoption of the ACTFL rubrics came after many years of using a previous generation of proficiency-based rubric that were not aligned with the ACTFL rubrics and John pursued adoption because the rubrics provided him with verbiage and a framework with which he could communicate and assess his students that aligned with his personal philosophy.

Victoria and Jennifer both were compelled to adopt the rubrics in the same school year when they worked together in the same district. Both Victoria and Jennifer shared that they were not expecting to be directed to adopt the rubrics until they returned to school in August and learned that their department head had unilaterally made this decision. Both teachers described this transition as difficult and Victoria, in particular, described the pace of adoption as "messy" and "contentious." Jennifer has since left this district.

Usage

All eight of the participants utilized the ACTFL Interpretive, Presentational, and Interpersonal Rubrics to assess students. However, these were not the only proficiency-based rubrics these teachers were using. Interestingly, four of the five teachers who chose to adopt the ACTFL proficiency-based rubrics on their own, also use additional rubrics from outside sources. Fred shared that he utilized proficiency-based rubrics from the Creative Language Classroom which published the *Adios* textbook that they use in his classroom. Madeline and Nancy shared that they also utilized the Can-Do Statements, published by ACTFL, and multiple teachers used

the Comprehensible Input/Output frameworks, also designed by ACTFL, as a curriculum progression to help develop proficiency. One of the three teachers who were compelled to adopt the ACTFL rubrics, Jennifer, is now also utilizing additional rubrics beyond the three ACTFL ones. As noted earlier, Jennifer has changed districts and currently has the flexibility to do so.

When asked about the role proficiency-based rubrics play when communicating with parents, most of the participants indicated that the rubrics play only a minimal role. Mary shared that "most parents don't understand what proficiency is" and that they ultimately care more about the student's grade. This sentiment was echoed by both Jennifer, Madeline, Victoria, and Fred. Conversely, John said that he was able to use the rubrics to shift from conversations about student grades and into skills and performances. Catherine noted that while parents often don't really understand what proficiency is, she is often able to use the rubrics to articulate the standard and provide "evidence (as to) whether or not they are meeting the standard." Nancy shared that being able to cite the rubrics when talking with parents is "huge" because their standards-based report card fields are directly linked to the descriptors of the ACTFL proficiency levels.

Nancy and John shared that they only use the ACTFL proficiency-based rubrics when assessing summative assignments. Mary, Victoria, Jennifer, Madeline, and Catherine stated that they modify or simplify the rubric when using it to assess students formatively. Catherine and Jennifer stated that they provide students with a greater volume of feedback and more specificity on formative assessments than on summative assessments. Catherine stated that in her opinion, feedback "should decrease over the unit as students grow and develop skills." Fred was the only participant to share that he approaches assessment using the rubrics the same way on both the formative and summative assessments.

The teachers all described using the rubrics as a tool that students can use to monitor their own performance. However, their descriptions and implications of the usage of the rubric varied greatly. For example, Mary said that she takes deliberate steps not to use the rubrics in situations where kids will feel like they have input on their grades because she feels her middle school students will feel pressure to give themselves an A. This contrasts with Madeline who noted that her middle school students are often more critical of themselves then she would be. Fred stated that he likes to set proficiency targets using the rubric and encourages his students to self-assess their own performance using accuracy ranges that he provides like between 60-80% or 80-100%.

When asked to describe how their expectations of student behaviors has evolved since using proficiency-based rubrics, the participant responses varied. For example, Jennifer noted that students in her lower-level classes find the environment a

> challenge, especially (when) doing anything interpersonal. The kids either don't like the language because they feel uncomfortable and embarrassed, or they don't like each other. However, I like that they do feel more comfortable participating because they know that they're not penalized heavily for making mistakes.

Victoria noted that there was a real improvement and a higher level of confidence when speaking Spanish by her students. Madeline and Fred expressed a similar statement. Fred stated that

> my classes have become much more communal and just social, in general, in the target language. I've found a lot of success in the empowerment of students and just in the empowerment of their ability and being able to show them places they're improving and where they would otherwise generally feel like they don't have skill. In my district, students have very low self-esteem in terms of their ability because they're always trying

to be perfect. And when I teach the kids, they don't have to be perfect in order to succeed, they are much more willing.

Only John noted that his expectations for student behavior hasn't changed because his goal was always language acquisition. What he has found in adopting the proficiency-based rubrics that that his students better understand the connection between their practice and their ability to acquire and develop their skills.

Assessing Students

The first question in this section asked the participants what "evaluating students" means to them and how proficiency-based rubrics help them to do so. The responses of the participants were all very different. For example, Mary noted that evaluating students means that you are constantly setting benchmarks and assessing them with formative and summative assessments. Victoria, noted that the rubrics

> give structure to what we're looking for students to do. Presentational rubrics allow me to assess what is the skill that you have that you can apply to this presentation that you're giving. Interpersonal rubrics allow me to assess what students can they do in the moment right. Do they understand what's being said to them and can they engage in the conversation? The interpretive rubrics allow me to assess what kids are taking in (listening and comprehending).

Nancy, who teaches in a standards-based middle school, stated that she prefers to think of evaluating students as meaning "giving feedback on how their performance is measure up to the performance targets for the course." Nancy went on to share that she hopes that her classroom environment and usage of the rubrics will help to empower her students to feel like they are in control of their learning. Fred described the rubrics as providing proof of what the students can

do and, like Nancy, he hopes that that his feedback can help his students to reflect on their

learning and growth. Finally, Catherine described evaluating students as

> letting them know where they are on that journey based on what my expectations were.
>
> The feedback (via the rubrics) is so much more important because it allows students to
>
> know what exactly they need to do next in order to move forward in proficiency.

All the participants noted that they use proficiency-based rubrics with both formative and

summative assignments. Victoria noted that she occasionally with use them for "check-ins along

the way." Fred shared that he assessed students with the rubrics across all modes of

communication, however he does not assess their homework.

As noted previously, all the participants assess student growth through the rubrics, only

Nancy reported out to students and families using rubric language on a report card. However,

Nancy noted that she also tracks

> growth in other ways also. I have my seating charts. I'm keeping notes on seating charts
>
> about student behaviors, usually negative ones that I want to address. So, I can see this is
>
> the third time you've come to class without your packet, or this is your fourth time
>
> speaking English this term where I've called you out for it.

In addition to the rubrics, seven of the eight participants also reported student

performance using traditional 100-point A/F grading practices and many included non-cognitive

categories as part of their grading scheme. For example, Victoria included both in class

participation and homework completion in students' final grades. Madeline, Nancy, and Jennifer

also included homework as well. Fred, Catherine, and John do not assign homework as they feel

that kids can really only practice and improve their language skills when they are together in

class. However, Catherine does include other non-cognitive assessments data within her students' final grades. Catherine shared that

> the students have their own growth goals and so they're tracking that and at the end of the quarter they give me a reflection on how they've grown in those ways. Sometimes they have nothing to do with proficiency, but sort of self-management skills…and if those are going to help (the student) in the future, whether it's in French class or not, I totally support that goal.

Instruction

All the participants shared that their usage of the rubrics has provided them with an opportunity to reflect upon and improve their practice. Fred noted that

> It's definitely made me reconsider what I look for when I ask a student to do a task. It has made me really consider; do I care that a student's pronunciation isn't perfect? Because even though it's comprehensible, there are things that I think in traditional rubrics that are very arbitrary, that proficiency-based rubrics, true proficiency-based rubrics do not take into account because it doesn't matter whether or not a student has perfect grammar or perfect pronunciation. Because if they are communicating and they are comprehensible, then what do I really care about.

John added that he feels it has changed the overall structure of his classes significantly. When he began teaching, he said that she focused on things like grammatical tenses and other "surface level knowledge." Now he finds that his students can focus on speaking and communicating.

> Nancy shared that adopting proficiency-based rubrics

> totally changed my practice. I wasn't even assessing my students interpersonal speaking or interpretive reading and listening skills before I knew about these rubrics. There was

no negotiation of meaning or spontaneity, and I never asked them to read or listen to an authentic resource and answer questions about it in English. Those things were completely absent from my practice. (The rubrics have) provided a much more well-rounded focus for the course so that students would really develop proficiency in all three modes. And it made a big impact on students' grades in the course. Once we moved to standards based with these rubrics, no one was getting a D or an F anymore.

Jennifer added that using the rubrics have helped to keep her grading practices "honest" and that students cannot hide when being assessed with the rubrics. Victoria shared that the adoption of more authentic resources in the classroom has helped her to be "very current in what's happening in the world today, and I love that part of it." Finally, Mary noted that she was able to free herself from a textbook and that she can "see (her) students enjoy class more," which makes her classroom more relaxing and engaging.

When asked about some of the ways that they differentiate their instruction for students using the rubrics, the answers were also quite varied. For example, Fred said that he uses a lot of meaningful grouping of like students which allows him to better tier his instruction to students in the classroom. John agreed with this sentiment, noting that he sets different output goals for students within the same assignment by using the rubrics. Catherine noted that she routinely adapts her students intended outcomes based on their IEPs and 504s and that this does not impact where her students are in the curriculum at the end of the grading period or by the end of the year. Victoria shared that her students are able to set and work at their pace and build confidence along the way.

Emergent Themes

The eight interviews in this study totaled approximately 4.35 hours in duration and produced 2,887 lines of data to be analyzed. Ravitch and Carl (2021) posit that themes do not simply emerge from the data gathered by the researcher but are instead actively constructed and developed by the researcher. After following Creswell and Creswell's (2018) method for analyzing the data and reflecting upon the initial codes created and the totality of the data, the following three themes emerged:

Theme #1. Performance feedback focuses on students' individual growth

Theme #2. The environment dictates the level of adherence to the ACTFL rubrics

Theme #3. The proficiency-based classroom creates a safer and more collaborative environment

Theme 1: Performance feedback focuses on students' individual growth

Each of the participants were asked multiple interview questions about how they rate students' overall performance, using the proficiency-based rubrics, their grading practices, and if they take into account students' self-assessments when making their summative ratings for students. Only one of the eight participants, Nancy, reported student progress using a standards-based scale aligned with the targets in the proficiency-based rubrics. Seven of the eight participants ultimately report student performance through more traditional grading practices with both a numeric and letter grade and therefore, have a parallel means of assessing student performance in addition to the rubrics.

There was also a wide variety in how and when the participants used the rubrics to assess students, specifically as it related to each mode (interpersonal, interpretive, presentational.) For example, Victoria only assessed with the rubrics for summative assessments across all three modes. Catherine noted that she uses the rubrics across all three modes for both formative and

summative assessments, however, she has modified her practices with the interpretive rubric because, essentially, she finds herself assessing is the answer correct or not, and not the elements of a student response. This contrasts with Marie who noted that she is focusing more recently on assessing her middle school students with the interpretive rubric and less frequently using the interpersonal and presentational rubrics.

Defining and Setting Expectations. For all the inconsistencies in usage between the participants, there were some common experiences that the participants articulated about how, specifically, by using the rubrics to assess their students, their conversations with students have become more about student growth and less about achievement. John shared that he has been able to shift his conversations with students and parents to be less about the letter grade on the assessment to being more about their vocabulary usage, their text type and what they can do to improve for next time.

Catherine shared that each fall her students complete an interpersonal speaking assessment in the lab. Part of this assignment required them to write a transcript of the recording and self-assess their performance with the rubric. Catherine shared that a major benefit of having students self-assess in this manner is that it provided time for students to learn how to use the rubric and to be able to better interpret the feedback she provides later in the year. Catherine noted that "Once we've done that once, the students are so much more conscientious about what they're doing."

Nancy, who teaches in a school that reports student performance using the ACTFL levels on their standards-based report cards, stated that she uses the performance indicators to set goals for students on how they can reach the next level of performance. Fred made a similar statement when noting that he is able to use the rubric to describe the expected level of proficiency,

intermediate mid, for example, and then describe what indicators (language, text type, etc.) would show him that a student is working at that level. Fred noted that a benefit of working in this manner is that he can then define exceeding expectations with even more specific expectations for language function and text type and introduce levels of accuracy such as between 80-100%.

John said that he often tries to provide feedback on assessments by saying "Here's where you were successful and here's where you can level up." Doing so, John notes, moves the conversation away from the things the student did wrong to and what prevented the student from getting an A, but more on how they can continue to improve. This, shared John, is important because it helps to foster a growth mindset in his students.

Self-Reflection. All of the participants stated that they collect self-reflection data from their students at regular intervals throughout the grading periods. However, the manner in which the self-reflection data is collected and used by the teacher varied differently. For example, the middle school teachers, Mary, Madeline, and Nancy all described their self-reflection rubrics as being three or four questions long with rating scales and a place where students could provide some more narrative feedback. Madeline stated that she will have her students pre-score their work and then she will ask them later to compare her ratings to theirs to find places of agreement and places where they disagree. Madeline went on to say that her students were usually much harder on themselves then she is.

Nancy shared that she takes this same approach a step further by asking students to use the feedback to outline what they will do going forward; she calls this "feed forward." This is a term Catherine used as well. In particular, Nancy requires that her students not only process the feedback she provided them through the creation of a narrative where they articulate their growth

since the last assignment and to set goals for their next performance. She will also have them

create self-assessments are the mid-point of the quarter where they will reflect on

> all the rubrics that they've received so far and try to think about, so what would your
>
> grade be? Are you happy with that? If not, what's your plan? What do you need from me?
>
> What do you need to do?

Victoria described a nearly identical process where she and the student fill out opposite sides of

the same rubric. Victoria said she tries to highlight both commendations and recommendations

for growth and the student is asked to do the same. Victoria's observation is that her students

know what their strengths and weaknesses are and they are not afraid to work on them once

identified.

> Catherine who teaches high school French, shared that she decided some years ago to

allow students to track their own growth goals in addition to the proficiency rubrics, which the

kids report to her on every quarter. Catherine noted that

> Sometimes they have nothing to do with proficiency, but sort of self-management skills
>
> and if those are going to help you in the future, whether it's in French class or not, I
>
> totally support that goal. I had a kid who (stated), "I don't analyze great. I need to do
>
> better on that. And so, my goal is going to be to read some people's analyses, understand
>
> what it takes to analyze and then apply that to my own work." And he did it, it was
>
> amazing.

> Jennifer, who teaches primarily students in level three and four classes, noted that one of

her real struggles is that kids don't move quickly through the proficiency scale levels. For

example, while a student may rapidly move from intermediate low to intermediate mid in the fall

of their ninth-grade year, it is not unrealistic to think they may remain at intermediate mid for

two years or more. As a result, she does not do a lot of self-reflection using the rubrics because she doesn't want them to feel like a lack of progress reflects poorly on them when it is the expected outcome. Instead, Jennifer focuses on having the students self-assess more often using Comprehensible Inputs which is a companion approach to skill and language development in proficiency-based world language classrooms (American Council of Teachers of Foreign Languages, 2023).

Nancy stated that she uses the self-reflection feedback a lot when conferencing with students. In her interview she stated that if a student didn't indicate that they studied well, or completed homework, etc. on their rubrics, that she will try to make this connection to her rating of their performance with the rubric. As an example of this, Nancy provided the following scenario about a student she asked to come and see her about study strategies they were using.

> We just checked in about the strategy you're using. Because really, if you're doing that and you're still spelling everything wrong, we probably should find a different strategy for you.

1:1 Conversations about Learning. Each of the participants were asked to describe their conversations with students about their learning since adopting proficiency-based rubrics. The participants uniformly responded that the quality of those conversations were more focused on student outcomes and them as learners and less on the transactional aspects of their performance when looked at through the prism of traditional grading practices. For example, John shared that it is no longer,

> Why did I get a 93 and this person got a 92? Or this one got a 94?" (I) can look at (the student's) writing and can say, your vocab is at this level, your text type is here, and the impactfulness of your errors is right here, and so you add up these scores and this is why

you got what you got. I have found that students actually end up taking ownership (when they) get an unfavorable score, then they come and seek me out and they say, "How can I improve on this skill? Because I noticed I haven't been doing too well.

Fred noted that his conversations have become more "realistic, positive, and constructive." Fred went on to suggest that these that he finds himself "giving pep talks, but at the same time, I find them telling me that they're doing better than they thought they would." Building the confidence of his as learners is something that Fred said he invests significant energy in now.

Nancy noted that by assessing students across all three modes, she feels as though she can always find where kids have strengths. For example, she noted that "some of her kids who really, really struggled to produce comprehensible output, might be very good interpreters of authentic resources." She went on to note that her conversations with students are "now much less evaluative and they're much more descriptive." These descriptions she said come directly out of the rubrics.

Victoria shared that she feels that she is "able to encourage them with the rubrics instead of saying, "No, that's not the way you say it.", or "This isn't the right conjugation." She noted that 10 or 15 years ago, this was a standard part of her practice. Instead, she now focuses on using each conversation as building blocks toward proficiency.

Theme 2: The Environment Dictates the Level of Adherence to the ACFTL Rubrics

In districts or schools where participants have the freedom to experiment with the rubrics, they do. Victoria and Mary, who both teach in schools where these practices are uniform across the department or grade level, do not currently have the freedom to do so. Fred shared while he uses

all three modes of the ACTFL rubrics (interpersonal, interpretive, presentational)…I only use what I feel are the most vital components of each because I don't agree with all of it. Sometimes, as Fred and Madeline noted, this is because their classroom or online texts have rubrics that better align with the tasks and curricula asked of their students. Catherine shared she has developed two interpersonal rubrics that she uses consistently across different grade levels. One of these rubrics focuses on interpersonal speaking and another one focuses on interpersonal writing.

Nancy shared that her department created and continually revises their three core rubrics annually. At this time her middle school students are

> using is a single point rubric where the target is whatever the established performance target is for the course…And then I'm breaking down the can do or I'm breaking down the performance target, which is intermediate low, which is stated up here into three key domains of what does it mean to be an intermediate low when it comes to presentational writing. The first is comprehensibility, accuracy, which is I can be understood by someone used to a language learner. The next one is vocabulary. I can use a variety of familiar and personalized words and expressions, give some details. And then the last one is text type. I can use simple strings of sentences, original sentences, and pose basic questions…Some people might also add in a piece about accuracy of the cultural information depending on the prompt. In presentational writing, we try to be very, very faithful to that language when we break down the components of the target. We're also very careful that we say performance target and not proficiency because just because you can do this at intermediate low, you're not really an intermediate low because I've totally coached you up to this and I've given you all the specific vocabulary for this task.

In her interview, Catherine shared that she has also developed her own interpretive speaking and writing rubrics because, as she describes it, if you don't take a more holistic approach you are only assessing if the answer is right or wrong. She shared that with her holistic rubric,

> It's not just if you got four out of five questions right, you have an 80%. It's more (that) I try to have a couple of questions on the main idea, a couple of questions on details, a couple of questions on the cultural influences, and then those sort of go to the rubric. And the rubric says, were you able to (identify) the main idea? Were you able (identify) details?

This way, Catherine notes, that if a student got one of those questions wrong, they still have the ability to demonstrate that they met expectations. As noted previously, 20% of the grade in Catherine's classes are based upon students' self-generated growth goals. Catherine noted that while it is most common to have these goals be based upon developing French skills, students often create goals based around "study habits, turning in classwork on time or getting enough sleep." Catherine has noticed that by including these goals within her proficiency-based and traditional grading formula, she has been able to better focus her students throughout the year.

Theme #3. The Proficiency-based Classroom Creates a Safer and More Collaborative Learning Environments

Throughout the data collected from the interviews, participants were repeatedly speaking about how collaborative their learning environments had become because of the massive increase in interpersonal communication between students and the realization, by students, that mistakes are part of the process of learning a language. These two prongs of the same theme emerged across all types of responses to the questions within the semi-structured protocol for all

participants. Mary succinctly observed that while this style of teaching is "more work," it (provides) a better environment for everyone.

Collaborative. Fred shared that he has 120 students this year and noted that although it is difficult to conference with them 1:1 every period, he does speak with all his students daily in the target language about something that is "tied to the learning target for that day." Fred shared that

> I think that my classes have become much more communal and just social, in general, in
>
> the target language. I've found a lot of success in the empowerment of students and just in
>
> the empowerment of their ability and being able to show them places they're improving
>
> and where they would otherwise generally feel like they don't have skill.

Jennifer shared that because of her class size, she is easily able to walk over to all of her students, and as they're working, chatting, doing whatever, check in with them individually every class. Jennifer went on to say that "There is no denying that kids feel more empowered and confident overall in a proficiency-based classroom, and definitely more so for the struggling learners. There are no failures."

Safe Learning Environment. Fred shared that his classroom is not based around arbitrary skills or vocabulary, but the goal that all of his students are comprehensible and can communicate with others in the target language. "It doesn't matter whether or not a student has perfect grammar or perfect pronunciation." Mary echoed that point by emphasizing that the goal is for students to create language. "(I) don't need to worry about the pace. (I) worry about whether students can actually understand what (I'm) trying to teach and whether they can use the language by themselves to express themselves."

Mary stated that she believes proficiency-based instruction and assessment is making the class environment better. She noted that

one way is (through) interpersonal speaking. It's talking about what you like, how is your weekend? How do you feel about this? So naturally, even though it's in a foreign language, it helps me to get to know my student better…I know their feelings, I know their emotions because we are focusing a lot on the interpersonal communication parts. The environment (allows) students and teacher guards down and then we actually communicate and build relationships.

Differentiation: Differentiation is a strategy used by teachers to personalize the learning experiences for students based on each student's individual needs (Marzano et al., 2017). As noted previously, every participant was asked how they differentiate for the needs of their students. For example, Fred shared that he uses the meaningful grouping strategy to help create teams of students with like or misaligned skills (depending upon the situation) in an effort to give more targeted feedback. John shared that he also uses meaningful grouping strategies to create different scaffold levels and pacing to get students to their desired proficiency level by the end of the year.

Nancy shared that she often tries to focus on a couple of domains from the rubric depending upon the needs of the student. For example, Nancy shared that

I might say, "For this one, I really just want you to focus on writing in full sentences and asking a couple questions." And I'm not as concerned about the spelling and the accuracy. It also helps if I'm distinguishing between accommodations and the other thing that's not accommodations. You're going to focus on these six can dos, that's modification, or I still want you to do all these nine can dos, but in order for you to meet intermediate low, you need this whole list of supports. Therefore, I'm going to give you a word bank. I'm going to give you sentence starters. You're going to have a card with your questions on it for the

interpersonal speaking. For the interpretive, I'm going to tell you which page these answers are found on, so you're not having to skim the whole text. I'm going to tell you what minute mark to look for so that you're not listening to the whole video to try to find it. So it's a little clearer, but it's a lot clearer to me what my job is when I'm differentiating.

Jennifer shared her students

do feel more comfortable participating because in a proficiency-based classroom, they know that they're not penalized heavily for making mistakes. However, because a proficiency-based classroom the ultimate goal is to get them comprehending and producing as much language as possible, with behavior challenged classrooms, (like hers) it's a lot of hard work.

Jennifer also stated she is able to differentiate for students by offering them alternate ways to provide evidence for the rubric. For example, she has allowed some students to describe pictures instead of using a lot of words to describe what is going on in a setting. Jennifer provided the example that in place of giving them a reading packet…they can draw pictures to reflect what they understand. Or they can record themselves and then talk to the microphone to show what they understand from the text. Victoria shared in her interview that she has made an effort to focus on the experience and sound of an English learner at each proficiency levels to provide them with a reference point in their own acquisition of language. This is something she and her students find to be extremely helpful.

Summary

The purpose of this qualitative transcendental phenomenological study was to explore the lived experience of Massachusetts public secondary (middle school/high school) world language

teachers who utilized proficiency-based rubrics to evaluate student performance. Eight participants participated in semi structured interviews that provided participants with the opportunity to share their experiences adopting the rubrics, how they were used and their impact on assessment and instruction. Doing so provided the researcher an opportunity to, as Bloomberg and Volpe (2019) described, utilize phenomenological research to listen to teachers and attempt "to develop patterns and relationships of meaning" (Bloomberg & Volpe, 2019, p. 54).

The data from the semi-structured interviews was coded after each interview and the themes were identified by looking at the responses of all eight of the participants. The emergent themes which evolved from this data were that performance feedback focuses on students' individual growth, that environment dictates the level of adherence to the ACFTL rubrics, and that the proficiency-based classroom creates a safer and more collaborative learning environment.

CHAPTER 5: CONCLUSION

Beginning with the publication of the ACTFL proficiency rubrics in 2012, world language teachers from across the United States have been adopting and implementing proficiency-based learning instructional and assessment practices, specifically the use of proficiency-based rubrics, into their classroom environments (American Council on the Teaching of Foreign Languages, 2012). In Massachusetts, the Department of Elementary and Secondary Education recently adopted new world language frameworks which now require teachers to implement proficiency-based learning environments (Massachusetts Department of Elementary and Secondary Education, 2021a). Proficiency-based assessment practices, like standards-based grading, were developed as an alternative approach to the traditional grading practices (100-poing A/F scale) that dominated secondary education environments in the 20th century and were known to have identified threats to their validity (Brookhart, 1993; Brookhart et al., 2016; Lipnevich et al., 2020; Vermont Agency of Education, 2018). Custable et al. (2019) argued that proficiency-based learning environments align well with the reality of learning a language. The American Council on the Teaching of Foreign Languages, which represents world language teachers from across the nation, has long been a proponent of adopting these practices, and the publication of their proficiency guidelines in 2012, helped to accelerate the adoption of these practices nationally (American Council of Teachers of Foreign Languages, 2012; Custable et al., 2019; Tigchelaar et al., 2017).

The purpose of this qualitative transcendental phenomenological study was to explore the lived experience of Massachusetts public secondary (middle school/high school) world language teachers who utilized proficiency-based rubrics to evaluate student performance. Proficiency-based rubrics are an assessment tool utilized by teachers who have adopted the proficiency-based

learning model of instruction (Custable et al., 2019). The most common proficiency-based assessment tools used by world language teachers are the ACTFL proficiency-based rubrics (Custable et al., 2019).

The researcher conducted eight one-on-one interviews that were transcribed, and the transcriptions were checked for accuracy by the participants. The data was then organized and coded to find emergent themes. The emergent themes described world language teachers' lived experiences while using proficiency-based rubrics to evaluate student performance and included (a) performance feedback focuses on students' individual growth, (b) the environment dictates the level of adherence to the ACTFL rubrics, and (c) proficiency-based classroom creates a safer and more collaborative environment.

Peoples (2021) posited that this chapter provides the researcher with an opportunity to "present creative solutions to problems that are based on their research findings" (p. 89). Additionally, this chapter provides an opportunity for the researcher to discuss the implications of this study and make recommendations for action and further study. To do so, the researcher will interpret the data within the context of the conceptual and theoretical frameworks and the literature review found in this study.

Interpretation and Importance of Findings

Denzin and Lincoln (2013) describe qualitative research as consisting "of a set of interpretive, material practices that make the world visible" (p. 6). Merriam and Tisdell (2016) posit that "qualitative researchers are interested in understanding how people interpret their experiences" (p. 5). Merriam and Tisdell (2016) further state that the "purpose of qualitative research are to achieve an understanding of how people make sense out of their lives" (p. 15).

This section outlines the interpretation and important findings of the research question that guided this study.

Interpretations for the Research Question

The research question for this study was "What are the lived experiences of secondary world language teachers who used Proficiency Based Rubrics to evaluate student performance?" The question was created to explore and understand the experiences of world language teachers in Massachusetts who have adopted the ACTFL proficiency guidelines and use the ACTFL proficiency-based interpretative, presentational, and interpersonal rubrics. Proficiency-based grading practices, also referred to as competency or mastery-based practices, are intended to develop a "deeper and more transformative approach to learning" where students are responsible for "co-constructing their learning with their teacher…and peers" (Custable et al., 2019, p. 370). Twadell et al. (2019) stated that within proficiency-based classroom environments, students collaborate with their teachers and peers to assess their own level of proficiency, establish goals, and achieve them through a process rooted in the student's self-reflection. This study utilized transcendental phenomenology and classroom assessment theory as a lens to assist the researcher in this question.

Inconsistent Implementation and Usage of the ACTFL Proficiency-Based Rubrics

Twadell et al. (2019) posited that the assessment of students is the central experience of a proficiency-based classroom environment. All eight participants shared that they had attended and participated in multiple proficiency-based assessment workshops, with five of the eight participants specifically mentioning that they attended the MaFLA Proficiency Academy. The MaFLA Proficiency Academy is an annual professional development opportunity for world language teachers put on by the Massachusetts Foreign Language Association (Massachusetts

Foreign Language Association, 2022). Catherine, Fred, Nancy, John, Madeline, Victoria, Jennifer, and Mary all teach in schools where student performance is ultimately reported using the 100-point A/F grading scale.

Victoria, Jennifer, and Mary all reported that while parents and guardians are exposed to how they assess students using proficiency-based rubrics annually at events like an open house, many of their parents still do not understand or value the rubrics and care more about the final letter grade. Of the eight participants, only Nancy teaches in a school that reports student progress using a standards-based report card of any kind. Nancy further noted that her school uses the ACTFL proficiency indicators as the actual elements of the standards-based report card shared with students and families.

The data suggested that the participants make a distinction between summative grades and summative proficiency ratings. In fact, all of the participants noted that student progress through the various levels of the proficiency scale (novice, intermediate, etc.) moves very slowly. Jennifer, Fred, John, and Catherine all noted in their interviews that there is an expected number of hours of practice required to progress from one level of proficiency to the next and, if a student uses only class time at the high school level, their progress could be very slow. To illustrate this point, Jennifer said that teachers in her school have now created subcategories of intermediate low into intermediate low-low and intermediate low-high just so students feel like they are making progress over the course of the school year.

Seven of the eight participants reported using traditional grading practices, 100-point A/F scale, in parallel with their tracking of proficiency. Traditional grading practices have been found to represent an amalgamation of values, biases, beliefs about learning, and the beliefs about students held by the teacher (Brookhart, 1993; Lipnevich et al., 2020). A key element of Andrade

and Brookhart's (2020) classroom assessment theory is that teachers include a wide variety of data sources, both formative and summative, when making summative determinations about what students know and are able to do.

In many ways, the dissonance of assessing the proficiency level of students in a foreign language through proficiency-based rubrics and reporting student performance through traditional grades should be expected. In fact, proponents of standards-based grading, like Vatterott (2015), posited that teachers could develop conversion charts that convert rubric scores into more traditional letter grades to help bridge adoption and make it more palatable. In fact, ACTFL promoted this type of adoption by launching its own conversion tool for teachers called a Grade-U-Later (Language Testing International, 2021).

John shared that he is a proponent of this approach. In his interview, John shared that his students have summative assessments per grading period which are each tied to one of the rubrics (interpretive, interpersonal, and presentational). These three assignments are the only grades in the quarter for his students, and his rationale for this is that he "wants their grade to reflect what they are able to do in Spanish…as accurately as we possibly can." John describes his approach to determining the 100-point A/F grade for students in this way

> The target for my students is set for the year, and I tell them it's their target, and that's
> what we are working to get to in June. And if you meet my target, you get a B. If you
> exceed it, you get an A. If you are partially meeting, it's a C. And if you're not quite there,
> it's a D. And so, it's an easy 95, 85, 75, 65, trying to marry standards-based grading into a
> traditional numeric system.

John reported that he treats class time like practice time. He wants them to take risks and engage with their peers and skill development without fear of negative grading repercussions, a

sentiment voiced by the other seven participants as well. Although he often provides formative assessments and activities to assist students in acquiring the target language and other non-cognitive tasks, John shared that it is possible that this work, if done well, can raise a student's overall grade from an A to an A +, or a B to a B +. This equates to an approximately 3% increase in a student's grade when a 95 is the highest possible score a student could have.

Victoria and Nancy both described how their grading formulas include three grade book categories that are each worth 30% and are aligned with the interpersonal, presentational, and intrapersonal rubrics. The final 10% of their grade is an amalgamation of homework and other non-cognitive classroom assignments. While their approach suggests that these two teachers would assess students in a similar manner, in reality, they have vastly different assessment strategies.

Nancy reported that student work on both formative and summative assessments is included in her grade book categories, while Victoria only uses summative assessment data. This allows Nancy's proficiency-based ratings of students to be based on multiple data points and opportunities for students to reperform a task. Because Victoria uses only summative assessment data in this category, her student's data is based on a single moment in time and the reperformance of a task over time does not factor into her determination of proficiency level. Importantly, reperformance of a skill or task is a key element of proficiency-based instruction as articulated by Twadell et al. (2019), and classroom assessment theory (Anandre & Brookhart, 2021). Finally, the most significant difference is that Victoria, ultimately, reports student progress and achievement using letter grades, while Nancy uses a standards-based report card with students and parents.

Similarly, Catherine noted that her summative ratings include the demonstration of sometimes non-cognitive tasks set by students that do not have to have any connection to the acquisition of a foreign language. Catherine explained that every quarter the students set their own goals,

> and, at the end of the quarter, they give me a reflection on how they've grown in those ways. Sometimes they have nothing to do with proficiency but sort of self-management skills that the kid learned. If those are going to help (them) in the future, whether it's in French class or not, I totally support that goal. (For example,) I had a kid who was like, "I don't analyze great. I need to do better on that. And so, my goal is going to be to read some people's analysis, understand what it takes to analyze (well) and then apply that to my own work." And he did it, it was amazing.
>
> (These assignments) aren't graded on a (proficiency) rubric that I've given them because it's so individual, but there is a rubric for that assignment (which focuses on) whether or not they encountered a challenge and how they handled that. If meeting expectations is that they saw the challenge, they named the challenge, and they sort of stopped, and exceeding expectations is they walked through that challenge.

Catherine, in her interview, highlighted how students had used this opportunity to focus on sleeping habits or nutritional practices as well.

These variations in the implementation of the proficiency-based rubrics are, it seems, why Coussens-Martin (2019) posits that secondary schools cannot simply adopt a standards-based curriculum and plug it into their existing traditional grading system. Guskey (2020) states that to do so introduces dissonance. Brookhart (2019) noted that while there had not been any research into the validity of grades constructed in this hybrid manner, the threats to validity that

standards-based and proficiency-based grading were intended to mitigate are present in some of the descriptions provided by participants in this study with their inclusion of non-cognitive assignments and the potential for teacher bias to factor into a student's summative grade.

Coussens-Martin (2019) noted that while standards-based assessment practices compare the performance of the student to established objective criteria, proficiency-based assessments compare the performance of the student to the expected outcomes (Custable et al., 2019). The data from this study suggest that the participants had a consistent understanding of how to implement proficiency-based instructional practices in their classrooms that foster a community where students are able to develop their language skills across multiple modes (interpretative, presentational, interpersonal) and experiment freely in the target language without fear of negative consequences. However, the data suggests that their assessment practices are more in line with standards-based grading because the participants compare their students to the descriptors in the proficiency-based rubrics or, as Mary, Jennifer, Fred, Nancy, and Madeline noted, the can-do statements and not the expected outcomes of the assignment or experience.

Only Some of the Expected Changes to Instruction Occurred as a Result of Adoption

Hoegh et al. (2019) stated that the usage of proficiency-based rubrics changes the instructional framework used by teachers. These changes in the literature include greater personalization and differentiated instruction, stronger one-on-one relationships between teacher and student and expanded conferencing with students, expanded use of small groups with like peers, and less direct instruction by the teacher (Hoegh et al., 2019; Marzano et al., 2019; Twadell et al., 2019). Seven of eight participants in this study reported that the adoption of the ACTFL proficiency-based rubrics for assessment resulted in significant changes occurring in their teaching and the classroom experience for students.

Only Nancy reported having a proficiency-based classroom environment prior to her adoption of the ACTFL rubrics, and she was the only participant to teach in a school where the proficiency-based indicators are included within the standards-based report card received by students and families; therefore, some of these practices were already in place. Six of the eight participants (Catherine, Fred, John, Madeline, Victoria, and Jennifer) experienced significant changes in their practices with the adoption of proficiency-based rubrics, and only one participant, Mary, began her teaching in a district where the adoption of the 2012 ACTFL proficiency guidelines was already underway.

Marzano et al. (2017) observed that while many teachers utilizing proficiency-based grading practices use the same instructional strategies as those using traditional grading practices, they do so differently. In particular, Marzano et al. (2017) argued that in a classroom where proficiency-based learning and instruction is occurring, there would be very little direct instruction as students will be working independently of each other or in small groups together. The data from this study suggests that this only partially occurred for the six participants who reported that significant change occurred in their classrooms.

Jennifer, Catherine, Fred, Victoria, Madeline, and John all expressed that the amount of peer-peer interactions between students increased significantly and that the amount of time spent providing teacher-led direct instruction has decreased in a corresponding fashion. However, each of these participants reported that their students and classes move together through the learning experiences. As a result, there are still periods of direct instruction and teacher-directed or facilitated transitions during every class and not the highly personalized and differentiated learning environments that Hoegh et al. (2019) suggested would occur. In fact, very few of the participants were able to describe how they differentiate their instruction beyond implementing

the accommodations on a student's IEP and/or providing choice in how a student will present their knowledge or skills. Further, the data suggest that students and small groups do not have the independence that Marzano et al. (2017) described.

When asked during the interview how often they conference with their students, a practice Marzano et al. (2017) and Hoegh et al. (2019) suggested would increase, only two of the participants, Jennifer and Fred, suggested that they conference with their students with any regularity. All of the participants said that they made a deliberate attempt to speak with students daily in the target language. Still, none suggested that their ability to conference with students increased, and three of the teachers cited class size as a barrier to doing so.

Implications

This section will discuss the implications of the data gathered in this study relative to the rationale and significance outlined in the first chapter. The implications of the data gathered for this study, which sought to understand the lived experience of teachers who utilize proficiency-based rubrics for instruction, may be significant for world language teachers and administrators in secondary schools who currently utilize proficiency-based rubrics to assess student learning and those in Massachusetts schools where adoption is imminent because of the recent adoption of the new curriculum frameworks for world language (Massachusetts Department of Elementary and Secondary Education, 2021a). Each participant in the study was able to describe how they adopted proficiency-based rubrics, how their usage of the rubrics impacted their planning, instructional environment, interactions with students about their learning, and their beliefs and practices relative to the assessment of students.

This study identified that among the participants interviewed, there was inconsistent implementation and usage of the ACTFL Proficiency-Based rubrics in secondary world language

classrooms. Specifically, seven of the eight participants in this study assessed students using both proficiency-based rubrics and reported their performance as both a summative rating and as a 100-point/A-F letter grade for a grading period because their school or district requires this type of grade to be reported to parents. Proficiency-based grading, like standards-based grading, was designed as an alternative to traditional grading practices where there had been widely acknowledged and demonstrated threats to the validity of final grades (Brookhart, 1993; Brookhart et al., 2016; Lipnevich et al., 2020). The approach taken by these participants and advocated by Vatterott (2015) to facilitate the adoption of the rubrics more broadly may actually facilitate the same types of bias and non-cognitive tasks in summative grades that proficiency-based assessments were designed to replace.

This study identified that among the participants interviewed, only some of the expected changes to instruction occurred through the adoption of proficiency-based assessment practices. Specifically, six of the eight participants observed that their students were communicating more with and between each other in one-on-one and in small groups. Because of this instructional shift, there was less direct instruction by these same teachers, in particular, on topics like grammar and cultural facts. However, the data did not suggest, as Marzano et al. (2017) had suggested it should, that the whole class direct instruction model would be unnecessary as the students worked independently through their tasks. In this study, all eight of the participants introduced and taught the new concepts of each unit to all of their students at one time.

Recommendations for Action

The purpose of this transcendental phenomenological study was to explore the lived experiences of secondary world-language teachers who used ACTFL proficiency-based rubrics for assessing student work. A review of the literature, data collected in one-one-one interviews

with eight participants, and the identification of emergent themes in the data provided the researcher with the opportunity to present the following recommendations for action. Importantly, the data gathered in this study came from eight participants who have worked in 14 different schools in Massachusetts since adopting proficiency-based grading practices.

The first recommendation is for school leaders, department heads, or school principals to assist world language teachers in the adoption of these practices by eliminating some of the barriers to successful adoption, specifically the preexisting grading policies which require teachers to create a parallel reporting system using traditional grading practices, to the ACTFL proficiency guidelines. The eight participants in this study had eight different sets of requirements for how they reported student work. Victoria stated that the grading formula was set by her department head, and every teacher in the department at the middle and high school was expected to use it; her system included keeping non-cognitive tasks like homework as a portion of her student's overalls grade because the proficiency-based ratings are not options for reporting student grades. Mary and Nancy shared that they must also be in alignment with their colleagues at the middle and high school level with how and when they use the ACTFL rubrics and set proficiency-based targets for students. All eight participants include non-cognitive tasks as part of their students' overall summative grades and to maintain student engagement. If schools were able to transition to a fully proficiency-based model instead of having to exist in both environments simultaneously, better student and staff outcomes would be realized.

The second recommendation is that world language teachers need to be cautioned not to alter their proficiency-based learning environments so drastically that the student experience in their classrooms is vastly different from that of their peers. John, Catherine, Fred, Madeline, and Jennifer all shared that they work in schools where their leadership provides them great leeway

in how they use the proficiency rubrics. While John has not substantially deviated from his approach since he adopted the proficiency rubrics, his approach to assessing students is more in line with standards-based grading because his students are continually compared to the expected behaviors and outputs of a student in that period of time. John is not assessing students relative to their proficiency in the task but instead to the desired long-term outcomes of the course. Catherine and Fred have made substantial changes to the rubrics they are assessing students with. Catherine has made a more "holistic" interpretative rubric that she uses in place of the ACTFL template and has developed a rubric to assess students on goals and tasks that may have little to nothing to do with the acquisition of a world language. Fred philosophically does not believe that there are "umbrella" rubrics that can be used universally. Therefore, Fred alters his rubrics as needed to make sure that they align with the task or assignment he is providing his students. Madeline and Fred both use a separate set of proficiency-based rubrics that align with the textbooks they have for class. Madeline and Mary noted that large summative assessments where the students are assessed with proficiency-based rubrics rarely occur because of the age of the students they teach. Teachers and school leaders should align their practices within their schools in districts in a way that builds the collective efficacy of their team, aligns with the inherent design of proficiency-based learning, and, most importantly, are good for their students.

Recommendations for Further Study

Bloomberg and Volpe (2019) noted that phenomenological research has many well-established limitations. A limitation of this study was the overall sample size of eight participants and the study design. As noted previously, the goal of phenomenological research is to understand the shared essence of the experience had by one or many individuals. As a result, there is not an ability to generalize the findings of this study for world language teachers.

Further exploration of the phenomena would benefit from a multiple-case study in secondary schools where the adoption and usage of proficiency-based rubrics and assessment practices have been uniformly adopted by the teachers. A similar multiple-case study could also be conducted in schools where proficiency-based rubrics have been adopted but teachers have the flexibility to choose where and how to utilize the proficiency-based rubrics and or the ability to customize the rubrics. A multiple-case study design could provide school leaders with more specific information as to how to prepare staff for adoption or sustain a culture of proficiency-based assessment in secondary world language classrooms through the use of focus groups, one-on-one interviews, and a review of documents.

Bloomberg and Volpe (2019) noted that the scope of a study includes the external conditions which act as boundaries for the researcher while the study is ongoing. While teachers in all subject areas can anchor their classroom experience in proficiency-based education, this study focused on the lived experience of eight world language teachers in the state of Massachusetts who were currently using proficiency-based rubrics in their classroom environment, which are based on the 2012 ACTFL Guidelines (American Council on the Teaching of Foreign Languages, 2012). States and communities across the country have adopted proficiency-based learning environments as their desired instructional model, with states like Massachusetts incorporating the ACTFL guidelines directly into their revised world language frameworks (Cover, 2018; Massachusetts Department of Elementary and Secondary Education, 2021; Stump et al., 2017; Vermont Agency of Education, 2018). The design and methodology of this study could serve as a model for researchers exploring the lived experience of additional secondary world language teachers in the state of Massachusetts or in other geographic regions of the country.

Finally, this study deliminated the data collection by limiting participants to a semi structured interview and not including a review of documents or re-interviews as part of the study design. This was an intentional choice by the researcher. However, if the researcher had the ability to reinterview participants after reviewing the transcripts of the semi structured interviews and after conducting a document review, in particular of the personalized rubrics developed by the study participants, the researcher may have been able to develop a richer understanding of the phenomena being studied.

Conclusion

For over a century, secondary teachers across the United States have adopted and utilized the 100-point/A-F grading scale as a means of reporting student achievement on assignments and in reporting end-of-quarter grades (Brookhart et al., 2016). Researchers have long known the shortcomings of the 100-point/A-F scale for grading student work and have attempted to address it with various attempts at reform (Brookhart et al., 2016; Carey & Carifio, 2012; Guskey & Brookhart, 2019; Starch & Elliott, 1912). One of those reform efforts was the introduction of proficiency-based learning. Advocates of proficiency-based learning argue that its emphasis on continuous improvement (learning) and not grading distinguishes it as an alternative to traditional grading practices (Twadell et al., 2019).

In 2012, the American Council on the Teaching of Foreign Languages (ACTFL) moved to integrate proficiency-based learning into the teaching of Foreign Languages across the United States by publishing their ACTFL Proficiency Guidelines Document (2012). Since the publication of these guidelines, world language teachers across the United States have been adopting and incorporating proficiency-based learning instructional practices and scales into their classrooms. The purpose of this qualitative transcendental phenomenological study was to

explore the lived experience of Massachusetts public secondary (middle school/high school) world language teachers who utilized proficiency-based rubrics to evaluate student performance. This study used Andrade and Brookhart's (2020) classroom assessment theory (CA) as the theoretical framework for this study and Husserl's (1964) transcendental phenomenology as a central component of the conceptual framework.

This literature review for this study explored the grading practices of teachers beginning in the 18th century through the present day. The review included the research and arguments that underpinned the need for grading reform and major reform movements in grading practices that include standards-based and proficiency-based grading. In exploring proficiency-based learning environments, the literature review focused on instructional and assessment strategies used by teachers, with particular attention paid to how world language teachers utilize the ACTFL proficiency-based rubrics.

The study utilized transcendental phenomenology as a methodology because of the study's focus on a shared phenomenon. As noted previously, this study examined how secondary world language teachers described their lived experiences using proficiency-based rubrics to evaluate student performance. Eight participants were recruited via email using the publicly available email address on school district websites where it was evident that world language teachers had already adopted proficiency-based assessments. Using a semi structured interview protocol, the participants were asked about their adoption of proficiency-based rubrics, as well as their usage in the classroom, how they assess students using the rubrics, and the impacts adoption and usage have had on their instruction.

The semi structured one-on-one interviews were then transcribed, deidentified, and then reviewed for accuracy by each participant. The interviews were then analyzed and coded using a

data procedure developed by Creswell and Creswell (2018). The emergent themes identified in the data were that performance feedback focuses on students' individual growth, that environment dictates the level of adherence to the ACFTL rubrics, and that the proficiency-based classroom creates a safer and more collaborative learning environment.

The findings of this study help to understand the lived experience of the participants by answering the research question. Specifically, the study found that there was an inconsistent implementation and usage of the ACTFL proficiency-based rubrics by the participants in the study. The inconsistencies were largely centered on whether the teachers had permission to modify the rubrics, how often and where the rubrics are used within a grading period, if and how rubric scores would be transferred into numerical or letter grades, and if other non-ACTFL rubrics could be used. The data suggest that the type of adoption (participant-led or school directed) experienced by the participant is an important factor. The study also found that only some of the expected changes to instruction occurred as a result of the adoption of proficiency-based rubrics for assessment. The data suggest that seven of the eight participants were able to only partially implement proficiency-based learning strategies (less direct instruction, increased one-on-one/small group instruction, self-directed learning outcomes, etc.) along with the proficiency-based rubrics and, therefore, were only able to realize some of the expected instructional benefits that come with full adoption.

References

Allal, L. (2020) Assessment and the co-regulation of learning in the classroom, *Assessment in Education: Principles, Policy & Practice*, 27(4), 332-349, https://www.doi.org/10.1080/0969594X.2019.1609411

American Council on the Teaching of Foreign Languages. (2012). *ACTFL proficiency guidelines 2012.* https://www.actfl.org/resources/actfl-proficiency-guidelines-2012.

American Council on the Teaching of Foreign Languages. (2022). *ACTFL performance descriptors.* https://www.actfl.org/resources/actfl-performance-descriptors-language-learners

American Council of Teachers of Foreign Languages. (2023). *Facilitate target language use.* Retrieved 3/29/23, from https://www.actfl.org/educator-resources/guiding-principles-for-language-learning/facilitate-target-language-use

Andrade, H. L., & Brookhart, S. M. (2020). Classroom assessment as the co-regulation of learning. *Assessment in Education: Principles, Policy & Practice*, 27:4, 350-372, https://www.doi.org/10.1080/0969594X.2019.1571992

Angelo, C., & Reis, A. B. (2021). Gender gaps in different grading systems. *Education Economics,* 29(1), 105-119, https://www.doi.org/10.1080/09645292.2020.1853681

Barnes, K. D., & Buring, S. M. (2012). The effect of various grading scales on student grade point averages. *American Journal of Pharmaceutical Education, 76*(3), 41. https://www.doi.org/10.5688/ajpe76341

Bloomberg, L. D., & Volpe, M. (2019). *Completing your qualitative dissertation* (Fourth edition ed.). SAGE.

Bonner, S. M., & Chen, P. P. (2021). Development and validation of the survey of unorthodox grading beliefs for teachers and teacher candidates. *Journal of Psychoeducational Assessment, 39*(6), 746-760. https://doi.org/10.1177/07342829211015462

Bonner, S. M., Torres Rivera, C., & Chen, P. P. (2018). Standards and assessment: Coherence from the teacher's perspective. *Educational Assessment, Evaluation and Accountability, 30*(1), 71-92. https://www.doi.org/10.1007/s11092-017-9272-2

Brookhart, S. M. (1993). Teachers' grading practices: Meaning and values. *Journal of Educational Measurement, 30*(2), 123-142. http://www.jstor.org/stable/1435460

Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., Stevens, M. T., & Welsh, M. E. (2016). A century of grading research: meaning and value in the most common educational measure. *Review of Educational Research, 86*(4), 803-848. https://www.doi.org/10.3102/0034654316672069

Bygren, M. (2020). Biased grades? Changes in grading after a blinding of examinations reform. *Assessment & Evaluation in Higher Education, 45*(2), 292-303. https://www.doi.org/10.1080/02602938.2019.1638885

Carey, T., & Carifio, J. (2012). The minimum grading controversy: Results of a quantitative study of seven years of grading data from an urban high school. *Educational Researcher, 41*(6), 201-208. https://www.doi.org/10.3102/0013189X12453309

Chen, P. P., & Bonner, S. M. (2017). Teachers' beliefs about grading practices and a

    constructivist approach to teaching. *Educational Assessment, 22*(1), 18-34.

    https://www.doi.org/10.1080/10627197.2016.1271703

Commission on Language Learning. (2017). *America's languages: Investing in language*

    education for the 21st century. American Academy of Arts and Sciences.

Coussens-Martin, S. R. (2019). *High school use of current-theory-based grading practices: A*

    multiple case study (Ed.D.). Available from ProQuest One Academic.

    (2239977140). https://une.idm.oclc.org/login?url=https://www.proquest.com/dissertations-

    theses/high-school-use-current-theory-based-grading/docview/2239977140/se-2

Cover, S. (2018, April 15,). Proficiency-based learning: State mandate stirs up parents, teachers,

    students . *Sun Journal* https://www.sunjournal.com/2018/04/14/proficiency-based-learning-

    state-mandate-stirs-up-parents-teachers-

    students/#:~:text=Many%20states%20created%20pilot%20programs%20or%20received%2

    0competitive,laws%20to%20require%20a%20switch%20to%20proficiency-

    based%20diplomas.

Creswell, J. W., & Creswell, J. D. (2018). *Research design* (Fifth edition ed.). SAGE.

Creswell, J. W., & Guetterman, T. C. (2019). *Educational Research* (6th ed.). Pearson.

Custable, W., Fisk, J., Grice, J., Jain, D., Lillydahl, D., Ramos, E., Reibel, A. R., Smith, B.,

    Twadell, E., & Wood, S. M. (2019). *Proficiency-Based Grading in the Content Areas:*

    Insights and Key Questions for Secondary Schools (Adapting Evidence-Based Grading for

    Content Area Teachers). Solution Tree

Denzin, N. K., & Lincoln, Y. S., (2013). *Collecting and interpreting qualitative materials* (4th ed.) SAGE

Ferlazzo, L. (2012, April 17,). Response: Several Kinds of Grading Systems. *Education Week.* https://www.edweek.org/teaching-learning/opinion-response-several-kinds-of-grading-systems/2012/04

Finn, K. E., Seymour, C. M., & Phillips, A. E. (2019). Weight bias and grading among middle and high school teachers. *British Journal of Educational Psychology, 90*(3), 635-647. https://www.doi.org/10.1111/bjep.12322

Fornaciari, C. J., & Lund Dean, K. (2013). I, s, t, and j grading techniques for es, ns, fs, and ps: Insights from the MBTI on managing the grading process. *Journal of Management Education, 37*(6), 828-853. https://www.doi.org/10.1177/1052562912461737

Griffin, R., & Townsley, M. (2021). Points, points, and more points: High school grade inflation and deflation when homework and employability scores are included. *Journal of School Administration Research and Development*, 6(1), 1-11–11. https://doi-org.une.idm.oclc.org/10.32674/JSARD.V6I1.3460

Guskey, T. (2020). *Go Set, Go*. Solution Tree Press.

Guskey, T. R. (1996). Reporting on student learning: Lessons from the past--prescriptions for the future. *Association for Supervision and Curriculum Development.Yearbook, ,* 13. https://une.idm.oclc.org/login?url=https://www-proquest-com.une.idm.oclc.org/trade-journals/reporting-on-student-learning-lessons-past/docview/195184264/se-2?accountid=12756

Guskey, T. R., & Brookhart, S. M. (ed.) (2019). *What We Know about Grading*. Association for

    Supervision & Curriculum Development.

Guskey, T. R., & Link, L. J. (2019). Exploring the factors teachers consider in determining

    students' grades. *Assessment in Education: Principles, Policy & Practice,* 26(3), 303-

    320. https://www.doi.org/10.1080/0969594X.2018.1555515

Hawe, E., Dixon, H., Murray, J., & Chandler, S. (2021). Using rubrics and exemplars to develop

    students' evaluative and productive knowledge and skill. *Journal of Further & Higher

    Education,* 45(8), 1033-1047. https://www.doi.org/10.1080/0309877X.2020.1851358.

Heidegger, M. (1962). *Being in time (J Macquarrie & E. Robinson, Trans)*. Oxford.

Hendry, G., and Anderson. J. (2013). "Helping students understand the standards of work

    expected in an essay: using exemplars in mathematics pre-service education classes."

    Assessment & Evaluation in Higher Education 38(6): 754–768.

    https://www.doi.org/10.1080/02602938.2012.703998

Hoegh, J. K., Heflebower, T., Warrick, P.B. (2019). *A handbook for developing and using

    proficiency scales in the classroom: (A clear, practical handbook for creating and utilizing

    high-quality proficiency scales)*. Marzano Research Lab.

Husserl, E. (1964). *The phenomenology of internal time-consciousness*. Indiana University Press.

Johnson, M. (2008). Grading in competence-based qualifications--is it desirable and how might

    it affect validity? *Journal of further and Higher Education, 32*(2), 175-184.

    https://une.idm.oclc.org/login?url=https://www.proquest.com/scholarly-journals/grading-

    competence-based-qualifications-is/docview/61951624/se-2?accountid=12756

Kenna, J. L., & Russell III, W. B. (2018). The Culture and history of standards-based educational reform and social studies in America. *Journal of Culture and Values in Education, 1*(1), 26-49. https://doi.org/10.46303/jcve.01.01.3

Knight, M. E. (2017). *A phenomenology examining high school teachers' perceptions of the effects of standards-based grading on planning, instruction, assessment, classroom environment, and students' characteristics and behaviors* (Order No. 10598892). Available from ProQuest One Academic. (1929217949). https://une.idm.oclc.org/login?url=https://www.proquest.com/dissertations-theses/phenomenology-examining-high-school-teachers/docview/1929217949/se-2

Language Testing International. (2021). *Scores.* https://www.languagetesting.com/aappl-scores#:~:text=The%20ACTFL%20Grade%2Du%2Dlator,to%20number%20and%20letter%20grades.&text=It%20allows%20you%20to%20customize,to%20best%20fit%20your%20needs.

Lee, B. D. (2020). *Taking advantage of second chances through reassessment* (Ed.D.). Available from ProQuest One Academic. (2437116283). https://une.idm.oclc.org/login?url=https://www-proquest-com.une.idm.oclc.org/dissertations-theses/taking-advantage-second-chances-through/docview/2437116283/se-2?accountid=12756

Lipnevich, A. A., Guskey, T. R., Murano, D. M., & Smith, J. K. (2020). What do grades mean? Variation in grading criteria in American college and university courses. *Assessment in Education: Principles, Policy & Practice, 27*(5), 480-500. https://www.doi.org/10.1080/0969594X.2020.1799190

Lleras-Muney, A. (2002). Were compulsory attendance and child labor laws effective? An

    analysis from 1915 to 1939. *The Journal of Law & Economics*, *45(2)*, 401–435.

    https://doi.org/10.1086/340393

Malouff, J. M., & Thorsteinsson, E. B. (2016). Bias in grading: A meta-analysis of experimental

    research findings. *Australian Journal of Education, 60*(3), 245-256.

    https://www.doi.org/10.1177/0004944116664618

Marzano, R. J., Norford, J. S., Finn, M., & Finn III, D. (2017). *A handbook for personalized*

    competency-based education. Marzano Research Lab.

Massachusetts Department of Elementary and Secondary Education. (2021a). *Massachusetts*

    curriculum framework for world languages

    2021. https://www.doe.mass.edu/frameworks/world-languages/2021.pdf

Massachusetts Department of Elementary and Secondary Education. (2021b). *2020-21 Teacher*

    by grade and subject report (DISTRICT) for all by full-time

    equivalents. https://profiles.doe.mass.edu/state_report/gradesubjectstaffing.aspx

Massachusetts Foreign Language Association. (2022). *MAFLA proficiency*

    academy, https://mafla.org/pdonline/proficiency-academy/

Merriam, S., & Tisdell, E. J. (2016). *Qualitative research* (fourth ed.). Jossey-Bass.

Merritt, R. D. (2021). *Grade point averages.* Salem Press Encyclopedia.

Moran, D. (2005). *Edmund Husserl: Founder of phenomenology*. Polity Press.

Moskal, B. M. (2002). Recommendations for developing classroom performance assessments

    and scoring rubrics. *Practical Assessment, Research & Evaluation, 8*, 14.

https://une.idm.oclc.org/login?url=https://www-proquest-com.une.idm.oclc.org/scholarly-journals/recommendations-developing-classroom-performance/docview/2366837818/se-2?accountid=12756

National Conference of State Legislatures. (2014). Overview of the common core state standards. https://www.ncsl.org/research/education/common-core-state-standards-overview.aspx. Retrieved 11/21/22

Olsen, B., & Buchanan, R. (2019). An investigation of teachers encouraged to reform grading practices in secondary schools. *American Educational Research Journal, 56*(5), 2004-2039. https://www.doi.org/10.3102/0002831219841349

Patten, M. L., & Newhart, M. (2018). *Understanding Research Methods* (10th ed.). Routledge. https://www.doi.org/10.4324/9781315213033

Patton, M. Q. (2015). *Qualitative research & evaluation methods* (4. ed.). Sage.

Peoples, K. (2021). *How to write a phenomenological dissertation: A step-by-step guide*. Sage Publications.

Percell, J. C. (2017). Lessons from alternative grading: Essential qualities of teacher feedback. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas, 90*(4), 111-115. https://www.doi.org/10.1080/00098655.2017.1304067

Quinn, D. M. (2020). Experimental evidence on teachers' racial bias in student evaluation: The role of grading scales. *Educational Evaluation and Policy Analysis, 42*(3), 375-392. https://www.doi.org/10.3102/0162373720932188

Randall, J., & Engelhard, G. (2010). Examining the grading practices of teachers. *Teaching and Teacher Education, 26*(7), 1372-1380. https://doi-org.une.idm.oclc.org/10.1016/j.tate.2010.03.008

Ravitch, S. M., & Carl, N. M. (2021). *Qualitative research* (Second Edition ed.). SAGE.

Roberts, C. M., & Hyatt, L. (2019). *The dissertation journey* (Kindle ed.). SAGE Publications.

Roulston, K. (2010). *Reflective interviewing: a guide to theory and practice*. Sage.

Starch, D., & Elliott, E. C. (1912). Reliability of the grading of high-school work in English. *School Review, 20*(7), 442-457. https://10.1086/435971

Stump, E., Fairman, J., Doykos, B., & Fink, P. (2017). Proficiency-based high school diploma systems in Maine: Implications for college and career access. *Center for Education Policy, Applied Research, and Evaluation*.

Tigchelaar, M., Bowles, R. P., Winke, P., & Gass, S. (2017). Assessing the validity of ACTFL can-do statements for spoken proficiency: A rasch analysis. *Foreign Language Annals, 50*(3), 584-600. https://www.doi.org/10.1111/flan.12286

Twadell, E., Onuscheck, M., Reibel, A. R., & Gobble, T. (2019). *Proficiency-based instruction: Rethinking lesson design and delivery (Your implementation strategy for proficiency-based instruction)*. Solution Tree.

Urhahne, D., & Wijnia, L. (2021). A review on the accuracy of teacher judgments. *Educational Research Review, 32*, 100374. https://doi-org.une.idm.oclc.org/10.1016/j.edurev.2020.100374

Vatterott, C. (2015). *Rethinking grading*. ASCD.

Vermont Agency of Education. (2018). *Research brief: Proficiency-based grading practices.* https://education.vermont.gov/sites/aoe/files/documents/edu-proficiency-based-learning-research-brief-proficiency-based-grading-practices.pdf

Warford, M. K. (2000). *Foreign language teacher educators and the diffusion of educational innovations* (Order No. 9985675). Available from ProQuest One Academic. (304618454). Retrieved from https://une.idm.oclc.org/login?url=https://www.proquest.com/dissertations-theses/foreign-language-teacher-educators-diffusion/docview/304618454/se-2

Weed, B. M. (2018). *Effective teacher-assigned grades: An analysis of three high school grading practices* (Ed.D). Available from ProQuest One Academic. (2047450348). https://une.idm.oclc.org/login?url=https://www-proquest-com.une.idm.oclc.org/docview/2047450348?accountid=12756

Welsh, M. (2019). Standards based grading. In Guskey, T. R., & Brookhart, S. M. (ed.) (2019). *What We Know about Grading.* (pp. 135-166) Association for Supervision & Curriculum Development.

Wetzler, E. L. (2019). How using a restricted grading range distorts GPAs and disproportionately penalizes low-performing students. *Frontiers in Education,* (29) https://doi.org/10.3389/feduc.2019.00023

Windschitl, M. (2002). Framing constructivism in practice as the negotiation of dilemmas: An analysis of the conceptual, pedagogical, cultural, and political challenges facing teachers. *Review of Educational Research, 72*(2), 131-175. http://www.jstor.org/stable/3516031

Appendix A: Recruitment Email

EMAIL INVITATION
Good Afternoon World Language Teachers,

My name is Jim Donovan, and I am a doctoral student at the University of New England in their Doctorate of Education program. I am also a high school principal in Massachusetts. My dissertation seeks to understand the lived experiences of secondary world language teachers who use Proficiency Based Rubrics to evaluate student performance. It is my hope that the information developed in this study may help teachers and administrators in secondary schools who currently utilize proficiency-based rubrics for evaluating student performance or those who may be adopting these practices. For teachers, the findings of the study could provide them with an understanding of how the adoption of proficiency-based rubrics in their classrooms may change how they evaluate their students. This study may also assist building principals, curriculum directors, and district administrators to help frame their internal discussions in planning for the adoption of proficiency-based learning environments and grading their students. This is particularly important as the underlying philosophies and practices of proficiency-based learning, specifically in grading, will likely be new for many teachers.

You are being asked to participate in this research project because you are a world language teacher in the state of Massachusetts who is 18 or older, and you are a current (full or part-time) teacher in public secondary schools in Massachusetts, including both middle and high schools, who utilize rubrics aligned with the 2012 ACTFL proficiency rubrics. Additionally, you possess, at minimum, your initial teacher license from the Massachusetts Department of Elementary and Secondary Education (DESE) and, at a minimum, a bachelor's degree. If you self-identify as meeting the criteria above, you are eligible to participate in this study. Your participation in this study is voluntary and will be limited to a one-hour recorded Zoom interview. Throughout the study, I will do my best to keep your personal information private and confidential, and all practices will be aligned with the requirements of the Office of Research Integrity at the University of New England.

If you are interested in participating in this study, please review the attached Participant Information Sheet and reply to this email. We will then schedule a time to meet at your earliest convenience. Thank you for your time and consideration for this research study.

Sincerely,

Jim Donovan

Appendix B: Participant Information Sheet

UNIVERSITY OF NEW ENGLAND

INNOVATION FOR A HEALTHIER PLANET

Participant Information Sheet

Version Date:	February 7, 2023
IRB Project #:	0123-23
Title of Project:	A Phenomenological Study of the Lived Experience of Secondary World Language Teachers Who Use Proficiency-Based Rubrics for Assessment.
Principal Investigator (PI):	James Donovan
PI Contact Information:	Jdonovan8@une.edu 781-405-2462

INTRODUCTION

- This is a project being conducted for research purposes. Your participation is completely voluntary.
- The intent of the Participant Information Sheet is to provide you with important details about this research project.
- You are encouraged to ask any questions about this research project now, during, or after the project is complete.
- The use of the word 'we' in the Information Sheet refers to the Principal Investigator and/or other research staff.

WHAT IS THE PURPOSE OF THIS PROJECT?

The general purpose of this research project is to explore the lived experiences of secondary world language teachers who use Proficiency Based Rubrics to evaluate student performance.

Eight participants will be invited to participate in this research as part of the principal investigator's dissertation research.

WHY ARE YOU BEING ASKED TO PARTICIPATE IN THIS PROJECT?

You are being asked to participate in this research project because you are a world language teacher in the state of Massachusetts who is 18 years of age or older, and you are a current (full or part-time) teacher in public secondary schools in Massachusetts, including both middle and high schools, who utilize rubrics aligned with the 2012 ACTFL proficiency rubrics. Additionally, you possess, at minimum, your initial teacher license from the Massachusetts Department of Elementary and Secondary Education (DESE) and, at a minimum, a bachelor's degree.

WHAT IS INVOLVED IN THIS PROJECT?

You will be asked to participate in one semi structured interview with the principal investigator that will last approximately one hour over Zoom.

· You will be given the opportunity to leave your camera on or off during the interview, and your interview will be audio recorded using Zoom.

· You will be emailed a copy of your interview transcript to review for accuracy. You will have five calendar days to respond, or the PI will assume that you have no comments, and the transcript will be assumed to be accurate.

WHAT ARE THE POSSIBLE RISKS OR DISCOMFORTS INVOLVED FROM BEING IN THIS PROJECT?

The risks involved with participation in this research project are minimal and may include an invasion of privacy or breach of confidentiality. This risk will be minimized by using pseudonym for each of the participants names and by eliminating any identifying information from the study. Participants will have the opportunity to review their transcripts for accuracy and will be given the choice to have their cameras off during the interview. Participants have the right to skip or not answer any questions, for any reason.

Please see the 'WHAT ABOUT PRIVACY & CONFIDENTIALITY?' section below for additional steps we will take to minimize an invasion of privacy or breach of confidentiality from occurring.

WHAT ARE THE POSSIBLE BENEFITS FROM BEING IN THIS PROJECT?

There are no likely benefits to being a participant in this research project. However, the information we collect may help teachers and administrators in secondary schools who currently utilize proficiency-based rubrics for evaluating student performance or those who may be adopting these practices.

WILL YOU BE COMPENSATED FOR BEING IN THIS PROJECT?

You will not be compensated for being in this research project.

WHAT ABOUT PRIVACY AND CONFIDENTIALITY?

We will do our best to keep your personal information private and confidential. However, we cannot guarantee absolute confidentiality. Your personal information may be disclosed if required by law. Additionally, your information in this research project could be reviewed by representatives of the University, such as the Office of Research Integrity and/or the Institutional Review Board.

The results of this research project may be shown at meetings or published in journals to inform other professionals. If any papers or talks are given about this research, your name will not be used. We may use data from this research project that has been permanently stripped of personal identifiers in future research without obtaining your consent.

- Data will only be collected during one on one participant interviews using Zoom, no information will be taken without participant consent, and transcribed interviews will be checked by participants for accuracy before they are added to the study.
- Pseudonyms will be used for all participants and any personally identifying information will be stripped from the interview transcript.
- All names and e-mails gathered during recruitment will be recorded and linked to a uniquely assigned pseudonym within a master list.
- The master list will be kept securely and separately from the study data and accessible only to the principal investigator.
- The interview will be conducted in a private setting to ensure others cannot hear your conversation.
- Participants are given the option to turn off their camera during Zoom interview.
- Participants will be able to review and approve the transcript of their interview before it can be used in the research study. Once approved, the recorded Zoom interview will be destroyed. Once all transcripts have been verified by the participants, the master list of personal information will be destroyed.

- All other study data will be retained on record for 3 years after the completion of the project and then destroyed. The study data may be accessed upon request by representatives of the University (e.g., faculty advisors, Office of Research Integrity, etc.) when necessary.
- All data collected will be stored on a password protected personal laptop computer accessible only by the principal investigator.

WHAT IF YOU WANT TO WITHDRAW FROM THIS PROJECT?

You have the right to choose not to participate, or to withdraw your participation at any time without penalty or loss of benefits. You will not be treated differently if you decide to stop taking part in this project. If you request to withdraw from this project, the data collected about you will be deleted when the master list is in existence, but the researcher may not be able to do so after the master list is destroyed.

WHAT IF YOU HAVE QUESTIONS ABOUT THIS PROJECT?

You have the right to ask, and have answered, any questions you may have about this research project. If you have questions about this project, complaints or concerns, you should contact the Principal Investigator listed on the first page of this document.

WHAT IF YOU HAVE QUESTIONS ABOUT YOUR RIGHTS AS A RESEARCH PARTICIPANT?

If you have questions or concerns about your rights as a research participant, or if you would like to obtain information or offer input, you may contact the Office of Research Integrity at (207) 602-2244 or via e-mail at irb@une.edu.

Appendix C: Interview Protocol

Interview Protocol Matrix for A Phenomenological Study of the Lived Experience of Secondary World Language Teachers who use Proficiency-Based Rubrics for Assessment.

Script prior to the interview:

Thank you again for being willing to participate in the interview aspect of my study. As I have mentioned to you before, the purpose of this study is to explore the lived experience of Massachusetts public secondary (middle school /high school) world language teachers who utilize proficiency-based rubrics to evaluate student performance. Specifically, this study seeks to understand the lived experiences of secondary world language teachers who use Proficiency Based Rubrics to evaluate student performance. Our interview today will last approximately one hour, and I will be asking you about your experience evaluating student performance through the usage of proficiency-based rubrics.

I will do my best to keep your personal information private and confidential. However, I cannot guarantee absolute confidentiality. Your personal information may be disclosed if required by law. Additionally, your information in this research project could be reviewed by representatives of the University, such as the Office of Research Integrity and/or the Institutional Review Board. Any personally identifiable information (e.g., name, e-mail, physical address, etc.) obtained for recruitment purposes will be secured in a password-protected Google Drive account, accessed via a password-protected computer, that only the researcher will have access to. This information will be destroyed once the recruitment phase has concluded for this study. The Zoom recording from

today's interview will be downloaded from Zoom and secured in a password-protected Google

Drive account, accessed via a password-protected computer, that only the researcher will have

access to.

I will then prepare a transcript for review by you, which will use pseudonyms and remove all identifying information. The list of pseudonyms that I create will be kept in a master list secured in a password-protected Google Drive account, accessed via a password-protected computer, that only I will have access to. The transcript will then be shared with you, and I ask that you please review and return it within five calendar days. If, after five calendar days, you have not returned it, the transcript will be considered accepted, and the master list and interview recordings will be destroyed. From that point forward, I will only use the newly deidentified transcripts when conducting the data analysis of this study. Throughout the study, all of the data and documents will be secured in a password-protected Google Drive, on a password-protected computer that only the researcher will have access to for the duration of the study. Finally, three years from the publication of this study, all of the study data secured in the password-protected Google Drive will be destroyed.

With that all said:
Would you like to participate in the study?
And finally, do you consent to record this interview under the terms stated above?

If yes: *Please know that you may decline to answer any question you wish and that you can request that the recording be stopped at any time.*

Before we begin the interview, do you have any questions? [Discuss questions]
If any questions (or other questions) arise at any point in this interview, you can feel free to ask them at any time. I would be more than happy to answer your questions.

Demographic Information

Background Information: To begin, I am going to ask you some questions about your professional history utilizing proficiency-based rubrics as a World Language Teacher

- First, for how long have you been a secondary world language teacher?
- When did you begin to utilize proficiency-based rubrics to evaluate student performance in world language?
- How was the decision made to use proficiency-based rubrics in your classroom?
- Can you tell me about the adoption process you went through to utilize proficiency-based rubrics in your classroom?

Usage

Thank you. This next set of questions focuses on your usage of proficiency-based rubrics in your classroom with students.

- Can you describe which proficiency-based rubrics you use (interpretative, interpersonal, presentational) and how you use them with students in your classroom?
- In your teaching, can you describe any differences in how you assess formative tasks as compared to summative tasks with proficiency-based rubrics?
- What role do the proficiency-based rubrics have when you communicate with parents/guardians about their student's overall growth?
- In your teaching, how have you encouraged students to use the rubrics to monitor and self-assess their own performance?
- In your teaching, how have your expectations for student behaviors, and classroom dynamics evolved since using proficiency-based rubrics?

Assessing Students

Thank you. This next set of questions focuses on how you assess students using proficiency-based rubrics.

- What does "evaluating students" mean to you, and how do proficiency-based rubrics help you to do so?
- With what types of assignments or tasks do you use proficiency-based rubrics, and how do students use the feedback you provide to them with the rubric?
- In your classroom, how do you use the ratings on the rubric to track student growth?
  - Are there any other ways you track student growth during a grading period, and how does this method align with your usage of the proficiency-based rubrics?
- When developing overall proficiency ratings for your students, do you expect that students would have reperformed a task or skill multiple times before earning a rating, for example, of proficient? What does this look like?
- How have you observed your students utilizing the learning targets or objectives within the rubric while they complete their tasks or assignments?
- How do you utilize the rubrics in the homework assignments you assign your students?

- How do you incorporate self-reflections generated by students using the proficiency rubrics about their work when preparing your summative ratings? What does this process look like?

Instruction

Thank you, this final set of questions focuses on proficiency-based rubrics and your instruction.

- In what ways do you feel the usage of proficiency-based rubrics informed your professional practice as a teacher?
- Since adoption, how has your usage of the whole-class instruction strategy evolved? When and where do you use it?
- In your classroom, how often are you able to conference 1:1 with your students? How has this changed since adopting proficiency-based learning?
- What are some of the ways you differentiate your instruction for students using the proficiency-based rubric?
- Can you describe for me the elements of a lesson built around the proficiency-based rubric, and in what ways have your planning routines changed since the adoption of the rubrics?
- How would you describe your conversations with students about their learning since your adoption of proficiency-based rubrics?

Conclusion:
I want to thank you for your time today. Before we end, I wanted to ask if you have any questions for me or any additional thoughts you would like to add to any of the topics we discussed today.

I would like to share with you my next steps briefly. When we finish today, the recording of this interview will be downloaded from Zoom and secured in a password-protected Google Drive account, accessed via a password-protected computer, that only I will have access to. The transcripts will then be prepared using pseudonyms, and I will remove all identifying information. The list of pseudonyms and all identifying information will be kept in a master list secured in a password-protected Google Drive account, accessed via a password-protected computer, that only I will have access to. I will then email you a copy of your transcript to review for accuracy. You will then have five calendar days to review the transcript and email me with your approval or provide clarifications. If, after five calendar days, I have not heard back from you, the transcript will be considered accepted, and the master list and interview recordings will be destroyed.

Thank you very much for your time today, I am most appreciative.

Appendix D: IRB Approval Letter

UNE UNIVERSITY OF NEW ENGLAND

INNOVATION FOR A HEALTHIER PLANET

Office of Research Integrity
Institutional Review Board

Biddeford Campus
11 Hills Beach Road
Biddeford, ME 04005
(207) 602-2244 T
(207) 602-5905 F

Portland Campus
716 Stevens Avenue
Portland, ME 04103

DATE OF LETTER: February 7, 2023

PRINCIPAL INVESTIGATOR: James Donovan
FACULTY ADVISOR: Andrea Disque, EdD

PROJECT NUMBER: 0123-23
RECORD NUMBER: 0123-23-01

PROJECT TITLE: A Phenomenological Study of the Lived Experiences of Secondary World Language Teachers Who Use Proficiency-Based Rubrics for Assessment

SUBMISSION TYPE: New Project
SUBMISSION DATE: 1/30/2023

ACTION: Determination of Exempt Status
DECISION DATE: 2/7/2023

REVIEW CATEGORY: Exemption Category # 2(ii)

The Office of Research Integrity has reviewed the materials submitted in connection with the above-referenced project and has determined that the proposed work is exempt from IRB review and oversight as defined by 45 CFR 46.104.

You are responsible for conducting this project in accordance with the approved study documents, and all applicable UNE policies and procedures.

If any changes to the design of the study are contemplated (e.g., revision to the research proposal summary, data collection instruments, interview/survey questions, recruitment materials, participant information sheet, and/or other approved study documents), the Principal Investigator must submit an amendment for review to ensure the requested change(s) will not alter the exempt status of the project.

If you have any questions, please send an e-mail to irb@une.edu and reference the project number as specified above within the correspondence.

Best Regards,

Bob Kennedy, MS
Director of Research Integrity