

APPLICATION OF THE REVISED USAGE RATING PROFILE-INTERVENTION TO
EXAMINE TEACHER ACCEPTABILITY OF INSTRUCTIONAL SUPPORTS FOR
ENGLISH LEARNERS

By

Sarina Roschmann

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

School Psychology – Doctor of Philosophy

2022

ABSTRACT

APPLICATION OF THE REVISED USAGE RATING PROFILE-INTERVENTION TO EXAMINE TEACHER ACCEPTABILITY OF INSTRUCTIONAL SUPPORTS FOR ENGLISH LEARNERS

By

Sarina Roschmann

English learners (ELs) are one of the fastest growing groups in U.S. public schools, making up approximately ten percent of the student population (National Center for Education Statistics [NCES], 2020, May). Several empirically-supported instructional supports exist to address the unique needs of this group (e.g., the supports included in the Sheltered Instruction Observation Protocol (SIOP; Echevarría et al., 2014); however, the extent to which these are used by classroom teachers is unclear. Social validity of these supports may play a critical role in the degree to which they are used. The purpose of the current study was to investigate the measurement qualities of an existing social validity measure (Usage Rating Profile-Intervention Revised [URP-IR]; Briesch et al., 2013) when applied to four instructional supports for ELs. In addition, the current study investigated potential predictors of URP-IR ratings, as well as differences in acceptability between four instructional supports. Lastly, the current study explored the social validity of EL supports beyond acceptability through qualitative teacher interviews. Results indicated that the existing factor structure of the URP-IR was not a strong fit for teacher ratings of four instructional supports for ELs. Further, only one of the predictors explored (teacher training) significantly predicted teacher URP-IR ratings, for two of the four supports (visual aids and alternate response format). Findings indicated that URP-IR ratings differed significantly between all four instructional supports. Lastly, qualitative findings provided a more complete understanding of social validity of instructional supports for ELs by

describing teachers' goals for ELs as well as perceptions of effectiveness of the four supports.

Implications for future research and practice are discussed.

Copyright by
SARINA ROSCHMANN
2022

To Craig, you are my world and inspire me every day.
To my parents and brothers, who taught me that the world is full of possibilities.

ACKNOWLEDGMENTS

I want to first thank, with all of my heart, my partner Craig and my family. Your unconditional encouragement and support means the world to me and keeps me moving forward. I would also like to express my gratitude to my advisor, Dr. Sara Witmer, for her unwavering support and guidance over the years. I would not be in this place today without her, and I truly appreciate the energy she invested in my growth and the patience she demonstrates as an advisor. Thank you as well to the rest of my dissertation committee members: Dr. Courtenay Barrett, Dr. Carrie Symons, and Dr. Martin Volker. Your expertise has been instrumental in moving this project forward. Lastly, I would like to thank my fellow research team member, Nathalie Marinho, for her help with qualitative data analysis, and Barry DeCicco at the MSU Center for Statistical Training and Consulting for his guidance with quantitative data analysis.

TABLE OF CONTENTS

LIST OF TABLES	x
LIST OF FIGURES	xii
CHAPTER I: INTRODUCTION.....	1
Background	1
Importance	1
Rationale	3
Purpose of the Current Study	5
CHAPTER II: LITERATURE REVIEW	8
Framework: Social Validity	8
Social Validity Measurement Approaches and Tools.....	14
Treatment Evaluation Inventory and Associated Revisions	15
Intervention Rating Profile and Associated Revisions	16
Behavior Intervention Rating Scale	17
Abbreviated Acceptability Rating Profile.....	18
Usage Rating Profile-Intervention	18
Factor 1: Acceptability.....	21
Factor 2: Understanding.....	22
Factor 3: Family-School Collaboration	22
Factor 4: Feasibility	22
Factor 5: System Climate.....	23
Factor 6: System Support.....	23
Predictors of Teacher Ratings of EL Supports	24
Teacher Training	26
Experience with ELs	27
Consultation with the ESL Teacher	28
Empirically-Supported Instructional Supports for ELs	30
The SIOP Model	30
Visual Aids.....	32
Vocabulary Supports.....	34
Incorporation of L1	36
Alternate Response Format.....	37
Summary	38
CHAPTER III: METHODS	41
Research Design.....	41
Participants.....	41
Measures	44
Acceptability	44
Social Validity Interview	45
Demographics and Participant Information	46

Experience Working with ELs.....	46
Prior Training.....	47
Role of the ESL Teacher.....	47
Procedures.....	48
Survey Pilot Study	48
Main Study.....	50
Recruitment.....	50
Survey Administration	51
Social Validity Interviews.....	52
Data Analysis	52
Research Question 1	52
Research Questions 2 and 3	53
Reverse Coding.....	54
Research Question 4	54
Sample Size and Power.....	54
Research Question 1	54
Research Question 2	55
Research Question 3	55
CHAPTER IV: RESULTS.....	57
Research Question 1	57
Data Description	57
Model Specification.....	59
Model Fit.....	61
Support 1	61
Support 2.....	63
Support 3.....	65
Support 4.....	68
Model Modifications and Fit	70
Support 1	70
Support 2.....	73
Support 3.....	76
Support 4.....	79
Single-Factor Model	82
Research Question 2	83
Experience.....	83
Training.....	84
Consultation	85
Research Question 3	86
Research Question 4	87
Range in Goals.....	88
Prioritized Goals	89
Barriers for ELs.....	90
Adverse Effects of Barriers on ELs	92
Positive Effects of Supports.....	93
Effects of Individual Supports	95

Contingencies and Shortcomings.....	97
Summary	98
CHAPTER V: DISCUSSION.....	99
Poor Model Fit	100
Potential Model Fit Improvements	103
Predictors of URP-IR Score.....	104
Teacher Training.....	104
Experience Working with ELs.....	105
Consultative Support.....	107
Differences in URP-IR Total Score	108
Implications of Qualitative Findings	110
Social Significance of the Goals	111
Social Importance of the Effects.....	112
Limitations	113
Implications for Future Research.....	115
Measurement Work.....	115
Predictors of Acceptability	116
Implications of the URP-IR Score	117
Implications for Practice	118
Summary and Conclusion	120
APPENDICES	121
Appendix A: Instructional Support Vignettes.....	122
Appendix B: URP-IR Items	126
Appendix C: Qualitative Interview Questions.....	128
Appendix D: Survey prenotification (deidentified).....	130
Appendix E: Terms and Definitions	131
REFERENCES	132

LIST OF TABLES

Table 1. Overview of the SIOP Model	30
Table 2. Demographics	42
Table 3. Overview of Research Questions.....	56
Table 4. Item-level Descriptives	57
Table 5. Fit Indices for Support 1	61
Table 6. Estimates of Regression Weights, Intercepts, Variance, and Covariances for Support 1	61
Table 7. Internal Reliability and Inter-Item Correlations by Factor for Support 1	62
Table 8. Fit Indices for Support 2	63
Table 9. Estimates of Regression Weights, Intercepts, Variance, and Covariances for Support 2	63
Table 10. Internal Reliability and Inter-Item Correlations by Factor for Support 2.....	65
Table 11. Fit Indices for Support 3	66
Table 12. Estimates of Regression Weights, Intercepts, Variance, and Covariances for Support 3	66
Table 13. Internal Reliability and Inter-Item Correlations by Factor for Support 3.....	67
Table 14. Fit Indices for Support 4	68
Table 15. Estimates of Regression Weights, Intercepts, Variance, and Covariances for Support 4	68
Table 16. Internal Reliability and Inter-Item Correlations by Factor for Support 4.....	69
Table 17. Fit Indices after Modifications for Support 1	72
Table 18. Estimates of Regression Weights, Intercepts, Variance, and Covariances for Support 1 after Modifications.....	72
Table 19. Fit Indices after Modifications for Support 2	75

Table 20. Estimates of Regression Weights, Intercepts, Variance, and Covariances for Support 2 after Modifications.....	75
Table 21. Fit Indices after Modifications for Support 3	78
Table 22. Estimates of Regression Weights, Intercepts, Variance, and Covariances for Support 3 after Modifications.....	78
Table 23. Fit Indices after Modifications for Support 4	81
Table 24. Estimates of Regression Weights, Intercepts, Variance, and Covariances for Support 4 after Modifications.....	81
Table 25. URP-IR Total Scores by Predictor Group	83
Table 26. Mean URP-IR Scores per Support and Pairwise Comparisons	87

LIST OF FIGURES

Figure 1. Path diagram of the six-factor model identified by Briesch et al. (2013)	60
Figure 2. Path diagram of the six-factor model with modifications for support 1.....	71
Figure 3. Path diagram of the six-factor model with modifications for support 2.....	74
Figure 4. Path diagram of the six-factor model with modifications for support 3.....	77
Figure 5. Path diagram of the six-factor model with modifications for support 4.....	80

CHAPTER I: INTRODUCTION

Background

Within the United States, ELs comprise a significant portion of public-school students at approximately ten percent (NCES, 2020). Across academic areas, ELs have consistently been found to have lower achievement than their non-EL peers (Genesee et al., 2005). Although such achievement differences may be expected in academic areas that heavily include English, it might be less suspected in areas whose core content does not involve English, such as math or science. However, scores from the National Assessment of Educational Progress (NAEP) reported by the NCES indicate otherwise. On the 2009 NAEP science assessment, fourth grade ELs were the lowest scoring group among all other student characteristic subgroups (i.e., sex, race/ethnicity, disability status, parental education level, free or reduced lunch, and urbanicity), placing them, on average, between the tenth and 25th percentile (NCES, 2017). Compared to non-EL students, fourth grade ELs scored, on average, over a full standard deviation below non-ELs. Such trends were found again on the 2015 NAEP assessment, as well as in middle and high school grades (NCES, 2017).

Importance

Given such differences in academic outcomes, particularly in non-English centric subjects such as science, it is critical to understand how ELs can best be supported in public school settings. One setting in which critical support to ELs can be provided is the general education classroom, as ELs tend to spend most of their time in these settings (Polat, 2010; Villegas et al., 2018). Specific language barriers are often present for ELs in these settings when non-English centric content is presented in English. Mainstream teachers are thus instrumental in ensuring that ELs have access to the instructional content presented. Elementary teachers may be

in a particularly good position to readily do so given that they often work with the same students across the entire day and therefore may become familiar with an individual student's needs for accommodation to address language barriers. However, it has been reported that teachers feel largely unprepared to meet the needs of ELs (Polat, 2010).

To ensure that instruction is accessible to ELs, it is critical to understand the specific language-related barriers ELs may face that interfere with their access to instruction. In general, any content presented in English, whether verbal or written, will be more cognitively taxing for ELs than non-ELs to access. However, certain features of this content may influence the extent to which it is additionally challenging for ELs. One of those features is linguistic complexity. For example, Lee et al. (2013) and Oliveira et al. (2014) report that instructional texts or other forms of presentation are often linguistically complex, both in vocabulary and grammar. This barrier may be addressed by reducing unnecessarily linguistically complex texts, as has been found in studies investigating access to test content (e.g., Martiniello, 2009). Another factor that can add to language-related barriers is the level of abstraction required in instruction (Lee et al., 2013). For example, abstracting everyday language into a scientific context (e.g., using the word "table" as a means of depicting information rather than the common household object) was highlighted as a common challenge for ELs by Lee et al. (2013) when evaluating the Next Generation Science Standards. Lastly, another language-related barrier is the extent to which language, written or spoken, is culturally loaded (Lee et al., 2013; Oliveira et al., 2014). For example, assuming familiarity with specific holidays or cultural practices may make instruction less accessible to ELs than their non-EL peers. Overall, given these barriers for ELs, mainstream teachers are being increasingly called on to evaluate and reduce the language demands of their instructional activities (Cummins, 2000). This may be particularly critical starting during mid- to

upper-elementary grades, when students are increasingly expected to engage in language-based tasks, such as deriving content from readings and demonstrating understanding through writing.

A considerable amount of research has focused on identifying and studying ways to adapt instruction to reduce language demands and meet the needs of ELs. A promising instructional model that incorporates a variety of instructional adaptations is the SIOP model (Echevarría et al., 2014). Use of this model has demonstrated effectiveness in increasing ELs' academic performance in various empirical studies (Echevarría et al., 2006; Echevarría et al., 2011; Short et al., 2011). Instructional adaptations that may be particularly effective in reducing language-related barriers for ELs during science instruction are visual aids, vocabulary supports, incorporation of the student's native language, and alternative response formats. However, given the persistent discrepancy in academic scores between ELs and non-ELs over time (NCES, 2017), it appears that more attention may be helpful on how to promote more widespread teacher use of empirically-supported instructional supports during mainstream content instruction (e.g., science), such as those provided in the SIOP model.

Rationale

When considering why research-based strategies may not be readily incorporated into school-based practice, one important area of focus is teachers' opinions of these strategies, given that they are ultimately the individuals implementing the strategies in practice. One framework that has been applied to help identify reasons why various efficacious intervention are not regularly implemented in practice is social validity. According to this framework, the opinions of providers (e.g., teachers) of interventions can provide valuable information to researchers as they seek to modify interventions for more widespread implementation. The framework includes three levels: social significance of the goals of treatment, social appropriateness of the treatment

procedures, and social importance of the effects of treatment (Wolf, 1978). Assessment of social validity contributes to intervention evaluation in two key ways. First, it provides researchers with an understanding of how effective interventions are perceived by those intended to implement them (i.e., consumers). Thus, researchers are better able to adapt interventions, if necessary, to maintain their effectiveness while also making them more socially valid to consumers. Second, assessment of social validity also supports those in consulting roles, such as school administrators, trainers, or school psychologists. Understanding of teachers' opinions of interventions can help consultants select empirically-supported interventions that have a higher likelihood of being implemented by those with whom they are consulting. Assessment of social validity is recognized as a critical factor in evaluation of interventions by several professional organizations (American Psychological Association [APA], 2002; National Association of School Psychologists [NASP], 2010).

Much empirical work has focused on developing measures of social validity, although primarily in the context of behavioral interventions. In their study, Briesch et al. (2013) developed a measure that involved the identification of six factors that may contribute to a teacher's actual use of an intervention or instructional strategy. These factors are Acceptability, Understanding, Family-School Collaboration, Feasibility, System Climate, and System Support. Briesch et al. (2013) developed this measure based on teachers' opinions of classroom-wide behavior management interventions; it appears to be a potentially promising measure to consider using in the context of teachers' opinions of instructional supports for ELs. However, further research is deemed necessary to better understand if, when applied to this new context, the same factor structure emerges (Briesch et al., 2013).

Investigation of the applicability of an existing measure to the context of instructional supports for ELs would have several benefits. First, assessment of teacher opinions of instructional supports for ELs would allow both researchers and consultants to understand whether some empirically-based supports are more acceptable to teachers than others. This would allow researchers or consultants to identify how to modify supports with lower ratings to increase acceptability, for example by increasing training on how to use the support or by decreasing the time required to implement the support. Second, researchers could investigate teacher-related characteristics and experiences and their relationship to acceptability of instructional supports for ELs. Of specific interest might be experiences that can be fostered, such as effective consultative relationships or training in effective instruction for ELs.

Purpose of the Current Study

The current study applied social validity as a framework to investigate teacher opinions of instructional supports for ELs. Given general education teachers' important roles in reducing language-related barriers for ELs and furthering their academic growth, it is critical to understand their opinions of empirically-based instructional supports that may best support ELs during general education instruction, as they are ultimately those responsible for implementing them. Assessing teacher opinions may encompass a variety of factors such as their own personal views of effectiveness of the instructional support, the extent to which they feel they understand how to use it, and the feasibility of providing the support, among others (Briesch et al., 2013). Knowledge of such factors could strengthen researchers' and consultants' ability to support teachers in selecting and implementing empirically-supported instructional supports for ELs. However, to develop such an understanding, there is a need for a validated measure, such as the URP-IR, that can reliably and accurately assess teacher opinions of instructional supports for

ELs. The current study therefore investigated the application of the URP-IR to the context of teacher opinions of instructional supports for ELs.

The current study surveyed elementary teachers' opinions of empirically-supported instructional supports for ELs during science instruction. Science was chosen due to the large differences in achievement between ELs and non-ELs in this content area and due to the potential of language-related barriers limiting ELs' access to instructional content. The study used an existing measure developed based on a social validity framework (i.e., URP-IR; Briesch et al., 2013) and investigated whether this measure can be effectively applied to measure teachers' opinions about instructional supports for ELs. To that end, the measure's psychometric properties and factor structure were explored. Additionally, the study investigated whether the URP-IR shows variability in teacher ratings of four different instructional supports for ELs during science instruction and whether URP-IR scores differed significantly based on selected teacher background experiences. Finally, beyond the teacher acceptability measured by the URP-IR, the study also examined the two other key aspects of social validity, the social significance of the goals and social importance of the effects, through qualitative follow-up. Thus, the current study addressed the following research questions:

1. Is the six-factor structure of the revised Usage Rating Profile-Intervention (URP-IR) a good model fit when using this measure for teacher ratings of empirically-supported instructional supports for English Learners during science instruction?
2. To what extent do the following teacher background experiences predict URP-IR ratings for the targeted EL supports?
 - a. Experience working with ELs
 - b. Teacher training on teaching ELs

- c. The role of the ESL teacher
3. What are elementary mainstream teachers' URP-IR ratings for four empirically-supported instructional supports for English Learners during science instruction? Are there significant differences in URP-IR scores across these instructional supports?
 4. What are teachers' opinions about the social significance of the goals and social importance of the effects of the selected EL instructional supports?

CHAPTER II: LITERATURE REVIEW

The literature relevant to the development of the current study will be reviewed in the following sections. First, social validity will be described as a broad framework for the study. Second, the evolution of social validity measurement tools will be discussed, and existing tools critiqued to highlight the need for further psychometric work in this area. The specific measurement tool selected for this study (i.e., the URP-IR), which examines teachers' opinions about supports for ELs, will be described in depth. Next, teacher characteristics that were anticipated to function as predictors of the associated opinions will be discussed. Lastly, research available on four empirically-supported instructional supports for ELs will be described and discussed in terms of how teacher acceptability ratings are anticipated to vary across supports. Altogether, this literature review will offer context as well as a rationale for the research questions posed as part of the current study.

Framework: Social Validity

Social validity has long been a consideration in empirical research when investigating interventions and the opinions of those tasked with implementing them. Social validity has expanded and evolved over time and has been recommended as part of intervention evaluation by several major professional organizations (APA, 2002; NASP, 2010). The framework originated in 1978, when Wolf argued for the utility of subjective measurement in the field of Applied Behavior Analysis (ABA). He called this subjective measurement "social validity," a construct that would describe the "social importance" of a specific intervention. Identifying ways to measure social validity, he argued, was critical in identifying research that was not only promising in terms of observable change in behavior, but also important and valid to those affected by that research. For example, Wolf reported of a stuttering intervention that did indeed

reduce stuttering significantly but resulted in monotone speech. He argued that while effective, it was not, in its current state, socially valid. He highlighted the potential negative outcomes, such as avoidance, that could come from situations in which consumers and implementers of objectively effective interventions do not find it socially valid.

Wolf, having identified the need for including social validity as a consideration in developing effective behavioral interventions, proposed a social validity framework to guide measurement of this construct. It included three levels on which society should validate behavioral interventions: social significance of the goals, social appropriateness of the procedures, and social importance of the effects. Social significance of the goals refers to identifying if “the specific behavioral goals [are] really what society wants.” Social appropriateness of the procedures refers to identifying if those involved with or affected by the intervention “consider the treatment procedures acceptable.” Social importance of the effects refers to identifying if consumers are satisfied with all of the results of the intervention, including any unpredicted ones (Wolf, 1978, p. 207).

Since Wolf’s (1978) influential paper on this topic, much work, theoretical and empirical, has incorporated social validity when studying interventions. Throughout this work, the primary focus has been placed on one of Wolf’s (1978) social validity levels: treatment acceptability. Little attention has been paid to the other two levels (social significance of the goals and social importance of the effects; Carter, 2010). One reason for this may be that, of the levels, treatment acceptability is perhaps the level that can provide researchers and consultants with the most directly useable information to promote practical implementation. Specifically, treatment acceptability generally refers to practical aspects of the treatment procedure (e.g., user understanding of how to implement the procedure, efficiency, costs, complexity), which can

generally be changed (Lennox & Miltenberger, 1990; Reimers et al., 1987). For example, if the intended user of an intervention or support has a limited understanding of the procedures of the support, a consultant can use this information to provide training to increase service providers' understanding of the use of the intervention or support. Or, if it is found that costs are reported to be high, researchers might further investigate ways to reduce associated costs of an intervention to make it more acceptable to service providers and yet maintain the intended effects on outcomes. This level of direct control for changing the intervention is generally not present in the other two levels of social validity (Carter, 2010).

Nonetheless, it is important to also assess teachers' opinions of the other two social validity levels (significance of the goals and importance of the effects). For example, general education teachers' opinion of what the instructional goals should be for ELs in general education science classes may influence the extent to which they provide instructional supports. Further, the extent to which instructional supports are effective may influence the likelihood that teachers provide instructional supports (Carter, 2010).

Given the extensive focus on the second level of social validity (treatment acceptability) and development of associated measurement instruments, this level was the primary focus of the current study. However, additionally, due to the importance of the other two social validity levels (significance of the goals and importance of the effects), the current study also evaluated these levels in the contexts of instructional supports for ELs.

Of the empirical work on social validity that has primarily focused on treatment acceptability, much has evaluated treatment acceptability as a potential predictor of treatment use and integrity (Witt & Elliot, 1985). Various models related to treatment acceptability have been developed, such as the Treatment Acceptability Model (Witt & Elliott, 1985), the Decision-

Making Model of Treatment Acceptability (Reimers et al., 1987), and the Expansive View of Treatment Acceptability (Lennox & Miltenberger, 1990). The Treatment Acceptability Model incorporates four sequential yet interactive elements of a treatment: acceptability, use, integrity, and effectiveness. Witt and Elliott (1985) highlighted that this model seems most applicable to experienced service providers and also acknowledged the limited empirical literature supporting their model. Reimers et al. (1987) expanded on this model by including the extent to which the service provider understands the treatment. In their Decision-Making Model of Treatment Acceptability, they posit that understanding precedes acceptability, such that any treatment that is expected to be maintained by a service provider is first expected to be well understood. Good understanding, in turn, then has the potential to lead to high acceptability and high implementation integrity. All of these components (understanding, acceptability, and integrity) must be in place to achieve high effectiveness and maintenance. Both of these models were critiqued by Lennox and Miltenberger (1990) in their Expansive View of Treatment Acceptability, who claimed that many more factors are needed to understand treatment acceptability and predict use of an intervention. Lennox and Miltenberger (1990) proposed 12 factors grouped into four sequential categories: efficacy considerations, secondary effects, legal and social implications, and practical considerations. Among the three models highlighted, there exists a trend toward expanding the theoretical concept of treatment acceptability to better predict a service provider's use and integrity in providing the treatment. Such models are helpful to guide the development of measurement tools of treatment acceptability.

To empirically support or refute whether there is evidence to suggest that higher acceptability ratings truly do predict higher intervention usage and/or integrity as proposed in each model highlighted above, several studies have been conducted. In terms of a correlation

between acceptability and treatment use, Krain et al. (2005) found inconsistent results. In their study of parents of children with Attention-Deficit/Hyperactivity Disorder (ADHD), the authors asked participants to rate the acceptability of both pharmacological and behavioral treatments, as well as their use of the treatment at a three-to-four-month follow-up. Results indicated that acceptability ratings only predicted use of pharmacological treatment but not behavioral treatment. Another study (McNeill, 2019) investigated the correlation between acceptability and use by surveying special education teachers about evidence-based practices for children with Autism. In this study, the author found results that did suggest a high correlation between treatment acceptability and use, such that those practices with high acceptability ratings were more likely to be used daily by teachers.

In terms of identifying whether treatment acceptability correlates with higher treatment integrity, two corresponding studies were identified. Sterling-Turner and Watson (2002) asked undergraduate students to rate a treatment plan to decrease tic behaviors using the Intervention Rating Profile-15 (IRP-15; Witt & Elliot, 1985) before and after implementing the plan themselves. Participants' adherence to the treatment plan was rated by the researchers. Results indicated very low, nonsignificant correlations between treatment integrity and acceptability at both time points, with $r = .001$ at pretreatment and $r = .13$ at posttreatment. The authors concluded that although valuable data, treatment acceptability measures like the IRP-15, may not accurately measure the likelihood that treatments will be used as intended. In the second study by Mautone et al. (2009), the authors asked special education teachers implementing reading interventions with students with ADHD to rate the interventions using the Behavior Intervention Rating Scale (BIRS; Von Brock & Elliot, 1987). Consultants collected integrity data four times throughout the intervention period using an integrity checklist. Results indicated that the

relationships between treatment integrity and treatment acceptability were significant with a small to moderate correlation coefficient ($r = .32$). These two studies suggest negligible to small relationships between treatment acceptability and treatment integrity.

The mixed results regarding treatment use and integrity suggests that there may be a disconnect between the measurement tools utilized and the theoretical conceptualization of treatment acceptability. Specifically, the measures utilized in the above studies are generally more reflective of the simpler Treatment Acceptability Model (Witt & Elliott, 1985) rather than the more expansive model proposed by Lennox and Miltenberger (1990). The Witt and Elliott (1985) conceptualization of treatment acceptability was criticized by both Chafouleas et al. (2009) and Briesch et al. (2013) in their development of the URP-I and URP-IR. In their development of this measure, the authors argued that, to predict usage (which they defined as both use and treatment integrity), acceptability should be assessed using a more expansive conceptualization. Although the authors did not reference a specific theoretical model, their conceptualization is more closely aligned with the Expansive View of Treatment Acceptability (Lennox and Miltenberger, 1990). Findings from the McNeill (2019) study, the only study to identify a strong correlation between acceptability ratings and use, are promising, as they utilized and adapted version of the URP-IR to measure acceptability.

In applying Wolf's (1978) social validity framework to the context of instructional supports for ELs, the current study further used the Expansive View of Treatment Acceptability as a guiding theoretical framework of the second level of social validity: treatment acceptability. This extends the existing social validity literature by considering a context and population that has received little attention in the literature. Further, it provides information on the extent to which a common factor structure emerges when applying the URP-IR, an instrument validated

for use in measuring acceptability of behavioral interventions, to a different set of supports, namely EL instructional supports.

Social Validity Measurement Approaches and Tools

Given the strong focus on treatment acceptability within social validity work, much empirical work has focused on developing valid measures of this construct. These include the Treatment Evaluation Inventory and (TEI; Kazdin, 1980), Intervention Rating Profile and associated revisions (Witt & Martens, 1983), the Behavior Intervention Rating Scale (BIRS; Von Brock & Elliot, 1987), the Abbreviated and Acceptability Rating Profile (AARP; Tarnowski & Simonian, 1992), and the Usage Rating Profile-Intervention and associated revisions (URP-I; Chafouleas et al., 2009; Briesch et al., 2013). Despite the development of all these measures, their use is not yet widespread, specifically in the school psychology literature. Silva et al. (2020) found that only approximately 40% of intervention studies published in major school psychology journals from 2005 to 2017 included measures of acceptability. Of these studies, only approximately 44% used one of eight validated measures of social validity and 44% used self-developed measures that did not undergo rigorous psychometric investigation. Further, of those using validated measures, 32% of researchers adapted the measure (e.g., modified item wording, selected only certain items for use, adapted the specific Likert scale used). This indicates that, although there is an abundance of available measures, these are either not well-known among researchers or not meeting the specific needs of researchers. Thus, in the following section some previously validated measures relevant to the current study will be discussed and their development, evolution, and potential shortcomings will be reviewed. A detailed review of these measures can also be found in Carter (2010).

Treatment Evaluation Inventory and Associated Revisions

The Treatment Evaluation Inventory (TEI) was developed by Kazdin (1980) as the first measure of treatment acceptability. Of the 16 items, 15 reportedly loaded highly on a single factor (loadings above .40), although exact loadings were not reported. The item with a small factor loading (.24) was removed from the measure. The single factor accounted for 51 percent of the variance and interitem correlations varied, ranging from .35 to .96 with a median of .67. It is unclear whether these reported inter-item correlations still included the item with the small factor loading, which may have reduced the median correlation. The TEI was later adapted into a short form (TEI-SF) by Kelley and colleagues (1989) and reduced to nine items. The TEI-SF demonstrated stronger internal consistency at .85.

For both of the TEI measures, there are several limitations to consider. First, the factor structure and internal consistency of the original TEI was tested only with a sample of 60 introductory psychology students, instead of those using the treatments. Second, Kazdin assumed treatment acceptability to be unidimensional and thus developed the items to reflect that assumption. The principal components analysis in his sample appeared to support the unidimensionality of the measure. However, in a follow-up study by Spirrison et al. (1992), the authors found that the factor structure of the TEI varied based on the treatment that was being evaluated. In their study of six different behavioral treatments, one yielded four factors, four others yielded two factors, and only one yielded a one-factor model. This calls into question the use of the TEI, as well as the TEI-SF as a unidimensional measure. The TEI-SF does not have reports of factor analytic work, which is concerning given its development from the TEI.

Intervention Rating Profile and Associated Revisions

Witt and Martens (1983) sought to design a tool that was applicable to more settings and interventions than just behavioral treatments (as had been the focus of the TEI). The Intervention Rating Profile (IRP) was therefore developed as a measurement tool intended specifically for educational interventions. According to the authors, the IRP provides a strong alternative to the TEI, as it is “a more global measure for assessing the perceptions of a variety of individuals concerning the acceptability of interventions designed for use in a wide array of applied settings” (Witt & Elliot, 1985, p. 278).

Evidence for the technical adequacy of the IRP was slightly stronger than that of the TEI. Witt and Martens (1983) recruited 180 preservice and student teachers, and each participant completed the IRP for six treatments. These treatments were presented through 18 different case studies to increase the generalizability and variability of the IRP ratings. Although Witt and Martens (1983) compared average IRP scores by treatment type, they did not disaggregate results of the factor analysis by treatment type. Results indicated that the IRP had one primary factor (41% of the variance) and four secondary factors (7-9% of the variance). Witt and Martens (1983) labeled this primary factor the “general concern that an intervention was appropriate and that it would help the child” and the secondary factors “acceptability,” “teacher time consumed by the intervention,” “effects of the intervention on other children,” and “amount of teacher skill required to implement the intervention.” Cronbach’s alpha of the overall measure was reported as .91.

The original 20-item IRP was revised later to only include 15 items (i.e., IRP-15; Martens et al., 1985), in an attempt to reduce the measure to a unitary “general acceptability” measure. This revision retained seven of the original IRP items, and added eight new ones. Martens et al.

(1985) administered the reduced scale to a small sample of 54 teachers and found item loadings between .82 and .95 onto one primary factor. Cronbach's alpha was reported as .98.

Behavior Intervention Rating Scale

Von Brock and Elliott (1987) developed the Behavior Intervention Rating Scale (BIRS) to include treatment effectiveness as a separate factor alongside acceptability, given their critique of existing measures (i.e., TEI, IRP) for excluding it. The authors argued that this was a critical covariant of acceptability, such that higher perceived effectiveness of a treatment increased teachers' ratings of acceptability. Thus, Von Brock and Elliott (1987) expanded the construct of social validity beyond acceptability and reintroduced level three of Wolf's social validity framework (social significance of the effects), which had previously received little attention in empirical development of measurement tools.

To develop the BIRS, Von Brock and Elliot used the IRP-15 items and added nine additional items intended to represent treatment effectiveness. They kept the term IRP for the IRP-15 items and called the nine treatment effectiveness items the Effectiveness Rating Profile (ERP). Together, the IRP and ERP composed the BIRS. This measure was administered to 216 teachers enrolled in a graduate program. Results of the factor analyses were mixed. Initially, Von Brock and Elliot (1987) ran two separate factor analyses: one for the IRP-15 and another for the ERP. Both scales yielded single factors. However, when analyzed as a single measure, the BIRS yielded a three-factor model. Two of the ERP items constituted a third factor (labeled "time of effectiveness"), while the IRP items remained a singular factor ("acceptability"), as did the remaining ERP items ("effectiveness"). Although only two of the nine ERP items constituted the "time of effectiveness" factor, it was retained by the authors. Cronbach's alpha of the BIRS was reported at .97. Although these results do provide support towards the unidimensionality of the

IRP-15, it puts into question the unidimensionality of the nine ERP items used as a measure of effectiveness.

Abbreviated Acceptability Rating Profile

Tarnowski and Simonian (1992) found in their work, primarily with mothers in medical settings, that the IRP-15 was often perceived as too long and complex. Thus, the authors developed the Abbreviated Acceptability Rating Profile (AARP) to address these limitations, by reducing the IRP-15 to eight items and changing the language. In their study, the authors administered the AARP to 60 mothers of children at an outpatient hospital clinic. They additionally cross-validated their findings with a second independent sample of 80 participants.

The AARP demonstrated adequate psychometric properties. Items loaded on a unitary factor that accounted for 85-90% of the variance, depending on the sample. Cronbach's alpha was reported at .97 in one study and at .98 in the cross-validation study. Although these findings are indicative of the technical adequacy of the AARP as a measure, it is limited in relevance to the focus of the current study, given that it was administered to parents rather than teachers. Additionally, findings are only applicable to medical settings and not necessarily to others, such as educational settings. It is unclear whether the items, as adapted for the AARP, adequately capture the considerations teachers would have when evaluating an intervention.

Usage Rating Profile-Intervention

Chafouleas et al. (2009) developed the Usage Rating Profile-Intervention (URP-I) to address the shortcomings of some existing measures, particularly the IRP. Similar to Von Brock and Elliot (1987), the authors wanted to move beyond acceptability, as they hoped to address the lack of evidence that suggested that higher acceptability ratings using existing tools (e.g., the IRP) truly correlated with higher use of or fidelity to an intervention. Thus, they called the

measure the “Usage” Rating Profile. To develop their measure, Chafouleas et al. (2009) adapted items from the original IRP and added their own to measure five hypothesized factors that would predict usage: Acceptability, Personal Enthusiasm, Understanding, Integrity, and Feasibility. The scale originally included 78 items, which were reduced to 55 following content validation by expert judges.

To evaluate the psychometrics of the URP-I, Chafouleas et al. (2009) administered it to 254 undergraduate and graduate education students. Participants used the measure to rate a self-management behavioral intervention. Results of the exploratory factor analysis indicated that a four-factor model was a better fit than the originally hypothesized five-factor model:

Acceptability, Understanding, Feasibility, and System Support. As a result of the factor analysis, as well as model fit indices, 20 items were removed from the URP-I, resulting in an overall measure including 35 items. Internal consistency reliability of all subscales was acceptable and ranged from .84 to .96.

Briesch et al. (2013) expanded on the original Chafouleas et al. (2009) study with the goal to effectively capture “information related to a myriad of factors hypothesized to influence the likelihood of school-based intervention usage.” Thus, Briesch et al. (2013) added additional items to strengthen the unsuspected System Support factor that emerged during the Chafouleas et al. (2009) study. The revised scale (URP-IR) included 75 items, which was reduced to 60 following expert content validation.

In their evaluation of the URP-IR’s psychometrics, the authors asked 1005 elementary school teachers to rate one of five class-wide behavior management interventions using the URP-IR. They conducted both an exploratory and confirmatory factor analysis. After the exploratory factor analysis, 31 more items were removed for various reasons (e.g., multicollinearity, low

communality, low pattern coefficient), thus resulting in a final measure of 29 items. The exploratory factor analysis indicated a six-factor model, which was also supported by the confirmatory factor analysis. Those six factors were Acceptability, Understanding, Feasibility, Family-School Collaboration, System Climate, and System Support. This model included the four from the original Chafouleas et al. (2009) study, with the addition of the Family-School Collaboration and System Climate factors.

The URP-IR (Briesch et al., 2013) has several strengths that make it a potentially useful social validity tool in educational settings. First, being developed more recently than any of the other scales reviewed above, the URP-IR was derived from already existing measures (particularly the IRP), but also addressed some of the limitations of those existing measures. Specifically, as attempted by the BIRS, the URP-IR was expanded to include more than just a unitary measure of acceptability. Items were added to address teacher concerns about implementation beyond their personal level of acceptability (e.g., enthusiasm, liking the intervention), such as systemic constraints (e.g., resources, administrative support) and level of understanding of the intervention. These additions go beyond the singular effectiveness factor added by the BIRS and are more closely aligned with later models of acceptability (e.g., Expansive View of Treatment Acceptability; Lennox & Miltenberger, 1990). Further, another strength of the URP-IR is that it was evaluated with more rigorous samples, both in size and generalizability. The development process of the scale items, including expert validation, is yet another advantage of the URP-IR that was not evident in many of the other measurement tools.

When considering the utility of the URP-IR especially for research in school psychology, it appears particularly advantageous. Similar to the IRP, the URP-IR was developed specifically for school-based interventions. However, unlike the IRP, Briesch et al. (2013) highlight that the

URP-IR is intended as an adaptable measure to be used across a variety of domains. This is promising, however, research on the URP-IR thus far has focused solely on interventions targeting behavioral concerns. Briesch and colleagues (2013) recommend further psychometric work be done on the utility of the URP-IR for academic interventions. This work is particularly warranted given that the Chafouleas et al. (2009) and Briesch et al. (2013) studies found different factor structures.

Given that the URP-IR demonstrates several strengths, each of the six factors identified by Briesch et al. (2013) will be more closely described in the following sections. This will provide a deeper understanding of the measure and its items. One aim of the current study was to build on this work by exploring whether the six-factor structure holds when applying the tool to examine acceptability of EL supports. Given findings that other measurement tools have shown varied factor structure depending on the intervention evaluated (Spirrison et al., 1992), it is critical to understand how the URP-IR functions specifically when examining teachers' opinions of instructional supports for ELs.

Factor 1: Acceptability. Briesch and colleagues (2013) define the Acceptability factor as “how acceptable the individual found the intervention to be and how enthusiastic the individual would be about implementing the intervention” (p. 89). This factor is comprised of nine items, such as “this intervention is a good way to handle the child’s behavior problem,” “I would implement this intervention with a good deal of enthusiasm,” and “I would be committed to carrying out this intervention” (p. 88). Overall, the factor measures the appropriateness of the intervention given the problem presented. The mean of inter-item correlations for this factor was .69 and Cronbach’s alpha was .95 (Briesch et al., 2013).

Factor 2: Understanding. The Understanding factor is defined as “the extent to which participants understood how to implement the target intervention” (Briesch et al., 2013, p. 89). It is comprised of three items: “I understand the procedures of this intervention;” “I understand how to use this intervention;” and “I am knowledgeable about the intervention procedures” (p. 88). This factor measures teachers’ understanding of the particular intervention in terms of implementation and use. The mean of inter-item correlations for this factor was .58 and Cronbach’s alpha was .79 (Briesch et al., 2013).

Factor 3: Family-School Collaboration. The Family-School Collaboration factor is defined as “the extent to which participants believed family-school collaborations were necessary for an intervention to be successfully utilized” (Briesch et al., 2013, p. 89). It is comprised of three items: “Parental collaboration is required in order to use this intervention;” “A positive home-school relationship is needed to implement this intervention;” “Regular home-school communication is needed to implement intervention procedures” (Briesch et al., 2013, p. 88). The mean of inter-item correlations for this factor was .55 and Cronbach’s alpha was .78 (Briesch et al., 2013). From a social validity perspective, additional requirements to make the support or intervention successful (i.e., family collaboration) would be expected to reduce the overall acceptability of the support or intervention. In their study, Briesch and colleagues (2013) found this subscale to be most weakly correlated with other subscales. It was most strongly correlated with the System Support subscale, another subscale on which higher scores indicative of lower likelihood of usage.

Factor 4: Feasibility. Briesch et al. (2013) define Feasibility as “whether or not the participant felt that implementing the intervention as described was feasible” (p. 89)” The factor is comprised of six items, including “material resources needed for this intervention are

reasonable,” “this intervention is too complex to carry out accurately,” and “I would be able to allocate my time to implement this intervention” (p. 88). Three of the questions refer to time, two refer to material resources, and one refers to the complexity of the intervention. The mean of inter-item correlations for this factor was .55 and Cronbach’s alpha was .88 (Briesch et al., 2013).

Factor 5: System Climate. Briesch et al. (2013) defined the System Climate factor as “whether the intervention was compatible with the school environment (e.g., sufficient staff support)” (p. 89). The factor is comprised of five items such as “my administrator would be supportive of my use of this intervention,” “these intervention procedures are consistent with the way things are done in my system,” and “my work environment is conducive to implementation of an intervention like this one” (Briesch et al., 2013, p. 88). The mean of inter-item correlations for this factor was .68 and Cronbach’s alpha was .91.

Factor 6: System Support. The System Support factor is defined as “whether respondents felt they would need external support to use the intervention” and is comprised of three items: “I would require additional professional development in order to implement this intervention;” “I would need consultative support to implement this intervention;” “I would need additional resources to carry out this intervention” (Briesch et al., 2013, pp. 88-89). This factor is distinct from the System Climate factor in that it refers to “practical aspects of supports” that are often provided through the system (e.g., professional development, consultation) rather than the overall culture of the system within which the particular support may or may not fit (Briesch et al., 2013). Further, it is distinct from the Feasibility factor as it refers to resources at the system-level rather than the teacher-level (e.g., time). The mean of inter-item correlations for this factor was .41 and Cronbach’s alpha was .67 (Briesch et al., 2013). Lower scores on this factor reflect a

“greater ability to independently implement the intervention” and thus this factor should be reverse scored when aggregating factors to derive an overall mean score (Briesch et al., 2013).

In summary, the URP-IR is a promising measurement tool of social validity, specifically for educational settings. It builds upon prior social validity measures, addressing some of their shortcomings by more comprehensively assessing factors that may influence usage of an intervention. Additionally, it was intended to be used for a variety of interventions. However, the URP-IR has so far been used only with behavioral interventions. Therefore, it is unclear whether, when used to assess the acceptability of other interventions or instructional supports, the same factor structure holds. Thus, the current study investigated the utility of this measure in the context of instructional supports for ELs.

Predictors of Teacher Ratings of EL Supports

Extensive work has been done to inform how various teacher background experiences may potentially relate to their ratings of instructional supports for ELs. Much of this work has been summarized by Pettit (2011) in their systematic review of the literature. Results indicated that teacher training specific to working with ELs, exposure to language diversity, and speaking another language were all predictors of teacher beliefs. In their review, Pettit (2011) defined the term belief as “encompassing many mental constructs such as knowledge, attitudes, and perceptions” (p. 126). Pettit (2011) chose this broad definition, as many authors in the EL literature use the terms knowledge, attitudes, and perceptions interchangeably and rarely provide definitions of these constructs. Additionally, it is important to note that Pettit (2011) focused on research of teacher beliefs as related to ELs in a broad sense, including topics such as instructional practices, inclusion of ELs in the mainstream classroom, and the role of ESL teachers. Thus, only some studies included in the review specifically investigated teacher

opinions of instructional supports for ELs in mainstream classrooms. For the purposes of the present study, attention was paid specifically to these studies in the subsections described below, as they most directly inform the current research questions.

In addition to Pettit's (2011) work, another body of literature that provides direction as to which predictors may be important to study is the social validity literature. Those who select or implement interventions (i.e., teachers) are often called consultants or consumers in the social validity literature. Carter (2010) identified the following consultant and consumer characteristics as potential predictors of social validity: training, history with treatment, personal ethical views regarding behavior principles, familiarity with recent research, knowledge of the client situation, presentation of treatment, assessments conducted, gender, knowledge of treatment, socioeconomic class, geographical area, experience parenting a child with a medical disorder, and marital distress. These are largely based off social validity research conducted in the ABA field and are representative of consumers who are often not mainstream teachers, such as special education teachers, behavioral interventionists, and parents.

Although the existing research does offer some ideas about which teacher background experiences may correlate with their acceptability of instructional supports for ELs, a validated measure has not been used to systematically examine these relationships. In the current study, three teacher characteristics were selected to examine as potential predictors of acceptability of EL instructional supports, as assessed by the URP-IR. Selections of teacher characteristics were based on past empirical work from both the broader research on teacher opinions of ELs and social validity research. The three characteristics selected were teacher training about teaching ELs, teacher experience working with ELs, and consultation from ESL teachers.

Teacher Training

Research on training for working with ELs has typically been divided into two categories: pre-service and in-service training. Pre-service training refers to training received during the formal training period before beginning teaching (i.e., university courses). In-service training can refer to several activities, such as workshops, conferences, and professional development. Training in the Pettit (2011) review was described as both professional development activities and university courses and focused on various topics such as English as a Second Language (ESL), foreign language (i.e., teachers having the experience of learning a foreign language), multicultural or bilingual education. The studies reviewed varied in their conceptualization of training.

From a social validity framework, training on instruction of ELs or related topics might be particularly influential on teacher opinions of various supports. Specifically, on the URP-IR measure, teachers with more training may rate supports more favorably on the Understanding subscale, as they may already have foundational knowledge that allows them to apply such a support more easily. Further, those teachers may also rate supports more favorably on the System Support subscale, as they require less support from consultants or administrators to implement the support than those with less training.

Results from the Pettit (2011) review indicated that teacher training about teaching ELs was the most consistent predictor of mainstream teachers' beliefs towards instructional practices for ELs. Specifically, those with more training tended to view strategies such as bilingual supports more positively, and also reported feeling more confident to implement them. For example, Karathanos (2009) reported more positive teacher perspectives about including ELs' L1 in mainstream instruction from those teachers whose university education included at least

nine credit hours of ESL-specific coursework. This coursework included instruction on theory and research-based methods to promote the success of ELs in schools, with a specific emphasis on appropriate use of ELs' L1 in mainstream instruction (i.e., allowing students to access content and demonstrate understanding in their first language). More recent studies also support the Karathanos (2009) finding (e.g., Huerta et al., 2019; Villegas et al., 2018). For example, Huerta et al. (2019) found that teachers who had received professional development specifically about working with ELs in science viewed providing instructional supports such as inclusion of L1, teaching academic vocabulary, integrating oral and written English into teaching, providing opportunities to develop written language skills, and providing small group instruction as more acceptable than teachers who had not received such training. Further, in a more recent review of the literature on teacher training regarding ELs, Villegas et al. (2018) identified eight studies that reported pre-service training about strategies for teaching ELs having a positive relationship with acceptability or positive attitudes towards adapting mainstream instruction for ELs. Given such findings, training experiences in both the university and in-service setting were a construct of interest in the current study. Based on the existing literature, it is likely that those teachers with more training on instructional strategies for ELs might rate empirically-supported instructional strategies for ELs more favorably than those teachers with less training.

Experience with ELs

Exposure to working with ELs was found to be a potential predictor of teacher beliefs by Pettit (2011) based on mixed results of various studies. However, many of the studies reviewed did not specifically study teacher beliefs regarding instructional supports for ELs; rather, they studied teacher beliefs more generally about inclusion of ELs in mainstream classrooms (e.g., Byrnes et al., 1997; Gandara et al., 2005; Shin & Krashen, 1996). These studies generally found

positive relationships between exposure to working with ELs and favorable views towards including ELs in mainstream classrooms. Two studies that specifically focused on teacher beliefs of an empirically-supported instructional strategy, namely inclusion of L1, were Karathanos (2009) and Lee and Oxelson (2006). Both studies reported no relationship between teacher experience working with ELs and their beliefs toward L1 inclusion. Neither study specifically reported the background characteristics of the ELs with whom teachers in the study reported having experience.

Overall, given some variation in the literature, it is unclear whether one can expect a difference in teacher ratings of instructional supports based on their experience working with ELs. Some studies have found differences (e.g., Byrnes et al., 1997; Gandara et al., 2005; Shin & Krashen, 1996). However, these studies did not have the specific focus on instructional supports whereas those that did not find a significant relationship indeed did focus on instructional supports (e.g., Karathanos, 2009; Lee & Oxelson, 2006). From a social validity framework, it was hypothesized that those teachers with more experience working with ELs would find the given instructional supports more favorable on the Feasibility and Acceptability subscales. Those teachers may have experience providing such supports already, thus reducing the amount of time and resources needed and increasing the extent to which they find the support valuable to address needs of ELs. Inclusion of experience with ELs as a variable in the current study was considered potentially valuable in providing further evidence supporting or refuting whether this variable contributed to teacher acceptability ratings of empirically-supported instructional strategies.

Consultation with the ESL Teacher

It is suggested in the available literature that there is significant variability in the roles of ESL teachers nationwide. Bell and Baecher (2012) reported that in a predominantly U.S.-based

ESL teacher sample ($n = 72$), pull-out, push-in, and coteaching services were part of their roles, with over half of teachers (67%) spending over half of their time providing pull-out services. Within these different roles, the extent to which ESL teachers might serve as consultants to mainstream teachers would be expected to vary significantly. For example, an ESL teacher who provides push-in services would have more opportunities to demonstrate strategies to the mainstream teacher than an ESL teacher who provides pull-out services.

Given that ESL teachers have such varied roles and that the extent to which consultation is provided has often been identified as a significant predictor of social validity ratings in the social validity literature (Carter, 2010), inclusion of this variable as a potential predictor of teacher URP-IR ratings was warranted. It was hypothesized that those teachers who experience more consultative support rate instructional supports more favorably on the Systems Support subscales, as they feel adequately supported to implement it.

Consultation provided by ESL teachers was not explored as a predictor of teacher beliefs by Pettit (2011). However, in another review by Khong and Saito (2014) on the barriers teachers experience when supporting ELs in mainstream education, the authors found the lack of effective relationships with ESL teachers to be a potential barrier to supporting ELs in mainstream education. This was due to two challenges. First, teachers reported lack of communication between ESL and mainstream teachers as a challenge (Khong & Saito, 2014). Therefore, it was hypothesized in the current study that those teachers receiving more consultation from ESL teachers may rate instructional supports for ELs more favorably, specifically on the System Support URP-IR items. A second challenge regarding ESL teachers identified by Khong and Saito (2014) was that some mainstream teachers do not view adapting their instruction for ELs as part of their role, but rather that of the ESL teacher. Thus, teachers' ratings on the Acceptability

URP-IR subscale may vary based on the consultative relationship with the ESL teacher, as this may shape their willingness to implement supports in their own instruction. For example, it was hypothesized in the current study that those receiving more consultation from ESL teachers may be more likely to view adapting instruction as part of their role and thus may find instructional supports more acceptable.

Empirically-Supported Instructional Supports for ELs

The SIOP Model

The research literature on effective instruction for ELs points to various ways to support ELs both in increasing their English proficiency and their content knowledge. One of the most well-known teacher training models for adapting mainstream instruction to meet the needs of ELs is the Sheltered Instruction Observation Protocol (SIOP; Echevarría et al., 2014). The SIOP model recommends adapting instruction through various strategies, called “features,” organized in eight components (see Table 1). SIOP has been found to increase ELs’ academic performance, as well as their English proficiency, in various studies (e.g., Echevarría et al., 2006; Echevarría et al., 2011; Short et al., 2011).

Table 1. *Overview of the SIOP Model*

Component (n = 8)	Feature (n = 30)
1. Lesson Preparation	1. Content objectives
	2. Language objectives
	3. Appropriate content concepts
	4. Supplementary materials
	5. Adaptation of content
	6. Meaningful activities
2. Building Background	7. Concepts linked to students’ backgrounds
	8. Links between past learning and new concepts
	9. Developing key vocabulary
3. Comprehensible Input	10. Appropriate speech
	11. Clear explanation of academic tasks
	12. A variety of techniques used
	13. Learning strategies
	14. Scaffolding techniques

Table 1 (cont'd)

5. Interaction	15. Higher-order questioning and tasks
	16. Frequent opportunities for interaction
6. Practice and Application	17. Grouping configurations
	18. Sufficient wait time
	19. Clarify concepts in L1
	20. Hands-on practice with new knowledge
	21. Application of content and language knowledge in new ways
7. Lesson Delivery	22. Integration of all language skills
	23. Support content objectives during lesson
	24. Support language objectives during lesson
	25. Promote student engagement
8. Review and Assessment	26. Pace lesson appropriately
	27. Key vocabulary
	28. Key content concepts
	29. Regular feedback on student output
	30. Assess student comprehension of objectives

Review of all components of the model is beyond the scope of this literature review; however, features of the model, along with findings from the empirical literature, were used to guide selection of four promising instructional supports for further exploration in the current study: use of visual aids, vocabulary supports, incorporation of student's L1, and allowing alternate response format. These four supports were selected for two reasons. First, supports were selected that were specifically linguistic accommodations meant to reduce language-related barriers, rather than simply instructional accommodations that may benefit all students. The SIOP model includes both of these types of supports; however, linguistic accommodations may ultimately be the most important to implement to ensure equal access to instruction for ELs. Thus, selection of supports was limited to linguistic accommodations. Second, the present study's focus is on content area instruction, specifically science. As such, supports were selected that might be particularly relevant to this domain of instruction.

In the following sections, each selected instructional support will be described in detail. It is important to note that although distinct, these four supports are not distinct categories or manualized interventions. Rather, they have somewhat fluid definitions and share certain features. Further, multiple of the selected supports could be used together or integrated to strengthen their effectiveness. However, for the purposes of the current study, the supports are described as isolated strategies. This allowed for investigation of teacher acceptability of the unique qualities of each support.

Visual Aids

Visual aids include non-linguistic and non-written representations of content and are usually provided alongside written text to allow for greater meaning-making for ELs (Martiniello, 2009; Serafini, 2012). Representations may be pictorial, depicting details of certain elements described in the written text, or schematic, depicting the connections between certain elements described in the written text (Martiniello, 2009). In the empirical research literature visual aids have been found effective in allowing ELs greater access to written text. In a study of the impact of linguistic complexity of mathematics items on the items' measurement comparability, Martiniello (2009) found that the measurement comparability between ELs and non-ELs was more often problematic for items with high linguistic complexity. However, some items on the fourth-grade test were accompanied by schematic visuals, while others were not. Martiniello (2009) found that the effect of linguistic complexity was more likely to be mitigated for those items that had accompanying visuals.

Visual aids aligned with the above definition are also included in the SIOP model, specifically in features four, nine, and 12. These features focus on providing supplemental materials (feature four) and various techniques (feature 12) to make content comprehensible.

Feature nine focuses on developing key academic vocabulary and visual aids can be used here in the form of a picture dictionary that serves as a support for contextualizing written definitions. Further, Echevarría et al. (2014) suggest that teachers utilize various sources like pictures from the Internet, hand-drawn images, or charts to help support instruction that is being given verbally. Such visual materials can then be incorporated into instruction through the use of PowerPoint or other visual technologies.

In the context of science instruction (the focus of the current study), visual aids can be used in a variety of ways. Following SIOP recommendations, a teacher may use visual aids to contextualize information provided both verbally and visually. For example, the teacher may provide students with a chart that visually depicts the water cycle as they are verbally explaining it to the class. Further, they can label cupboards where relevant materials are kept both in writing and picture format. Additionally, when giving instructions for an activity, they may provide relevant visuals alongside those instructions. To develop these visuals, teachers should follow guidelines as laid out by Solano-Flores and colleagues (2014): represent constituents (words, phrases, idiomatic expressions) likely to pose a challenge to [ELs]; represent only one or two of those constituents; represent concrete concepts (e.g., objects, actions); not be complex. This list was slightly reduced, as Solano-Flores et al. (2014) intended this list as guidance for developing visual aids during testing.

The provision of visualizations may address some of the language-related barriers ELs experience, by providing visuals alongside the verbal or written input. It was hypothesized that use of visual aids as an instructional support would be highly socially valid to most teachers. Based on the social validity framework, one might expect this support to be rated quite favorably, as it requires relatively few resources, preparation, and external support. Further, this

support potentially has high face validity, as visual aids are typically a common instructional strategy used by teachers for all students.

Vocabulary Supports

Vocabulary supports include provisions of translations or explanations of key vocabulary. When identifying vocabulary supports, the key vocabulary is often referred to as “academic vocabulary.” Academic vocabulary is defined as “the language for reading and writing, English grammar, prosody, oral academic discourse, English syntax, and self-talk that promotes thinking and knowing” (Echevarría et al., 2014, p. 69). Academic vocabulary can be divided into three sections: content vocabulary, general academic vocabulary, and word parts (Echevarría et al., 2014). Content vocabulary includes those words relevant to the specific subject or lesson that students may come across as they are learning the content. General academic vocabulary are those words that are not specific to one subject or lesson, but rather are those that relate to the learning or task process. For example, these might include “data table,” “argue,” “conclusion,” or “furthermore.” Often, these words are not explicitly taught; however non-EL students are typically familiar with them, whereas ELs are not (Echevarría et al., 2014). Word parts vocabulary are comprised of roots and affixes, such as “photo” often being an affix meaning “light” (Echevarría et al., 2014).

The value of providing vocabulary support to ELs to increase their access to content has been demonstrated particularly in the test accommodation literature. In their meta-analysis of test accommodations, Pennock-Roman and Rivera (2011) found that the provision of English dictionaries or glossaries was the most effective accommodation, given that there were little to no time constraints placed on the students. For example, Abedi et al. (2001) randomly assigned eighth grade ELs to either receive or not receive a customized dictionary, which including only

words on the test, during a science test. Students who received the customized dictionary scored significantly higher at 1.5 points (of 20 possible points) above the students in the standard condition.

Within the SIOP model, Echevarría et al. (2014) provide several strategies for providing vocabulary support during instruction. First, vocabulary itself is represented by an entire feature, number nine: key vocabulary emphasized. Although vocabulary instruction is primarily done through targeted ESL services, Echevarría et al. (2014) include it in the SIOP model to foster ELs' content learning. The authors argue that teachers must become conscious of the vocabulary used in their instruction and teach or accommodate the academic vocabulary that is necessary for successful comprehension of the lesson. They provide various activities to facilitate understanding of key vocabulary. For example, they recommend Cunningham's (2004) word wall activity, which was developed as a method for supporting sight word learning for all students. For each lesson, teachers select key vocabulary and present them on a large poster that is reviewed with all students prior to the lesson and throughout as needed. Using the word wall mirrors the idea of the customized dictionary that was used by Abedi et al. (2001) by selecting key vocabulary relevant to the content presented.

The word wall activity can be applied in the context of science instruction for elementary EL students. Using this support during instruction would require a teacher to pre-select key vocabulary from the lesson, specifically the general academic vocabulary that is important for the lesson. Although the content vocabulary is also critical to teach, it is likely already part of the lesson. Inclusion of general academic vocabulary is what would differentiate the word wall as a support for all students from a word wall that specifically supports ELs. Use of the word wall support would likely reduce the language-related barriers ELs would otherwise experience.

Incorporation of L1

L1 is often used in the EL literature to refer to the native language of the student. Incorporation of L1 as a support means allowing the use of the L1 to access content. Use of L1 to support ELs has been found effective in the empirical literature and Cummins (2007), as well as August and Shanahan (2006) posit that use of L1 can facilitate development of L2 (English) and is often promoted as a scaffolding strategy (Settlage et al., 2014). Under some circumstances, translation of tests into L1 can allow greater access for ELs to demonstrate their knowledge of the content; specifically, this is effective when ELs have proficiency in their L1 or have received instruction in their L1 (Pennock-Roman & Rivera, 2011). Further, provision of L1 support through means such as bilingual glossaries during test taking has been found to be one of the most effective accommodations to increase students' access to test content (Pennock-Roman & Rivera, 2011).

The incorporation of ELs' L1 is represented in the SIOP model in two features: feature five (adaptation of content) and feature 19 (clarify concepts in L1). In feature five, the SIOP model calls for integration of content in the student's L1, for example through reading, listening, or watching relevant sources. Feature 19 focuses on allowing for clarification beyond the lesson given by the teacher in English. Here, SIOP advocates for pairing ELs who speak the same L1 to talk about any points of confusion together; however, if not possible, teachers are also encouraged to use online translation tools or reference materials to clarify key vocabulary and concepts.

In the context of science instruction for elementary students, teachers may incorporate an EL student's L1 in a variety of ways. One example is for the classroom teacher to provide ELs with materials in their L1 (or with available subtitles or translations) using media. They may find

videos in the EL's L1 explaining content and provide those as resources to ELs. Further, they can allow ELs to use laptops or tablets to utilize electronic translators and encourage the students to use those translators. This would reduce the language-related barrier ELs may experience during English-only instruction and thus increase their understanding of the material. From a social validity framework, varying degrees of teacher endorsement of this support was hypothesized. Despite its empirical support, teachers may not feel comfortable with students accessing content that they themselves potentially cannot understand and thus cannot assess its quality. Hesitancy towards use of L1, as well as misconceptions about its effectiveness, have in fact been found in several studies (Pettit, 2011).

Alternate Response Format

Providing this support means to allow students to demonstrate their knowledge or understanding in a way other than what is typically expected for a task. For example, on a worksheet that requires students to write responses, an alternate response format would allow students to give responses orally, through pictures, or in their native language. SIOP advocates for allowing ELs to demonstrate their understanding of content in various formats in several features within component eight (Review and Assessment). Echevarría et al. (2014) encourage teachers to use multidimensional assessments such as verbal responses or artwork to assess ELs' knowledge of lesson content. Specifically, in feature 30 (assess student comprehension of objectives) it is recommended that teachers ask students to draw pictures of or act out key vocabulary terms. Additionally, in feature 26 (pace lesson appropriately), it is suggested that teachers adapt instructional tasks to fit the student's English proficiency level. Overall, it is part of the SIOP model, as demonstrated throughout some of its features to "offer multiple pathways for children to demonstrate their understanding of the content" (p. 21).

Applying this instructional support for upper elementary ELs during science instruction could be done in several ways. One example is for the classroom teacher to allow students to demonstrate their understanding of the concept through something other than written or spoken English. For example, the teacher may allow the student to draw concepts, while non-ELs may compose a written explanation of the same concept. Using alternate response formats not only reduces the language barrier ELs may experience when expressing their understanding of a concept, it also provides greater access to the instructional content by maximizing their ability to engage with the material (Symons, 2020). It is also related to concepts from inquiry-based science instruction, a promising instructional model (Furtak et al., 2012) in which students form and demonstrate understanding through student-led activities. From a social validity framework though, it was hypothesized that teachers may have some concerns about the acceptability of alternate response format. It may require additional preparation to identify alternate response formats appropriate for a given lesson. Additionally, teachers may need support in how to use work products in alternate response formats as a tool to capture student's understanding of the material.

Summary

In summary, the current study applied the social validity framework to measuring teacher opinions of empirically-supported instructional supports for ELs. The relevant social validity literature was reviewed and pointed to the value of the UPR-IR (Briesch et al., 2013) as a measurement tool for this context. However, the technical adequacy and measurement characteristics of this measure have not yet been studied outside of classroom-wide behavior interventions. Thus, it is unclear whether the factor structure of the measure would apply to the context of the current study. Further, both the social validity and EL literature have pointed to

potential predictors of mainstream teachers' opinions of instructional supports for ELs. Thus, the current study investigated whether teacher training specific to teaching ELs, experience with ELs, and consultation from the ESL teacher predicted teacher URP-IR ratings of instructional supports for ELs. Although there are some inconsistencies in the literature, it was expected that each of these variables would be positively correlated with URP-IR ratings. Teachers rated four empirically-supported instructional supports for ELs: visual aids, vocabulary supports, inclusion of L1, and alternate response format. These supports have all been found to reduce language-related barriers for ELs and are included in the SIOP model, an evidence-based model for adapting mainstream instruction for ELs. Each support has distinct features that may influence the extent to which teachers find it acceptable. Lastly, the available social validity literature and existing measurement tools largely focus on Wolf's (1978) second level of social validity, social appropriateness of the treatment procedures (or, treatment acceptability). However, the other two levels – social significance of the goals of treatment and social importance of the effects of treatment – are also important to assess to yield a more complete understanding of the social validity of instructional supports for ELs. Thus, the current study incorporated assessment of these two levels in addition to use of the URP-IR to study treatment acceptability.

The current study will address the following research questions:

1. Is the six-factor structure of the revised Usage Rating Profile-Intervention (URP-IR) a good model fit when using this measure for teacher ratings of empirically-supported instructional supports for English Learners during science instruction?
2. To what extent do the following teacher background experiences predict URP-IR ratings for the targeted EL supports?
 - a. Experience working with ELs

- b. Teacher training on teaching ELs
 - c. Role of the ESL teacher
- 3. What are elementary mainstream teachers' URP-IR ratings for four empirically-supported instructional supports for English Learners during science instruction? Are there significant differences in URP-IR scores across these instructional supports?
- 4. What are teachers' opinions about the social significance of the goals and social importance of the effects of the selected EL instructional supports?

CHAPTER III: METHODS

Research Design

The current study utilized a dominant status, sequential design (Johnson & Onwuegbuzie, 2004) for data collection. The study first included a quantitative phase, in the form of a survey collecting only quantitative data, and then a qualitative phase in the form of follow-up interviews with select teachers. The quantitative data collected during phase one is the primary data of interest and is thus the dominant source of data for the study. The qualitative data serves as supplemental data to enrich the interpretations of the quantitative data. For a comprehensive overview of the research questions, associated variables, and data analysis, please see Table 3 at the end (pg. 56) of this *Methods* section.

Participants

14,999 elementary teachers across the United States were invited to participate in the current study. Of these invited participants, 3,000 were identified through Market Data Retrieval (MDR), an education database that provides contact information for educators across the country. Due to low response rate from this sample (see the *Recruitment* section below), additional participants were identified from elementary school websites across the country.

In total, 738 participants started the survey, and 295 participants fully completed the survey, including answering demographic questions. This left 443 participants with partially complete data. Of those, 81 participants filled out at least one of the four sets of 29 support questions. The rest (n=362) opened the survey but did not answer any questions. For those 81 participants who completed at least one set of support questions (n=81), imputation methods were utilized to impute responses for the remaining sets of support questions, rendering a final N of 376. Imputation was conducted using multiple imputation methods through SPSS, for all

items for each of the four supports. Constraints were placed on each item such that the imputed value had to be an integer and fall in the Likert scale range of one through six of the URP-IR. Additionally, the following demographic variables were utilized as predictors when possible: age, gender, race, highest level of education, and years of teaching experience.

Demographics of the study sample are outlined in Table 2 below. It is important to note that even though imputation was utilized to impute missing data for the URP-IR items, it was not utilized to impute missing demographic data. Thus, Table 2 below shows the demographic information of the 295 participants who answered demographic questions. Additionally, it is important to note that responses to demographic questions were not required to be provided by participants, and thus, N's vary due to participants skipping questions if they chose to do so.

Table 2. *Demographics*

	Percentage (n)
Age (n = 279)	
30 or under	20.4 (57)
31-40	19.4 (54)
41-50	25.4 (71)
51-60	28.3 (79)
61 or over	6.5 (18)
Gender (n = 289)	
Female	93.1 (269)
Male	5.2 (15)
Prefer to self-describe	1.7 (5)
Race (n = 288) <i>*percentages add up to greater than 100 due to participants selecting more than one race</i>	
White	87.5 (252)
Black or African American	4.5 (13)
American Indian or Alaska Native	2.4 (7)
Asian	3.0 (9)
Native Hawaiian or Pacific Islander	0.3 (1)
Other	5.9 (17)
Languages spoken other than English (n=69) <i>*percentages add up to greater than 100 due to participants selecting more than one language</i>	
Spanish	66.7 (46)
Chinese	2.9 (2)
Tagalog	2.9 (2)

Table 2 (cont'd)

	Other	37.7 (26)
Highest level of education (n = 290)		
	Associate degree	0.3 (1)
	Bachelor's degree	29.3 (85)
	Master's degree	67.9 (197)
	Doctoral degree	1.4 (4)
	Professional degree (JD, MD)	1.0 (3)
Years of teaching experience (n = 287)		
	5 years or less	18.1 (52)
	6-15 years	35.2 (101)
	16-25 years	32.1 (92)
	26 years or more	14.6 (42)
Grade levels taught (n = 208) <i>*percentages add up to greater than 100 due to participants selecting more than one grade level</i>		
	Kindergarten	38.5 (80)
	1 st	50.5 (105)
	2 nd	48.6 (101)
	3 rd	48.1 (100)
	4 th	47.6 (99)
	5 th	41.8 (87)
Grade levels they have taught science for (n = 204) <i>*percentages add up to greater than 100 due to participants selecting more than one grade level</i>		
	Kindergarten	37.3 (77)
	1 st	51.0 (104)
	2 nd	47.5 (97)
	3 rd	49.0 (100)
	4 th	47.1 (96)
	5 th	38.2 (78)
Location of school (n = 290)		
	Rural	16.6 (48)
	Suburban	53.1 (154)
	Urban	30.3 (88)
Geographical location of school (n = 290)		
	Midwest	26.9 (78)
	Northeast	14.8 (43)
	Southeast	19.7 (57)
	Southwest	4.1 (12)
	West	34.5 (100)
Percent of students eligible for free and reduced lunch (n = 284)		
	Less than 25%	20.4 (58)
	25-49%	16.9 (48)
	50-74%	21.5 (61)
	75% or more	41.2 (117)

Measures

The survey in the current study consisted of multiple sections: (1) consent and confidentiality statement, (2) a vignette description and set of 29 URP-IR questions for each of the four supports, (3) demographics and other background information, and (4) participant contact information for incentive distribution and question about interest in participating in the follow-up interview. The survey was administered online and distributed via Qualtrics.

Acceptability

Acceptability of each instructional support was measured by the 29 URP-IR items (Briesch et al., 2013). Although Briesch et al. (2013) did not refer to the construct assessed by the URP-IR as “acceptability,” it is referred to as this in the current study, as it is meant to assess Wolf’s (1978) second level of social validity. Each instructional support (i.e., visual aids, vocabulary supports, incorporation of L1, and alternate response format) was described in detail before participants were asked to rate it using the URP-IR. Each vignette followed the same structure, in which the support was first defined, then how to implement the support was described in multiple steps, and an applied specific science lesson example was provided. This structure was modeled after the sample provided in Briesch et al. (2013) and provided all participants with a common understanding of the instructional support prior to rating it. For the full vignettes, please see Appendix A. Participants were asked to envision incorporating this support in all of their science lessons and rate it using the URP-IR items. Order of the vignettes was counterbalanced to control for potential ordering effects.

The 29 URP-IR items followed the presentation of each of the 4 support vignettes. The items were kept identical across the supports and were identical to the existing URP-IR items – with the exception that slight wording changes were applied to reflect the context of the current

study. For example, “this intervention is a good way to handle the child’s behavior problem” was modified to “this instructional support is a good way to handle the child’s language barrier.” Participants were asked to rate each item on a Likert-type scale ranging from “Strongly disagree” to “Strongly agree.” For all 29 items, please see Appendix B.

In the Briesch et al. (2013) study, exploratory and confirmatory factor analyses identified six factors: Acceptability, Understanding, Feasibility, Family-School Collaboration, System Climate, and System Support. Results indicated that the URP-IR has adequate psychometric properties, with Cronbach’s alpha values as follows for each subscale: Acceptability $\alpha=.95$; Understanding $\alpha=.80$; Feasibility $\alpha=.84$; Family-School Collaboration $\alpha=.79$; System Climate $\alpha=.91$; System Support $\alpha=.72$.

Social Validity Interview

The structure of the interviews was guided by the semi-structured social validity interview developed by Gresham and Lopez (1996) which covers all three levels of social validity. For the purpose of the follow-up interview, which was to collect more information about the two levels other than treatment acceptability, the interviewer covered only topics addressing the social significance of the goals and social importance of the effects (see Appendix C for the full list of questions). Questions were substantially adapted to fit the context of the current study, as the original Gresham and Lopez (1996) interview was intended for a practical consultation purpose, rather than a research purpose. Despite these contextual differences, it was clearly the best available measure identified in the literature to guide development of interview questions for the current study. In the adapted version of this interview, the social significance of the goals was conceptualized as the teachers’ perceptions of the goals for ELs in mainstream science instruction and how the selected instructional supports align with those goals. The social

importance of effects was conceptualized as teachers' anticipated levels of satisfaction with the outcomes of the instructional supports.

Demographics and Participant Information

Participants were asked to report on the following demographics and personal information when completing the survey: age, race, gender, highest level of education, languages spoken other than English, location of their school (rural, urban, suburban), the state in which they teach, years of teaching experience, grades levels they have taught, grade levels in which they specifically provided science instruction, number of different languages spoken by ELs that they have taught, and percent of students eligible for free or reduced lunch at their school. Descriptive analyses were run to describe the final study sample (see ***Participants*** section above).

Experience Working with ELs

Participants reported how many ELs they had in their classroom in the last non-COVID-19 schoolyear. This variable was used as a predictor in the analysis for research question two. On average, teachers (n = 290) reported that they had six EL students in their classroom, with a median of four and standard deviation of 6.2. Having low experience working with ELs was defined as having one or less ELs in their classroom, having medium experience was defined as having between two and five ELs, and having high experience was defined as having six or more ELs. Based on these definitions, 23% of teachers (n = 68) were considered as having low experience, 39% (n = 112) were considered as having medium experience, and 38% (n = 110) were considered as having high experience.

Prior Training

Participants' prior training on instructional supports for ELs was measured by asking participants to report how much pre- and in-service training they received on "teaching students who are limited English proficient (LEP) or English-language learners (ELLs)." Phrasing of this question mirrors language used in the National Teacher and Principal Survey administered by the NCES (2020, p. 19). Pre-service training includes university-level course work and in-service training includes several activities, such as "online or web-based professional development," "workshops," and "conferences" (NCES, 2020, p. 24). Participants were asked to estimate how many hours of training, pre- and in-service, they have received in total. As part of the survey, participants were provided with guidance on how to estimate hours (e.g., one semester-long 3-credit university course counts as 45 hours). Based on the variation in responses reported by teachers in the current study, those with zero hours of training were considered as having no training, those who reported between one and 45 hours of training (i.e., the equivalent of one semester-long 3-credit course or less) were considered to have some training, and those with more than 45 hours were considered to have the most training. In the sample of the current study, 15% (n=42) reported no training, 44% (n=122) reported some training, and 41% (n=113) reported most training.

Role of the ESL Teacher

The role of the ESL teacher was assessed by asking participants to report how much consultative support they receive from the ESL teacher (if they had one) in the last non-COVID-19 school year. Teachers were asked to estimate the amount in hours. Consultative support was defined as the ESL teacher providing guidance on using instructional supports for ELs in the mainstream classroom. This consultation may occur via email, one-on-one meetings, or grade-

level meetings (Bell & Baecher, 2012). Based on the variation in responses reported by teachers in the current study, those with zero hours of consultation were considered as having no consultative support, those who reported between one and ten hours of consultation were considered as having some consultative support, and those who reported more than ten hours of consultation were considered as having the most consultative support. In the sample of the current study, 37% (n=104) reported receiving no consultative support, 48% (n=134) reported receiving some consultative support, and 15% (n=41) reported receiving the most consultative support.

Procedures

Survey Pilot Study

A pilot study was conducted to help with finalizing the survey in several ways. First, the pilot study was intended to facilitate collection of critical feedback about phrasing of questions and vignettes to ensure clarity. Second, participant feedback helped determine whether items such as demographics and teacher background information had appropriate answer options. Third, the pilot study offered an approximate length of time required to complete the survey. Lastly, data from the pilot survey indicated whether there would likely be sufficient variation in responses to the items on the URP-IR to conduct rigorous and meaningful analyses on the measure when used with a larger sample.

Participants for the pilot study included nine general education elementary teachers. For ease of recruitment, the pilot sample included teachers from any elementary grade level.

Participants were recruited through convenience sampling by reaching out to personal contacts of the author and recruiting via social media. As a token of appreciation for their time, participants each received a \$10.00 Visa gift card. Feedback was provided in writing and reviewed after

survey completion. Feedback was elicited after each major section of the survey (i.e., (1) consent and confidentiality statement, (2) a vignette description of a classroom teacher and the ELs in her classroom, (3) each instructional support vignette accompanied by URP-IR items (repeated four times for each support), and (4) demographics and other background information. Participants were asked to provide feedback by answering the following question(s): “Was the information/questions in this section of the survey clear?” and “If not, please note what was unclear and any suggestions you have to help improve the clarity of the survey.”

The average age of participants was 33.9 years (n=8), two-thirds (n=6) identified as female, and all participants identified as White (n=9). Teaching experience among the participants ranged from three to 12 years, with an average of 5.7 years. Out of five opportunities to provide suggestions for improving the survey, only two participants gave suggestions, once each. One suggestion related to the instructional support itself and how to improve it. No changes were made to the vignette based on this feedback, as it was not related to the clarity of the support itself. The other participant who indicated that something was unclear provided no further elaboration and did not respond to follow-up questions. In terms of time needed to complete the survey, participants in the pilot study ranged from eight to 28 minutes, after removing outliers. The average completion time was 15.5 minutes (n=7). Lastly, regarding variability in URP-IR items, pilot survey data indicated that sufficient variation to run meaningful analyses could be expected. Of the 116 items (i.e., 29 items per each of the four supports), 61% (n=71) had a standard deviation of over one and 39% of items (n=45) had a standard deviation between .5-1.

Main Study

Recruitment. Due to low response rates, recruitment of participants was completed in multiple phases and used different methods. First, a random list of 1500 teachers was retrieved from MDR. Teachers were included in the list if they were general education teachers of grades three or four and taught at public elementary schools in the United States. The contact list from MDR was selected such that 750 teachers came from high percentage EL districts and 750 teachers came from low percentage EL districts. This was intended to provide variability in the experience teachers have had working with ELs, a key variable of interest in the current study. High percentage EL districts were defined as 20% or more students classified as EL. This cutoff was chosen based on a 2017 report by the U.S. Department of Education (U.S. DoE), which used the same classification for districts.

Following low response rates from the initial 1500 MDR sample (5% started the survey and 2% completed the survey), several adaptations were made to increase sample size. First, restrictions on the grade levels of teachers were removed to include all elementary level teachers. Second, teachers were not required to have taught science. Lastly, recruitment was broadened beyond MDR recruitment. Using the database search of the National Center for Education Statistics (NCES), all K-5 elementary schools in the United States were identified. Of those, schools were randomly selected. If the school website publicly provided teacher emails, all general education teachers were emailed and invited to participate in the survey. Through this method, 11,020 teachers were contacted to participate. Briefly, the survey was shortened to only include one support vignette, as it was hypothesized that a shorter survey would increase response rates. The shorter survey was sent to 2,010 of the 11,020 teachers but yielded no improved response rates (4% started the survey and 2% completed it), and the original survey

was sent to the rest of the teachers. In addition to the NCES recruitment efforts, a second 1,500 sample was retrieved from MDR. Thus, 13,520 teachers were invited with the broadened participation criteria (i.e., any elementary grade level and science instruction experience was not required). Throughout these recruitment efforts, response rates remained consistent (5% started the survey and 2% completed the survey). Overall, 738 participants started the survey and 295 completed it fully.

Across the various recruitment methods and samples, recruitment occurred via email. Slight adjustments were made after the first MDR sample. The first MDR sample received a prenotification email informing them of the survey, its purpose, and its procedures. For a copy of this prenotification email, see Appendix D. Participants from the second MRD sample and the NCES sample did not receive this prenotification email in an effort to reduce recruitment time. Participants who did not complete the survey following the initial invitation were sent up to three reminder emails weekly. All participants who responded to the survey could enter into a drawing to win one of five \$100.00 Visa gift cards. Additionally, all participants who completed the survey received a resource guide (see Appendix E) for working with ELs in the general education setting compiled by the principal investigator.

Survey Administration. Participants completed the survey online via Qualtrics. Before starting the survey, participants were informed that their participation was voluntary and that their data would only be shared in deidentified form. At the end of the survey, participants indicated whether they would like to be entered into the drawing for one of five \$100 Visa gift cards and whether they would like to receive the EL resource guide (see Appendix E). Participants who wished to be entered or receive the resource guide were redirected to a second

survey, where they could provide their email address. This ensured that the email addresses of participants were not connected to their survey responses.

Social Validity Interviews. At the end of the survey, participants were asked whether they were willing to be contacted for a 30-minute follow-up interview. The teachers who consented were divided into three groups based on their experience working with ELs: low experience (one or less ELs in their classroom during their last non-COVID-19 school year), medium experience (two to five ELs), and high experience (six or more ELs). Teachers were randomly selected and invited to participate. If those who were randomly selected did not reply or chose not to participate, another teacher was randomly selected to take their place. In total, nine teachers participated in the interviews, three with low experience, three with medium experience, and three with high experience. Each teacher who participated received a \$10.00 Visa gift card. All interviews were held via video conferencing technology and were recorded.

Data Analysis

Research Question 1

The first research question was: Is the six-factor structure of the revised Usage Rating Profile-Intervention (URP-IR) a good model fit when using this measure for teacher ratings of empirically-supported instructional supports for English Learners during science instruction? To address RQ1, confirmatory factor analyses (CFA) were conducted for each instructional support.

To test whether the six-factor model identified by Briesch et al. (2013) was a good fit, CFA analyses were conducted using SPSS Statistics software with the AMOS add-on module. The model for each support was evaluated for goodness-of-fit using three measures: chi-square test, root mean square error of approximation (RMSEA), and comparative fit index (CFI). The model was considered a good fit for the data if the chi-square test was non-significant, the

RMSEA value was below .10 (Byrne, 2009), and the CFI was above .95 (Brown, 2006). If these measures indicated a poor fit, the model was further evaluated for possible modifications using modification indices (MI) provided by AMOS.

Research Questions 2 and 3

The second research questions was: To what extent do the following teacher background experiences predict URP-IR ratings for the targeted EL supports: (1) Experience working with ELs; (2) Teacher training on teaching ELs; (3) The role of the ESL teacher. The third research question was: What are elementary mainstream teachers' URP-IR ratings for four empirically-supported instructional supports for English Learners during science instruction? Are there significant differences in URP-IR scores across these instructional supports?

Analyses for these research questions was completed using analysis of variance (ANOVA). The total URP-IR score was used as the dependent variable given that results of the EFA did not support use of individual subscales. For research question two, regarding predictors of URP-IR scores based on teacher experiences, three ANOVAs were conducted for each support, one for each predictor (for a total of 12 analyses). All predictors are categorical variables and were the independent variable, with each category representing a level. These ANOVA models were a between-groups design, as each participant had only one value for each predictor. Due to the between-groups nature of the analyses, caution in interpreting the results is warranted due to potential confounding factors.

For research question three, regarding differences in teacher ratings of the four different instructional supports, the URP-IR total score for each support was the dependent variables. The instructional supports were the independent variable with each support as a level. This model was a repeated measures design, as all participants have data for each instructional support.

Reverse Coding

Items for which a higher rating indicates more negative opinions were reverse coded when computing the URP-IR total score. The URP-IR total score was used in analyses for research questions 2 and 3. The items that were reverse coded were: 6, 11, 13, 19, 24, 25, 28, 29. These items were not reverse coded in the factor analyses for research question 1.

Research Question 4

Research question four was: What are teachers' opinions about the social significance of the goals and social importance of the effects of the selected EL instructional supports?

Qualitative data gained from the follow-up interviews were transcribed and analyzed using thematic analysis (Braun & Clarke, 2006). Two independent coders first familiarized themselves with the data through transcribing interviews and reading transcripts closely. Then, each coder developed initial codes from each transcript. Each coder reviewed the codes derived from all transcripts to develop larger themes and subsequently reviewed those to ensure that the themes related meaningfully to the research question. Each coder then gave each theme a concise title. The coders then compared their findings and met to discuss and finalize the themes.

Sample Size and Power

Research Question 1

The final sample size for the CFA was $N=376$. Given that the URP-IR has 29 items and six factors, the variables-to-factors ratio is 4.8. According to Mundfrom et al. (2005), a sample of 250 is sufficient to achieve good-level criterion (.92) at a variables-to-factors ratio of five, regardless of communality.

Research Question 2

Power analyses were conducted prior to the study, under the assumption that use of individual subscale scores would be possible and with a conservative estimate of four categories for each predictor. Under this assumption, research question 2 involved three MANOVA computations to assess the relationship between the predictor variables and URP-IR scores. Each predictor was set at four categories for a more conservative estimate. To account for the multiple analyses, the alpha level was adjusted to .01 using the Bonferroni correction. Power analyses indicated that N=32 would suffice to detect large effects, N=68 would suffice to detect medium effects, and N=384 would suffice to detect small effects. This assumed an alpha level of .01 and 95 percent power. Given this, the current sample was sufficient to detect large and medium effects and just below the number needed to detect small effects.

Research Question 3

As noted previously, this power analysis was run under assumptions set prior to data collection and analysis. Given these assumptions, research question 3 involved one MANOVA computation to compare the four selected instructional supports. Power analyses conducted indicated that N=28 would suffice to detect large effects, N=60 would suffice to detect medium effects, and N=336 would suffice to detect small effects. This assumed an alpha level of .05 and 95 percent power. Given this, the current sample was sufficient to detect large, medium, and small effects.

Table 3. *Overview of Research Questions*

Research Question	Variables/Data Analysis	Hypotheses
1. Is the six-factor structure of the revised Usage Rating Profile-Intervention (URP-IR) a good model fit when using this measure for teacher ratings of empirically-supported instructional supports for English Learners during science instruction?	<u>Variables:</u> URP-IR items <u>Analyses:</u> Confirmatory Factor Analysis (CFA)	The six-factor structure will be an adequate fit
2. To what extent do the following teacher background experiences predict URP-IR ratings for the targeted EL supports? 1. Experience working with ELs 2. Teacher training on teaching ELs 3. The role of the ESL teacher	<u>Dependent Variables:</u> URP-IR total score <u>Independent Variable:</u> predictor (experience, training, ESL role) <u>Independent Variable Levels:</u> (<i>varies by predictor</i>): predictor categories <u>Analyses:</u> between-subjects ANOVAs	Each of the following will positively correlate with more favorable URP-IR scores: • Experience with ELs • EL-related training • Access and consultation with ESL teachers
3. What are elementary mainstream teachers' URP-IR ratings for four empirically-supported instructional supports for English Learners during science instruction? Are there significant differences in URP-IR scores across these instructional supports?	<u>Dependent Variables:</u> URP-IR total score <u>Independent Variable:</u> instructional support <u>Independent Variable Levels (4):</u> visual aids, vocabulary supports, incorporation of L1, alternate response format <u>Analyses:</u> within-subjects ANOVA	There will be significant differences in URP-IR ratings between instructional supports
4. What are teachers' opinions about the social significance of the goals and social importance of the effects of the selected EL instructional supports?	<u>Measure:</u> Semi-Structured Interview for Social Validation <u>Analyses:</u> Thematic analysis	Findings will contextualize the quantitative analyses by providing information on the other two levels of social validity

CHAPTER IV: RESULTS

Research Question 1

Data Description

The final total number of respondents was 379. Table 4 below displays the mean and standard deviation for each item for each of the four supports. The possible range for each item was one (“strongly disagree”) through six (“strongly agree”). In addition, skew and kurtosis was assessed for each item. Data were slightly skewed but did not violate assumptions of normality, except for three items for support 1 (visual supports). The skewed items were items one, three, and ten, and were all skewed negatively. The item/support combination with the most favorable average rating was item 16 (“Administrator would be supportive of this support”) for support 1 (visual aids). The item/support combination with the least favorable average rating was item 28 (“Would need additional resources to carry out”) for support 3 (inclusion of L1).

Table 4. *Item-level Descriptives*

		Mean (sd)			
Item #	Stem	Support 1	Support 2	Support 3	Support 4
Acceptability subscale					
1	Support is a good way to handle language barrier	5.31 (1.04)	4.29 (1.76)	4.94 (1.03)	4.79 (1.23)
2	I would implement this support with a good deal of enthusiasm	5.28 (.82)	4.85 (1.22)	4.54 (1.24)	4.77 (1.27)
3	This support would not be disruptive to other students	5.15 (1.19)	4.98 (1.25)	4.29 (1.58)	4.74 (1.26)
4	Procedures easily fit in with my current practices	5.35 (.67)	4.63 (1.53)	4.23 (1.47)	5.03 (.95)
6*	Would not be interested in implementing this support	2.48 (1.68)	2.58 (1.49)	2.77 (1.49)	2.33 (1.50)
8	I would have positive attitudes about this support	5.20 (.84)	4.80 (1.19)	4.56 (1.31)	4.95 (1.14)
10	Fair way to handle language barrier	5.32 (.96)	4.51 (1.46)	4.91 (1.05)	4.76 (1.31)

Table 4 (cont'd)

Understanding subscale	12	Effective choice for addressing a variety of language barriers	4.90 (1.13)	4.50 (1.48)	4.53 (1.21)	5.03 (1.06)
	17	I would be committed to carrying out this support	5.37 (.75)	4.43 (1.60)	4.14 (1.41)	4.86 (1.12)
	18	I understand the procedures	5.08 (.86)	4.49 (1.75)	4.71 (1.07)	5.05 (1.19)
	22	I understand how to use this support	5.11 (.86)	4.57 (1.73)	4.33 (1.37)	5.10 (.93)
	27	I am knowledgeable about the support procedures	5.29 (.78)	4.55 (1.59)	4.47 (1.41)	4.86 (.93)
Family-School subscale	13*	Parental collaboration is required	2.55 (1.61)	2.84 (1.78)	2.86 (1.52)	2.56 (1.38)
	25*	Positive home-school relationship is needed	2.91 (1.45)	2.63 (1.48)	3.05 (1.58)	2.97 (1.49)
	29*	Regular home-school communication is needed	2.98 (1.49)	2.95 (1.63)	2.98 (1.52)	2.98 (1.50)
Feasibility subscale	5	Total time required would be manageable	4.82 (1.07)	4.32 (1.60)	4.18 (1.28)	5.07 (.88)
	9	Material resources are reasonable	5.06 (.93)	4.88 (1.27)	4.17 (1.31)	4.84 (1.26)
	11*	Too complex to carry out	2.00 (1.17)	2.01 (1.25)	2.66 (1.44)	2.53 (1.71)
	14	Would be able to allocate my time	4.66 (1.18)	4.49 (1.36)	4.01 (1.28)	5.21 (.94)
	15	Time required for record keeping would be reasonable	4.49 (1.41)	4.58 (1.19)	3.98 (1.22)	4.74 (1.37)
	20	Preparation of materials would be minimal	4.21 (1.39)	4.45 (1.39)	3.70 (1.47)	4.43 (1.43)
System Climate subscale	7	Support would be consistent with the mission of my school	5.02 (1.09)	4.69 (1.34)	4.68 (1.35)	4.71 (1.57)
	16	Administrator would be supportive of this support	5.44 (.69)	5.10 (.96)	4.69 (1.30)	5.16 (1.01)

Table 4 (cont'd)

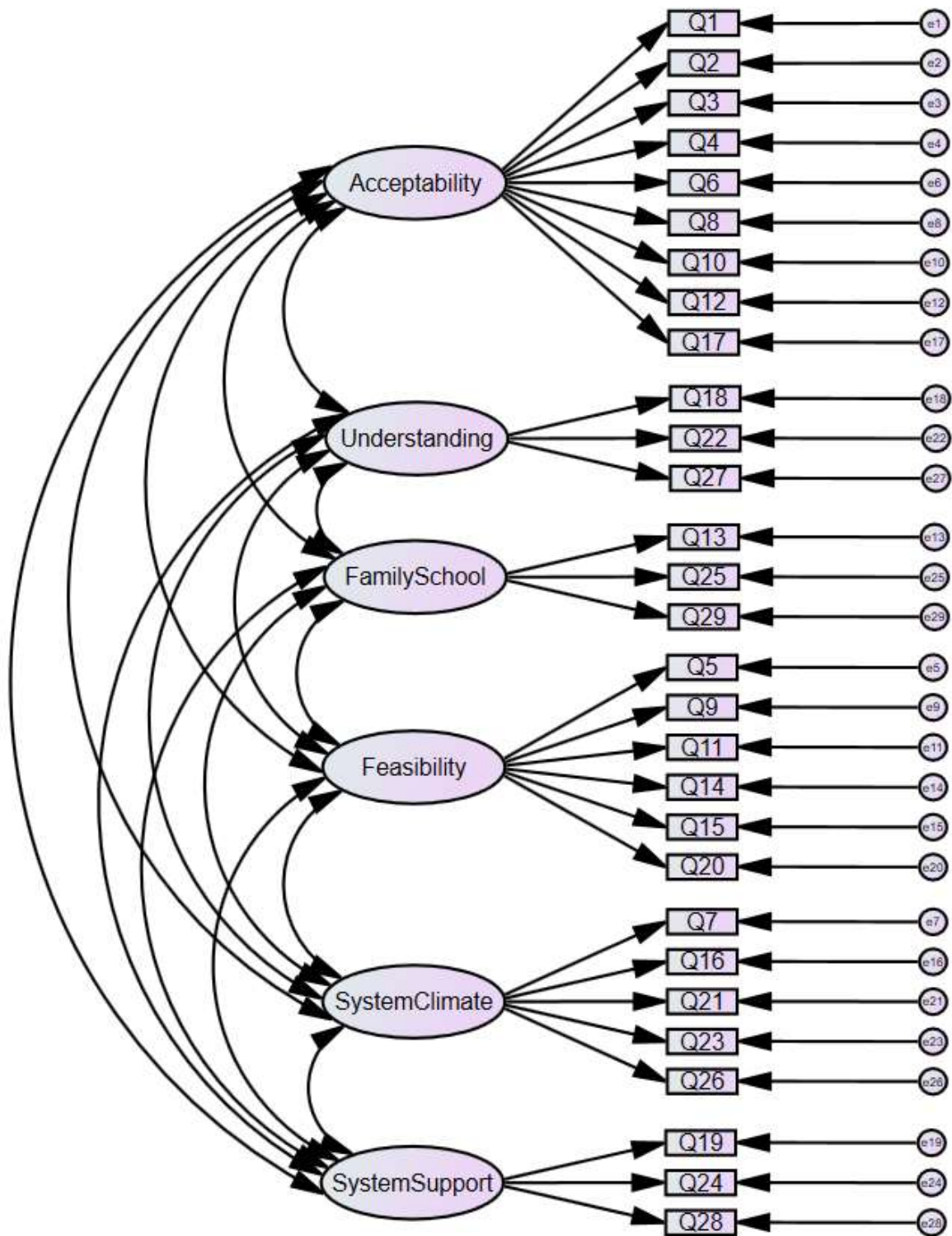
System Support subscale	21	Procedures are consistent with the way things are done in my system	4.77 (1.28)	4.68 (1.20)	3.92 (1.37)	4.57 (1.46)
	23	Work environment is conducive to the implementation	4.81 (1.30)	4.77 (1.32)	4.27 (1.43)	4.77 (1.25)
	26	Support is well matched to what is expected in my job	5.11 (.93)	4.79 (1.17)	4.38 (1.33)	4.71 (1.43)
	19*	Would require additional professional development	2.85 (1.62)	2.05 (1.26)	3.42 (1.56)	2.88 (1.71)
	24*	Would need consultative support	2.74 (1.79)	2.52 (1.59)	3.16 (1.66)	2.32 (1.37)
	28*	Would need additional resources to carry out	3.29 (1.60)	2.53 (1.42)	3.79 (1.68)	2.79 (1.52)

*lower scores on these items indicate more positive views towards supports. These items were reverse coded when calculating the total scale score. For factor analyses, items were not reverse coded.

Model Specification

The model tested was the six-factor model identified by Briesch et al. (2013). The factors tested were: Acceptability, Understanding, Family-School Collaboration, Feasibility, System Climate, and System Support. This model was tested for each individual support first, with intent to examine model fit across all four supports combined if individual fit was adequate. Figure 1 below shows the model diagram. Model fit was analyzed using SPSS Amos. Normality of data was assessed, and findings indicated normality was not violated. Maximum likelihood estimation was utilized to obtain model estimates. Covariances were put in place between each of the six factors, following the model outlined by Briesch et al. (2013).

Figure 1. Path diagram of the six-factor model identified by Briesch et al. (2013)



Model Fit

This section will first provide information on the model fit for each of the four supports. Because poor model fit was identified for each support, no multi-group model combining all four supports was conducted. The following section briefly describes modification procedures used in an effort to identify models with improved model fit. Ultimately, modification procedures lead to model fit improvement, although fit indices still were not adequate to signify a good model fit.

Support 1. Model fit for support 1, visual aids, was poor as demonstrated by the fit indices displayed in Table 5 below. The chi-square test was significant ($p < .01$, $\chi^2(362) = 2501.14$), the CFI value was well below .95 (CFI = .70), and the RMSEA was above .10 (RMSEA = .13). Regression weights of items onto respective factors was adequate, with each item demonstrating statistical significance. Table 6 below shows the regression weight estimates, intercept estimates, variance estimates, and covariance estimates for support 1. Cronbach's alpha for the six factors for support 1 were acceptable and ranged from .74 to .85. Internal reliability values and inter-item correlations for each factor are displayed in Table 7 below.

Table 5. Fit Indices for Support 1

Chi-square	CFI	RMSEA	AIC
2501.14	.70	.13	2705.14

Table 6. Estimates of Regression Weights, Intercepts, Variance, and Covariances for Support 1

Item	Regression weight estimate	Intercept estimate	Variance estimate
Acceptability			.36
1	1.00	5.31	.73
2	1.04	5.30	.28
3	.62	5.15	1.26
4	.83	5.35	.20
6	-1.36	2.48	2.16
8	1.21	5.20	.18
10	.70	5.32	.75
12	1.51	4.90	.45
17	.54	5.37	.46
Understanding			.62

Table 6 (cont'd)

	18	1.00	5.08	.12
	22	.94	5.11	.19
	27	.38	5.29	.51
Family-School				1.22
	13	1.00	2.55	1.36
	25	1.12	2.91	.57
	29	1.19	2.98	.46
Feasibility				.53
	5	1.00	4.82	.60
	9	.54	5.06	.71
	11	-.42	2.00	1.27
	14	1.26	4.66	.54
	15	1.51	4.49	.77
	20	.81	4.21	1.58
System Climate				.93
	7	1.00	5.02	.26
	16	.40	5.44	.32
	21	1.00	4.77	.70
	23	1.20	4.81	.35
	26	.56	5.11	.57
System Support				1.44
	19	1.00	2.85	1.19
	24	1.30	2.75	.77
	28	.91	3.29	1.35
Covariance estimate				
Acceptability – Understanding			.39*	
Acceptability – Family-School			-.14	
Acceptability – Feasibility			.30*	
Acceptability – System Climate			.50*	
Acceptability – System Support			-.32*	
Understanding – Family-School			-.26*	
Understanding – Feasibility			.39*	
Understanding – System Climate			.63*	
Understanding – System Support			-.59*	
Family-School – Feasibility			-.04	
Family-School – System Climate			-.33*	
Family-School – System Support			.87	
Feasibility – System Climate			.49*	
Feasibility – System Support			-.38*	
System Climate – System Support			-.77*	

*statistically significant at $p < .001$

Table 7. Internal Reliability and Inter-Item Correlations by Factor for Support 1

Subscale	Cronbach's alpha	Inter-item correlations									
Acceptability	.81	Item	1	2	3	4	6	8	10	12	17
		1	1.00								
		2	.59**	1.00							
		3	.26**	.28**	1.00						
		4	.42**	.64**	.28**	1.00					
		6	-.22**	-.33**	-.10	-.30**	1.00				

Table 7 (cont'd)

		8	.50**	.63**	.24**	.61**	-.46**	1.00				
		10	.52**	.54**	.30**	.40**	-.01	.29**	1.00			
		12	.42**	.58**	.17**	.53**	-.45**	.71**	.37**	1.00		
		17	.33**	.53**	.24**	.51**	-.05	.30**	.55**	.20**	1.00	
Understanding	.77	Item	18	22	27							
		18	1.00									
		22	.78**	1.00								
		27	.33**	.46**	1.00							
Family-School	.84	Item	13	25	29							
		13	1.00									
		25	.60**	1.00								
		29	.58**	.77**	1.00							
Feasibility	.74	Item	5	9	11	14	15	20				
		5	1.00									
		9	.59**	1.00								
		11	-.24**	-.37**	1.00							
		14	.50**	.24**	-.15**	1.00						
		15	.48**	.21**	-.17**	.64**	1.00					
		20	.43**	.31**	-.04	.39**	.30**	1.00				
System Climate	.85	Item	7	16	21	23	26					
		7	1.00									
		16	.47**	1.00								
		21	.67**	.44**	1.00							
		23	.81**	.42**	.66**	1.00						
		26	.51**	.45**	.49**	.47**	1.00					
System Support	.81	Item	19	24	28							
		19	1.00									
		24	.64**	1.00								
		28	.55**	.59**	1.00							

Support 2. Model fit for support 2, vocabulary supports, was poor as demonstrated by the fit indices displayed in Table 8 below. The chi-square test was significant ($p < .01$, χ^2 (362) = 2371.33), the CFI value was well below .95 (CFI = .77), and the RMSEA was above .10 (RMSEA = .12). Regression weights of items onto respective factors was adequate, with each item demonstrating statistical significance. Table 9 below shows the regression weight estimates, intercept estimates, variance estimates, and covariance estimates for support 2. Cronbach's alpha for the six factors for support 2 were acceptable and ranged from .75 to .96. Internal reliability values and inter-item correlations for each factor are displayed in Table 10 below.

Table 8. *Fit Indices for Support 2*

Chi-square	CFI	RMSEA	AIC
2371.33	.77	.12	2575.33

Table 9. *Estimates of Regression Weights, Intercepts, Variance, and Covariances for Support 2*

Item	Regression weight estimate	Intercept estimate	Variance estimate
------	----------------------------	--------------------	-------------------

Table 9 (cont'd)

Acceptability				1.78
	1	1.00	4.29	1.30
	2	.51	4.85	1.02
	3	.40	4.98	1.28
	4	.91	4.63	.85
	6	-.43	2.58	1.88
	8	.71	4.80	.51
	10	.81	4.51	.98
	12	.63	4.50	1.47
	17	1.05	4.43	.61
Understanding				2.84
	18	1.00	4.49	.20
	22	.95	4.57	.41
	27	.86	4.55	.40
Family-School				1.39
	13	1.00	2.84	1.75
	25	.86	2.63	1.14
	29	1.22	3.00	.58
Feasibility				2.37
	5	1.00	4.32	.18
	9	.50	4.88	1.03
	11	-.12	2.01	1.54
	14	.76	4.49	.46
	15	.60	4.58	.56
	20	.38	4.45	1.59
System Climate				1.34
	7	1.00	4.69	.46
	16	.55	5.10	.50
	21	.77	4.68	.64
	23	.85	4.77	.79
	26	.73	4.79	.66
System Support				.63
	19	1.00	2.05	.95
	24	1.62	2.52	.87
	28	1.13	2.53	1.19
Covariance estimate				
Acceptability – Understanding				1.70*
Acceptability – Family-School				-.37*
Acceptability – Feasibility				1.59*
Acceptability – System Climate				1.52*
Acceptability – System Support				-.27*
Understanding – Family-School				-.91*
Understanding – Feasibility				2.30*
Understanding – System Climate				1.41*
Understanding – System Support				-.64*
Family-School – Feasibility				-.58*
Family-School – System Climate				-.28
Family-School – System Support				.78*
Feasibility – System Climate				1.35*

Table 9 (cont'd)

Feasibility – System Support	-.53*
System Climate – System Support	-.25*

*statistically significant at $p < .001$

Table 10. *Internal Reliability and Inter-Item Correlations by Factor for Support 2*

Subscale	Cronbach's alpha	Inter-item correlations										
Acceptability	.88	Item	1	2	3	4	6	8	10	12	17	
		1	1.00									
		2	.60**	1.00								
		3	.31**	.17**	1.00							
		4	.56**	.31**	.37**	1.00						
		6	-.28**	-.27**	-.15**	-.27**	1.00					
		8	.66**	.58**	.30**	.57**	-.39**	1.00				
		10	.54**	.57**	.33**	.60**	-.32**	.29**	1.00			
		12	.59**	.59**	.18**	.25**	-.29**	.71**	.37**	1.00		
17	.63**	.44**	.36**	.73**	-.33**	.30**	.55**	.20**	1.00			
Understanding	.96	Item	18	22	27							
		18	1.00									
		22	.90**	1.00								
		27	.88**	.86**	1.00							
Family-School	.77	Item	13	25	29							
		13	1.00									
		25	.38**	1.00								
		29	.56**	.66**	1.00							
Feasibility	.80	Item	5	9	11	14	15	20				
		5	1.00									
		9	.61**	1.00								
		11	-.12*	-.09	1.00							
		14	.83**	.46**	-.15**	1.00						
		15	.74**	.42**	-.12*	.76**	1.00					
		20	.37**	.12*	-.16**	.49**	.49**	1.00				
System Climate	.86	Item	7	16	21	23	26					
		7	1.00									
		16	.58**	1.00								
		21	.65**	.57**	1.00							
		23	.63**	.50**	.53**	1.00						
		26	.61**	.46**	.54**	.55**	1.00					
System Support	.75	Item	19	24	28							
		19	1.00									
		24	.47**	1.00								
		28	.60**	.47**	1.00							

Support 3. Model fit for support 3, inclusion of L1, was poor as demonstrated by the fit indices displayed in Table 11 below. The chi-square test was significant ($p < .01$, $\chi^2(362) = 1713.55$), the CFI value was well below .95 (CFI = .75), and the RMSEA was at .10. Regression weights of items onto respective factors was adequate, with each item demonstrating statistical significance. Table 12 below shows the regression weight estimates, intercept estimates, variance estimates, and covariance estimates for support 3. Cronbach's alpha for the six factors for

support 3 were acceptable and ranged from .70 to .84. Internal reliability values and inter-item correlations for each factor are displayed in Table 13 below.

Table 11. *Fit Indices for Support 3*

Chi-square	CFI	RMSEA	AIC
1713.55	.75	.10	1917.55

Table 12. *Estimates of Regression Weights, Intercepts, Variance, and Covariances for Support 3*

Item		Regression weight estimate	Intercept estimate	Variance estimate
Acceptability				.29
	1	1.00	4.94	.78
	2	1.75	4.54	.63
	3	1.52	4.29	1.83
	4	1.53	4.23	1.47
	6	-1.36	2.77	1.69
	8	1.73	4.56	.84
	10	.98	4.91	.83
	12	1.43	4.54	.86
Understanding	17	2.13	4.14	.68
				.68
	18	1.00	4.71	.46
	22	1.36	4.33	.61
	27	.92	4.47	1.41
Family-School				.58
	13	1.00	2.86	1.72
	25	1.85	3.05	.53
	29	1.79	2.98	.45
Feasibility				.90
	5	1.00	4.18	.75
	9	.83	4.17	1.08
	11	-.73	2.67	1.60
	14	1.01	4.01	.71
	15	1.05	3.98	.50
	20	1.06	3.70	1.14
System Climate				.70
	7	1.00	4.68	1.12
	16	.95	4.69	1.05
	21	1.24	3.92	.81
	23	1.32	4.27	.82
	26	1.04	4.38	1.00
System Support				1.04
	19	1.00	3.42	1.40
	24	1.16	3.16	1.35
	28	1.04	3.79	1.70
Covariance estimate				

Table 12 (cont'd)

Acceptability – Understanding	.27*
Acceptability – Family-School	.04
Acceptability – Feasibility	.36*
Acceptability – System Climate	.39*
Acceptability – System Support	-.06
Understanding – Family-School	-.06
Understanding – Feasibility	.47*
Understanding – System Climate	.35*
Understanding – System Support	-.44*
Family-School – Feasibility	-.01
Family-School – System Climate	.06
Family-School – System Support	.34*
Feasibility – System Climate	.60*
Feasibility – System Support	-.48*
System Climate – System Support	-.20

*statistically significant at $p < .001$ Table 13. *Internal Reliability and Inter-Item Correlations by Factor for Support 3*

Subscale	Cronbach's alpha	Inter-item correlations									
Acceptability	.84	Item	1	2	3	4	6	8	10	12	17
		1	1.00								
		2	.42**	1.00							
		3	.08	.43**	1.00						
		4	.34**	.45**	.28**	1.00					
		6	-.23**	-.44**	-.36**	-.24**	1.00				
		8	.40**	.56**	.34**	.39**	-.32**	1.00			
		10	.45**	.37**	.19**	.22**	-.20**	.37**	1.00		
		12	.41**	.52**	.34**	.25**	-.32**	.37**	.55**	1.00	
		17	.43**	.61**	.39**	.47**	-.40**	.62**	.33**	.48**	1.00
Understanding	.73	Item	18	22	27						
		18	1.00								
		22	.66**	1.00							
		27	.41**	.40**	1.00						
Family-School	.80	Item	13	25	29						
		13	1.00								
		25	.45**	1.00							
		29	.45**	.80**	1.00						
Feasibility	.83	Item	5	9	11	14	15	20			
		5	1.00								
		9	.38**	1.00							
		11	-.48**	-.20**	1.00						
		14	.52**	.49**	-.27**	1.00					
		15	.60**	.47**	-.34**	.65**	1.00				
		20	.62**	.39**	-.39**	.46**	.56**	1.00			
System Climate	.81	Item	7	16	21	23	26				
		7	1.00								
		16	.54**	1.00							
		21	.45**	.41**	1.00						
		23	.46**	.48**	.63**	1.00					
		26	.33**	.29**	.52**	.51**	1.00				
System Support	.70	Item	19	24	28						
		19	1.00								
		24	.46**	1.00							
		28	.43**	.44**	1.00						

Support 4. Model fit for support 4, alternate response format, was poor as demonstrated by the fit indices displayed in Table 14 below. The chi-square test was significant ($p < .01$, $\chi^2(362) = 2423.90$), the CFI value was well below .95 (CFI = .76), and the RMSEA was above .10 (RMSEA = .12). Regression weights of items onto respective factors was adequate, with each item demonstrating statistical significance. Table 15 below shows the regression weight estimates, intercept estimates, variance estimates, and covariance estimates for support 4. Cronbach's alpha for the six factors for support 4 were acceptable and ranged from .77 to .87. Internal reliability values and inter-item correlations for each factor are displayed in Table 16 below.

Table 14. *Fit Indices for Support 4*

Chi-square	CFI	RMSEA	AIC
2423.90	.76	.12	2627.90

Table 15. *Estimates of Regression Weights, Intercepts, Variance, and Covariances for Support 4*

Item		Regression weight estimate	Intercept estimate	Variance estimate
Acceptability				1.22
	1	1.00	4.79	.30
	2	1.01	4.77	.37
	3	.43	4.75	1.35
	4	.60	5.03	.47
	6	-.38	2.33	2.06
	8	.87	4.96	.36
	10	.96	4.76	.59
	12	.41	5.03	.91
	17	.87	4.86	.33
Understanding				1.03
	18	1.00	5.05	.39
	22	.82	5.10	.17
	27	.58	4.86	.52
Family-School				1.09
	13	1.00	2.56	.80
	25	1.16	2.97	.74
	29	1.13	2.98	.83
Feasibility				.29
	5	1.00	5.07	.49
	9	1.81	4.84	.65
	11	-2.44	2.53	1.20

Table 15 (cont'd)

	14	.65	5.21	.77
	15	1.55	4.74	1.19
	20	2.14	4.43	.71
System Climate				1.51
	7	1.00	4.71	.93
	16	.29	5.16	.90
	21	1.00	4.57	.63
	23	.93	4.77	.23
	26	.94	4.71	.68
System Support				1.40
	19	1.00	2.88	1.51
	24	.78	2.32	1.02
	28	1.05	2.79	.78
Covariance estimate				
Acceptability – Understanding			.73*	
Acceptability – Family-School			-.28*	
Acceptability – Feasibility			.51*	
Acceptability – System Climate			1.19*	
Acceptability – System Support			-.47*	
Understanding – Family-School			-.34*	
Understanding – Feasibility			.44*	
Understanding – System Climate			1.08*	
Understanding – System Support			-.68*	
Family-School – Feasibility			-.24*	
Family-School – System Climate			-.35*	
Family-School – System Support			.87*	
Feasibility – System Climate			.60*	
Feasibility – System Support			-.42*	
System Climate – System Support			-.64*	

*statistically significant at $p < .001$ Table 16. *Internal Reliability and Inter-Item Correlations by Factor for Support 4*

Subscale	Cronbach's alpha	Inter-item correlations										
Acceptability	.87	Item	1	2	3	4	6	8	10	12	17	
		1	1.00									
		2	.83**	1.00								
		3	.31**	.32**	1.00							
		4	.56**	.60**	.44**	1.00						
		6	-.22**	-.21**	-.03	.20**	1.00					
		8	.80**	.73**	.34**	.61**	-.21**	1.00				
		10	.78**	.71**	.30**	.49**	-.22**	.73**	1.00			
		12	.41**	.38**	.08	.34**	-.27**	.33**	.23**	1.00		
17	.73**	.74**	.27*	.64**	-.31**	.72**	.65**	.46**	1.00			
Understanding	.81	Item	18	22	27							
		18	1.00									
		22	.79**	1.00								
		27	.48**	.52**	1.00							
Family-School	.83	Item	13	25	29							
		13	1.00									
		25	.61**	1.00								
		29	.58**	.67**	1.00							
Feasibility	.82	Item	5	9	11	14	15	20				
		5	1.00									
		9	.45**	1.00								

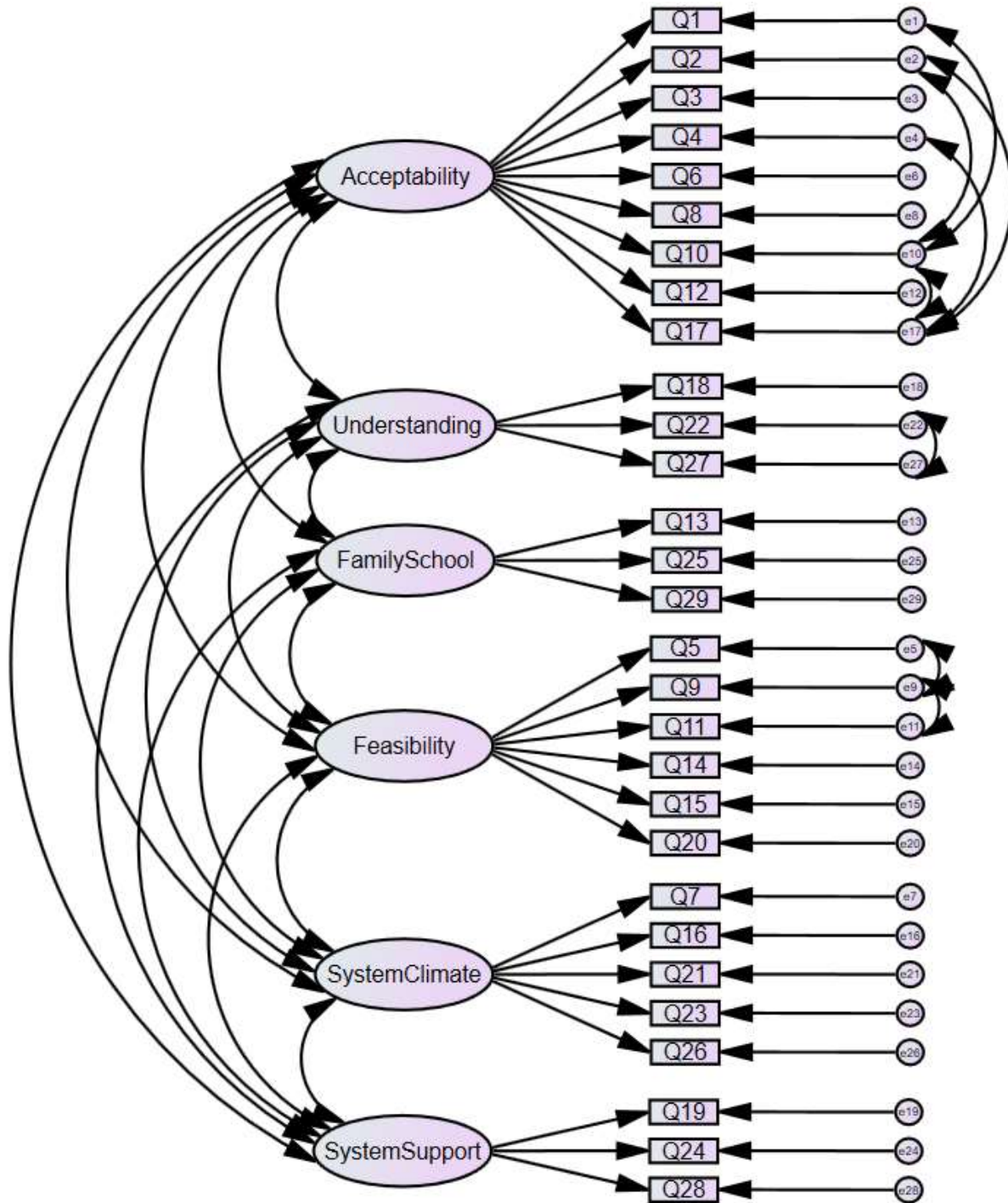
Table 16 (cont'd)

			11	-.33**	-.57**	1.00		
			14	.60**	.26**	-.16**	1.00	
			15	.59**	.40**	-.39**	.38**	1.00
			20	.56**	.64**	-.60**	.35**	.53** 1.00
System Climate	.86	Item	7		16	21	23	26
			7	1.00				
			16	.16**	1.00			
			21	.71**	.17**	1.00		
			23	.71**	.36**	.77**	1.00	
			26	.65**	.32**	.70**	.75**	1.00
System Support	.77	Item	19		24	28		
			19	1.00				
			24	.46**	1.00			
			28	.56**	.56**	1.00		

Model Modifications and Fit

Support 1. Modification indices provided by Amos were evaluated for impact on overall model fit and theoretical applicability. For support 1, the modification indices that were applied were putting covariances in place between the residuals of following item pairs: 1 and 10, 2 and 10, 2 and 17, 4 and 17, 10 and 17, 5 and 9, and 9 and 11. All of these items, with the exception of items 5, 9, and 11 are part of the Acceptability subscale. Items 5, 9, and 11 are part of the Feasibility subscale. Figure 2 below shows the model with modifications added.

Figure 2. Path diagram of the six-factor model with modifications for support 1



Despite these modifications, model fit for support 1, visual aids, remained poor as demonstrated by the fit indices displayed in Table 17 below. The chi-square test was significant ($p < .01$, $\chi^2(362) = 2049.22$), but yielded a lower value at 2049.22 than the value of 2501.14

without the modifications. The CFI value was well below .95 (CFI = .76) but was higher than the value of .70 without modifications. The RMSEA was slightly above .10 (RMSEA = .11) but was lower than the value of .13 without modifications. The AIC value also decreased with modifications from 2705.14 to 2269.22. Regression weights of items onto respective factors were adequate, with each item demonstrating statistical significance. Table 18 below shows the regression weight estimates, intercept estimates, variance estimates, and covariance estimates for support 1 after modifications.

Table 17. *Fit Indices after Modifications for Support 1*

Chi-square	CFI	RMSEA	AIC
2049.22	.76	.11	2269.22

Table 18. *Estimates of Regression Weights, Intercepts, Variance, and Covariances for Support 1 after Modifications*

Item	Regression weight estimate	Intercept estimate	Variance estimate
Acceptability			.32
1	1.00	5.31	.77
2	1.04	5.28	.33
3	.60	5.15	1.29
4	.83	5.35	.23
6	-1.55	2.48	2.05
8	1.30	5.20	.16
10	.58	5.32	.75
12	1.64	4.90	.42
17	.43	5.37	.47
Understanding			.62
18	1.00	5.08	.12
22	.93	5.11	.20
27	.32	5.29	.54
Family-School			1.22
13	1.00	2.55	1.35
25	1.12	2.91	.56
29	1.19	2.98	.47
Feasibility			.43
5	1.00	4.82	.70
9	.42	5.06	.76
11	-.39	2.00	1.29
14	1.44	4.66	.50
15	1.76	4.49	.66
20	.83	4.21	1.63
System Climate			.94

Table 18 (cont'd)

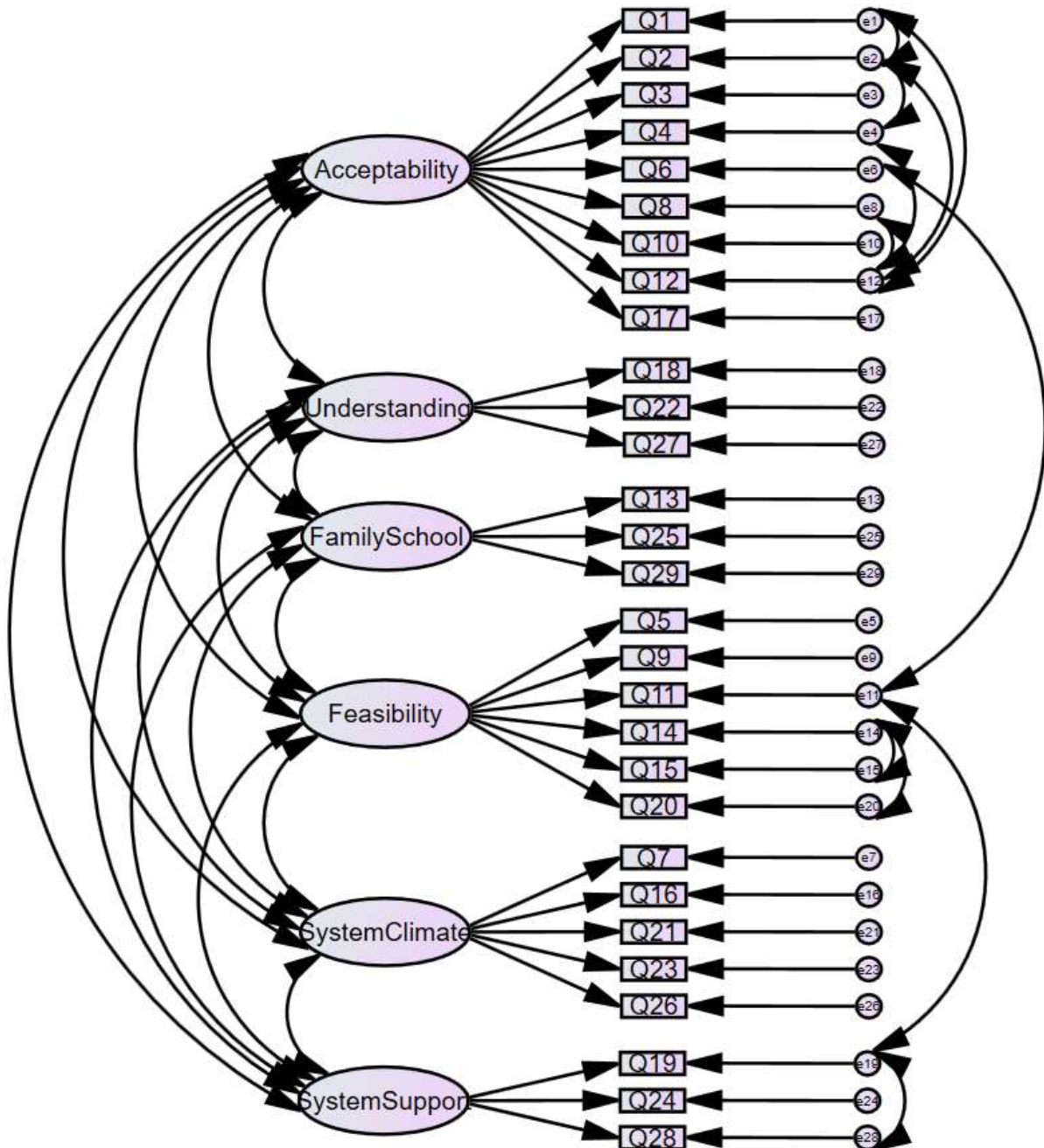
	7	1.00	5.02	.25
	16	.39	5.44	.33
	21	.99	4.77	.70
	23	1.20	4.81	.35
	26	.56	5.11	.58
System Support				1.44
	19	1.00	2.85	1.19
	24	1.30	2.75	.77
	28	.91	3.29	1.35
Covariance estimates				
Acceptability – Understanding			.38*	
Acceptability – Family-School			-.14*	
Acceptability – Feasibility			.25*	
Acceptability – System Climate			.50*	
Acceptability – System Support			-.34*	
Understanding – Family-School			-.27*	
Understanding – Feasibility			.35*	
Understanding – System Climate			.65*	
Understanding – System Support			-.60*	
Family-School – Feasibility			-.06	
Family-School – System Climate			-.33*	
Family-School – System Support			.87*	
Feasibility – System Climate			.46*	
Feasibility – System Support			-.37*	
System Climate – System Support			-.77*	
e5 – e9			.38*	
e10 – e17			.25*	
e2 – e10			.18*	
e1 – e10			.22*	
e2 – e17			.15*	
e4 – e17			.09*	
e22 – e27			.12*	
e9 – e11			-.26*	

*statistically significant at $p < .001$

Support 2. Modification indices provided by Amos were evaluated for impact on overall model fit and theoretical applicability. For support 2, the modification indices that were applied were putting covariances in place between the residuals of the following item pairs: 1 and 2, 1 and 12, 2 and 4, 2 and 12, 4 and 12, 6 and 11, 8 and 12, 11 and 19, 14 and 15, 14 and 20, 19 and 28. Items 1, 2, 4, 6, 8, and 12 are part of the Acceptability subscale. Items 11, 14, 15, and 20 are

part of the Feasibility subscale. Items 19 and 28 are part of the System Support subscale. Figure 3 below shows the model with modifications added.

Figure 3. Path diagram of the six-factor model with modifications for support 2



Despite these modifications, model fit for support 2, vocabulary supports, remained poor as demonstrated by the fit indices displayed in Table 19 below. The chi-square test was

significant ($p < .01$, $\chi^2 (362) = 1926.83$), but yielded a lower value at 1926.83 than the value of 2371.33 without the modifications. The CFI value was below .95 (CFI = .82) but was higher than the value of .77 without modifications. The RMSEA was slightly above .10 (RMSEA = .11) but was lower than the value of .12 without modifications. The AIC value also decreased with modifications from 2575.33 to 2152.83. Regression weights of items onto respective factors was adequate, with each item demonstrating statistical significance. Table 20 below shows the regression weight estimates, intercept estimates, variance estimates, and covariance estimates for support 2 after modifications.

Table 19. *Fit Indices after Modifications for Support 2*

Chi-square	CFI	RMSEA	AIC
1926.83	.82	.11	2152.83

Table 20. *Estimates of Regression Weights, Intercepts, Variance, and Covariances for Support 2 after Modifications*

Item	Regression weight estimate	Intercept estimate	Variance estimate
Acceptability			1.66
1	1.00	4.29	1.42
2	.50	4.85	1.05
3	.42	4.98	1.27
4	.98	4.63	.76
6	-.44	2.58	1.88
8	.72	4.80	.55
10	.83	4.51	.99
12	.62	4.50	1.42
17	1.09	4.43	.60
Understanding			2.84
18	1.00	4.49	.20
22	.95	4.57	.41
27	.86	4.55	.41
Family-School			1.34
13	1.00	2.84	1.80
25	.88	2.63	1.13
29	1.26	2.96	.53
Feasibility			2.47
5	1.00	4.32	.08
9	.50	4.88	1.01
11	-.17	2.01	1.50
14	.73	4.49	.51
15	.57	4.58	.62

Table 20 (cont'd)

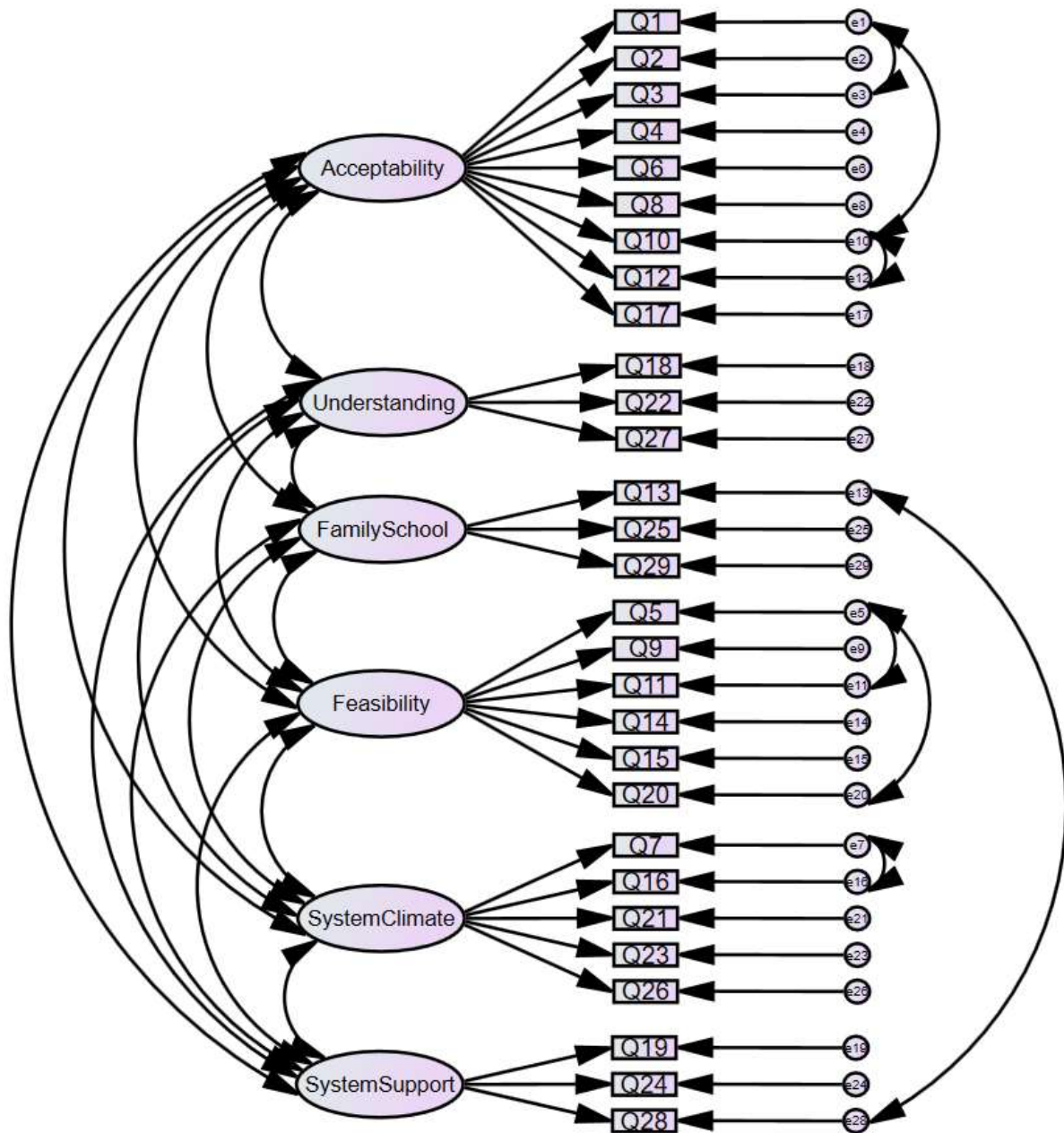
	20	.34	4.45	1.65
System Climate				1.35
	7	1.00	4.69	.45
	16	.54	5.10	.51
	21	.77	4.68	.65
	23	.85	4.77	.77
	26	.72	4.79	.67
System Support				.28
	19	1.00	2.05	1.13
	24	2.88	2.52	.19
	28	1.31	2.53	1.51
Covariance estimates				
Acceptability – Understanding			1.67*	
Acceptability – Family-School			-.37*	
Acceptability – Feasibility			1.60*	
Acceptability – System Climate			1.49*	
Acceptability – System Support			-.29*	
Understanding – Family-School			-.88*	
Understanding – Feasibility			2.34*	
Understanding – System Climate			1.42*	
Understanding – System Support			-.53*	
Family-School – Feasibility			-.58*	
Family-School – System Climate			-.27	
Family-School – System Support			.43*	
Feasibility – System Climate			1.37*	
Feasibility – System Support			-.46*	
System Climate – System Support			-.25*	
e4 – e12			-.36*	
e2 – e12			.44*	
e19 – e28			.52*	
e8 – e12			.20*	
e1 – e2			.42*	
e2 – e4			-.22*	
e1 – e12			.39*	
e14 – e15			.16*	
e14 – e20			.23*	
e6 – e11			.44*	
e11 – e19			.44*	

*statistically significant at $p < .001$

Support 3. Modification indices provided by Amos were evaluated for impact on overall model fit and theoretical applicability. For support 3, the modification indices that were applied were putting covariances in place between the residuals of the following item pairs: 1 and 3, 1 and 10, 10 and 12, 13 and 28, 5 and 11, 5 and 20, and 7 and 16. Items 1, 3, 10, and 12 are part of

the Acceptability subscale. Item 13 is part of the Family-School subscale. Items 5, 11, and 20 are part of the Feasibility subscale. Items 7 and 16 are part of the System Climate subscale. Item 28 is part of the System Support subscale. Figure 4 below shows the model with modifications added.

Figure 4. Path diagram of the six-factor model with modifications for support 3



Despite these modifications, model fit for support 3, inclusion of L1, remained poor as demonstrated by the fit indices displayed in Table 21 below. The chi-square test was significant ($p < .01$, $\chi^2(362) = 1511.51$), but yielded a lower value at 1511.51 than the value of 1713.55 without the modifications. The CFI value was well below .95 (CFI = .79) but was higher than the value of .75 without modifications. The RMSEA was below the cutoff of .10 (RMSEA = .09) and was lower than the value of .10 without modifications. The AIC value also decreased with modifications from 1917.55 to 1729.51. Regression weights of items onto respective factors was adequate, with each item demonstrating statistical significance. Table 22 below shows the regression weight estimates, intercept estimates, variance estimates, and covariance estimates for support 3 after modifications.

Table 21. *Fit Indices after Modifications for Support 3*

Chi-square	CFI	RMSEA	AIC
1511.51	.79	.09	1729.51

Table 22. *Estimates of Regression Weights, Intercepts, Variance, and Covariances for Support 3 after Modifications*

Item	Regression weight estimate	Intercept estimate	Variance estimate
Acceptability			.29
1	1.00	4.94	.77
2	1.76	4.54	.63
3	1.59	4.29	1.78
4	1.56	4.23	1.56
6	-1.38	2.77	1.68
8	1.74	4.56	.85
10	.89	4.91	.86
12	1.39	4.54	.90
17	2.16	4.14	.66
Understanding			.67
18	1.00	4.71	.47
22	1.39	4.33	.57
27	.91	4.47	1.43
Family-School			.59
13	1.00	2.86	1.71
25	1.80	3.05	.59
29	1.80	2.98	.39
Feasibility			.77

Table 22 (cont'd)

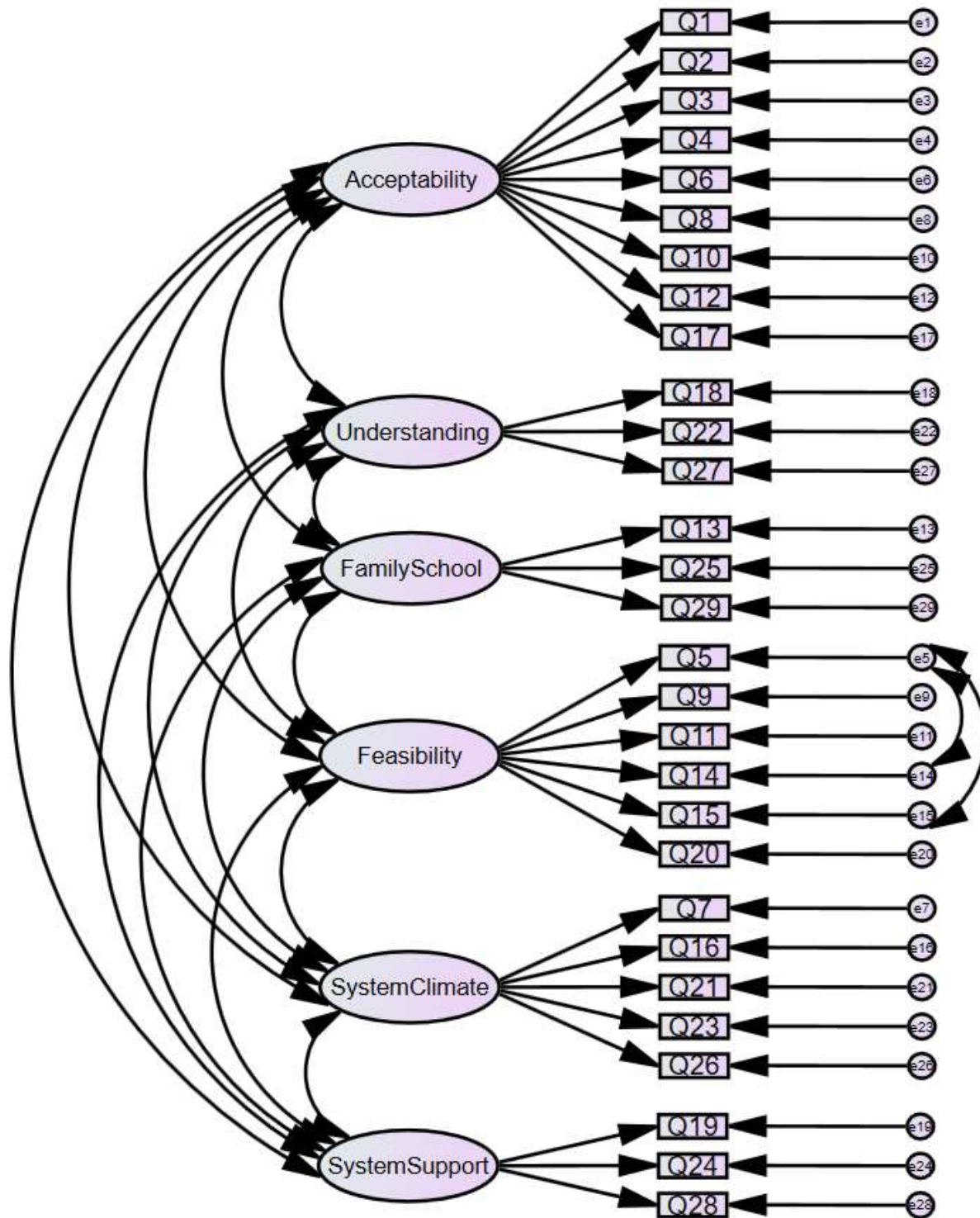
	5	1.00	4.18	.85
	9	.92	4.17	1.05
	11	-.70	2.67	1.69
	14	1.13	4.01	.65
	15	1.14	3.98	.48
	20	1.07	3.70	1.27
System Climate				.62
	7	1.00	4.68	1.20
	16	.96	4.69	1.12
	21	1.34	3.92	.78
	23	1.43	4.27	.78
	26	1.12	4.38	.98
System Support				1.15
	19	1.00	3.42	1.29
	24	1.06	3.16	1.45
	28	1.02	3.79	1.68
Covariance estimates				
Acceptability – Understanding			.28*	
Acceptability – Family-School			.04	
Acceptability – Feasibility			.35*	
Acceptability – System Climate			.36*	
Acceptability – System Support			-.08	
Understanding – Family-School			-.06	
Understanding – Feasibility			.45*	
Understanding – System Climate			.32*	
Understanding – System Support			-.46*	
Family-School – Feasibility			.01	
Family-School – System Climate			.06	
Family-School – System Support			.35*	
Feasibility – System Climate			.54*	
Feasibility – System Support			-.45*	
System Climate – System Support			-.24*	
e7 – e16			.35*	
e10 – e12			.31*	
e1 – e3			-.30*	
e1 – e10			.17*	
e5 – e20			.29*	
e5 – e11			-.29*	
e13 – e28			-.48*	

*statistically significant at $p < .001$

Support 4. Modification indices provided by Amos were evaluated for impact on overall model fit and theoretical applicability. For support 4, only few modification indices that would greatly impact overall model fit made theoretical sense. Therefore, the modification indices that were applied were putting covariances in place between the residuals of only two item pairs: 5

and 14, and 5 and 15. These items are all part of the Feasibility subscale. Figure 5 below shows the model with modifications added.

Figure 5. Path diagram of the six-factor model with modifications for support 4



Despite these modifications, model fit for support 4, alternate response format, remained poor as demonstrated by the fit indices displayed in Table 23 below. The chi-square test was significant ($p < .01$, $\chi^2(362) = 2271.32$), but yielded a lower value at 2271.32 than the value of 2423.90 without the modifications. The CFI value was well below .95 (CFI = .77) but was higher than the value of .76 without modifications. The RMSEA was above .10 (RMSEA = .12) and when rounded, was equal to the value of .12 without modifications. The AIC value also decreased with modifications from 2627.90 to 2479.32. Regression weights of items onto respective factors was adequate, with each item demonstrating statistical significance. Table 24 below shows the regression weight estimates, intercept estimates, variance estimates, and covariance estimates for support 1 after modifications.

Table 23. *Fit Indices after Modifications for Support 4*

Chi-square	CFI	RMSEA	AIC
2271.32	.77	.12	2479.32

Table 24. *Estimates of Regression Weights, Intercepts, Variance, and Covariances for Support 4 after Modifications*

Item	Regression weight estimate	Intercept estimate	Variance estimate
Acceptability			1.22
1	1.00	4.79	.29
2	1.01	4.77	.37
3	.42	4.75	1.35
4	.60	5.03	.47
6	-.37	2.33	2.06
8	.87	4.96	.36
10	.96	4.76	.59
12	.41	5.03	.91
17	.87	4.86	.33
Understanding			1.03
18	1.00	5.05	.39
22	.82	5.10	.17
27	.58	4.86	.52
Family-School			1.10
13	1.00	2.56	.80
25	1.16	2.97	.74
29	1.13	2.98	.83
Feasibility			.26

Table 24 (cont'd)

	5	1.00	5.07	.49
	9	1.90	4.84	.65
	11	-2.61	2.53	1.16
	14	.61	5.21	.79
	15	1.58	4.74	1.24
	20	2.22	4.43	.76
System Climate				1.52
	7	1.00	4.71	.92
	16	.28	5.16	.90
	21	.99	4.57	.63
	23	.93	4.77	.23
	26	.94	4.71	.68
System Support				1.43
	19	1.00	2.88	1.47
	24	.77	2.32	1.04
	28	1.03	2.79	.79
Covariance estimates				
Acceptability – Understanding			.73*	
Acceptability – Family-School			-.28*	
Acceptability – Feasibility			.49*	
Acceptability – System Climate			1.20*	
Acceptability – System Support			-.48*	
Understanding – Family-School			-.34*	
Understanding – Feasibility			.42*	
Understanding – System Climate			1.08*	
Understanding – System Support			-.69*	
Family-School – Feasibility			-.22*	
Family-School – System Climate			-.35*	
Family-School – System Support			.88*	
Feasibility – System Climate			.59*	
Feasibility – System Support			-.40*	
System Climate – System Support			-.65*	
e5 – e14			.29*	
e5 – e15			.22*	

*statistically significant at $p < .001$

Single-Factor Model

Due to poor model fit of the six-factor model identified by Briesch et al. (2013) both with and without modification indices applied, a single-factor model was tested. This model, with all 29 items loading onto one overall factor was run for each of the four supports. Model fit was inadequate for all four supports, as indicated by significant chi-square tests, CFI values well

below .95, and RMSEA values above .10. Due to these findings, a single-factor model solution does not appear more adequate than the six-factor model identified by Briesch et al (2013).

Research Question 2

Research question two aimed to examine whether certain teacher characteristics predicted overall URP-IR scores. Due to poor model fit, which did not support using subscale scores, only the total URP-IR score for each support was used as the dependent variable. Despite poor model fit for a single-factor solution, internal consistency for the entire measure was strong for all four supports ($\alpha = .90-.94$). The predictor variables were (1) teacher experience working with ELs, (2) teacher training on working with ELs, and (3) teacher consultative support. A total of twelve ANOVA analyses were conducted, one for each predictor for each of the four supports. Table 25 below shows a summary of the mean URP-IR scores and standard deviations for each predictor group and each support.

Table 25. *URP-IR Total Scores by Predictor Group*

	URP-IR Total Score (SD)			
	S1: Visual Aids	S2: Vocabulary Supports	S3: Inclusion of L1	S4: Alternate Response Format
Experience				
None (n=68)	140.78 (14.82)	136.60 (22.24)	125.35 (19.48)	139.46 (20.23)
Low (n=112)	143.29 (18.73)	141.11 (21.24)	124.06 (22.78)	141.05 (20.71)
High (n=110)	145.24 (17.03)	141.63 (19.49)	125.96 (22.05)	144.92 (21.24)
Training				
None (n=42)	140.07 (15.66)	137.31 (22.55)	122.38 (24.05)	140.98 (21.94)
Low (n=122)	141.71 (16.04)	139.71 (19.62)	124.91 (18.94)	138.83 (20.96)
High (n=113)	147.21 (17.98)	142.98 (21.29)	126.53 (22.81)	146.37 (20.01)
Consultation				
None (n=104)	143.18 (15.26)	138.70 (20.30)	123.10 (23.65)	143.13 (19.66)
Low (n=134)	144.04 (17.99)	141.74 (22.13)	124.47 (20.17)	142.52 (21.50)
High (n=41)	144.44 (17.77)	141.66 (17.26)	132.29 (19.14)	141.10 (21.32)

Experience

Teacher experience working with ELs was not a significant predictor of the total URP-IR score for any of the four supports ($p > .05$). For support 1 (visual aids), the ANOVA determined

that average URP-IR scores did not differ significantly for teachers with different levels of experience ($F(2, 287) = 1.411, p = .246$). For support 2 (vocabulary supports), the ANOVA determined that average URP-IR scores did not differ significantly for teachers with different levels of experience ($F(2, 287) = 1.376, p = .254$). For support 3 (inclusion of L1), the ANOVA determined that average URP-IR scores did not differ significantly for teachers with different levels of experience ($F(2, 287) = .218, p = .804$). For support 4 (alternate response format), the ANOVA determined that average URP-IR scores did not differ significantly for teachers with different levels of experience ($F(2, 287) = 1.699, p = .185$).

Despite non-significant findings, across all four supports, the average total URP-IR score was higher for teachers with high experience than for teachers with no experience. The teachers with medium experience had an average total score in between those with no and high experience for all supports, except for support 3 (inclusion of L1). For this support, teachers with medium experience had an average total score lower than both teachers with no and high experience.

Training

Teacher training on working with ELs was a significant predictor for supports 1 (visual aids) and 4 (alternate response format) and was not a significant predictor for support 2 (vocabulary supports) and 3 (inclusion of L1). For support 1 (visual aids), the ANOVA determined that average URP-IR scores differed significantly for teachers with different levels of training ($F(2, 274) = 4.299, p = .015$). For support 2 (vocabulary supports), the ANOVA determined that average URP-IR scores did not differ significantly for teachers with different levels of training ($F(2, 274) = 1.379, p = .253$). For support 3 (inclusion of L1), the ANOVA determined that average URP-IR scores did not differ significantly for teachers with different

levels of training ($F(2, 274) = .595, p = .552$). For support 4 (alternate response format), the ANOVA determined that average URP-IR scores differed significantly for teachers with different levels of training ($F(2, 274) = 3.976, p = .020$).

For support 1, URP-IR scores increased from the no training (140.07 ± 15.66), to the low training (141.71 ± 16.04), to the high training group (147.21 ± 17.98), respectively. Tukey post hoc tests indicated that the difference between the no training and the high training group, as well as the difference between the low training and the high training group was statistically significant ($p = .05; p = .03$). The difference between the no training group and the low training group was not found to be statistically significant.

For support 4, URP-IR scores increased from the low training (138.83 ± 20.96), to the no training (140.98 ± 21.94), to the high training group (146.37 ± 20.01), respectively. Tukey post hoc tests indicated that the difference between the low training and the high training group was statistically significant ($p = .02$). The difference between the no training group and the low training group, as well as the difference between the no training group and the high training group was not found to be statistically significant.

Consultation

The average hours of consultative support teachers receive was not a significant predictor of the total URP-IR score for any of the four supports. For support 1 (visual aids), the ANOVA determined that average URP-IR scores did not differ significantly for teachers with different levels of consultation ($F(2, 276) = .110, p = .895$). For support 2 (vocabulary supports), the ANOVA determined that average URP-IR scores did not differ significantly for teachers with different levels of consultation ($F(2, 276) = .686, p = .504$). For support 3 (inclusion of L1), the ANOVA determined that average URP-IR scores did not differ significantly for teachers with

different levels of consultation ($F(2, 276) = 2.830, p = .061$). For support 4 (alternate response format), the ANOVA determined that average URP-IR scores did not differ significantly for teachers with different levels of consultation ($F(2, 276) = .140, p = .870$).

For supports 1 (visual aids) and 3 (inclusion of L1), average scores increased as the amount of consultative support increased, although this increase was nonsignificant. For support 2 (vocabulary supports), teachers with no consultative support had a lower URP-IR score than those teachers with some or most consultative support, although this difference in score was nonsignificant. For this support, teachers with some or most consultative support had an effectively equal average total URP-IR score. For support 4 (alternate response format), the total average URP-IR score decreased as consultative support increased, although this relationship was nonsignificant.

Research Question 3

Research question three aimed to examine whether significant URP-IR score differences were found between the four instructional supports. A repeated measures ANOVA with a Greenhouse-Geisser correction revealed that there was a statistically significant difference between the instructional supports ($F(2.775, 1040.46) = 70.283, p < .05$). Post-hoc analysis, with the Bonferroni correction applied to correct for multiple comparisons, revealed that differences in total URP-IR scores were statistically significant among all four supports ($p < .05$). Table 26 below shows the mean scores for the four supports as well as the pairwise comparisons between the supports. Support 1 (visual aids) had the highest overall URP-IR mean score, followed by support 4 (alternate response format), support 2 (vocabulary support), and support 3 (inclusion of L1) with the lowest overall URP-IR mean score. The largest mean score difference was between support 1 and support 3, with a 17.04 score difference. The smallest mean score difference was

between support 1 and support 4, with a 3.13 mean score difference. Effect sizes were calculated to determine the practical significance of the differences in mean scores. Small effect sizes were identified for most comparisons and medium effect sizes were identified for the difference in mean scores between support 1 and support 3 ($d_{RM}=.71$) and between support support 3 and support 4 ($d_{RM}=.57$).

Table 26. *Mean URP-IR Scores per Support and Pairwise Comparisons*

	Mean URP-IR total score (SD)	Mean difference in URP- IR total score	Effect size (d_{RM})
Support 1 (visual aids)	139.95 (18.55)		
Support 2		6.72*	.28
Support 3		17.04*	.71
Support 4		3.13*	.13
Support 2 (vocabulary supports)	133.24 (24.90)		
Support 3		10.32*	.41
Support 4		-3.59*	.15
Support 3 (inclusion of L1)	122.92 (20.77)		
Table 26 cont'd			
Support 4		-13.90*	.57
Support 4 (alternate response format)	136.82 (23.24)		

*Significant at the .05 level

Research Question 4

Research question four aimed to explore the other two aspects of social validity - social significance of the goals and social importance of the effects. Qualitative data were collected through interviews with nine teachers to address this question. Thematic analysis of the data by two independent coders yielded seven main themes: (1) range in goals; (2) prioritized goals; (3) barriers for ELs; (4) adverse effects of barriers; (5) positive effects of supports; (6) effects of individual supports; (7) contingencies and shortcomings. The following sections will describe each theme in detail and discuss how the theme describes social significance of the goals and social importance of the effects of the four supports.

It is important to note that although efforts were made to reduce researcher bias throughout the process of collecting and analyzing data, absolute objectivity can never be truly achieved. As such, the researchers' (both the principal investigator and the secondary coder) background and identity may have, in part, shaped the lens through which data was interpreted. When evaluating the findings of the current study, it is critical that this be considered. Both researchers are graduate students in school psychology graduate program in the Midwest, are interested and have conducted research in the area of school-based supports and accommodations for English Learners, and identify as English Learners themselves.

Range in Goals

When asked about what their goals for ELs were during science instruction, it was notable how widely teachers' answers varied. Broadly, responses could be characterized as emphasizing academic and social-emotional goals. The academic goals mentioned for ELs were mainly goals related to comprehension of content, the ability to demonstrate or communicate knowledge, increase in science knowledge specifically, and increase in vocabulary. For example, one teacher mentioned that "for English Learners specifically, I would say another goal for them is just to acquire vocabulary," highlighting that ELs have unique goals from other students in the classroom. Another teacher mentioned that for ELs vocabulary is "huge in their understanding" and is therefore a goal that is emphasized. Other teachers hoped that ELs would be able to "articulate their thinking," which may involve vocabulary but also expresses the broader goal of communicating understanding and knowledge.

In contrast, some teachers were more focused on social-emotional goals or goals more adjacent to academic skills for ELs. These included goals such as feelings of belonging, enjoyment of science and learning, and engagement and participation. Specifically, many

teachers felt strongly that belonging is a critical goal for ELs. One teacher described that she focuses on increasing feelings of belonging for her ELs, so that “they just know that if they pronounce something differently that they won't, they won't be laughed at or, or somehow singled out.”

Overall, it seems that there is no consensus as to which goals teachers think of when asked about ELs in science instruction. Some of the academic goals, such as comprehension of the content, might be the same as the goals for all students, ELs and non-ELs. However, others, such as acquisition of vocabulary and feelings of belonging, seemed to be more specifically targeting the unique needs of ELs.

Prioritized Goals

Although teachers provided varying responses when asked about goals open-endedly, when provided five different goals for ELs during science instruction and asked to rank them in order of importance, teachers answered quite similarly. Teachers tended to prioritize goals like “feelings of belonging” or “engagement” over goals like “access to instruction,” “increase in science knowledge,” and “increase in English skills.” Most teachers felt that the former goals were prerequisite goals for the latter goals. Only two teachers rated neither “feelings of belonging” or “engagement” among their top two goals. “Access to instruction” was most often rated as third, and “increase in science knowledge” and “increase in English skills” tended to be rated as last. Some teachers struggled with the idea of putting science knowledge last, as demonstrated by this quote: “science knowledge is important, but at the end of the day, these all build; science knowledge won't happen if all these other things don't happen.” Others felt that science knowledge was not critical as a goal: “the last thing really is science knowledge. I want

to get them more excited about it than I really care about them taking understanding away with them.”

Barriers for ELs

When asked about what barriers ELs experience in science instruction, teachers provided a range of answers. Broadly, their answers could be sub-themed into academic and social-emotional barriers. Understanding the barriers teachers view as prevalent and disruptive to ELs’ success in school is important from a social validity perspective because it informs consultants of which problems teachers will likely feel most motivated to address through supports or interventions.

Examples of academic barriers identified by teachers were things such as not having sufficient vocabulary or academic language, the uniqueness of the science process, and overall comprehension. For example, one teacher explained: “a lot of my ELs struggle with comprehension; making meaning of the text and like, vocabulary.” Other teachers added that science often has tricky vocabulary, more so than other subject areas. Additionally, other teachers stated that science, in addition to the vocabulary, is different from other subject areas. One teacher described it as: “this is a 'what do you think' learning.” For example, the scientific process may be unfamiliar to some ELs, which may impede comprehension. Additionally, one teacher expanded on the idea of vocabulary as the barrier and discussed that this may not get at the true barrier:

if I was talking about the Moon's cycle and I looked up all the Spanish words for half-moon, full moon, and the student’s family doesn't go out and look at the moon or ever reference the moon, or when they do, they reference it in a way as being one of the deities

of the sky or something, and all I do is reference it by using a word, the word is not a barrier. It's, it's the concept that's a barrier.

Teachers also discussed barriers beyond academic ones, broadly barriers related to social emotional factors. Many teachers discussed that they have observed ELs experience being singled out, isolated, or as though they do not belong in the classroom:

Sometimes feeling like they belong because they just come from different cultures sometimes. And now, I've learned a lot as a teacher, that sometimes the way that we interact with the mainstream students, like White students, Black students that we're, you know, that we are more accustomed to in the rural areas here in North Carolina; you know, a lot of the family dynamics we already know but when, you know, English language Learners come from a different country or just they have different cultural things that we have to learn about to, to help them and, and like teach our whole class; those things, so that the children do feel they belong. And, I feel like if you build a class of community, that that addresses a lot of that but they had to, they have to feel a sense of belonging before they're going to engage fully.

Another barrier raised by three of the teacher participants was teacher knowledge, preparation, or willingness to implement supports. For example, one teacher described that some teachers' mindsets could create barriers for ELs. She described this mindset as "Why should I have to, you know, learn, you know, my one student's language just so that I can help them understand my content. That's not my job." Other teachers described that even though there might be a willingness to support ELs, teachers do not always receive the training necessary to do so: "I did not have any classes in ELL in college. So, I think teacher preparation is probably a big part of that too." Discussions of teacher preparation or willingness fall into the second aspect

of social validity, acceptability, which was not technically the focus of this interview.

Nonetheless, several teachers raised it as a potential barrier to ELs in science instruction, so it was included in the qualitative analysis.

Adverse Effects of Barriers on ELs

When asked how barriers experienced by ELs affect them, other students, and the teachers, teacher participants' responses mainly focused on the effects on ELs themselves. Additionally, the effects they discussed were exclusively adverse, and teachers provided answers that could be sub-themed into academic and social-emotional effects. Understanding which effects are most prevalent for teachers is critical from a social validity perspective, as supports that address those adverse effects are more likely to appeal to teachers tasked with implementing them.

In terms of the academic effects barriers have for ELs in mainstream science instruction, teachers expressed that without barriers, ELs would experience more growth more quickly and would have an overall better understanding of the content being taught. One teacher explained that without barriers, ELs may have better access to instruction: "it wasn't the concept at all, that he couldn't understand about energy. That he could get, but it was like, what to then do to demonstrate his understanding." In other words, removing the linguistic or other barriers would allow ELs to express their understanding better. Teachers also pointed out that decreased barriers would benefit the overall teaching environment. For example, one teacher explained: "when you're going with the flow and you're communicating, then you can get a lot more done. You can go at a quicker pace."

In terms of the social-emotional effects barriers have for ELs in mainstream science instruction, teachers focused on aspects such as engagement, confidence, and trust. One teacher

expressed that reduced barriers might give ELs “more confidence or more trust in me and in our school that, like, we're going to take care of them and make sure their needs are met.”

Additionally, one teacher saw engagement as a major benefit of reducing barriers: “as long as you address those issues early on (...) you just open up a whole new avenue for students (...) and that increases their engagement.”

Positive Effects of Supports

When asked about the effects of the four supports, teachers often discussed positive effects. The positive effects most often noted by teachers related to an improved ability to communicate, improved learning, positive social-emotional effects, and the universality of the supports. First, teachers felt that the instructional supports proposed in the current study would help ELs communicate. For example, vocabulary supports might “give them the words to respond to us.” Teachers often identified that ELs experience barriers that interfere with their ability to articulate their thinking. When asked about the four supports, most felt that they would address this barrier.

Second, teachers felt that the supports improve ELs’ learning experience. One teacher explained that the supports foster comprehension for ELs and level the playing field:

“They're able to then apply it to their own new learning and that, that's a big game changer for a lot of kids. That means, hey my friend over here, who was born in the United States speaks English at home, I can still converse with that person, whether I speak Spanish at home, whether I speak Somali at home, you know. It's, it's really important. So, like I said it evens the playing field for a lot of those kids.”

With regard to improved learning, teachers generally expressed that the supports help overcome the barriers ELs may experience by providing more ways to access the content

presented, which in turn allows ELs to form a deeper comprehension of the concept than they would have otherwise had.

Third, given the large number of social-emotional barriers teachers discussed, many of the identified positive effects of the four supports also related to social-emotional factors. These included things such as an increase in student confidence and self-esteem, an improved teacher-student relationship, and an increase in engagement. One teacher expressed that she believed the supports would give ELs “more confidence and they would participate probably even more in what we were doing in class, and, and feel, like, feel good about themselves.” Another teacher also highlighted that the supports may make ELs feel better not only about themselves as learners, but also about their relationship with their teacher:

“You're saying that their L1 is important; that that cultural component is important. And that you're, you're showing that to them because you're willing to use that in the classroom and say this is important.”

Lastly, regarding social-emotional effects of the four supports, many teachers noted the increase in engagement they have or would expect to see. For example, one teacher said:

“you just open up a whole new avenue for students to express themselves and to bounce ideas off of each other and that increases their engagement, that increases the teachers engagement.”

Many teachers also discussed that coupled with the increase in access and comprehension that the supports afford, engagement also plays a key role in improved learning for ELs.

Lastly, another positive effect of the supports discussed was whether the supports were effective for all ELs. Most teachers (n=5) said yes; however, two said no, and two were unsure. Among those who thought that the supports would work well for all ELs, most indicated that

they would have to be adapted to fit each EL individually. However, two teachers felt that the diversity within ELs was too great to address with just these four supports. In addition, many teachers thought that the four supports would help all students, not just ELs.

Effects of Individual Supports

When asked about the effects of the four supports, teachers often singled out certain supports. Often, they discussed positive effects of the single supports, but sometimes, they also expressed concerns about single supports. Due to teachers sometimes talking about the four supports generally, and sometimes talking about one specific support, this theme is described separately from the previous one.

Consistently, teachers emphasized the positive effects of vocabulary and visual supports more than those of inclusion of L1 and alternate response format. Additionally, when raising concerns, teachers typically raised them about inclusion of L1. Regarding vocabulary and visual supports, teachers often expressed that they thought these supports would have positive effects for ELs. It appeared that teachers had a lot of prior experience using these supports and felt that in addition to helping ELs access instruction, these supports are beneficial to all students in their classroom. Generally, teachers expressed that use of vocabulary and visual supports would increase access to instruction for ELs: “with the visual aids, it gives them a better understanding. And so, they're on that, they're on a playing field that they might not have been before.” Some of the examples teachers gave were that visual supports help create meaning, decrease language barriers, and are enjoyable to kids and that vocabulary supports help with recollection, meaning making, and expression of thinking.

Regarding inclusion of L1, teachers at times questioned the utility and/or feasibility of these supports. One teacher asked about inclusion of L1: “do you find that messes up their

phonetic process?” Additionally, this teacher expressed that she did not feel prepared to implement L1 supports on her own: “I would not know how to go about doing that, the incorporation. I just really need someone to help me with that, but I think it would probably be helpful for them.” However, some teachers also expressed value of L1 as a support for ELs. Specifically, what appeared pertinent to teachers was that inclusion of L1 emphasized ELs’ cultural background, which in turn may increase feelings of belonging and engagement in the material: “You’re saying that their L1 is important; that that cultural component is important. And that you’re, you’re showing that to them because you’re willing to use that in the classroom and say this is important.”

Regarding alternate response format, teachers expressed mixed opinions about whether it would be a helpful support for ELs. Some teachers highlighted that alternate response format allows for assessment of skills without confounding barriers: “the important thing is that they’re able to show that they understand, that they’ve grasped the content, as opposed to showing that they know how to write a paper.” Additionally, others expressed that alternate response formats increase comfort and confidence in ELs: “alternate response format can help kids feel more comfortable, or more confident, expressing their ideas.” Those teachers who expressed hesitations about alternate response format appeared more unsure of it rather than critical: “the alternative response format, again, it could work but it might, it might be, it might not as well.” Some teachers may have not had prior experience providing this support, which may have been a reason for the hesitation or lack of positive response when asked about the effectiveness of the four supports.

Contingencies and Shortcomings

In addition to the positive effects that teachers anticipated or have experienced as a result of the four supports, many teachers also discussed shortcomings of the supports or expressed certain conditions under which the supports may be most effective.

In terms of shortcomings and concerns, teachers discussed that the four supports may not be sufficient to address all of the barriers ELs may experience. Additionally, teachers expressed concerns that the supports would not work for all ELs, and thus, it would be difficult to provide individualized supports with a diverse group of ELs. One teacher explained: “that becomes a challenge of pulling in all those different languages and being able to connect their native language with English language, and then pictures that make sense to their particular culture.”

Aside from concerns, many teachers also highlighted some circumstances under which the supports may work best for ELs. A common one was the concept of teacher intent or mindset when using the supports. One teacher explained:

“if the teacher is aware of their biases and, like, you know, knows their identity, and know the students' identity, and they're open to, like, using these strategies, yeah they're going to work. But, like, I can give teachers all day and I can coach them and mentor them on these but if they themselves have these like deficit mindsets or you know, or like “no it's because they can't learn”, then I can give you these but they won't work”.

Lastly, according to many participants, another critical aspect to achieve positive effects of the four supports, is individualization of the supports. Teachers explained that ELs' needs should be assessed on a case-by-case basis, that supports should be tweaked as needed, and that the supports should match each ELs' level of need.

Summary

In summary, the seven main themes yielded by the interview data were (1) range in goals; (2) prioritized goals; (3) barriers for ELs; (4) adverse effects of barriers; (5) positive effects of supports; (6) effects of individual supports; (7) contingencies and shortcomings. Each theme was described and discussed above. Further implications of these findings and integration of these findings with quantitative results will be provided in the discussion section.

CHAPTER V: DISCUSSION

The current study investigated teacher opinions regarding instructional supports for English Learners using a social validity framework. Results suggested that teachers had favorable but varying opinions of all four supports presented in the current study. Confirmatory factor analyses indicated that a previously identified factor structure of the URP-IR is not a strong model fit for the data of the current study. Further results indicated that teacher training was a significant predictor of teacher opinions of some of the supports. In the following section, the findings associated with each research question and how those findings relate and contribute to the existing empirical literature will first be discussed. Then, important limitations to consider when interpreting findings of the current study will be highlighted. Lastly, implications for future research and practice will be discussed.

Prior to discussing findings associated with each individual research question, it is important to highlight some important context that should be considered during interpretation of the findings. More specifically, in the current study, overall URP-IR scores were quite high. Across all supports, average responses for all questions (with just one exception) were in the direction of favoring each support. This is in contrast to findings from Briesch et al. (2013), where average responses more frequently indicated non-favorable views of the proposed intervention. Although average responses were predominantly favorable in both the current study and in the Briesch et al. (2013) study, average ratings were often minimally favorable in the Briesch et al. (2013) study, whereas average ratings were often strongly favorable in the current study. This is a notable finding as it may indicate important differences between teachers' opinions of behavioral interventions versus their opinions of supports for ELs. It is also

important to note that the reduced variability in ratings in the current study may have also impacted the ability confirm a similar factor structure.

Poor Model Fit

Results of the CFA indicated poor model fit for all four supports when testing the six-factor model, even following adjustments made based on modification indices. This contrasts with the findings of Briesch et al. (2013) who obtained model fit when applying the URP-IR to measure teacher opinions of a variety of behavior interventions. These contrasting findings raise the question of whether the URP-IR allows for similar measurement of teacher opinions across different types of interventions and supports. There are a variety of potential explanations of these different findings, including difference in supports studied, item skew, and methods of data collection.

First, a potential explanation for the difference in findings is the nature of the supports/interventions studied, as there may be different factors that better represent teacher opinions of EL academic supports than those that represent teacher opinions of behavioral interventions. Based on the current findings, it appears that the factors to consider potentially vary between behavioral interventions and instructional supports for ELs. Across the social validity literature, acceptability is generally thought of as a combination of individual factors (e.g., personal views of the intervention), intervention factors (e.g., feasibility, complexity), and environmental factors (e.g., fit and systemic support; Briesch et al., 2013). Development of the URP-IR identified six more nuanced factors; however, perhaps reduction to three factors instead of six would be a better approach to measuring acceptability for EL instructional supports. The social validity literature has largely focused on behavioral interventions, often not in school-

based contexts; the current study appears to suggest that the factors derived from this line of literature may not apply to measurement of acceptability of EL supports.

Relatedly, findings of poor model fit may also speak to the difference in constructs of behavior and language. In particular, there may be an important difference in how these constructs are generally conceptualized. Student behavioral challenges are typically viewed as a problem that needs solving and can be improved, whereas language, and in particular multilingualism, is increasingly viewed as individual differences that can be fostered to support students' learning. This difference in constructs is particularly evident in the wording of URP-IR items, which refer to "behavior problems." This wording was changed to "language barriers" in the current study, but nonetheless reflect the solution-focused approach often associated with behavioral concerns.

Another important difference to note when aiming to interpret the lack of model fit identified in the current study compared to the adequate model fit identified in the Briesch et al. (2013) study is the particularly positive perceptions held by most respondents about the supports presented in the current study. This may have created a ceiling effect and, in turn, may have negatively affected the ability to detect distinct factors of the URP-IR. As noted previously, the descriptive data of the current study revealed much more positively skewed average responses to each item across the four supports compared to descriptives provided by Briesch et al. (2013). Although this skew did not violate normality, this nonetheless may have implications for measurement differences between the two studies, as Briesch et al. (2013) may have had more variation in scores that allowed for better detection of item loadings onto factors.

Lastly, a small difference between the Briesch et al. (2013) study and the current study that may have affected findings is the methods of data collection. Briesch et al. (2013) collected

URP-IR responses from teachers through phone interviews rather than online data collection. Participants answering via phone interviews may have been more engaged in carefully considering and answering questions than participants in the current study who answered via an online survey. This is only a small difference between the two studies, but one that could have potentially provided more rigorous data to Briesch et al. (2013) for the purposes of assessing measurement qualities of the URP-IR. In other words, participants in the Briesch et al. (2013) study may have provided more nuanced responses that more accurately reflected their true opinions, which in turn also allowed for detection of the six factors. A study that has investigated measurement invariance across phone- and web-based survey administration is Wang et al. (2017). Authors found that although most of the 15 measures studied demonstrated measurement invariance across administration modes, three did not. Authors argued that even though differences in administration mode may not affect measurement invariance under most circumstances, it is still an important aspect to consider when assessing measurement qualities of a measure because there may be circumstances under which administration mode makes a difference.

Although findings of the current study did not offer evidence for a factor structure similar to that identified by Briesch et al. (2013), results did appear to support the notion that the URP-IR offered a reliable indicator of teacher acceptability. Specifically, the coefficient alpha for all four supports was acceptable (.90-.94 across the four supports), indicating that the URP-IR has strong internal consistency and items may be appropriate to use to derive a total score. Additionally, item-level response patterns were largely similar between the current study and the Briesch et al. (2013) study. Specifically, those items expected to require reverse coding (i.e., items on the Family-School Collaboration factor, items on the System Support factor, and items

6 and 11) all functioned as expected. Due to these findings, it was determined that the URP-IR in the context of the current study was a reliable indicator of overall acceptability, despite poor model fit results. Thus, predictor analyses and analyses comparing ratings for the four supports were correspondingly completed using the total URP-IR score.

Potential Model Fit Improvements

Although findings of the current study did not support adequate fit to the hypothesized six-factor model, findings do provide some insight into what could be done in the future to potentially improve model fit. Because the focus of the current study was to determine whether the same factor structure was identified for the URP-IR across different types of interventions, and not to explore what a particularly good fitting model would be for a different intervention, no further analyses investigating fit of a different model were conducted. However, the data do offer some guidance for future work that might continue to explore how to best measure teacher acceptability of EL supports. A key theme from the data from the current study is that reducing redundancy among items would likely be helpful in improving the measurement characteristics of the measure. Examples of this can be found from the data on the Feasibility factor and on the Acceptability factor.

First, regarding the Feasibility factor, modification indices for each support called for covariances between a set of three items, indicating that these three items were likely measuring a construct that could be measured with a single item. These three questions were not the same for each support; however, it was always three within the factor. Upon investigation of the items, some questions that may be redundant are those asking about time involved in implementing the support or those asking about materials required.

Similar concerns regarding redundancy of questions arose on the Acceptability factor, indicating that the items on this factor may not be measuring aspects of acceptability distinct enough to justify multiple items. This factor has nine items in total and upon review of the suggested modifications and respective questions, it appears that redundancy of items is most problematic for items relating to whether the teacher thinks the support will benefit the targeted student(s) and those items relating to the teacher's motivation to implementing the support.

Predictors of URP-IR Score

Of the 12 ANOVAs run, only two revealed a significant relationship between the predictor variable and the overall URP-IR score. These two significant relationships were between the training for working with ELs and acceptability for support 1 (visual aids) and acceptability for support 4 (alternate response format). Additionally, the relationship between consultative support and acceptability for support 3 (inclusion of L1) was marginally significant. These findings are somewhat consistent with findings of existent researcher literature. However, the existing literature has found teacher training to be a more consistent predictor of teacher beliefs than is indicated by the current study, where it was a predictor only for some supports but not others. In the following sections, the findings for each predictor studied and relevance to existing literature will be expanded upon.

Teacher Training

Although teacher training was only found to significantly predict URP-IR scores in a couple of instances, those instances reflected a pattern similar to that identified in other studies. Specifically, it appears that a certain amount of teacher training may be necessary to identify those with significantly more favorable views of supports. For example, for support 1 (visual aids), it appeared that only teachers with more than the equivalent of one three-credit university

course rated visual supports significantly more favorably. Similarly, for support 4 (alternate response format), those teachers also rated alternate response format significantly more favorably. This pattern aligns with findings by Karathanos (2009) who reported that teachers with at least nine credit hours of instruction on working with ELs had significantly more favorable views towards inclusion of L1. Although the specific EL supports for which this pattern was found were different across these studies (visual aids and alternate response format vs. inclusion of L1), the general pattern held that teachers with at least a certain amount training (between three to nine or more credit hours' worth of training) were more likely to view the supports favorably. This pattern supports the importance of further exploring the role of teacher training as it relates to acceptability of instructional supports for ELs. At the same time, it is also critical to consider that both the current findings and findings by Karathanos (2009) were correlational. Thus, it is unclear whether more teacher training causes more positive opinions, or whether there may exist an underlying relationship, such that those teachers with highly positive opinions seek out additional training and are also more eager to implement supports for ELs.

Experience Working with ELs

Contrary to what was hypothesized, prior experience working with ELs was not found to be a predictor of URP-IR scores. To some degree, this appears at odds with both the social validity framework and literature that has found a significant relationship between teacher experience and their opinions about teaching ELs in mainstream classrooms. Multiple studies (e.g., Byrnes et al., 1997; Gandara et al., 2005; Shin & Krashen, 1996) have identified that teachers with more experience working with ELs hold more positive opinions towards including ELs in mainstream classroom, as well as their ability to support ELs in mainstream education. The social validity framework also posits that those who have more experience working with a

certain population would find proposed supports more acceptable (Carter, 2010). However, other studies specific to examining teacher opinions of instructional supports for ELs are in line with the findings of the current study (e.g., Karathanos, 2009; Lee & Oxelson, 2006). Specifically, as was found in the current study, both Karathanos (2009) and Lee and Oxelson (2006) found no correlation between teachers' experience working with ELs and their opinions towards inclusion of L1. One potential explanation for why a relationship was not found in the current study is that the overall strongly positive ratings teachers provided of the supports, in which there was somewhat limited variation, may have made it harder to detect a relationship. Another potential explanation is that the supports in the current study (with the exception of inclusion of L1) may be similar to supports teachers already provide to non-EL students, and thus the extent to which teachers find them acceptable may not depend on whether they have experience working with ELs or not. For example, the word wall strategy adapted in this study as an example of a vocabulary support for ELs was originally developed as an intervention to increase sight word learning for all students (Cunningham, 2004). Further, the SIOP model, which was used to identify supports for the current study, does not limit instructional supports solely to those that may benefit ELs. Rather, it aims to develop strategies that support all students, with specific focus on ensuring that strategies take into consideration the unique needs of ELs (Echevarría et al., 2014). Lastly, findings from the qualitative component of the study indicated that many teachers commented especially on visual aids and vocabulary supports as supports that they view as helpful for all students, not just ELs. This suggests that these teachers may have favorable views of the supports regardless of whether they have previously worked extensively with ELs or not.

Consultative Support

Contrary to what was hypothesized, teacher report of consultative support was not found to predict URP-IR scores. Although the relationship between consultative support and support 3 (inclusion of L1) was marginally significant, and in the expected direction, the majority of findings indicated no relationship between consultative support and URP-IR scores. This conflicts with the social validity framework, which posits that more consultative support is related to higher social validity of a support (Carter, 2010). To date, we are unaware of further prior research that has specifically explored the relationship between consultative support and teacher opinions of EL instructional supports. However, other work outside of a narrow focus on consultation does suggest a relationship between social networks and social validity. Specifically, Neal et al. (2020) found that principals with more support within their district had more favorable views towards a system-wide drop-out prevention intervention. Given these findings, consultative support, or perhaps social support, more broadly, continues to be an important predictor to study.

A possible explanation for non-significant findings in the current study is that participants overwhelmingly reported low amounts of consultative support received. Those teachers categorized as receiving the most consultation in the current study reported receiving ten or more hours in one school year and only made up 15% of the sample. The skew towards limited to no consultative support for mainstream teachers working with ELs that was identified in the current study may have limited the degree to which it was possible to detect a relationship. Limited consultative support has been documented in prior research, indicating that mainstream teachers report feeling “on their own” when it comes to supporting ELs in their classrooms (Snyder, 2020). Further, Bell and Baecher (2012) reported that although extensive (i.e., frequent and

consistent) and formal (e.g., scheduled meetings) consultation may be most effective at supporting ELs' needs in mainstream settings, ESL teachers rarely engage in this type of consultation.

Lastly, another challenge to be aware of with this predictor variable is that consultation support varies considerably in terms of its nature and quality (Bell and Baecher, 2012), which may make it difficult to accurately measure. For example, ESL teachers may consult with teachers in various formats, both informally (e.g., email, stopping by the classroom, talking in the hallway) or formally (e.g., consultation on lesson planning, participation in grade-level meetings, consulting on long-term student goals). Bell and Bacher (2012) found that ESL teachers engage in informal consultation much more frequently, which may make this a difficult construct to quantitatively measure for the purposes of predictor analyses. The current study defined consultative support as the number of hours received, on average, in a single year. It did not ask about the type of consultation received or consultative experience over the course of participants' teaching careers. More rigorous measurement of consultative support may be a valuable focus of future study to further explore this construct and its relationship to teacher views of instructional supports for ELs.

Differences in URP-IR Total Score

Although the lack of model fit presents concerns regarding the quality of this measure, the fact that the internal reliability of the total measure was strong for each support (.90-.94) and the substantial expert vetting done during initial development and revision of the instrument (Chafouleas et al., 2009; Briesch et al., 2013) suggests it is a reasonable measure for examining general acceptability of EL supports. Therefore, differences in scores between each support are discussed using the total score in the following section. Comparison of the URP-IR across

different supports is a valuable addition to the current literature on the URP-IR, as previous studies of this measure focused predominantly on its measurement qualities. Although the URP-IR does not yet appear to be strong measure of teacher opinions of EL supports and would benefit from additional measurement work, it appears to be sensitive enough to detect significant differences in teacher opinions. More specifically, of the four supports studied in the current study, support 1 (visual aids) was rated most favorably, followed by support 4 (alternate response format), support 2 (vocabulary supports), and support 3 (inclusion of L1).

There are a few potential explanations for the differences in average ratings by specific support. First, the supports varied to some extent in their level of complexity and resources required. For example, support 1 (visual aids), which received the highest overall ratings, requires less time for preparation than a support like inclusion of L1, which received the lowest overall scores. In the vignettes provided to teachers in the current study, visual aids required the teacher to find visuals and include them in instructional materials. As was found in the qualitative interviews of the current study, this may already be a practice teachers regularly engage in, as many reported using visual aids for all students. This is also supported by responses to item 4 on the URP-IR (“the support procedures fit easily in with my current practices”), which were considerably lower for inclusion of L1, on average, than for visual aids. Inclusion of L1, as it was presented to teachers in the current study, required teachers to identify an online translator and reference materials in the ELs’ L1 ahead of the lesson. This may take much more preparation time for teachers, especially if it is not something that is part of their existing practices. Item level responses appear to support this notion, as teachers, on average, rated items related to feasibility, such as “material resources are reasonable” or “the support is too complex to carry out”, as more favorable for support 1 than for support 4. In the social validity literature,

the time required to prepare or implement an intervention has consistently been found to be an important aspect in how overall acceptable consumers find an intervention (e.g., Elliott et al., 1984; Martens et al., 1985; Witt & Martens, 1983).

Further, supports with higher ratings, such as visual aids and alternate response format may be perceived as more effective for ELs to access instructional content, thus increasing their general acceptability to teachers. In contrast, teachers may be more unsure about the effectiveness of a support such as inclusion of L1 and may thus rate it lower in terms of acceptability. As indicated by item-level responses, teachers in the current study reported lower overall enthusiasm for and commitment to (e.g., items 2 and 17) inclusion of L1 compared to visual aids. Lowest ratings towards inclusion of L1 as an instructional support for ELs appears consistent with previous literature. In their review of the literature, Pettit (2011) identified several studies that reported teachers' hesitancy towards use of L1. For example, teachers have been found to believe that continued use of L1 interferes with English language acquisition (Reeves, 2006; Walker et al., 2004) and is thus detrimental to the academic achievement of ELs. This belief has been found to expand even beyond the classroom setting, as studies have found that many teachers believe that any use of L1 (whether in the classroom or at home) slows English language acquisition (Pettit, 2011).

Implications of Qualitative Findings

Considerable research has focused on treatment acceptability, Wolf's (1978) second level of social validity and the quantitative focus of the current study. Limited empirical work has focused on the other two levels of social validity proposed by Wolf (1978): social significance of the goals and social importance of the effects. These two levels were the focus of the qualitative component of the current study. In order to offer a more complete understanding of the social

validity of EL supports among teachers according to Wolf's (1978) social validity framework, implications of the qualitative findings in light of quantitative findings will be discussed in the following section.

Social Significance of the Goals

Regarding the social significance of the goals, qualitative results indicated that teachers reported a variety of different goals they may consider when selecting a support for ELs. Most often emphasized were goals such as increasing sense of belonging and engagement in instruction. The SIOP model, which was used to identify supports for the current study, focuses strongly on increasing ELs' access to content instruction. Given these findings, there may be a potential misalignment between the goals of the supports proposed in the current study and the goals most valued by teachers. Social validity theory would posit that when there is an alignment between the goals of a support and the goals of the person tasked to implement it, social validity and subsequent implementation increases (Carter, 2010). However, when teachers in the current study were asked to rank five given goals – where accessibility was one of the options – accessibility was consistently ranked in the middle. Meanwhile, other goals such as increase in engagement or sense of belonging were most often ranked as most important. These findings are in slight contrast with quantitative findings, which demonstrated high acceptability scores for all four supports. It therefore appears critical that consultants proposing supports assess both the goals that are important to the teacher, as well as acceptability of the proposed supports, as this will provide a more complete understanding of how motivated the teacher might be to implement the given support.

Another consideration regarding the social significance of goals and its effect on social validity is that teachers may vary greatly in what they find important. This finding emerged both

when asking teachers about their goals for ELs in mainstream science instruction and when asking them about which barriers ELs experience in this setting. Among the nine teachers who participated in the interviews, a wide range of answers were provided. In analyzing patterns in responses, their answers were sub-themed into academic (e.g., lack of vocabulary, difficulties with comprehension) and social-emotional (e.g., sense of not belonging). These findings highlight the need for consultants to carefully assess consultee goals as part of their consultative work. In fact, Erchul and Martens (2010) highlight goal setting as a critical component of their problem-solving approach to school-based consultation, and state that this should occur prior to intervention generation and implementation. This assures that the intervention ideas generated will address the problem identified by the consultee (Erchul and Martens, 2010).

Social Importance of the Effects

Regarding the effects of the four supports, most teachers generally reported feeling as though the proposed supports would help ELs overcome barriers, both academic and social-emotional. This is an important finding because, according to the social validity framework, if teachers find proposed supports both highly acceptable and believe they will work, implementation and high fidelity are much more likely (Carter, 2010). In the current study, how effective teachers believed the supports to be was largely consistent with how acceptable they found the supports to be based on the high URP-IR ratings.

Despite generally high perceptions of effectiveness for the four supports, in the instances in which concerns were raised about the effectiveness of the supports, it was typically about support 3 (inclusion of L1). Similarly, support 3 received the lowest URP-IR ratings overall. In qualitative interviews, concerns were raised about whether inclusion of L1 interferes with students' English acquisition. Additionally, concerns were raised about teachers' abilities to

implement the support in a way that is helpful for ELs. This is also consistent with previous research, which has indicated that teachers often have misconceptions about the effectiveness of incorporating L1 despite the evidence supporting its effectiveness (Pettit, 2011). This may be important for consultants to consider when proposing supports to teachers. An important area for future research may therefore be identifying avenues to increase teacher understanding of the effectiveness of incorporating L1.

Lastly, regarding effectiveness, there was some inconsistency among teachers regarding whether the proposed supports can be effective for all ELs (and potentially non-ELs) or not. Some felt that supports were highly universal while others felt that it was important that supports be adjusted on an individual-needs basis. Although both of these groups largely agreed that supports could be beneficial for ELs, those that felt supports needed individualization reported more hesitancy about implementing supports. These teachers were concerned about the feasibility of individualizing supports, especially in classrooms with a high degree of linguistic and cultural diversity. Based on this finding, it may be critical for consultants to assess teachers' opinion on this, as they may be unlikely to implement a support that they feel does not work well for multiple students or that will take a lot of work to tailor to every student's individual needs.

Limitations

Several limitations of this study are important to highlight. A major limitation is the use of the overall score of the URP-IR despite poor model fit of a single-factor model. Minor limitations include the low response rates and corresponding generalizability of the participant sample. Additionally, analyses exploring the measurement qualities of the URP-IR were limited to only a CFA. Lastly, a reliance on self-report measures. Each of these limitations is further described below.

First, findings from research question one indicated poor model fit for the six-factor model identified by Briesch et al. (2013); this subsequently limited the analyses possible for research questions two and three. Had the six-factor model been a strong fit within the current study, analyses for research questions two and three could have been split by subscale scores rather than involving only the total score. This could have added additional nuance to findings and aided with implications for future research and practice. Additionally, a subsequent major limitation of the current study is use of the total URP-IR score despite analyses of a single-factor model indicating poor model fit. Due to this, results of analyses utilizing the total URP-IR score should be interpreted with caution.

Second, a low response rate was obtained despite numerous adjustments made to recruitment efforts. A limitation of this study is therefore that the teachers who did participate likely are not a strong representation of the general teacher population. Indeed, the demographics of the study sample were slightly different from the general population of teachers in the United States. Of participants for whom demographic data were available, 93% identified as female, 88% identified as White, and 68% reported having a master's degree. Comparatively, the NCES reported that in the 2017-18 school year approximately 89% of elementary teachers identified as female, 79% of teachers identified as White, and 55% of teachers held a master's degree (NCES, 2021). The current study therefore predominantly describes the views of teachers who are disproportionately White and have higher levels of education, which should be taken into account when interpreting findings. In addition, due to the low response rate, imputation methods were utilized to increase the number of responses for analysis. Imputation methods are limited in their ability to reflect true response patterns and should also be taken into account when interpreting findings.

Another limitation of the current study is that teachers were asked to self-report the predictor variables of interest. Due to the length of the survey, simplicity in measuring these questions was prioritized. However, this may have potentially weakened the measurement quality of the variables. This was a particular limitation for measurement of teacher experience with ELs. The number of ELs in the teacher's classroom during the course of one school year was used as a proxy to measure this variable. A stronger measurement approach could have been to measure teacher experience more thoroughly through multiple questions that better capture experience; for example, assessing number of ELs worked with over the course of their career, level of need of ELs worked with, and experience utilizing instructional supports targeted at ELs. Similarly, for the other two predictor variables, participants may not have accurately remembered all training hours and university courses they took on working with ELs or may not have been able to accurately estimate the number of consultation hours they received. Given these limitations, future research may look to measure these variables more rigorously.

Implications for Future Research

Measurement Work

Given the findings of the current study, additional measurement work would be helpful to further establish the URP-IR as a quality measure of teacher acceptability of academic interventions or supports. Specifically, exploratory work should be conducted to identify whether another factor structure may be more appropriate or if reduction of items would improve model fit. The Acceptability and Feasibility subscales, in particular, are in need of further attention as most of the modifications were applied to these two subscales. Additionally, future measurement work should continue to evaluate the factor structure of multiple supports or interventions separately, as was done in the current study. This is because in the current study, all four

supports showed poor model fit, yet each support varied in which modifications were recommended. Further, future measurement work should also consider differentiating supports or interventions for different content areas. The current study focused specifically on instructional supports during science instruction, which may not generalize to considerations teachers may have for supports or interventions for other content areas. Science has, more so than other content areas, evolved towards inquiry-based instruction, and supporting ELs in such a setting may differ from supporting them in other content areas.

Lastly, regarding additional measurement work, the current study assessed academic supports for English Learners. Most social validity work has focused on behavioral interventions for individual students (Silva et al., 2020). The URP-IR was developed with the intention to be used across various interventions/supports and populations (Briesch et al., 2013). More research of the measurement qualities of the URP-IR will be necessary with other supports (e.g., academic supports for non-EL students) or populations (e.g., students with learning disabilities) to determine if the six-factor structure identified by Briesch et al. (2013) holds in these contexts.

Predictors of Acceptability

In addition to measurement work, future studies should also continue to assess the relationship between teacher characteristics and ratings of acceptability. The most promising predictor to continue investigating is the effect of teacher training on ratings of acceptability, as it was the only significant predictor in the current study. Specifically, focusing on rigorous assessment of pre-service training (i.e., university coursework) appears promising, as both the current study and results from Karathanos (2009) indicated that a certain amount of coursework (perhaps more than a 3-credit course) is most closely related to more positive views of supports for ELs. More rigorous assessment could help more clearly define the type of training teachers

are receiving in terms of content covered and the amount of coursework that is dedicated to coverage of supporting ELs. This may help more clearly indicate how to potentially improve teacher training beyond simply providing more of it. Additionally, experimental studies that provide training to randomly selected groups of university students would be a strong addition to the literature, as both the current study and Karathanos (2009) were purely correlational studies and can make no interpretation of whether it was training that affected teacher opinions or whether teachers with already favorable opinions seek out additional training in this area.

Implications of URP-IR Score

Despite some measurement limitations identified in the current study, the URP-IR continues to be a promising tool for the measurement of social validity. It builds upon prior measures and comprehensively assesses aspects of social validity most likely to affect usage of a support based on theory and prior research. Additional measurement work would strengthen it as a measurement tool for interventions beyond behavioral ones. Further, future research work should also look towards identifying the practical implications of the URP-IR score. Results of the current study showed that teacher opinions regarding the four supports significantly varied, despite all four receiving high average ratings. Future research should focus on assessing the relationship between the URP-IR score and key implementation variables, such as uptake of the intervention and implementation fidelity. Based on the results of the current study, one might expect that support 1 (visual aids) is most likely to be used by teachers, as the support with the highest URP-IR score; in comparison, support 3 (inclusion of L1) may be the support least likely to be used by teachers, as it received the lowest average URP-IR score. However, until further research is conducted examining these relationships, this remains unclear. Additionally, if research does identify this relationship to be supported, further research should also focus on

ways to increase acceptability of effective supports. Specifically, increasing acceptability of supports such as inclusion of L1 may be particularly helpful given that there may exist some teacher hesitancy despite it being a critical piece of linguistically-responsive instruction (August & Shanahan, 2006; Cummins, 2007; Settlage, 2014).

Implications for Practice

Given the findings of poor model fit of the URP-IR for academic supports for ELs, there currently is not enough evidence supporting the use of subscale scores for decision-making. Therefore, when using the UPR-IR for academic supports, consultants should utilize the total score only or evaluate responses on an item-by-item basis to identify areas of potential concern. The URP-IR could thus be used as a starting point for discussion between consultants and teachers. For behavioral interventions similar to those studied by Briesch et al. (2013), it seems appropriate to continued using the subscale scores for the URP-IR to inform decision-making.

Another implication for school-based practitioners to consider, given the findings of the current study, is that all four proposed supports appear highly acceptable to teachers working with ELs. School-based consultants such as school psychologists are in a position to further encourage use of these supports, if they are not being used in their schools already. Further, school psychologists can then consult on appropriate use of the supports by helping teachers decide when they are necessary and when they may need to be adjusted to fit a student's individual needs. Lastly, school psychologists can support teachers in evaluating the effectiveness of the supports by collecting progress monitoring data. All four of the proposed supports in the current study appear to be strong candidates for school psychologists to recommend and/or monitor, given the strong favorable views towards them.

Additionally, although all four supports are strong candidates for school psychologists or other school-based consultants to promote, the current study indicates that teachers may be more hesitant to utilize inclusion of L1 than the other supports studied. This hesitancy conflicts to an extent with the strong evidence base indicating that inclusion of L1 is an effective and culturally responsive practice (CRP) to support ELs. Given this potential disconnect, school psychologists or other school-based consultants may be in a particularly unique position to utilize the URP-IR to understand teacher concerns and to identify ways to better support teachers in implementing supports such as inclusion of L1. Use of inclusion of L1 or other CRPs broadly allow teachers to demonstrate value and respect for students' diverse abilities in the ways they approach instructional planning, use of language in their classroom, relationships with students, and connection with their students' lived cultural experiences (Linan-Thompson et al., 2018). More specifically, studies have indicated that when inclusion of L1 (also called "translanguaging") is embraced, celebrated, and utilized as means to enhance meaning making, students' academic performance, confidence in English skills, and identity development are enhanced (García and Sylvan, 2011; Palmer et al., 2014).

Lastly, the qualitative findings of the current study highlighted the importance of all three parts of social validity instead of sole focus on acceptability. Qualitative results provided insight that teachers may have varying goals for ELs, which may not directly align with how acceptable they find the recommended support to be. Social validity theory would posit that if acceptability is high but importance of the goals is low, implementation of the support would be negatively affected. Similarly, if teachers do not believe the recommended support will work, yet they find it highly acceptable, implementation could also be negatively affected. Thus, consultants should strive to assess teacher opinions of all three levels of social validity. The URP-IR primarily

assesses acceptability, and consultants may need to supplement assessment of social validity with face-to-face discussion. As a guideline to formulating questions for discussion, consultants can use the semi-structured interview by Gresham and Lopez (1996), or the adapted questions utilized in the current study (appendix C).

Summary and Conclusion

A variety of empirically-supported instructional supports exist to increase access to instruction for ELs in mainstream classrooms. However, the extent to which these supports are socially valid and used by teachers is unclear. The current study indicated that the six-factor structure previously identified for an existing measure (URP-IR; Briesch et al. 2014) was not a strong fit when evaluating acceptability of four instructional supports for ELs in mainstream science instruction. Despite these measurement limitations, total scores indicated that teachers found all four supports highly acceptable, with visual aids as the most acceptable support (followed by alternate response format, vocabulary supports, and inclusion of L1). Variation in teacher training was found to be significant predictor of URP-IR ratings for two of the supports: visual aids and alternate response format. Prior experience working with ELs and consultative support were not found to be significant predictors of URP-IR ratings. Given the findings of the current study, future research may consider additional measurement work investigating the factor structure of the URP-IR for different interventions/supports. Additionally, school-based consultants may consider inclusion of social validity assessment in their consultative work.

APPENDICES

Appendix A: Instructional Support Vignettes

All names are pseudonyms

Visual aids:

Definition: Visual aids include non-linguistic and non-written pictorial or graphical representations of content and are usually provided alongside written text to allow for greater meaning-making for ELs.

There are many possible ways to use visual aids during science instruction; here is just one example application: During a water cycle lesson, Ms. Smith provided several visual aids to her EL students. She provided a figure showing the water cycle with pictures for her ELs to follow along with as she taught. On the associated worksheet activity, Ms. Smith also provided visuals to clarify certain words, phrases, or idioms. To create these visual, Ms. Smith followed these steps:

She picked words, phrases, or idioms likely to pose a challenge to ELs In her visuals, she only represented one or two of these words, phrases, or idioms In her visual, she represented concrete concepts (e.g., objects, actions) instead of complex ideas

INSTRUCTIONS: Please respond to the following items specifically about the use of *Visual Aids* for ELs in your classroom during all of your own science lessons (during a regular, non-COVID-19 year). If you do not have any ELs in your classroom, imagine providing this support if you had an EL in your classroom. If you have not taught science for 3rd or 4th graders, please imagine doing so as you are answering the questions.

Vocabulary support:

Definition: Vocabulary supports include provision of definitions or explanation of academic vocabulary that is necessary for successful comprehension of the lesson or activity.

There are many possible ways to use vocabulary support during science instruction; here is just one example application: During a water cycle lesson, Ms. Smith utilized a Word Wall as vocabulary support for her EL students. To do so, she followed these steps:

She pre-selected key vocabulary from the water cycle lesson. She chose general academic vocabulary, which are those words that are not specific to one subject or lesson, but rather those that relate to the learning or task process. In Ms. Smith's case, she selected "cycle," "process," "rise," "drops," "heats," "lake," "river," "vapor," and "cloud." She put these words on a large poster hung in her classroom. Each word is accompanied by brief explanations. Before beginning the water cycle lesson, Ms. Smith reviewed the key vocabulary with the whole class. She encouraged students to ask for reminders during the lesson if they forgot the meaning of a word.

INSTRUCTIONS: Please respond to the following items specifically about the use of *Vocabulary Support* for ELs in your classroom during all of your own science lessons (during a regular, non-COVID-19 year). If you do not have any ELs in your classroom, imagine providing this support if you had an EL in your classroom. If you have not taught science for 3rd or 4th graders, please imagine doing so as you are answering the questions.

Inclusion of L1:

Definition: L1 is an EL's native language. Incorporation of L1 means allowing students to use their L1 to access content.

There are many possible ways to use incorporation of L1 during science instruction; here is just one example application: During an independent activity about the water cycle, Ms. Smith's students were instructed to write about and explain each phase of the water cycle. Students were supposed to use tablets to access a website (provided by Ms. Smith) that gives more information about the water cycle. To make this activity accessible to her ELs, Ms. Smith followed these steps:

She provided her EL students with a website for an online translator that allowed them to translate words from English to their L1 and vice versa. She showed her EL students how to use the translator at the beginning of the activity. Ahead of the lesson, she found alternative reference materials (e.g., websites, videos) in the L1 of the student. She provided these to her ELs and encouraged them to use them.

INSTRUCTIONS: Please respond to the following items specifically about the use of *Incorporation of L1* for ELs in your classroom during all of your own science lessons (during a regular, non-COVID-19 year). If you do not have any ELs in your classroom, imagine providing this support if you had an EL in your classroom. If you have not taught science for 3rd or 4th graders, please imagine doing so as you are answering the questions.

Alternate Response Format:

Definition: Allowing alternate response format means to allow students to demonstrate their knowledge or understanding in a way other than what is typically expected for a task and that does not require substantial English language proficiency.

There are many possible ways to use alternate response format during science instruction;

here is just one example application: Ms. Smith instructed her students to demonstrate their understanding of the water cycle by writing the steps out in narrative form. For her EL students, Ms. Smith modified the activity by following these steps:

She considered alternate format options that would allow her to assess her students' understanding of the key concepts. She instructed her EL students to draw the water cycle and label with key vocabulary, rather than write it out in full narrative form.

INSTRUCTIONS: Please respond to the following items specifically about the use of *Alternate Response Format* for ELs in your classroom during all of your own science lessons (during a regular, non-COVID-19 year). If you do not have any ELs in your classroom, imagine providing this support if you had an EL in your classroom. If you have not taught science for 3rd or 4th graders, please imagine doing so as you are answering the questions.

Appendix B: URP-IR Items

	Strongly disagree	Disagree	Slightly disagree	Slightly agree	Agree	Strongly agree
1. This support is a good way to handle the student's language barrier						
2. I would implement this support with a good deal of enthusiasm						
3. This support would not be disruptive to other students						
4. The support procedures easily fit in with my current practices						
5. The total time required to implement the support procedures would be manageable						
*6. I would not be interested in implementing this support						
7. Use of this support would be consistent with the mission of my school						
8. I would have positive attitudes about implementing this support						
9. Material resources needed for this support are reasonable						
10. This support is a fair way to handle the student's language barrier						
*11. This support is too complex to carry out accurately						
12. This support is an effective choice for addressing a variety of language barriers						
*13. Parental collaboration is required in order to use this support						
14. I would be able to allocate my time to implement this support						

15. The amount of time required for record keeping would be reasonable						
16. My administrator would be supportive of my use of this support						
17. I would be committed to carrying out this support						
18. I understand the procedures of this support						
*19. I would require additional professional development in order to implement this support						
20. Preparation of materials needed for this support would be minimal						
21. These support procedures are consistent with the way things are done in my system						
22. I understand how to use this support						
23. My work environment is conducive to the implementation of a support like this one						
*24. I would need consultative support to implement this support						
*25. A positive home-school relationship is needed to implement this support						
26. Implementation of this support is well matched to what is expected in my job						
27. I am knowledgeable about the support procedures						
*28. I would need additional resources to carry out this support						
*29. Regular home-school communication is needed to implement support procedures						

*lower scores on these items indicate more positive views towards supports. These items were reverse coded when calculating the total scale score. For factor analyses, items were not reverse coded.

Appendix C: Qualitative Interview Questions

1. Social Significance of Goals

1. What are the primary goals for an EL in mainstream science instruction? To what degree would [support] address these goals?
2. Please rank order the following by level of importance:
 - a. Access to instruction
 - b. Increase in science knowledge
 - c. Increase in English skills
 - d. Increase engagement in instruction
 - e. Increase feelings of belonging
3. What are the main barriers for ELs in mainstream education? How do these cause difficulties?
4. How would your EL students be affected if barriers were decreased? How would other students be affected? How would your teaching be affected?

2. Social Importance of Effects

1. How well do you think [support] would work?
2. What changes do you think you would observe as a result of [support]? Would the [support] make a difference in an EL's level of access to instruction?
3. Would an EL's level of access to instruction be similar to that of the average student in your classroom as a result of [support]? Why or why not?
4. Do you think you would be satisfied with the outcomes of [support]? Why?
5. Do you think [support] would work for all ELs? Why or why not?

6. Would you recommend [support] to other teachers? Why or why not? What aspects of [support] would you change before recommending it to other teachers?

Appendix D: Survey prenotification (deidentified)

Dear [first name],

I am a graduate student in the school psychology program at Michigan State University. For my dissertation, I am interested in how general education teachers think about supporting students who are English Language Learners in their classrooms. You have been randomly selected through Market Data Retrieval to participate in my research study. Participation is voluntary. This email is to inform you about the procedures of the study and the link to the study will come in a separate email.

You are being asked to participate in an online survey that will take about 20-30 minutes to complete. As a token of my appreciation for your participation, you will be entered into a drawing to receive one of five \$100.00 Visa gift cards. The first 500 participants will be eligible to enter the drawing. I will send a follow-up email when the drawing has closed.

Thank you ahead of time for your consideration. You will receive the link to the survey on XX/XX/XX. If you have any questions, please let me know.

Sincerely,

Sarina Roschmann (she/her/hers)
Doctoral Student, School Psychology
Department of Counseling, Educational Psychology & Special Education
Michigan State University
roschma1@msu.edu

Appendix E: Terms and Definitions

English Learners: English Learners (ELs) are students who are in the process of learning English. Many similar terms exist, including English Language Learners (ELLs), Dual-language learners, Bilingual/multilingual students, non-English-proficient, limited-English proficient (LEP), and English as a second language (ESL) students. Each of these terms brings certain connotations and modified and new related terms are continuously emerging. EL is the most commonly used term used in the literature currently and is also used by the National Clearinghouse for English Language Acquisition (NCELA), which is why it was selected for use in the current study.

Instructional Supports: Instructional supports are defined as practices, accommodations, or supports provided by teachers that increase access to content or instruction.

Mainstream: Mainstream is defined as a general education classroom setting in which content is typically presented primarily in English. It is not an ideal term due to its connotation of these settings as “regular” and anything else as “irregular.” However, it is most frequently used currently in the literature and a better term has not yet been identified (Pettit, 2011).

Mainstream teachers: Mainstream teachers are defined as those who teach in mainstream settings and are not direct providers of other support services (e.g., ESL, special education).

Non-ELs: Non-ELs are defined as those students whose primary language is English. These students are thus not expected to experience difficulties accessing content presented in English due to limited English proficiency.

Opinions: Opinions, among multiple possible terms (beliefs, attitudes, perceptions) was chosen because it is used by Carter (2010) in *The Social Validity Manual*. Opinions refers to what is obtained from “those receiving, implementing, or consenting to a treatment” and what is “then used to make decisions about current or future uses of the treatment” (Carter, 2010, p. 2).

REFERENCES

REFERENCES

- Abedi, J., Lord, C., Boscardin, C. K., & Miyoshi, J. (2001). *The Effects of accommodations on the assessment of Limited English Proficient (LEP) students in the National Assessment of Educational Progress (NAEP): CSE technical report 537: (300662004-001)* [Data set]. American Psychological Association. <https://doi.org/10.1037/e300662004-001>
- American Psychological Association (APA). (2002). Criteria for evaluating treatment guidelines. *American Psychologist*, 57, 1052–1059. <https://doi.org/10.1037/0003-066X.57.12.1052>
- August, D. E., & Shanahan, T. E. (2006). *Developing literacy in second-language learners: Report of the National Literacy Panel on Language-Minority Children and Youth*. Lawrence Erlbaum Associates Publishers.
- Bell, A. B., & Baecher, L. (2012). Points on a continuum: ESL teachers reporting on collaboration. *TESOL Journal*, 3(3), 488–515. <https://doi.org/10.1002/tesj.28>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Briesch, A. M., Chafouleas, S. M., Neugebauer, S. R., & Riley-Tillman, T. C. (2013). Assessing influences on intervention implementation: Revision of the Usage Rating Profile-Intervention. *Journal of School Psychology*, 51(1), 81–96. <https://doi.org/10.1016/j.jsp.2012.08.006>
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press.
- Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming (multivariate applications series)*. New York: Taylor & Francis Group.
- Byrnes, D., Kiger G., & Manning, M. L. (1997). Teachers' attitudes about language diversity. *Teaching and Teacher Education*, 13, 637–644.
- Chafouleas, S. M., Briesch, A. M., Riley-Tillman, T. C., & McCoach, D. B. (2009). Moving beyond assessment of treatment acceptability: An examination of the factor structure of the Usage Rating Profile—Intervention (URP-I). *School Psychology Quarterly*, 24(1), 36–47. <http://dx.doi.org.proxy2.cl.msu.edu/10.1037/a0015146>
- Carter, S.L. (2010). *The social validity manual: A guide to subjective evaluation of behavior interventions in applied behavior analysis*. Elsevier Inc.
- Cummins, J. (2000). *Language, power and pedagogy: Bilingual children in the crossfire* (Vol. 23). Multilingual Matters.
- Cummins, J. (2007). Rethinking monolingual instructional strategies in multilingual classrooms. *Canadian Journal of Applied Linguistics*, 10(2), 221–240.
- Cunningham, P.M. (2004). *Phonics they use: Words for reading and writing* (4th ed.). New York, NY: Harper-Collins College Press.

- Echevarría, J., Richards-Tutor, C., Changes, R., & Francis, D. (2011). Using the SIOP Model to promote the acquisition of language and science concepts with English learners. *Bilingual Research Journal*, 34(3), 334-351.
- Echevarría, J., Short, D., & Powers, K. (2006). School reform and standards-based education: An instructional model for English language learners. *Journal of Educational Research*, 99(4), 195-211.
- Echevarría, J., Vogt, M., & Short, D. J. (2014). *Making content comprehensible for elementary English learners: The SIOP model*. Pearson Education, Inc.
- Elliott, S. N., Witt, J. C., Galvin, G. A., & Peterson, R. (1984). Acceptability of positive and reductive behavioral interventions: Factors that influence teachers' decisions. *Journal of School Psychology*, 22, 353-360.
- Erchul, W. P., & Martens, B. K. (2010). *School consultation: Conceptual and empirical bases of practice*. Springer-Verlag.
- Furtak, E. M., Seidel, T., Iverson, H., & Briggs, D. C. (2012). Experimental and quasi-experimental studies of inquiry-based science teaching: A meta-analysis. *Review of Educational Research*, 82(3), 300-329.
- Gandara, P., Maxwell-Jolly, J., & Driscoll, A. (2005). *Listening to teachers of English language learners: A survey of California teachers' challenges, experiences, and professional development needs*. Santa Cruz, CA: Center for the Future of Teaching and Learning.
- García, O., & Sylvan, C. E. (2011). Pedagogies and practices in multilingual classrooms: Singularities in pluralities. *The Modern Language Journal*, 95(3), 385-400.
- Genesee, F., Lindholm-Leary, K., Saunders, W., & Christian, D. (2005). English language learners in US schools: An overview of research findings. *Journal of Education for Students Placed at Risk*, 10(4), 363-385.
- Gresham, F. M., & Lopez, M. F. (1996). Social validation: A unifying concept for school-based consultation research and practice. *School Psychology Quarterly*, 11(3), 204.
- Huerta, M., Garza, T., Jackson, J. K., & Murukutla, M. (2019). Science teacher attitudes towards English learners. *Teaching and Teacher Education*, 77, 1-9.
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33(7), 14-26.
<https://doi.org/10.3102/0013189X033007014>
- Karathanos, K. (2009). Exploring US mainstream teachers' perspectives on use of the native language in instruction with English language learner students. *International Journal of Bilingual Education and Bilingualism*, 12(6), 615-633.
<https://doi.org/10.1080/13670050802372760>
- Kazdin, A. E. (1980). Acceptability of alternative treatments for deviant child behavior. *Journal of Applied Behavior Analysis*, 13(2), 259-273. <https://doi.org/10.1901/jaba.1980.13-259>

- Kelley, M. L., Heffer, R. W., Gresham, F. M., & Elliott, S. N. (1989). Development of a modified treatment evaluation inventory. *Journal of psychopathology and behavioral assessment*, 11(3), 235-247.
- Khong, T. D. H., & Saito, E. (2014). Challenges confronting teachers of English language learners. *Educational Review*, 66(2), 210–225.
<https://doi.org/10.1080/00131911.2013.769425>
- Krain, A. L., Kendall, P. C., & Power, T. J. (2005). The role of treatment acceptability in the initiation of treatment for ADHD. *Journal of Attention Disorders*, 9(2), 425–434.
<https://doi.org/10.1177/1087054705279996>
- Lee, J. S., & Oxelson, E. (2006). It's not my job: K–12 teacher attitudes toward students' heritage language maintenance. *Bilingual Research Journal*, 30, 453–477.
- Lee, O., Quinn, H., & Valdés, G. (2013). Science and language for English language learners in relation to next generation science standards and with implications for Common Core State Standards for English language arts and mathematics. *Educational Researcher*, 42(4), 223–233. <https://doi.org/10.3102/0013189X13480524>
- Lennox, D. B., & Miltenberger, R. G. (1990). On the conceptualization of treatment acceptability. *Education and Training in Mental Retardation*, 25(3), 211–224.
- Linan-Thompson, S., Lara-Martinez, J. A., & Cavazos, L. O. (2018). Exploring the intersection of evidence-based practices and culturally and linguistically responsive practices. *Intervention in School and Clinic*, 54(1), 6-13.
- Martens, B. K., Witt, J. C., Elliott, S. N., & Darveaux, D. X. (1985). Teacher judgments concerning the acceptability of school-based interventions. *Professional Psychology: Research and Practice*, 16(2), 191–198.
<http://dx.doi.org.proxy1.cl.msu.edu/10.1037/0735-7028.16.2.191>
- Martiniello, M. (2009). Linguistic complexity, schematic representations, and differential item functioning for English language learners in math tests. *Educational Assessment*, 14(3-4), 160-179. doi: 10.1080/10627190903422906
- Mautone, J. A., DuPaul, G. J., Jitendra, A. K., Tresco, K. E., Junod, R. V., & Volpe, R. J. (2009). The relationship between treatment integrity and acceptability of reading interventions for children with Attention-Deficit/Hyperactivity Disorder. *Psychology in the Schools*, 46(10), 919–931. <https://doi.org/10.1002/pits.20434>
- McNeill, J. (2019). Social validity and teachers' use of evidence-based practices for autism. *Journal of autism and developmental disorders*, 49(11), 4585-4594.
- Mundfrom, D. J., Shaw, D. G., & Ke, T. L. (2005). Minimum sample size recommendations for conducting factor analyses. *International Journal of Testing*, 5(2), 159–168.
https://doi.org/10.1207/s15327574ijt0502_4
- National Association of School Psychologists (NASP). (2010). Model for comprehensive and integrated school psychological services.

- National Center for Education Statistics (2020). Characteristics of public and private elementary and secondary school teachers in the united states: Results from the 2017–18 National Teacher and Principal Survey.
- National Center for Education Statistics (2020, May). *English language learners in public schools*. https://nces.ed.gov/programs/coe/indicator_cgf.asp
- National Center for Education Statistics (2017). Table 223.10: Average National Assessment of Educational Progress (NAEP) science scale score, standard deviation, and percentage of students attaining science achievement levels, by grade level, selected student and school characteristics, and percentile: 2009, 2011, and 2015. https://nces.ed.gov/programs/digest/d16/tables/dt16_223.10.asp
- Neal, J. W., Neal, Z. P., Barrett, C. A., & Brutzman, B. (2020). Are Principals' Social Networks Associated with Interventions' Social Validity?. *School Mental Health*, 12(4), 812-825.
- Oliveira, A. W., Meskill, C., Judson, D., Gregory, K., Rogers, P., Imperial, C. J., & Casler-Failing, S. (2015). Language Repair Strategies in Bilingual Tutoring of Mathematics Word Problems. *Canadian Journal of Science, Mathematics and Technology Education*, 15(1), 102–115. <https://doi.org/10.1080/14926156.2014.990173>
- Palmer, D. K., Martínez, R. A., Mateus, S. G., & Henderson, K. (2014). Reframing the debate on language separation: Toward a vision for translanguaging pedagogies in the dual language classroom. *The Modern Language Journal*, 98(3), 757-772.
- Pennock-Roman, M., & Rivera, C. (2011). Mean effects of test accommodations for ELLs and non-ELLs: A meta-analysis of experimental studies. *Educational Measurement: Issues and Practice*, 30(3), 10-28. doi: <https://doi.org/10.1111/j.1745-3992.2011.00207.x>
- Pettit, S. K. (2011). Teachers' beliefs about English language learners in the mainstream classroom: A review of the literature. *International Multilingual Research Journal*, 5(2), 123–147. <https://doi.org/10.1080/19313152.2011.594357>
- Polat, N. (2010). A comparative analysis of pre-and in-service teacher beliefs about readiness and self-competency: Revisiting teacher education for ELLs. *System*, 38(2), 228-244.
- Reeves, J. R. (2006). Secondary teacher attitudes toward including English-language learners in mainstream classrooms. *Journal of Educational Research*, 99(3), 131–142.
- Reimers, T. M., Wacker, D. P., & Koepl, G. (1987). Acceptability of behavioral interventions: A review of the literature. *School Psychology Review*, 16(2), 212–227. <https://doi.org/10.1080/02796015.1987.12085286>
- Serafini, F. (2012). Expanding the four resources model: Reading visual and multi-modal texts. Pedagogies: An International Journal, 7(2), 150–164. <https://doi.org/10.1080/1554480X.2012.656347>
- Settlage, J., Gort, M., & Ceglie, R. J. (2014). Mediated language immersion and teacher ideologies: Investigating trauma pedagogy within a "Physics in Spanish" course activity. *Teacher Education Quarterly*, 41(3), 47-66.

- Shin, F. H., & Krashen, S. (1996). Teacher attitudes toward the principles of bilingual education and toward students' participation in bilingual programs: Same or different? *Bilingual Research Journal*, 20, 45–53.
- Short, D., Echevarría, J., & Richards-Tutor, C. (2011). Research on academic literacy development in sheltered instruction classrooms. *Language Teaching Research*, 15(3), 363–380.
- Silva, M. R., Collier-Meek, M. A., Coddling, R. S., & DeFouw, E. R. (2020). Acceptability assessment of school psychology interventions from 2005 to 2017. *Psychology in the Schools*, 57(1), 62–77. <https://doi.org/10.1002/pits.22306>
- Snyder, E. (2020). *English Language Learners in Multi-Tier System of Supports (MTSS) Reading Implementation: An Exploratory Study of Inclusion and Teacher Perceptions*. Michigan State University.
- Solano-Flores, G., Wang, C., Kachchaf, R., Soltero-Gonzalez, L., & Nguyen-Le, K. (2014). Developing testing accommodations for English language learners: Illustrations as visual supports for item accessibility. *Educational Assessment*, 19(4), 267–283. <https://doi.org/10.1080/10627197.2014.964116>
- Spirrison, C. L., Noland, K. A., & Savoie, L. B. (1992). Factor structure of the Treatment Evaluation Inventory: Implications for measurement of treatment acceptability. *Journal of Psychopathology and Behavioral Assessment*, 14(1), 65–79. <https://doi.org/10.1007/BF00960092>
- Sterling-Turner, H. E., & Watson, T. S. (2002). An analog investigation of the relationship between treatment acceptability and treatment integrity. *Journal of Behavioral Education*, 12.
- Symons, C. (2020). Instructional practices for scaffolding emergent bilinguals' comprehension of informational science texts. *Pedagogies: An International Journal*, 0(0), 1–19. <https://doi.org/10.1080/1554480X.2020.1738938>
- Tarnowski, K. J., & Simonian, S. J. (1992). Assessing treatment acceptance: The abbreviated acceptability rating profile. *Journal of Behavior Therapy and Experimental Psychiatry*, 23(2), 101–106. [https://doi.org/10.1016/0005-7916\(92\)90007-6](https://doi.org/10.1016/0005-7916(92)90007-6)
- U.S. Department of Education (2017). *Our nation's English learners*. <https://www2.ed.gov/datastory/el-characteristics/index.html>
- Villegas, A. M., Saiz de La Mora, K., Martin, A. D., & Mills, T. (2018). Preparing Future Mainstream Teachers to Teach English Language Learners: A Review of the Empirical Literature. *The Educational Forum*, 82(2), 138–155. <https://doi.org/10.1080/00131725.2018.1420850>
- Von Brock, M. B., & Elliott, S. N. (1987). Influence of treatment effectiveness information on the acceptability of classroom interventions. *Journal of School Psychology*, 25(2), 131–144. [https://doi.org/10.1016/0022-4405\(87\)90022-7](https://doi.org/10.1016/0022-4405(87)90022-7)

- Walker, A., Shafer, J., & Iiams, M. (2004). "Not in my classroom": Teacher attitudes towards English language learners in the mainstream classroom. *National Association for Bilingual Education Journal of Research and Practice*, 2, 130–160.
- Wang, M., Chen, R. C., Usinger, D. S., & Reeve, B. B. (2017). Evaluating measurement invariance across assessment modes of phone interview and computer self-administered survey for the PROMIS measures in a population-based cohort of localized prostate cancer survivors. *Quality of Life Research*, 26(11), 2973-2985.
- Witt, J., & Elliott, S. (1985). Acceptability of classroom intervention strategies. In T. Kratochwill (Ed.), *Advances in school psychology* (Vol. 4, pp. 251–288). Erlbaum.
<https://interlib.lib.msu.edu/remotearch/illiad.dll?Action=10&Form=75&Value=1245701>
- Witt, J. C., & Martens, B. K. (1983). Assessing the acceptability of behavioral interventions used in classrooms. *Psychology in the Schools*, 20(4), 510–517.
- Wolf, M. M. (1978). Social validity: the case for subjective measurement or how applied behavior analysis is finding its heart 1. *Journal of applied behavior analysis*, 11(2), 203-214.