

PILOT STUDY OF A “QUALITY OF USE” SCALE WITH AN ELEMENTARY
READING PROGRAM

A Dissertation

by

ZELMA JANE GRAGG

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2011

Major Subject: Educational Psychology

PILOT STUDY OF A “QUALITY OF USE” SCALE WITH AN ELEMENTARY

READING PROGRAM

A Dissertation

by

ZELMA JANE GRAGG

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Richard I. Parker
Committee Members,	Laura Stough
	Kimberly Vannest
	Luana Zellner
Head of Department	Victor Willson

May 2011

Major Subject: Educational Psychology

ABSTRACT

Pilot Study of a “Quality of Use” Scale with an Elementary Reading Program.

(May 2011)

Zelma Jane Gragg, B.A., University of Texas at El Paso;

M.Ed., University of Texas at El Paso

Chair of Advisory Committee: Dr. Richard I. Parker

This study developed a summative scale that could be administered in a short time period to determine the Quality of Use (QOU) of an intervention used by teachers. The scale can be completed in less than an hour using easily attainable information. The QOU scale was applied to an elementary reading program to determine if the program results were dependent upon the quality of the fidelity of teacher use.

The study focused on use of the Linguistic Pattern Series (LPS) portion of the Integrated Skills Method (ISM) Reading Program by 20 special education teachers in 13 elementary schools in San Antonio, Texas. Progress is measured by the use of the Decoding Skills Test (DST) (ISM Teaching Systems, Inc., 2004-b). To determine each teacher’s QOU, a summative scale was developed composed of five items (Initial Placement, Frequency of Direct Instruction, Materials – LPS, Materials – Literature/Test Prep, and Scheduling) that possessed low-moderate cohesiveness of $\text{Alpha}=.71$.

Results of the study showed a correlation between QOU summary scores and residualized DST Raw Score Grade Equivalent (RSGE) gains. The QOU could predict

$.771^2 = 50\%$ of score variance. This is a strong prediction for a non-student external measure in education.

DEDICATION

I would like to dedicate this to my grandfather, George Seth Barnum, who made sure, in the middle of the Great Depression, all three of his children, including his daughter, would complete college degrees. I also dedicate this to my father, Darold P. Barnum, who continued the ideals of higher education for all his children, female as well as male and to my older brother, Darold T. Barnum, who led the way obtaining his doctorate many years ago and always encouraged his little sister that she could attain any degree of education she wanted.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Richard Parker, and my committee members, Dr. Laura Stough, Dr. Kimberly Vannest, Dr. Luana Zellner, and retired professor, Dr. Linda Parrish, for their guidance and support throughout the course of this research.

Thanks also go to my friends, colleagues, and the department faculty and staff for making my time at Texas A&M University a great experience. I want to extend my gratitude to Dr. Ellis Richardson, who gave me access to his Integrated Skills Method Linguistic Pattern Series program data, and to San Antonio ISD administrators and teachers, who were willing to participate in the study.

I also would like to thank my husband, Wendell Gragg, for his love and encouragement and for being willing to move almost 700 miles so his wife could work on her doctorate. And most of all, I would like to thank my Lord Jesus Christ, without whom none of this would have been possible.

NOMENCLATURE

AYP	Adequate Yearly Progress
DST	Decoding Skills test
EPISD	El Paso Independent School District
FOI	Fidelity of Implementation
IDEA 2004	Individuals with Disabilities Education Act reauthorization of 2004
ILGE	Instructional Level Grade Equivalent
ISM	Integrated Skills Method
LDAA	Locally-Determined Alternate Assessments
LEA(s)	Local Education Agency(ies)
LEP	Limited English Proficient
LoU	Level(s) of Use
LPS	Linguistic Pattern Series
NCLB	No Child Left Behind Act
QOU	Quality of Use
RSGE	Raw Score Grade Equivalent
RPTE	Reading Proficiency Tests in English
RtI	Response to Intervention
SAISD	San Antonio Independent School District
SLD	Specific Learning Disability

SDAA II	State-Developed Alternative Assessment II
TAKS	Texas Assessment of Knowledge and Skills
TAKS-A	TAKS Accommodated
TAKS-M	TAKS Modified
TAKS-Alt	TAKS Alternative
TEA	Texas Education Agency
WWC	What Works Clearinghouse

TABLE OF CONTENTS

	Page
ABSTRACT	iii
DEDICATION	v
ACKNOWLEDGEMENTS	vi
NOMENCLATURE	vii
TABLE OF CONTENTS	ix
LIST OF FIGURES	xi
LIST OF TABLES	xii
CHAPTER	
I INTRODUCTION	1
Context	1
Statement of Problem	4
Statement of Need	7
Statement of Purpose	8
Research Questions	11
II REVIEW OF LITERATURE	12
Quality of Use Concept	12
Quality of Use Measures	14
Quality of Use Need	21
III METHOD	22
Overview	22
Context	24
Respondents	25
Intervention	27
Instrumentation	28
Design	34

CHAPTER	Page
Analysis by Research Question	36
Procedures	38
IV RESULTS.....	48
Preliminary Study Results	48
Research Question #1	51
Research Question #2.....	56
Research Question #3	57
V DISCUSSION AND CONCLUSIONS.....	63
Overview	63
Research Question #1	66
Research Question #2.....	71
Research Question #3	72
Implications	73
Limitations	74
Conclusions	76
REFERENCES.....	82
APPENDIX A	94
APPENDIX B	95
APPENDIX C	99
APPENDIX D	103
APPENDIX E.....	106
APPENDIX F	109
APPENDIX G	110
APPENDIX H	113
VITA	114

LIST OF FIGURES

	Page
Figure 1 Quality of Use Study Design.....	35

LIST OF TABLES

	Page
Table 1 Means and Standard Deviations of Scale Items	55

CHAPTER I

INTRODUCTION

Context

On January 8th in 2002, the President signed the No Child Left Behind Act of 2001 (Texas Education Agency, 2005-a). This legislation (NCLB), with its accountability requirements, was enacted to assist in creating systemic school change. One long-term goal of NCLB is student proficiency in both mathematics and reading/language arts by 2013-14 school year (No Child Left Behind Act of 2001, 2002; Texas Education Agency, 2002; U. S. Department of Education, 2002). Per the Act requirement that each state establish an Adequate Yearly Progress (AYP) definition (U.S. Department of Education, 2002), Texas implements NCLB through the use of a single AYP definition for all state schools. AYP in Texas includes the testing, with baseline performance standards, of all students in mathematics and reading/language arts. The baseline performance standards will increase until they reach 100% by 2013/14. The campuses and districts must also meet participation standards. Both performance and participation apply to all students as well as various student groups: LEP (Limited English Proficient), special education, economically disadvantaged, African American, Hispanic and White student groups. Other measures, such as high schools meeting a graduation rate standard, are also included (Texas Education Agency, 2005-a).

This dissertation follows the style of *Behavior Modification*.

All schools are reviewed, whether they are Title I, Part A (Title I) schools or not. NCLB extended accountability provisions which formerly applied just to campuses and districts receiving Title I funds (Texas Education Agency, 2005-a). Title I is a program, designed to help all children reach state academic standards, that provides financial assistance to Local Education Agencies (LEAs) and their schools who have high numbers or percentages of poor children (U.S. Department of Education, 2006). Non-Title I schools are not necessarily held to the same school improvement interventions as the Title I campuses; though they are still subject to amending their individual school improvement plan for not meeting AYP targets (Texas Education Agency, 2005-a). Title I school campuses, failing to meet AYP for at least two consecutive years, must revise or develop a two-year plan to address NCLB requirements. One of the requirements of this plan is that the campus offer supplemental educational services for students that come from low-income families. These services include tutoring and academic enrichment which must be high quality and research based (Texas Education Agency, 2002).

High quality and research based instruction is part of the Response to Intervention (RtI) model addressing student needs by using a continuum of services (Texas Education Agency, 2007-a). RtI is part of the Individuals with Disabilities Education Act reauthorization of 2004 (IDEA 2004). This general education includes assessment and intervention with learners who are struggling (NASDSE & CASE, 2006). It also becomes part of the process for determining if a student has a specific learning disability (SLD). IDEA 2004 states that, as part of evaluation procedures, a

process may be used that determines response to scientific, research-based intervention (Individuals with Disabilities Education Improvement Act of 2004, 2004).

RtI uses an intervention model that is tiered. The first tier, Tier I, is focused on group interventions. These are designed for use with all students as preventive and proactive measures. This curriculum is expected to be efficacious with about 80% to 85% of students. Tier II interventions are still group targeted interventions, but are designed to help about 15% of students. These Tier II interventions are in addition to Tier I curriculum. Tier III brings intensive interventions to about 5% of students. These are individualized and will be adjusted as students reach targeted skills. RtI, from IDEA 2004, may have a major influence on the monitoring of children's progress. It can be used to address NCLB's challenges and improve the outcomes for all students (NASDSE & CASE, 2006).

The requirements of NCLB are also being felt by district reading programs such as those funded under Reading First (Stewart, 2004). Reading First is a federal formula grant giving support to districts implementing programs and assessment tools based on scientifically based reading research (U. S. Department of Education, 2009). NCLB sets a new focus for this instruction, requiring the use of scientific, research based reading instruction. There are five components of reading that NCLB focuses on: phonics, phonemic awareness, vocabulary, comprehension and fluency (Stewart, 2004). One of the programs in Texas adopted by several large districts in that state, that has these components, is the Integrated Skills Method (ISM) Reading Program. This research based program includes the use of the Decoding Skills Test (DST), developed by Dr.

Ellis Richardson and Dr. Barbara DiBenedetto-Corona and cited by TEA as a valid reading development measure. The phonic/linguistic portion of program is the Linguistic Pattern Series (LPS). The ISM Reading program, as part of Fordham University's Interdependent Learning Model, was part of a nation-wide research program, Project Follow Through. This research program, sponsored by the U.S. Office of Education, for two years assessed the program effects in Atlanta, GA. The seven research program schools' first through third graders' results were compared to all of the Atlanta first through third graders' results after two years of the program. At the research program schools, there was a 20% increase in the number of students scoring at or above 50th percentile. The improvement at the comparison schools was only about 3% (ISM Teaching Systems, Inc., 2003). Yet, even in a program showing positive results, ISM Teaching Systems, Inc. (2004-a) noted in their San Antonio ISD Annual Dyslexia Program Report for 2003-04 school year that the program produced uneven effects across schools. They suggested that this could be due to the implementation level of the program at each school.

Statement of Problem

When a school adopts a new program, it is important to determine whether a program works and, if so, why it works. In order to determine the viability of a program, the researcher must first know whether a program is actually used, and at what level, i.e. its Quality of Use (QOU). QOU of a program logically would impact student learning results. In the case of the Integrated Skills Method (ISM) program, per the

designer/publisher, there was a dramatic difference between groups of students that gained the most and the least from the program in the San Antonio school system. The highest group (Catching Up group) gained an average of more than two years between pretest and posttest. This is versus the lowest group (No Effect group) where the gain was an average of seven months which was less than the predicted gain without the program (ISM Teaching Systems, Inc., 2004-a). It would be very important to the district and the program to know whether it was the quality of program implementation, program internal weakness, or other factors that made the difference in the various student groups. Possible contributing factors affecting the student groups would include school demographics, teacher demographics, or student demographics. Identifying causes of the noted differences would help districts know whether the program could be used, if well implemented, with most students experiencing reading difficulties or if the program should be matched to school, teacher or student demographics.

To determine how well a program is being implemented, researchers have several options. Direct observation is often considered desirable as observers can report what they have directly seen. Informal observations can give an evaluator context for the program. However, for evaluating a program, they are not sufficient. Evaluators need to determine what they need to observe and when. Then how they are going to record it must be planned. The observers must be trained to use these recording methods (King, Morris & Fitz-Gibbon, 1987). One difficulty with observation is the need to observe many aspects of the use of the program for long time periods (Loucks, Newlove, and Hall, 1975). These long term observations can last for months, as in a participant-

observer study in which observations continued for 18 months (Hull & Zacher, 2007). Other observations have continued for three years, with the researcher observing every week for two to four days during school hours (Brinegar, 2010). The activities observed could be changed just by the act of having an observer present (Loucks, Newlove, and Hall, 1975).

An alternative to observations is collecting data on use level by using interviews (Loucks, Newlove, and Hall, 1975). The use of interviews is one form of self-reporting by personnel using a program. It involves in person discussions of experiences in program use (King, Morris & Fitz-Gibbon, 1987). This often may require a number of interviewers, each of whom will need to be trained and supervised. It does permit the most communication between the interviewer and the personnel responding. It tends to gain and retain respondent cooperation. Nonresponse bias is very low for interviews (Alreck & Settle, 1995). However, interviews can be time consuming so the other form of self-reporting, using questionnaires, can allow more efficient information collection (King, Morris & Fitz-Gibbon, 1987).

Questionnaires may include questions about frequency and the length of activities as well as just confirming these activities occurred if important to program implementation. Other important information to obtain would include which students participated, which students were non-involved and even if students were not paying attention. However, questionnaires alone will not provide everything one would need to know about a program (King, Morris & Fitz-Gibbon, 1987). Both these survey type methods can be open to response bias such as that of acquiescence. If the personnel

being interviewed or responding to questions feel a particular response is desired, they will answer that way rather than give their real response (Alreck & Settle, 1995).

Records of the program may constitute a more credible source than self-reports. These records give evidence of what has occurred in the program. Examples of these records could include completed workbooks, student drawings, progress charts, teacher logs and state standardize test scores. Records alone, however, may lack details which only those using the program could provide (King, Morris & Fitz-Gibbon, 1987).

Statement of Need

A method of determining the quality of program implementation needed to be developed that would incorporate strengths and minimize weaknesses of the various methods. This Quality of Use (QOU) determination method needed to involve carefully planned, documented, short-term observations using a summative scale. A demographic questionnaire could be developed to determine if there were other threats to internal validity present. The summative scale would provide QOU levels in multiple areas that could be combined to produce an average QOU for a particular teacher. This could then be applied to a program, such as the Integrated Skills Method (ISM) program, to determine if the progress, or lack of progress, could be attributed this reading program.

Statement of Purpose

The purpose of this study was to develop a summative scale that could be administered in a short time period to determine the Quality of Use (QOU) of a particular intervention used by teachers. This scale should not only be able to be completed in a matter of hours, or at the most, a few days; it should be based on easily attainable information. One use of this QOU scale would be to determine if the results of the program were depended upon the quality of the fidelity of teacher use or upon some other variable. Other potential uses could be those of determining the need for: accountability, frequent monitoring, self-monitoring, remedial training, new training and program emphasis change.

The Quality of Use scale, hereafter referred to as QOU, would have a range of potential indicators. Some indicators to be considered would include: use or non-use of reading levels for initial placement (Putnam, 1996), use or non-use of direct instruction (Hunt, 1996; Putnam, 1996), frequency of instruction (Roberts, Jurgens & Burchinal, 2005), records of program materials (King, Morris & Fitz-Gibbon, 1987), lesson plans (Frudden, 1984), teacher experience (Treka, 1994), presence of other reading programs/tutorials (Cobb, 2001), Speech services (Nathan, Stackhouse, Goulandris & Snowling, 2004), socioeconomic status (SES) (Battle & Pastrana, 2007), ethnic makeup (Baker, Keller-wolff & Wolf-wendel, 2000) and language proficiency (Baker, Keller-wolff & Wolf-wendel, 2000). However, it was not yet known how these indicators relate to one another and exactly how they co-exist in a particular classroom. It is important to know which indicators, or groups of indicators, influence student outcomes.

Measuring the QOU of the application and interaction of these influential indicators could determine what actually contributes to the success, or lack of success of a particular classroom intervention.

Multiple indicators can be measured together in a single additive scale, a summated rating scale, and each indicator is considered an individual item measuring a particular dimension of a construct (Spector, 1992). The items on this scale should be related. If they are unrelated, they should not be combined because they may not be measuring the same variable (Babbie, 1990). Similarity of measurement by multiple items can be summarized as inter-item correlation, or as distances on a 2-dimensional cluster graph. High inter-item relationships permit the overall score to be judged as indicating a single factor or trait (StatSoft, Inc., 2010) and helps obtain high reliability between raters (Spector, 1981). Reliability, the ability of a measure to give the same result every time, is related to validity, which is how well a measure reflects the concept being examined (Babbie, 1990). How reliable a measure is can limit its validity (Spector, 1981). Items included on the QOU scale should measure, with validity, concepts that are important to the design of the program to show good student gains.

Because of the number of potential QOU indicators, it is desirable to reduce as many as possible to a single scale score, if they have a monotonic relationship, to produce a composite measure of the variable of QOU. In a monotonic relationship, the indicators tend to increase or decrease together (Webster, 1913). It is noted that the resulting summative scale score will have only ordinal properties (Babbie, 1990).

Individual indicator items can be combined in a summative scale through various algorithms. A simple combination method is to sum their individual subscale scores together—this is a simple additive approach (DeVellis, 2003). A more complex method is to pre-determine their relative importance and combine them in such a way that some are weighted more heavily than others. An example of this would be if two thirds of the items reflect on aspect studied and one third relates to a different aspect. If both aspects need to be equally represented in the index, then a different weight should be assigned to give equal importance to each aspect (Babbie, 1990). A third method is to define the summative score by Bayesian combinations of items. Bayesian pertains to statistical methods in which population parameters are random variables that have known probability distributions (Dictionary.com, n.d.). The final QOU scale is a result of Bayesian logic statements, rather than arithmetic addition.

This QOU summative scale would then be used to evaluate the effect of different quality of implementation levels of the Integrated Skills Method (ISM) Reading Program on student reading levels, over a one year time period. As per Mills and Ragan (2000), a program's performance will be influenced by the extent it is used as intended. ISM Teaching Systems, Inc. (2004-a) reported that their highest group gained an average of more than two years between pretest and posttest. This is in contrast to the lowest group where the gain was an average of seven months which was less than the predicted gain without the program. It would be valuable to know whether it was the quality of program implementation or another factor that made the difference in results in various student groups.

Research Questions

This study addresses the following questions: 1) Can an efficient, reliable, and valid Quality of Use (QOU) summative scale be developed with strong internal consistency (that is inter-item correlation) based on multiple sources to produce an average QOU for a particular teacher/classroom? 2) Do student results from the Integrated Skills Method (ISM) Linguistic Pattern Series (LPS) reading program, as measured by changes in the Raw Score Grade Equivalent (RSGE) scores during a one year time period, vary significantly depending on the quality of teacher use of the program? 3) What is the relationship between Quality of Use (QOU) as a predictor of student achievement, and other potential determinants of changes in student results such as: teacher experience (Trcka, 1994), presence of other reading programs/tutorials (Cobb, 2001), Speech services (Nathan, Stackhouse, Goulandris & Snowling, 2004), socioeconomic status (SES) (Battle & Pastrana, 2007), ethnic makeup (Baker, Keller-wolff & Wolf-wendel, 2000) and language proficiency (Baker, Keller-wolff & Wolf-wendel, 2000)?

CHAPTER II

REVIEW OF LITERATURE

Quality of Use Concept

Whether or not a new program is being used is only weakly assessed by determining whether the required materials were physically in the classroom. That assessment leaves open the question of whether the materials were used or were left in the closet. Hall and Hord (2001) go on to state the assumption formerly was that materials and training lead to use. Witt, Noell, LaFluer and Mortenson (1997) found, in their study of teacher intervention use, although the teachers started with complete treatment adherence, the adherence decreased after training. None of the teachers in this study continued the treatment adherence above 80% longer than 2 days after their training. Even in a study of childhood literacy programs, that began with an 82% adherence at first observation and maintain at a 79% adherence during the second and third observations, the percentages of various intervention components varied from 17% to 100% (Zvoch, 2009). Decreased intervention integrity, in a study of a mathematics intervention, generally showed poorer student response (Noell, Gresham & Gansle, 2002).

Inconsistent treatment adherence can cause more experimental variability which can then make it difficult for an intervention study to have valid statistical conclusions (Gresham, MacMillan, Beebe-Frankenberger & Bocian, 2000). Gresham et al. (2000) found that few learning disability studies actually measure and report intervention

integrity. In the 5 years before the 2000 article, only 18.5% of the articles reviewed measured how the interventions were implemented. By 2005, a review of intervention studies conducted in schools over a 15 year period showed only 30% of the studies reviewed reported implementation data (McIntyre, Gresham, DiGennaro and Reed, 2007). O'Donnel (2008) reviewed over 120 documents, of which 23 primary studies, the newest of which was dated in 2006, measured quantitatively the relationship between implementation fidelity and outcomes. Only 5 of these studies, or about 22 %, met all of the criteria: primary intervention research, implementation fidelity to K-12 interventions, efficacy of interventions for core school subjects, statistical quantitative fidelity measures, fidelity effectiveness correlation and reporting of sample size. O'Donnel states, that after reviewing fidelity literature, "there is a shortage of K-12 core curriculum intervention studies that empirically measure fidelity of implementation and its relationship to outcomes" (p. 51).

With the current need for Response to Intervention (RtI) to provide research-based instruction (Texas Education Agency, 2007-a), organizations like What Works Clearinghouse (WWC), established 2002, have developed to determine scientifically what actually works in the field of education (What Works Clearinghouse, 2010). Although this organization has very specific requirements of studies that are included in their review of interventions (What Works Clearinghouse, 2008), in a 2008 response to Stockard (2010), they state that their process of review may downplay the fidelity of implementations. WWC further states, in WWC Procedures and Standards Handbook under Corrections and Adjustments, that "The WWC makes no adjustments or

corrections for variations in implementation of the intervention...” (What Works Clearinghouse, 2008). Without information on intervention integrity, it is not possible to tell how close the intervention treatment was to the intended treatment (McIntyre, Gresham, DiGennaro and Reed, 2007).

The differences in an intervention could then impact efficiency of learning, especially if differences are maintained over time. If this continued for a full academic year, in multiple areas, fewer educational objectives would be taught (Grow, et al., 2009). Findings that intervention use is related to outcomes indicate the teacher’s implementation of a program is important for it to be successful. More adherence to the use of a program, i.e. high QOU, by teachers also could see an increase in effectiveness (Biggs, Vernberg, Twemlow, Fonagy, & Dill, 2008). Various methods for measuring program implementation have been suggested. Two systems of measurement reviewed are Levels of Use (LoU) and Fidelity of Implementation (FOI). Quality of Use (QOU) of a program would incorporate aspects of both these intervention use measures.

Quality of Use Measures

Levels of Use

Studies and observations by Hall and Hord (2001) showed that there were several different patterns of behavior for users and non-users of an intervention. Hall, Loucks, Rutherford, & Newlove (1975) present a Levels of Use (LoU) chart that defines use levels. Of the eight use levels, three levels describe non-use and five levels describe use levels. The non-users are split into: 1) Nonuse (LoU 0) – a user with limited/no

knowledge of the intervention or innovation, 2) Orientation (LoU I) – user actively learns about the intervention, and 3) Preparation (LoU II) – the user is getting ready to use the intervention and has a time planned to begin. The user levels are: 1) Mechanical Use (LoU III) – the user is focused on the daily use of the intervention and changes are made to enable the user rather than the learner, 2) Routine (LoU IVA) – the user has stabilized his use of the intervention and few changes are being put into practice, 3) Refinement (LoU IVB) – the user is refining the intervention to increase the student impact both short and long-term, 4) Integration (LoU V) – the user is working with colleagues to have a collective impact on students, and 5) Renewal (LoU VI) – the user looks for major modifications of the intervention, or replaces it with an alternative, to increase student impact (Anderson, 1997; Gray, 1997; Hall & Hord, 2001; Hall, Loucks, Rutherford, & Newlove, 1975). LoU is one of the reasons that many program evaluations show no significant differences between subjects using a new intervention and the control group. The group using the intervention may be implementing it only on a low LoU, especially if it is measured during the first period of use. The lack of gains may be due to LoU, not just effect of the program (Hall, Loucks, Rutherford, & Newlove, 1975). Measuring LoU then becomes a critical piece in an evaluation study (Hall & Hord, 2001). For example, using the CBAM-LoU (Concerns-based Adoption Model-Levels of Use), based on the eight levels of LoU, as one of the measures of the level of technology adoption, a study by Hancock and Knezek (2007) showed a strong relationship between the technology adoption level and the frequency of use of a free course management system.

Measurement of LoU has presented problems. Hall & Hord (2001) note that researchers have attempted to measure LoU through a paper and pencil self-report. Hall & Hord feel this “is like trying to decipher semaphore signals by listening to a radio” (p. 86). This is due to possible response bias that can be a weakness of survey type instruments. Among other causes, response bias can enter through the respondent’s desire to enhance their image or to answer in a socially acceptable manner (Alreck & Settle, 1995). This could also be an issue in the interview methods Hall & Hord (2001) recommend; even with extensive training. According to Loucks, Newlove, and Hall (1975), the LoU interviews were to be used instead of long-term observations as an alternative to assess levels (Hall & Hord, 2001). Interviews alone would still be susceptible to this threat of social desirability response bias as well as other types of response bias such as that of acquiescence where the interviewee gives responses they perceive as desirable by the researcher (Alreck & Settle, 1995).

Fidelity of Implementation

An associated concept that is related to how a program is used is Fidelity of Implementation (FOI). FOI can be defined as content and instructional strategies delivered both accurately and consistently as they were intended (Fullan & Pomfret, 1977; Gresham et al., 2000; National Research Center on Learning Disabilities, 2006, 2007). Mills and Ragan (2000) state that designers anticipate their program’s performance will be influenced by the extent it is used in the intended way. Dane and Schneider (1998) found various procedures to ascertain fidelity in the studies they

reviewed. Five aspects that studies have measured to determine FOI are: adherence – extent of delivery of program components as prescribed, exposure – quantity of program content given to participants, quality of program delivery – qualitative aspects of delivering program content, participant responsiveness – engagement of participants, and program differentiation – determination that subjects receive only planned, unique interventions (Dane & Schneider, 1998; Ruiz-Primo, 2006; Dusenbury, Brannigan, Falco & Hansen, 2003).

The first aspect of FOI, adherence, refers to delivering the program as designed. In research, programs are supported to achieve fidelity. In actual implementation, fidelity is not expected to be maintained to the same degree (Dusenbury, et al., 2003). If a program is implemented with undocumented variations from the original design, it would be difficult to interpret results. If results are negative, this could be seen as a problem with the program itself rather than a problem with program delivery (Dane & Schneider, 1998). Many intervention failures could be attributed to lack of appropriate implementation (Gresham, 1989).

Second, the aspect of exposure refers to the amount of content of the program each participant receives. This can be the number, the length or the frequency of program interventions (Ruiz-Primo, 2006). Programs seem to be less effective when subjects have attended fewer of planned sessions (Dane & Schneider, 1998). Greater use of a program could lead to an increase in its effectiveness (Biggs, et al., 2008).

The third aspect, quality of program delivery, refers to how well the program providers use the program processes and deliver the program content (Ruiz-Primo,

2006). These are not related directly to content implementation; but qualitative aspects such as leader preparedness, enthusiasm and attitude toward the program (Dane & Schneider, 1998). Techniques involved in some interventions rely on the program implementers to use interactive techniques requiring this type of facilitating or coaching skills (Dusenbury, et al., 2003).

Participant responsiveness, the fourth aspect of FOI, refers to the engagement of participants (Ruiz-Primo, 2006). This aspect would measure how well a participant responds to sessions in amount of enthusiasm and participation (Dane & Schneider, 1998). It can be further defined as how engaged participants are in the various aspects of a program (Dusenbury, et al., 2003).

The last aspect, program differentiation, determines that subjects receive only planned interventions (Dane & Schneider, 1998). These interventions should be unique and distinguishable from other programs (Ruiz-Primo, 2006; Dusenbury, et al., 2003).

Besides these five factors, there are other treatment integrity related factors presented by Gresham (1989). Gresham's treatment integrity is very similar to FOI as he defines it as referring "to the degree to which a consultation plan is implemented as intended" (Gresham, 1989, p. 37). These factors are: "(a) complexity of treatments, (b) time required to implement treatments, (c) materials/resources required for treatments, (d) number of treatment agents required, (e) perceived and actual effectiveness of treatments, and (f) motivation of treatment agents" (Gresham, 1989, p. 38).

Complexity of treatments becomes important as the more complex treatments have lower treatment integrity (Gresham, 1989). Even a difference in the requirement of

more or less paperwork can affect teachers' willingness to participate (Lane, Kalberg, Bruhn, Mahoney & Driscoll, 2008). Major difficulties are encountered in trying to setup and maintain extremely complex treatments. Time becomes a factor due to the obvious interaction between how complex a treatment is and how long such a treatment will take to implement. An intervention that requires a teacher to dedicate a great deal of time is probably not going to have good treatment integrity (Gresham, 1989).

Another factor associated with how easily a teacher can implement an intervention is that of how many extra material and resources are required. If the intervention needs material and resources that are not usually found in school classrooms, then this will also result in lower treatment integrity (Gresham, 1989).

Poor treatment integrity is also a possible result of an intervention that requires many treatment agents. This factor is likely to be combined with the complexity issue as an intervention requiring many agents is probably more complex than one only needing a single agent. The difficulty appears to lie in the fact that any of the various agents can fail to follow the treatment design and cause poor treatment integrity (Gresham, 1989).

Interventions judged to be feasible and highly effective, were more likely to be used by educators (Carter & Pesko, 2008). Further, if an intervention is considered effective, it may have higher treatment integrity than an intervention considered ineffective. This also applies to interventions that quickly produce change. These interventions may be more likely to be continued with higher treatment integrity than interventions that take longer to show results. The intervention treatment integrity appears to be reinforced by quick, positive results (Gresham, 1989).

Finally, the motivation of the intervention agent may have an effect on treatment integrity. Teachers may refer a child primarily for testing and removal from the classroom rather than for intervention in the regular class (Gresham, 1989; Ysseldyke, Christenson, Pianta & Algozzine, 1983). If motivation is for removal, rather than remediation, than interventions may have poor treatment integrity. It is unlikely the teacher will implement accurately an intervention to help the child stay in the regular class if the teacher's desire is removal (Gresham, 1989).

Methods suggested, by Gresham et al. (2000), for assessing FOI include both direct and indirect assessment. Direct assessment involves direct observation or videotaping for later evaluation. Indirect assessment includes such instruments as: interviews, self-reports, rating scales and permanent products. Each method has potential issues. Observer reactivity is a problem with using observation methods as the intervention may only be correctly implemented when observers, live or videotaping, are present. As with LoU, the survey type instruments are susceptible to response bias (Alreck & Settle, 1995). Permanent products, which are produced as artifacts during use of intervention components, are less likely to be affected by problems such as social desirability, reactivity and they require less time to assess (Gresham et al., 2000). Records alone, however, may lack details which only those using the program could provide (King, Morris & Fitz-Gibbon, 1987).

Both Levels of Use (LoU) and Fidelity of Implementation (FOI) are trying to get at the same concept. They both are trying to determine how close implementation of an intervention is to the way it was designed to be implemented. Hall, Loucks, Rutherford,

& Newlove (1975) note that what happens when an intervention, or innovation, is actually implemented can vary tremendously and they present a LoU chart that defines use levels. FOI can be defined as content and instructional strategies delivered both accurately and consistently as they were intended (Fullan & Pomfret, 1977; Gresham et al., 2000; National Research Center on Learning Disabilities, 2006, 2007). Without this control of how an intervention is implemented, conclusions cannot be drawn as to the effect of the independent variable, the planned intervention (Gresham et al., 2000).

Quality of Use Need

A method of determining the quality of program implementation needs to be developed that would incorporate strengths and minimize weaknesses of the various methods. This Quality of Use (QOU) determination method would need to involve carefully planned, documented, short-term observations of intervention artifacts using a summative scale. A demographic questionnaire could be developed to determine if there were other threats to internal validity present. The summative scale would provide QOU levels in multiple areas that could be combined to produce an average QOU for a particular teacher or other interventionist. This could then be applied to an intervention, to determine if the progress, or lack of progress, could be attributed this intervention.

CHAPTER III

METHOD

Overview

This study was designed to develop a summative scale that could be used to determine the Quality of Use (QOU) of teachers using a program to determine if the results of the program were depended upon the quality of the fidelity of teacher use or upon some other variable. This scale used documented, short observations of intervention artifacts using the carefully developed summative scale. A demographic questionnaire was used to determine if there were other threats to internal validity present. The scale and questionnaire were designed to be completed in less than an hour. The questionnaire was completed while the artifact observation was taking place during a live classroom visit. By having the subjects complete the questionnaire, rather than interviewing, there was less chance of subjects giving answers they thought were desirable. At the same time, the observer was available to clarify questions and ask follow up questions concerning the artifacts. Since the questionnaires were turned in at the end of the observation, the response rate was higher than that expected if they were sent out as a survey.

Subjects of this study were special education teachers in Linguistic Pattern Series (LPS) programs in 13 elementary schools in the San Antonio ISD that represented high and low levels of overall student progress in the Integrated Skills Method (ISM) reading program observed in the spring semester of 2008. A small preliminary study, consisting

of two teachers in one school, was undertaken to examine the QOU scale quality including determining whether the scale was working as designed, or needed further development. The full study, consisting of 20 teachers in 13 schools, was designed to answer if an adequate, multi-item Quality of Use (QOU) summative scale could be developed to include classroom-level information from multiple sources summing to an average QOU score for a particular teacher and classroom.

Also examined was whether student results from the Integrated Skills Method (ISM) Linguistic Pattern Series (LPS) reading program varied significantly depending on the quality of teacher use of the program. Then the Quality of Use (QOU) effect was examined to determine if it was independent of the other possible determinants of changes in student results. The developed summative scale was examined to determine if it was cohesive, or covaried, using a Classical Item Analysis and use of Cronbach's Alpha to measure overall inter-item correlation. Residualized gain scores were used to examine the final scale that was composed of five items that possessed low-moderate cohesiveness.

Gain scores, or simple change scores, can be problematic, because they often correlate substantially with pretest scores (Rogosa & Willett, 1983; Gardner & Neufeld, 1987). So much of the relationship between an external variable and a gain score may have existed a priori in the relationship between external variable and pretest scores. For this reason, there is general agreement that some statistical remedy be applied, one of the more effective being a residualized difference (Williams, Zimmerman, Rich & Steed, 1984) i.e. residualized gain scores. Residualized gain scores are the residuals (or

leftovers) of the gain scores, after pretest prediction has been partialled out of them. By residualizing, we ensure that the ordering of gain scores across classes is unrelated to their pretest order (Hough & Piper, 1982).

Context

This study took place in the state of Texas. Texas implements No Child Left Behind (NCLB) through the use of a single definition for all schools to reach adequate yearly progress (AYP). Even non-Title I schools are required to make amendments to school improvement plans if they fail to make this progress for two consecutive years. Title I schools failing to make progress for a particular indicator must fulfill some Title I School Improvement requirements. These could include such activities as offering supplemental education services or providing school choice. These requirements are progressive and are based on how many years the AYP standard is not met. Reading is one of the areas assessed and included in the AYP calculation. The following assessments have been part of this calculation: Texas Assessment of Knowledge and Skills (TAKS), State-Developed Alternative Assessment II (SDAA II), Locally-Determined Alternate Assessments (LDAA) and Reading Proficiency Tests in English (RPTE) (Texas Education Agency, 2005-a). In 2007, per NCLB requirements, SDAA II and LDAA are no longer allowed options. Besides the opportunity to take the regular TAKS, the options for students with disabilities are changed to: TAKS Accommodated (TAKS-A), TAKS Modified (TAKS-M) and TAKS Alternative (TAKS-Alt). These are designed to line up with the federal requirements for assessing students who have

disabilities. TAKS and TAKS-A both meet the requirement for a general assessment that may be given with or without accommodations. TAKS-M has standards that measure modified grade-level achievement on this alternate assessment. TAKS-Alt is an alternative assessment whose standards are alternate rather than grade-level (Texas Education Agency, 2007-b).

The Texas district participating in the study, the San Antonio Independent School District (San Antonio ISD), is a large district in Texas serving 56,371 students (Texas Education Agency, 2006-b). It is currently the ninth largest Texas district (SAISD, 2006). San Antonio ISD ethnic breakdown is: African American (8.9%), Hispanic (87.3%), White (3.7%), Native American (0.0%), and Asian/Pacific Islander (0.0%). Their economically disadvantaged percentage in the '05/'06 profile stood at 92.2% (Texas Education Agency, 2006-b). As a district, San Antonio ISD met AYP (Texas Education Agency, 2005-b), although two of the district's high schools are listed as requiring school improvement in both Reading and Math by the Texas Education Agency (2005-c).

Respondents

Teachers

Subjects of this study were special education teachers in Linguistic Pattern Series (LPS) programs in 13 elementary schools in the San Antonio ISD. Teachers were from 13 elementary schools in the district representing high and low levels of overall student progress in the Integrated Skills Method (ISM) program. These schools were from a

pool of schools (Appendix A) with 9 or more students divided into high (Instructional Level Grade Equivalent [ILGE] gains greater than or = to 0.7 between '05 and '06) and low (ILGE gains less than or = to 0.3 between '05 and '06) gains (E. Richardson, personal communication, March 15, 2007) and paired (7 high and 6 low) representing schools with similar demographic data (Texas Education Agency, 2006-a). It is noted all 18 schools in the pool were given the choice to participate or not in the study, as were the teachers at the 13 schools that chose to participate.

Depending upon the school, teachers are responsible for different numbers of students. The average number of students taught, per teacher, is about 10 (E. Richardson, personal communication, April 24, 2006). The teachers were divided into two teacher types: 1) teachers new to the program (less than 2 years of experience) and 2) teachers experienced with the program (2 or more years experience). Teachers were further categorized by their Quality of Use of the program.

In order to teach in the ISM Program, teachers take a full-day workshop, five to six hours in length, of preparation training. Trainers also meet with program teachers to discuss new developments, provide program information, and enrich teaching techniques as well as teaching strategies. These meetings take place at least twice a year and can also provide a time that teachers can consult with each other concerning the program (ISM Teaching Systems, Inc., 2004-b).

Students

The students of the teachers studied were elementary special education students currently in Resource or Special Education Behavior classes. Special education students are registered in the program when ISM has received and entered basic student information and their Decoding Skills Test (DST) scores (ISM Teaching Systems, Inc., 2004-b). These students attend the schools that were chosen from a pool of schools (Appendix A) with 9 or more students divided into high (ILGE gains greater than or = to 0.7 between '05 and '06) and low (ILGE gains less than or = to 0.3 between '05 and '06) gains (E. Richardson, personal communication, March 15, 2007) and paired (7 high and 6 low). In the 13 schools chosen, the ethnic breakdown is: African American (0.3 to 72.1%), Hispanic (25.2 to 99.3%), White (0.0 to 20.5%), Native American (0.0 to 0.3%), and Asian/Pacific Islander (0.0 to 0.9%). Their economically disadvantaged percentage in the '05/'06 profile varied from 62.3 to 100.0% (Texas Education Agency, 2006-a).

Intervention

The Integrated Skills Method (ISM) Reading Program is composed of three program components: 1) Linguistic Component which uses the Linguistic Pattern Series (LPS) to teach the students reading and writing skills, 2) Literature Component which teaches the students reading enjoyment and how to learn through teacher selection of independent reading level literature and 3) Test Prep Component which uses the implementer's choice of a structured, sequenced workbook reading program, such as the Barnell Loft Specific Skill Series, to prepare students for reading tests. This study

focused on the LPS portion of the ISM Reading Program that was developed for the El Paso Independent School District (EPISD) for compliance with Texas Education Agency (TEA) dyslexia intervention program rules (ISM Teaching Systems, Inc., 2003).

The LPS program is organized into achievement level sequences that are color coded and divided into three or four sub-levels. Decoding words is taught through letter sounds. Then students apply this skill to sentences, stories and other academic challenges. All levels of the LPS program emphasize reading skills such as comprehension, composition and problem solving. The program uses carefully structured sequence of the linguistic patterns to teach reading and writing. For each target pattern, the students first learn to read and then comprehend both sentences and stories followed by performance of exercises that stress that pattern. This gives the students reinforcement in the use of the pattern in writing as well as reading and providing the full range of development of reading skills (ISM Teaching Systems, Inc., 2003).

Instrumentation

Dependent Measure - DST

The Decoding Skills Test (DST) consists of three subtests, which measure various aspects of decoding, and is designed to be individually administered. The first subtest is the basal vocabulary subtest. This subtest has 110 items arranged in groups of 10 words. These words are on various reader levels and chosen to represent basal vocabulary taught in standard reading programs. The second subtest covers phonic

patterns and measures aspects of student use of letter sounds. The 120 items are divided into 24 groups and consist of real word monosyllabic and polysyllabic items and corresponding nonsense word items. The third subtest is contextual reading. It is used for clinical diagnostics and is rarely used as part of the Integrated Skills Method (ISM) program (E. Richardson & DiBenedetto, 1999). When the DST was correlated to the Iowa Test of Basic Skills, the New York City Wide Reading Test, the Gates-MacGinitie Reading Tests, the Gilmore Oral Reading Test and teacher data for program placement; the strongest correlations were with the decoding subtests scores, which were all above .70, and with program placement which all showed above .80. Interrater agreement, for each DST subtest, was between 84% and 100%. The split-half coefficients were .95 to .99 for each subtest (K. Richardson, 1987).

Before students enter the program, they take the DST. The resulting test scores are sent to ISM Teaching Systems and entered into the data base. The schools receive student lists reporting individual scores and recommendations for program placement. Teachers report lesson placements when instruction begins. They also report mastery tests scores at the mid-point of the school year and at the school year's end. Annually, the DST is administered again to provide an independent progress evaluation and current achievement level. Each student's accumulated data base information is printed and returned to the school at beginning as well as at the end of every school year for teacher and administrator use (ISM Teaching Systems, Inc., 2004-b).

Independent Measures - Quality of Use Scale

An adaptation of the Innovation Configuration Map, described by Hall and Hord (2001), has been developed to formulate a summative scale (Appendix C). This scale was used to determine each teacher's QOU of the program. The summative scale include these areas: initial placement, direct instruction, frequency of direct instruction, frequency of independent/interdependent related activities, linguistic pattern series materials, literature and test prep materials, ISM class scheduling and lesson plans. Although not part of the scale, background demographic information was also examined to determine other possible influences on student success.

Initial Placement

Level of appropriateness of initial placement was determined by examination of each teacher's records, Integrated Skills Method (ISM) records, teacher questionnaire and general questions asked during the classroom observation. Initial placement was considered most appropriate if placed at the ISM recommended level based on the Decoding Skills Test (DST) Instructional Level Grade Equivalent/Raw Score Grade Equivalent (ILGE/RSGE). This could vary to some of the least appropriate placements, such as being placed only by general availability.

Direct Instruction

Level of direct instruction was determined by examination of each teacher's records, teacher questionnaire and general questions asked during the classroom

observation. Levels can vary from the highest level of direct instruction, indicated by direct instruction by the teacher in one to three small groups at a time, with one to six students working on approximately the same reading level to low level of students reading to each other without a teacher present.

Frequency of Direct Instruction

Frequency of direct instruction was determined by examination of each teacher's records, teacher questionnaire and general questions asked during the classroom observation. Highest frequency would involve direct instruction of all groups 3 or more times per week and the lowest would involve only zero to one time per week.

Frequency of Independent/Interdependent Related Activities

Frequency of independent/interdependent related activities was determined by examination of each teacher's records, teacher questionnaire and general questions asked during the classroom observation. The highest QOU would be use of independent/interdependent literature and/or test preparation activities by all groups two or more times per week. The lowest QOU would be use of these activities one time or less every two weeks.

Linguistic Pattern Series Materials

Use of linguistic pattern series materials was determined by examination of each teacher's record and evidence in materials or artifacts from related assignments. For the

highest QOU, the teacher would need to have the LPS current assigned book and mastery test available for all program students and showing some student work in the majority of them. Also there would be some LPS games/strategy materials present and put together for use for each LPS level assigned. The lowest QOU would show LPS current assigned book and/or Mastery test not available or showing no student work in the majority of them.

Literature and Test Prep Materials

Use of literature and test preparation materials was determined by examination of each teacher's record and evidence in materials or artifacts from related assignments. The highest level QOU would involve literature and test preparation materials at all LPS assigned levels present and evidence of use of both types present. The lowest level would involve no literature or test preparation materials at LPS assigned levels present or no evidence of use present.

ISM Class Scheduling

Frequency of ISM classes was determined by examination of each teacher's records, teacher questionnaire and general questions asked during the classroom observation. The highest level would involve ISM classes scheduled at least 45 mins five times a week. This could vary to the lowest level of ISM classes scheduled less than 30 mins two times a week or less than 60 mins total per week.

Lesson Plans

The presence of lesson planning was determined by examination of each teacher's records. The highest level here would show individual lesson plans present for all students for ISM (LPS), literature and test preparation. The lowest level would be no consistent lessons plans present.

Data Collection

Data was collected by: 1) physical classroom observation, including notating evidence of program use in lesson plans and filled-in consumable materials, 2) responses to observer's questions, 3) responses to demographic questionnaire and 4) program data provided by program providers.

A small preliminary trial of the study, using these data collection methods, was undertaken to verify the appropriateness of the scale and its reliability. As in the full study, areas that were examined, to determine a teacher's QOU, are: initial placement, direct instruction, frequency of direct instruction and independent/interdependent related activities, linguistic pattern series materials, literature and test prep materials, scheduling and lesson plans. Background demographic information was examined. The preliminary study was designed to determine if the scale is appropriate and works as designed or needs to be further developed. Reliability was examined two ways. First, the preliminary study school was visited twice to determine if the measures, as design, have re-test/repeat rater reliability over time. Second, to determine if there is inter-rater reliability, the preliminary study school information was documented, as well as rated

using the scale. A second rater was asked to look at this material and rate it using the scale to determine if similar scores are obtained. This material was obtained by documenting for each of the eight areas examined. Findings of the preliminary study are included in the Results section.

Demographics

Demographics pertaining to teachers' general experience, ISM experience, training, job duties/programs, and assigned student characteristics were collected by: 1) reviewing TEA statistical information, ISM program information, district information, and school information and 2) reviewing teachers' responses to a demographic questionnaire and general questions asked during classroom observation. (Appendixes D & E) Relevant demographic information is included in the Results section.

Design

Figure 1 describes the study's design. Figure 1 is the design figure timeline and shows the independent variable, Quality of Use (QOU), in relation to the dependent variable, Linguistic Pattern Series (LPS) program-Decoding Skills Test (DST). (See Figure 1)

Teachers QOU at High PreT/LPS Begins-----LPS Continues---Prelim-----Study-----PstT

Performing Schools

(10 teachers from 6-7 schools teaching a total of 70 to 120 students*)

Teachers QOU at Low PreT/LPS Begins-----LPS Continues---Prelim-----Study-----PstT

Performing Schools

(10 teachers from 6-7 schools teaching a total of 70 to 120 students*)

Fall (2007)

Spring (2008)

PreT - Pretest - Linguistic Pattern Series (LPS) – Decoding Skills Test (DST)

LPS - LPS program implemented in 20 classrooms in 13-14 elementary schools during the school year.

Prelim - Preliminary Study (02-27-08 and 03-12-08)

Study - Main Study (03-24-08 to 04-04-08)

PstT - Post Test (Annual) Linguistic Pattern Series (LPS) – Decoding Skills Test (DST)

* - Number of students based on '05/'06 statistics (Texas Education Agency, 2006-a)

Figure 1 - Quality of Use Study Design

Analysis by Research Question

Can an Efficient, Reliable, and Valid Quality of Use (QOU) Summative Scale be Developed with Strong Internal Consistency (that is Inter-Item Correlation) Based on Multiple Sources to Produce an Average QOU for a Particular Teacher/Classroom?

For research question one, a summative scale was developed to ascertain if critical components of a particular reading program were actually being used and, if used, in a manner reflecting the quality of the fidelity to the program design. In order to ameliorate the quality of results of a short term study, without the susceptibility to response bias of surveys or interviews alone, this scale relied on short-term observation of artifacts, with follow up questions, in addition to descriptions of program use. Teachers had a QOU on each of the eight scales (0-4) with an overall average QOU score computed from combining the individual scores. The QOU scale used Bayesian logic statements, rather than arithmetic addition. Scaling of QOU was done to permit that variable to be used efficiently in analyses with other variables, such as student achievement. In this case, the scaling must be done prior to looking at student achievement scores, but after obtaining QOU data from teachers and classrooms, to avoid personal biases that may affect selection (Babbie, 1990). A preliminary study was designed to determine re-test/repeat rater reliability and inter-rater reliability. These were measured using Cohen's Kappa.

Do Student Decoding Results Vary by Teacher Quality of Use (QOU) of Program?

For research question two, Decoding Skills Test (DST) pretest and posttest score difference, interval data, was the dependent variable. The independent variable of QOU, rank order data, was calculated for each individual teacher studied. Teachers had a QOU on each of the eight scales (0-4) with an overall average QOU score computed from combining the individual scores. DST pretest and posttest scores were used from Integrated Skills Method (ISM) Teaching Systems' records for all students active in the program. QOU was determined by teacher program observation using the summative scale. For this analysis, Multiple Regression procedures were used to analyze the relationship between these variables.

What Is the Relationship between QOU as a Predictor of Student Achievement, and Other Potential Determinants of Changes in Student Decoding Results?

For research question three, current measurement was taken of other independent variables available from TEA, district, school records and demographic questionnaire responses such as: teacher experience (Trcka, 1994), presence of other reading programs/tutorials (Cobb, 2001), Speech services (Nathan, Stackhouse, Goulandris & Snowling, 2004), socioeconomic status (SES) (Battle & Pastrana, 2007), ethnic makeup (Baker, Keller-wolff & Wolf-wendel, 2000) and language proficiency (Baker, Keller-wolff & Wolf-wendel, 2000). Multiple Regression procedures were used to measure multiple variables when more than one variable may affect the dependent variable.

Procedures

Preliminary Study

In the preliminary study, teachers' classrooms in the preliminary study school were observed. During and after this observation, the teachers were asked for responses, by filling out a questionnaire (Preliminary Study - Appendix D), to ascertain their Quality of Use (QOU). This was to mitigate the tendency of respondents to respond to sensitive questions with socially desirable answers. It is noted that mail surveys have better performance on more sensitive questions, but face to face interview have a lower non-response (de Leeuw, 1992). Since this questionnaire does not deal with known sensitive topics, such as income (de Leeuw, 1992); the lower non-response was desirable. Filling out the questionnaire, rather than only being asked oral questions, should help the performance level of the instrument. The higher response rate was maintained by the teachers being asked to complete it by the end of the observation. Also during the observation, teachers were allowed to ask the observer questions to clarify what information was being requested.

The observations took place in each classroom at the school site. A period of approximately 45 minutes was spent in each teacher's classroom. Areas examined, to determine a teacher's QOU, were: initial placement, direct instruction, frequency of direct instruction and independent/interdependent related activities, linguistic pattern series materials, literature and test prep materials, scheduling and lesson plans. Background demographic information was collected. QOUs that were most consistent

with the program designed were given points based on a summative scale (Appendix C) per the following methods and guidelines.

Initial placement. Level of appropriateness of initial placement, the first QOU scale item, was determined by examination of each teacher's records, Integrated Skills Method (ISM) records, teacher questionnaire and general questions asked during the classroom observation. Initial placement was considered most appropriate if placed at the ISM recommended level based on the Decoding Skills Test (DST) Instructional Level Grade Equivalent/Raw Score Grade Equivalent (ILGE/RSGE). As seen in the Preliminary Study Quality of Use Innovation Configuration Map (Appendix B), this can vary to the point that teachers only used non-DST scores or general availability to group students and did not verify any placement above the ISM recommended level by use of a mastery test.

Direct instruction. The second QOU scale item, quality of direct instruction, was determined by examination of each teacher's records, teacher questionnaire and general questions asked during the classroom observation. The highest QOU level would be direct instruction by the teacher in one to three small groups, at one time, of one to six students with approximately same reading level. As seen in the Preliminary Study Quality of Use Innovation Configuration Map (Appendix B), this can vary to the point that the lowest QOU would involve no direct instruction, with students grouped with approximately same reading level reading to each other.

Frequency of direct instruction. Frequency of direct instruction, the third QOU scale item, was determined by examination of each teacher's records, teacher questionnaire and general questions asked during the classroom observation. The highest QOU would involve direct instruction of all groups three or more times per week. As seen in the Preliminary Study Quality of Use Innovation Configuration Map (Appendix B), this can vary to the point that the lowest QOU would involve direct instruction of all groups zero-one times per week.

Frequency of independent/interdependent related activities. Next, the frequency of independent/interdependent related activities QOU scale item was determined by examination of each teacher's records, teacher questionnaire and general questions asked during the classroom observation. The highest QOU here would involve use of independent/interdependent literature and/or test preparation activities by all groups two or more times per week. As seen in the Preliminary Study Quality of Use Innovation Configuration Map (Appendix B), this can vary to the point that the lowest QOU would involve use of independent/interdependent literature and/or test preparation activities by all groups one time or less every two weeks.

Linguistic pattern series material. The QOU scale item of use of linguistic pattern series (LPS) materials was determined by examination of each teacher's record and evidence in materials or artifacts from related assignments. For the highest QOU, the teacher would need to have the LPS current assigned book and mastery test available

for all program students and showing some student work in the majority of them. Also there would be some LPS games/strategy materials present and put together for use for each LPS level assigned. As seen in the Preliminary Study Quality of Use Innovation Configuration Map (Appendix B), this can vary to the point that the QOU level would have LPS current assigned book and/or Mastery test not available or showing no student work in the majority of them.

Literature and test preparation materials. Use of literature and test preparation materials QOU scale item was determined by examination of each teacher's record and evidence in materials or artifacts from related assignments. The highest level QOU would involve literature and test preparation materials at all LPS assigned levels present and evidence of use of both types present. As seen in the Preliminary Study Quality of Use Innovation Configuration Map (Appendix B), this can vary to the point that the QOU would involve no literature or test preparation materials at LPS assigned levels present or no evidence of use present.

ISM class scheduling. Next, the QOU scale item of frequency of ISM classes was determined by examination of each teacher's records, teacher questionnaire and general questions asked during the classroom observation. At the highest QOU, ISM classes would be scheduled at least 45 mins five times a week. As seen in the Preliminary Study Quality of Use Innovation Configuration Map (Appendix B), this can

vary to the point that the lowest level would show ISM classes scheduled less than 30 mins two times a week or less than 60 mins total per week.

Lesson plans. The presence of lesson planning, the last QOU scale item, was determined by examination of each teacher's records. The highest level of QOU would show individual lesson plans present for all students for ISM (LPS), literature and test preparation. As seen in the Preliminary Study Quality of Use Innovation Configuration Map (Appendix B), this can vary to the point that the final QOU level, there are no consistent lesson plans present.

Background demographic information. Background demographic information was then collected to determine other possible influences on student success. This was gathered by reviewing TEA statistical information, ISM program information, district information, and/or school information. Also, teachers' responses to a demographic questionnaire and general questions asked during classroom observation were reviewed. Progress of the students was examined to determine if QOU of the program is an accurate predictor student success. Experience level of the teachers was obtained from district information, teacher questionnaire and/or by examining the Integrated Skills Method (ISM) records of teacher training.

Full Study

In the full study, teachers' classrooms in the various schools were observed. During and after this observation, the teachers were asked for responses, by filling out a questionnaire (Full Study - Appendix E), to ascertain their Quality of Use (QOU). This was to mitigate the tendency of respondents to respond to sensitive questions with socially desirable answers. It is noted that mail surveys have better performance on more sensitive questions, but face to face interview have a lower non-response (de Leeuw, 1992). Since this questionnaire does not deal with known sensitive topics, such as income (de Leeuw, 1992); the lower non-response is desirable. Filling out the questionnaire, rather than only being asked oral questions, should help the performance level of the instrument. The higher response rate was maintained by the teachers being asked to complete it by the end of the observation. Also during the observation, teachers were allowed to ask the observer questions to clarify what information was being requested.

The observations took place at each school site. A period of approximately 45 minutes was spent in each teacher's classroom. Areas examined, to determine a teacher's QOU, were: initial placement, direct instruction, frequency of direct instruction and independent/interdependent related activities, linguistic pattern series materials, literature and test prep materials, scheduling and lesson plans. Background demographic information was collected. QOUs that were most consistent with the program designed were given points based on a summative scale (Appendix C) per the following methods and guidelines.

Initial placement. Level of appropriateness of initial placement, the first QOU scale item, was determined by examination of each teacher's records, Integrated Skills Method (ISM) records, teacher questionnaire and general questions asked during the classroom observation. Initial placement was considered most appropriate if placed at the ISM recommended level based on the Decoding Skills Test (DST) Instructional Level Grade Equivalent/Raw Score Grade Equivalent (ILGE/RSGE). As seen in the Full Study Quality of Use Innovation Configuration Map (Appendix C), this can vary to the point that teachers only used non-DST scores or general availability to group students and did not verify any placement above the ISM recommended level by use of a mastery test.

Direct instruction. The second QOU scale item, quality of direct instruction, was determined by examination of each teacher's records, teacher questionnaire and general questions asked during the classroom observation. The highest QOU level would be direct instruction by the teacher in one to three small groups, at one time, of one to six students with approximately same reading level. As seen in the Full Study Quality of Use Innovation Configuration Map (Appendix C), this can vary to the point that the lowest QOU would involve no direct instruction, with students grouped with approximately same reading level reading to each other or direct instruction with groups of students, of any size, on various reading levels reading at the same time.

Frequency of direct instruction. Frequency of direct instruction, the third QOU scale item, was determined by examination of each teacher's records, teacher questionnaire and general questions asked during the classroom observation. The highest QOU would involve direct instruction of all groups three or more times per week. As seen in the Full Study Quality of Use Innovation Configuration Map (Appendix C), this can vary to the point that the lowest QOU would involve direct instruction of all groups zero-one times per week.

Frequency of independent/interdependent related activities. Next, the frequency of independent/interdependent related activities QOU scale item was determined by examination of each teacher's records, teacher questionnaire and general questions asked during the classroom observation. The highest QOU here would involve use of independent/interdependent literature and/or test preparation activities by all groups two or more times per week. As seen in the Full Study Quality of Use Innovation Configuration Map (Appendix C), this can vary to the point that the lowest QOU would involve use of independent/interdependent literature and/or test preparation activities by all groups one time or less every two weeks.

Linguistic pattern series material. The QOU scale item of use of linguistic pattern series (LPS) materials was determined by examination of each teacher's record and evidence in materials or artifacts from related assignments. For the highest QOU, the teacher would need to have the LPS current assigned book and mastery test available

for all program students and showing some student work in the majority of them. Also there would be some LPS games/strategy materials present and put together for use for each LPS level assigned. As seen in the Full Study Quality of Use Innovation Configuration Map (Appendix C), this can vary to the point that the QOU level would have LPS current assigned book and/or Mastery test not available or showing no student work in the majority of them.

Literature and test preparation materials. Use of literature and test preparation materials QOU scale item was determined by examination of each teacher's record and evidence in materials or artifacts from related assignments. The highest level QOU would involve literature and test preparation materials at all LPS assigned levels present and evidence of use of both types present. As seen in the Full Study Quality of Use Innovation Configuration Map (Appendix C), this can vary to the point that the QOU would involve no literature or test preparation materials at LPS assigned levels present or no evidence of use present.

ISM class scheduling. Next, the QOU scale item of frequency of ISM classes was determined by examination of each teacher's records, teacher questionnaire and general questions asked during the classroom observation. At the highest QOU, ISM classes would be scheduled at least 45 mins five times a week. As seen in the Full Study Quality of Use Innovation Configuration Map (Appendix C), this can vary to the point

that the lowest level would show ISM classes scheduled less than 30 mins two times a week or less than 60 mins total per week.

Lesson plans. The presence of lesson planning, the last QOU scale item, was determined by examination of each teacher's records. The highest level of QOU would show individual lesson plans present for all students for ISM (LPS), literature and test preparation. As seen in the Full Study Quality of Use Innovation Configuration Map (Appendix C), this can vary to the point that the final QOU level, there are no consistent lesson plans present.

Background demographic information. Background demographic information was then collected to determine other possible influences on student success. This was gathered by reviewing TEA statistical information, ISM program information, district information, and/or school information. Also, teachers' responses to a demographic questionnaire and general questions asked during classroom observation were reviewed. Progress of the students was examined to determine if QOU of the program is an accurate predictor student success. Experience level of the teachers was obtained from district information, teacher questionnaire and/or by examining the Integrated Skills Method (ISM) records of teacher training.

CHAPTER IV

RESULTS

Preliminary Study Results

A small preliminary study was undertaken to examine the QOU scale quality, specifically, its reliability, stability, and cohesiveness of its 8 items (initial placement, direct instruction, frequency of direct instruction and independent/interdependent related activities, linguistic pattern series materials, literature and test prep materials, scheduling and lesson plans). The preliminary study sample consisted of 2 teachers in one school. The preliminary study was designed to determine whether the scale was appropriate and works as designed or needed further development.

Two types of reliability were examined: inter-rater reliability and stability over time. First, to determine whether there was inter-rater reliability, one teacher's information for each of the QOU's 8 items was rated and then documented. A second rater was asked to look at this documented material and rate it using the scale to determine if similar scores would be obtained.

The second form of reliability assayed was stability over time. The preliminary study school was visited twice, 2 weeks apart, to determine if the QOU scores would have repeat rater reliability over time. After the first visit, 3 items (Initial Placement, Direct Instruction, Materials - Literature, Test Prep) were rewritten to account for unforeseen options. Initial Placement was change to show "Use of DST ILGE to

initially place students” instead of “Use of DST ILGE/RSGE to initially place students”. This change was made to be consistent in evaluating the teachers.

Next, the areas addressing direct instruction were amended. Direct Instruction was changed to “Direct instruction by teacher/aide” from “Direct instruction by teacher” when it was noted that, as is frequent practice in Special Education classes, both the teacher and her aide were involved in providing direct instruction for the students. Also, the final category, that was designed to show no direct instruction from a teacher or aide, was amended to add “or direct instruction with groups of students, of any size, on various reading levels reading orally at the same time from different books” as this variation was not anticipated as being used by teachers.

Finally, the Material – Literature, Test Prep was changed to “evidence of use at assigned grade levels of both types present”. This wording replaced “evidence of use of both types present” to clarify that use had to be shown on all assigned grade levels instead of just use of both types of materials on any level.

For all these items, repeat reliability was not able to be checked due to changes in collection procedure. Items that could be checked were: Frequency of Direct Instruction, Frequency of Independent/Interdependent Related Activities, Materials – ISM (Linguistic Pattern Series), Scheduling and Lesson Plans.

Inter-rater Reliability

Inter-rater reliability was checked by pertinent materials copied (see Appendix G for samples) and one classroom described verbally by the primary observer to an offsite

materials reviewer after the first rating. The reviewer has a background of work experience in school computer labs, in school libraries, in substitute teaching and is in an MLIS program. The reviewer's rating was completed on one classroom consisting of 14 students. Materials copied consisted of sample pages from ISM (Linguistic Pattern Series) books, attendance sheet (which also served as partial lesson plans) and a copy of teacher's printed schedule for students. The reviewer observed the materials and asked questions of the primary observer concerning classroom setting and items present in the classroom to make his ratings.

Stability Over Time - QOU Revisions

The preliminary study was designed to determine if the scale was appropriate and worked as designed or needed to be further developed. During the phase of testing and re-testing for repeat rater reliability, it was determined more data needed to be examined on 3 items (Initial Placement, Direct Instruction, Materials - Literature, Test Prep) during the 2nd visit. Because these 3 items were rewritten, the repeat rater reliability and stability findings for the early versions no longer applied. There was not time for re-testing the new items. Items that could be checked for the repeat rater reliability and stability were: Frequency of Direct Instruction, Frequency of Independent/Interdependent Related Activities, Materials – ISM (Linguistic Pattern Series), Scheduling and Lesson Plans.

Research Question #1

Can an efficient, reliable, and valid Quality of Use (QOU) summative scale be developed with strong psychometric features based on multiple sources to produce an average QOU for a particular teacher/classroom? Strong psychometric features include: Reliability, Stability, and Inter-item Cohesiveness or Scalability. A scale was developed to ascertain the average QOU for each teacher. Information from the preliminary study was used to examine the reliability and stability of the scale, as well as refine the guidelines of each area examined. Results from the main study were used to determine inter-item cohesiveness or scalability. This question will be answered for all three separate criteria: Reliability, Stability, and Inter-item Cohesiveness or Scalability.

Reliability

The first of the three criteria for Research Question #1, Reliability, was examined during the preliminary study by use of an offsite materials reviewer as described in the preliminary study review. Inter-rater reliability was checked by pertinent materials copied and the classroom described to an offsite materials reviewer after the first rating. This rating was completed on one classroom consisting of 14 students. Materials copied consisted of sample pages from ISM (Linguistic Pattern Series) books, attendance sheet (which also served as partial lesson plans) and a copy of teacher's printed schedule for students. The reviewer examined the materials and asked questions of the primary observer concerning classroom setting and items present in the classroom to make his ratings. He also asked to have terms defined, such as ILGE/RSGE (Instructional Level

Grade Equivalent/ Raw Score Grade Equivalent). Oral information given by the primary observer was told to the reviewer without elaboration that would influence his scores.

The scores were compared between the initial, onsite rating of the teacher and the reviewer's rating of the teacher offsite. The ratings results did not show variance when the same guidelines were used to examine the data by both raters. The results showed 100% agreement on all items.

Stability

The second criteria for Research Question #1, stability of the 8 variables in the QOU scale, was to be measured when the preliminary study school was visited twice by the primary observer, 2 weeks apart, to determine if the QOU scores would have repeat rater reliability over time. Data was noted on a copy of the Quality of Use Innovation Configuration Map with Summative Scale. After the first visit, some of the items (Initial Placement, Direct Instruction, Materials - Literature, Test Prep) were examined and it was determined more information was needed than had been obtained during the first visit of the preliminary study.

Initial Placement was rewritten to show "Use of DST ILGE to initially place students" instead of "Use of DST ILGE/RSGE to initially place students". This change was made to be consistent in evaluating the teachers.

Direct Instruction was amended to "Direct instruction by teacher/aide" from "Direct instruction by teacher". This change was made when it was noted that both the teacher and her aide were involved in providing direct instruction for the students. Also,

the final category of Direct Instruction, that was designed to show no direct instruction from a teacher or aide, was changed to add “or direct instruction with groups of students, of any size, on various reading levels reading orally at the same time from different books”. This change was made due to a teacher action that was not anticipated before the preliminary study was begun.

The Material – Literature, Test Prep prompt was also rewritten. “Evidence of use at assigned grade levels of both types present” was used instead of “evidence of use of both types present”. This change clarified that use had to be shown on all assigned grade levels instead of just use of both types of materials on any level.

As noted previously, repeat reliability was not able to be checked due to changes in collection procedure before the second visit. Some items could be checked for repeat reliability. These items were: Frequency of Direct Instruction, Frequency of Independent/Interdependent Related Activities, Materials – ISM (Linguistic Pattern Series), Scheduling and Lesson Plans.

When the items were examined; it was noted that there were varying amounts of consistency shown. Two items, Frequency of Direct Instruction and Frequency of Independent/Interdependent Related Activities were very consistent for both teachers across both preliminary study visits and were in 100% agreement. On the items of Scheduling and Lesson Plans, ratings were very consistent between the first and second measure with 2nd teacher (100%).

Items that were not consistent included Materials – ISM (Linguistic Pattern Series) which did show change with both teachers; but this was due more to refining

what questions needed to be asked of the teachers to interpret their artifacts, then actual changes in measurement. As noted, on the items of Scheduling and Lesson Plans, ratings were very consistent between the first and second measure with 2nd teacher (100%); but were not consistent with the 1st teacher. In the area of Scheduling, it was discovered, on the second visit, that the teacher had only given the schedule for oral time on the ISM lessons. As the ISM program includes additional components, the second visit results included schedule results that showed oral reading and non-oral reading scheduled time. Also, since the San Antonio ISD school district does not require formalized lesson plans in these classes, those teachers that do have some type of lesson plans appear to not be consistent with whether they address individual plans or general group work across time. This was reflected in the 2nd teacher's ratings. Information gleaned from having repeat visits helped refine the collecting of data for the final study, but did not reflect the stability of the instrument.

Scalability

The third criteria of Inter-item Cohesiveness or Scalability, for Research Question #1, was examined as to whether the eight items on the Quality of Use (QOU) summative scale were cohesive, or covaried across the observed 20 teachers from 13 schools. For this examination a Classical Item Analysis was conducted. Preliminary descriptive data includes the Means and Standard Deviations of Scale Items. (See Table 1).

Table 1 - Means and Standard Deviations of Scale Items

	Mean	Standard Deviation	Low/High Scores
Initial Placement	0.600	1.314	0/4
Direct Instruction	3.000	1.654	0/4
Freq - Direct Instruc	3.750	0.910	0/4
Freq - Related Activities	4.000	0.000	4/4
Materials – LPS	2.850	1.663	0/4
Materials – Lit/Tst Prep	2.250	1.743	0/4
Scheduling	3.000	1.414	0/4
Lesson Plans	1.050	1.669	0/4

Except in the area of Frequency of Related Activities, all the items showed the full range of zero to four. The Frequency of Related Activities was affected by an administrative directive that added a specific additional program that all classes were required to use in addition to the ISM materials. This caused all teachers to earn a score of 4 as all teachers were using this required independent activity.

The Classical Item Analysis itself showed that, as a group, the eight items covaried or “hung together” poorly as a scale. Cronbach’s Alpha, which measures overall inter-item correlation, was only .490. A newly constructed scale should show an Alpha at least in the .70’s, to permit a meaningful, single summative score (Hinkin,

1995; Nunnally, 1978; Spector, 1991). After deleting the fourth item, Frequency of Related Activities, for its lack of variance; the scale gives an Alpha of .50.

Individual item analysis showed item #8, Lesson Plans, with a negative correlation (-.291) with the rest of the items as a group, and item #2, Direct Instruction, with a near-zero (+.025) correlation with the other items. Those two items were dropped from the scale, leaving a five item total from the original eight. It is noted that deleting items from the scale could potentially change the meaning of QOU. These remaining items, as a group, possessed a Cronbach's Alpha of .709. Thus, the final scale was composed of five items (#1- Initial Placement, #3- Freq - Direct Instruc, #5- Materials – LPS, #6- Materials – Lit/Tst Prep, and #7- Scheduling) that possessed low-moderate cohesiveness of Alpha=.71.

Research Question #2

Do student results from the Integrated Skills Method (ISM) Linguistic Pattern Series (LPS) reading program, as measured by changes in Raw Score Grade Equivalent (RSGE) scores during a one year time period, vary significantly depending on the quality of teacher use of the program?

This predictive validity question asks about the relationship between QOU scores and student gain in RSGE scores over a one-year period. On the Decoding Skills Test (DST) there are two possible scores: the ILGE (Instructional Level Grade Equivalent) and the RSGE (Raw Score Grade Equivalent). The ILGE score comes from the grade level before the list on which a student has difficulty reading the words. It indicates the

highest level at which he/she can comfortably read and determines the level for instruction to begin. The RSGE score is determined by totaling the number of words read. This score measures a student's reading vocabulary and can indicate progress of word recognition skills. The RSGE is a more discerning measure of student gains (ISM Teaching Systems, Inc., 2008). Analyses therefore will be based on RSGE scores. Because QOU is a class-level variable, student scores were summarized by classroom and class means used for the main correlational analysis.

The correlation, between QOU summary scores (based on five questions) and residualized RSGE gains, was examined. The QOU could predict $.771^2 = 50\%$ of score variance. This is a strong prediction for a non-student external measure in education. (Please see Appendix H – Pearson Correlations Section.)

Research Question #3

What is the relationship between Quality of Use (QOU) as a predictor of student achievement, and other potential determinants of changes in student results such as: teacher experience (Treka, 1994), presence of other reading programs/tutorials (Cobb, 2001), Speech services (Nathan, Stackhouse, Goulandris & Snowling, 2004), socioeconomic status (SES) (Battle & Pastrana, 2007), ethnic makeup (Baker, Keller-wolff & Wolf-wendel, 2000) and language proficiency (Baker, Keller-wolff & Wolf-wendel, 2000)?

The substantial relationship between QOU and student residualized gain scores on RSGE may have been moderated by, or even substantially caused by, a completely

unrelated variable. This variable is termed a moderator. The moderator is an extrinsic variable, quantitative or qualitative, that affects the predictive relationship between the independent and the dependent, or criterion, variables (Statistics Solutions, 2009). Of the potential moderators, 3 were examined. Because QOU is a classroom-level variable, a moderator variable also must be measured at or aggregated to the classroom level. The analytic method used was to categorize each extrinsic variable into a two- or three-level variable, and then include that categorical variable along with QOU as predictors of RSGE. Simple 3-category variables were created for each potential moderator, because three categories, if data were balanced, would provide a sufficient sample size of four to six per category from the small total N of only 16.

The final analysis was a series of Multiple Regression procedures, from which the interaction effect could answer the question of differential relationship between QOU and RSGE for different levels of the moderating variable. Within a full multiple regression model, there are main effects for each predictor, plus an interaction effect between predictors. Both R and R^2 , from multiple regression, are used as indices of predictor strength. R is squared to yield R^2 (coefficient of determination), which shows the amount of explained variance in a particular variable (Gall, Borg & Gall, 1996). Three of the other possible determinants of change (Ethnicity, Teacher Experience in ISM and Socioeconomic Status) in student results were examined individually as independent predictors and as potential moderators of the primary relationship between QOU and student score gains. Moderator variables are variables affecting the strength and/or direction of a relationship that exists between a predictor variable and the

dependent variable (Baron & Kenny, 1986). Valentine and Cooper refer to Cohen's assertion that effect sizes in the social sciences can be generally be defined as small if $r = .10$, medium if $r = .30$ and large if $r = .50$ (Valentine & Cooper, 2003; Cohen, 1992).

To determine R and R^2 , Multiple Regression was conducted, with RSGE residualized gain scores as the dependent measure, standardized QOU (based on five items) scores as the main predictor, and Ethnicity (three-category variable) as a second predictor, and potential moderator. The three categories of performance by Ethnicity shows: Level 1 - less than or equal to 97% (N=4, M=.250, SD=.173), Level 2 – greater than 97% to less than 98.8% (N=6, M=.483, SD=.360) and Level 3 – greater than or equal to 98.8% (N=6, M=.433, SD=.225).

Alone, QOU predicts $R^2=.594$ or $R=.771$ and Ethnicity predicts $R^2= .117$ or $R=.342$. Therefore, Ethnicity is only a weak independent predictor of student skill gain. With the Ethnicity variable removed, $R^2=.534$ or $R=.731$. Therefore, most of the predictive strength of QOU is independent of (does not depend on) Ethnicity. If QOU and Ethnicity effects are combined as two independent predictors, results do improve: $R^2=.651$ or $R=.807$. When their interaction effect is added to these two independent main effects, $R^2=.697$ or $R=.835$. Therefore, Ethnicity added to QOU does enhance the latter's predictive power. However, Ethnicity is a weak predictor alone ($R=.342$) and acts only slightly as a moderator (adds .028 to the main effects R).

The next predicted moderator, Teacher Experience in ISM, was examined as a three-category variable, with standardized QOU (five-item) scores as predictors, and RSGE residualized gain scores as dependent measure. The three categories of

performance by Teacher Experience in ISM shows: Level 1 - less than or equal to 2 years experience (N=5, M=.580, SD=.311), Level 2 – greater than 2 years to less than 4 years experience (N=5, M=.200, SD=.255) and Level 3 – greater than or equal to 4 years experience (N=6, M=.433, SD=.151).

As presented earlier, QOU predicts $R^2=.594$ or $R=.771$. Teacher Experience in ISM alone is a moderately strong predictor of student skill gain: $R^2=.324$ or $R=.569$. When the Teacher Experience is removed from the main prediction by QOU, results do substantially drop: $R^2=.368$ or $R=.607$. If QOU and Teacher Experience Main effects are combined, predictive power increases: $R^2=.692$ or $R=.832$. When their interaction effect is added to this, $R^2=.774$ or $R=.880$. Teacher Experience is therefore a substantial predictor alone ($R=.569$), whose main effect adds considerable to QOU predictive power (R increases from .771 to .832). However, beyond this main effect, Teacher Experience has only a small moderator role, adding .048 R , pointing to its Main effect role.

The final predicted moderator, Socioeconomic Status (SES), was examined as a three-category variable, with standardized QOU (five-item) scores as predictors, and RSGE residualized gain scores again as the dependent measure. The three categories of performance by Socioeconomic Status shows: Level 1 - less than or equal to 89.1% (N=5, M=.340, SD=.288), Level 2 – greater than 89.1% to less than 93.1% (N=5, M=.540, SD=.152) and Level 3 – greater than or equal to 93.1% (N=6, M=.350, SD=.339).

As a reminder, in the primary relationship, QOU predicts $R^2=.594$ or $R=.771$ and SES predicts $R^2=.106$ or $R=.326$. When SES is removed from this primary relationship,

the predictive strength of QOU drops somewhat: $R^2=.501$ or $R=.708$. This means that QOU does not depend largely on SES predictive strength. The combined main effects of QOU and SES effects are combined then $R^2=.607$ or $R=.779$. When their interaction effect is added to this, $R^2=.740$ or $R=.860$. Socioeconomic Status thus has only a small additive effect to QOU ($R=.771$ to $.779$) but a larger interactive or moderator effect ($R=.779$ to $.86$).

In summary, none of the three extrinsic variables (Ethnicity, Teacher Experience and Socioeconomic Status) show sizeable moderating effects; although, per Cohen (1992) Teacher Experience in ISM appeared to have some effect as a predictor alone. The first variable examined, Ethnicity, proved to be only a weak independent predictor of student skill gain. Most of the predictive strength of QOU is independent of Ethnicity. When Ethnicity added to QOU, it does enhance the latter's predictive power; however, Ethnicity is a weak predictor alone and acts only slightly as a moderator.

The second variable examined, Teacher Experience in ISM, by itself is a moderately strong predictor of student skill gain. When the Teacher Experience is removed from the main prediction by QOU, results do substantially drop. Teacher Experience's main effect does add considerable to QOU predictive power. However, beyond this main effect, Teacher Experience has only a small moderator role.

The last variable examined, Socioeconomic Status (SES), when removed from this primary relationship does drop the predictive strength of QOU somewhat; but to such a small degree that QOU does not depend largely on SES predictive strength. Socioeconomic Status thus has only a small additive effect but it does have a larger

interactive or moderator effect. Overall, though, the substantial relationship between QOU and student residualized gain scores on RSGE does not appear to have been moderated by one of these variables.

CHAPTER V

DISCUSSION AND CONCLUSIONS

Overview

The purpose of this study was to develop a Quality of Use (QOU) summative scale with good psychometric properties to answer research questions about the QOU of teachers using a specific program. One intended application of such a scale is to determine whether QOU moderates the effects of a novel intervention on student performance. Noell, Gresham & Gansle (2002) found that decreased intervention integrity generally showed poorer student response. Inconsistent treatment adherence can cause more experimental variability which can then make it difficult for an intervention study to have valid statistical conclusions (Gresham, MacMillan, Beebe-Frankenberger & Bocian, 2000). Various methods for measuring intervention implementation have been suggested. Two of these systems of measurement are Levels of Use (LoU) and Fidelity of Implementation (FOI). Both these systems are trying to determine how close implementation of an intervention is to the way it was designed to be implemented. Hall, Loucks, Rutherford, & Newlove (1975) note that what happens when an intervention, or innovation, is actually implemented can vary tremendously and they present a LoU chart that defines use levels. FOI can be defined as content and instructional strategies delivered both accurately and consistently as they were intended (Fullan & Pomfret, 1977; Gresham et al., 2000; National Research Center on Learning Disabilities, 2006, 2007). Without this control of how an intervention is implemented,

conclusions cannot be drawn as to the effect of the independent variable, the planned intervention (Gresham et al., 2000).

Measurement of LoU has presented problems. Hall & Hord (2001) note that researchers have attempted to measure LoU through a paper and pencil self-report. Hall & Hord feel this “is like trying to decipher semaphore signals by listening to a radio” (p. 86). This is due to possible response bias that can be a weakness of survey type instruments. Among other causes, response bias can enter through the respondent’s desire to enhance their image or to answer in a socially acceptable manner (Alreck & Settle, 1995). This could also be an issue in the interview methods Hall & Hord (2001) recommend; even with extensive training. According to Loucks, Newlove, and Hall (1975), the LoU interviews were to be used instead of long-term observations as an alternative to assess levels (Hall & Hord, 2001). Interviews alone would still be susceptible to this threat of social desirability response bias as well as other types of response bias such as that of acquiescence where the interviewee gives responses they perceive as desirable by the researcher (Alreck & Settle, 1995).

Methods suggested, by Gresham et al. (2000), for assessing FOI include both direct and indirect assessment. Direct assessment involves direct observation or videotaping for later evaluation. Indirect assessment includes such instruments as: interviews, self-reports, rating scales and permanent products. Each method has potential issues. Observer reactivity is a problem with using observation methods as the intervention may only be correctly implemented when observers, live or videotaping, are present. As with LoU, the survey type instruments are susceptible to response bias

(Alreck & Settle, 1995). Permanent products, which are produced as artifacts during use of intervention components, are less likely to be affected by problems such as social desirability, reactivity and they require less time to assess (Gresham et al., 2000).

A method of determining the quality of program implementation needed to be developed that incorporated strengths and minimize weaknesses of the various methods. The Quality of Use (QOU) determination method involves carefully planned, documented, short-term observations of intervention artifacts, with follow up questions, using a summative scale. A demographic questionnaire was developed to determine if there were other threats to internal validity present. The summative scale provides QOU levels in multiple areas that are combined to produce an average QOU for a particular teacher or other interventionist.

The newly developed QOU scale was applied to the classroom teaching of 20 Special Education teachers who were using the curriculum, Integrated Skills Method (ISM) Linguistic Pattern Series (LPS), in 13 schools in the San Antonio Independent School District (SAISD). The ISM program was chosen as it has a consistent pre-test/post-test, the Decoding Skills Test (DST), that is given routinely to all program students when they first enter the program and annually thereafter (ISM Teaching Systems, Inc., 2004-b). The classrooms in the SAISD were chosen as the ISM program is a well researched, established program in the district.

An unanticipated circumstance in the SAISD was the required use of an additional reading program in all Special Education reading classrooms. As there was no instance of these programs not being present; the effect of additional reading program

on the QOU of the ISM program could not be determined. This item was removed from the scale as there was no difference in scores across classroom concerning this issue.

Another circumstance, found in Special Education classes in SAISD, was that teachers were not required to maintain formal lesson plans, so teachers who had any type of lesson plans beyond basic scheduling was unusual. One teacher had a locked file cabinet that had not been accessible for some time. While she stated some records were there, they were not accessible for the observation and she was scored on that basis. This was done for two reasons: 1) she did not have access to the materials so it would be similar to not having them and 2) other teachers also claimed to have similar materials but could not produce them when asked.

Research Question #1

Can a multi-item Quality of Use (QOU) summative scale be developed to include classroom-level information from multiple sources summing to an average QOU score? A scale was developed to ascertain the average QOU for each teacher. The original scale was reduced from eight to five items because the original eight items covaried poorly as a scale. It is noted that item analysis is done to discover those items forming a scale that is internally consistent and to be able to delete items that are not consistent (Spector, 1992). Deleting items from the scale could potentially change the meaning of QOU. It could, however, indicate that some of the dropped items were not good indicators of the QOU construct. Spector (1992) states that internal consistency implies

that items measure the same construct. If they do not intercorrelate, they may not be representing a common construct.

Individual item analysis showed Lesson Plans, with a negative correlation (-.291) with the rest of the items as a group, and Direct Instruction, with a near-zero (+.025) correlation with the other items. A third item, Frequency of Related Activities, was deleted for its lack of variance. It is noted that during the observations, teachers noted that they were not required to have lesson plans. Teachers with lesson plans may have less consistency with district policy, which could explain the negative correlation in the Lesson Plan item. Similarly, it was district policy for all classes to have a particular related activity in use with the LPS program, which explains why no variance was found in Frequency of Related Activities. The Direct Instruction item did not appear to correlate well with the other items. This item may not be measuring an aspect of direct instruction that contributes to the QOU. Rather than focusing on how the instruction was delivered, this item focused more on number of students in a group and number of groups being taught at one time. While this element may affect some aspect of instruction, it did not appear to have a strong effect on the QOU of the program. The remaining items (#1- Initial Placement, #3- Freq - Direct Instruc, #5- Materials – LPS, #6- Materials – Lit/Tst Prep, and #7- Scheduling) possessed low-moderate cohesiveness.

The multi-item Quality of Use (QOU) summative scale was developed to include classroom-level information from multiple sources summing to an average QOU score for a particular teacher and classroom. The summative inter-rater reliability for this

scale was checked by copying pertinent materials from the preliminary study and describing the classroom to an offsite materials reviewer after the first rating. When scores are compared between the initial rating of the teacher and the offsite reviewer's rating of the teacher offsite, it is seen that the ratings did not change when the same guidelines were used to examine the data by both raters.

The preliminary study was designed originally so that the 8 variables in the QOU summative scale could be reviewed for stability. To do this, the two teachers at the preliminary school were visited twice by the primary observer, 2 weeks apart, to determine if the QOU scores would have repeat rater reliability over time. However, after the first visit, some of the items (Initial Placement, Direct Instruction, Materials - Literature, Test Prep) were examined and it was determined more information was needed than had been obtained during the first visit of the preliminary study. These items were rewritten to address the information needed and could no longer be checked for stability during the preliminary study.

Items that could be checked for repeat reliability were: Frequency of Direct Instruction, Frequency of Independent/Interdependent Related Activities, Materials – ISM (Linguistic Pattern Series), Scheduling and Lesson Plans. When the items were examined; it was noted that there were varying amounts of consistency shown. Two items, Frequency of Direct Instruction and Frequency of Independent/Interdependent Related Activities were very consistent for both teachers across both preliminary study visits and were in complete agreement. On the items of Scheduling and Lesson Plans, ratings were consistent between the first and second measure with 2nd teacher. On the

items of Scheduling and Lesson Plans, ratings were consistent between the first and second measure with 2nd teacher. Items that were not consistent included Materials – ISM (Linguistic Pattern Series) which did show change with both teachers due more to refinement of what questions needed to be asked of the teachers to interpret their artifacts, then actual changes in measurement.

On the items of Scheduling and Lesson Plans were not consistent with the 1st teacher. In the area of Scheduling, it was discovered, on the second visit, that the teacher had only given the schedule for oral time on the ISM lessons. Also, since the San Antonio ISD school district does not require formalized lesson plans in these classes, those teachers that do have some type of lesson plans appear to not be consistent with whether they address individual plans or general group work across time as reflected in the 2nd teacher's ratings. Information gleaned from having repeat visits helped refine the collecting of data for the final study, but did not reflect the stability of the instrument.

One of the advantages of this scale is the limited amount of time needed to observe classrooms; while long term observations can last for months or years (Hull & Zacher, 2007; Brinegar, 2010). It was anticipated that it would take about 45 mins to observe a classroom using this scale. To complete the scale did take approximately 45 mins if the time for the teachers to complete the additional demographic questionnaire is not included. It was found that practice with the scale did make it easier to use and would reduce the time needed to complete.

A second advantage would be that an observer not familiar with the ISM program would be able to complete the scale after a short introduction to the various components. During the preliminary study, an offsite reviewer who was not familiar with the ISM program was able to obtain the same ratings on a teacher's materials and class description as the original rater.

A third advantage of this scale was that the participants did not seem to be concerned about the scale once they understood that the results would be confidential and that their supervisors would not be judging their performance based on those results. Also, minimal time requirements on the part of the participant were required for scale completion.

Practically, QOU measurement could be instrumental in implementing a Response to Intervention (RtI) program. This general education practice is a tool to assess and work with learners who are struggling (NASDSE & CASE, 2006). It also becomes part of the process for determining if a student has a specific learning disability (SLD). For this evaluation to be appropriate, it is imperative that the RtI interventions are implemented and used as designed. Variations, that are not documented, may cause interpretation difficulties. If the intervention is implemented as designed, then the effectiveness of the intervention can be evaluated. A measure, such as the QOU scale, could be used to see if the intervention was being implemented as designed and effectiveness of the intervention could be appropriately measured.

Research Question #2

Do student results from the Integrated Skills Method (ISM) Linguistic Pattern Series (LPS) reading program vary significantly depending on the quality of teacher use of the program? Student results from the Integrated Skills Method (ISM) Linguistic Pattern Series (LPS) reading program, as measured by changes in the Raw Score Grade Equivalent (RSGE) scores during a one year time period, do vary significantly depending on the quality of teacher use of the program. Because QOU is a class-level variable, student scores were summarized by classroom and class means used for the main correlational analysis. The correlation, between QOU summary scores (based on five questions) and residualized RSGE gains, was examined. The QOU could predict $.771^2 = 50\%$ of score variance. This is a strong prediction for a non-student external measure in education.

The QOU (Quality of Use) results appear to be consistent with Fidelity of Implementation (FOI) concepts. FOI can be defined as content and instructional strategies delivered both accurately and consistently as they were intended (Fullan & Pomfret, 1977; Gresham et al., 2000; National Research Center on Learning Disabilities, 2006, 2007). Five aspects that studies have measured to determine FOI are: adherence – extent of delivery of program components as prescribed; exposure – quantity of program content given to participants; quality of program delivery – qualitative aspects of delivering program content; participant responsiveness – engagement of participants and program differentiation – determination that subjects receive only planned, unique interventions (Dane & Schneider, 1998; Ruiz-Primo, 2006). The Quality of Use (QOU)

of the ISM program addressed three FOI concepts: adherence, exposure and quality of program delivery. From the study results, it appears that QOU scale shows strong prediction ability for a non-student external measure.

Research Question #3

What is the relationship between Quality of Use (QOU) as a predictor of student achievement, and other potential determinants of changes in student results such as: teacher experience (Trcka, 1994), socioeconomic status (SES) (Battle & Pastrana, 2007), and ethnic makeup (Baker, Keller-wolff & Wolf-wendel, 2000)? Quality of Use (QOU) effect was examined to determine if it is independent of the other possible determinants of changes in student results. This QOU effect was shown to be independent of the possible determinants of change in the student results in the areas of: ISM Teacher Experience, SES and Ethnicity.

In reviewing the effects of these areas, Ethnicity and Socioeconomic Status have only a weak independent predictor of student skill gain. Ethnicity added to QOU does enhance the latter's predictive power; however, it is a weak predictor alone and acts only slightly as a moderator. Socioeconomic Status has only a small additive effect to QOU but a larger interactive or moderator effect. QOU does not depend largely on SES predictive strength. Teacher Experience in ISM alone is a moderately strong predictor of student skill gain. Its main effect adds considerable to QOU predictive power. However, beyond this main effect, Teacher Experience has only a small moderator role.

Other potential causes of student improvement that were considered were: Speech services (Nathan, Stackhouse, Goulandris & Snowling, 2004), language proficiency (Baker, Keller-wolff & Wolf-wendel, 2000) and presence of other reading programs/tutorials (Cobb, 2001). Although total number of students with Speech services or possible language proficiency issues could be determined per classroom; the identification of which student did receive Speech services or had language proficiency issues could not be identified from the data collected. Also, if the same student had both these issues would not be able to be determined. The presence of other reading programs could not be examined as SAISD has put in place a policy that requires at least one other reading program be included in all Special Education reading classes; therefore, all classes examined had at least one other program present and no comparison was possible.

Implications

Measurement of how programs are implemented has presented problems. Long-term observations are necessarily expensive and take time. Survey type instruments are prone to response bias, such as the respondent's desire to enhance their image or to answer in a socially acceptable manner (Alreck & Settle, 1995). This could also be an issue in the interview methods Hall & Hord (2001) describe; even with the extensive training component. Interviews alone would still be susceptible to these threats of response bias as well as that of acquiescence where the interviewee gives responses they perceive as desirable by the researcher (Alreck & Settle, 1995). The Quality of Use

(QOU) summative scale, which involves carefully planned, documented, short-term observations, as well as interview questions, address these issues. The summative scale provides QOU levels in multiple areas that combine to produce an average QOU for a particular teacher.

It is possible that the five item QOU scale could be used with a variety of programs. Some areas, such as the Frequency of Direct Instruction, could be used with very little change. Other portions would need to be amended to directly address similar components of other programs. The QOU scale areas of Initial Placement, as designed for the LSP program, refers to the use of the DST and/or mastery tests to initially place students. If a different test is used to place students, it should be referred to instead of the DST and mastery tests. Results of this study indicate that the following five items are most related to student success: Initial Placement, Frequency of Direct Instruction, Materials (both program materials and supplementary materials) and Scheduling. These areas of the final five item scale should be addressed; as how the students are placed, how frequently for how long they are instructed and documented use of all materials appear critical to the success of the program. More study is needed to determine if this scale can be with a variety of programs with similar results.

Limitations

The major limitations to this study are: a) Assessment of QOU score stability, b) Limited measure of inter-rater reliability, c) Unknown effect of receive Speech services and/or had language proficiency issues, d) Presence of other reading programs and e)

Limited population studied. The first limitation of this study is in the area of assessment of stability of QOU scores. Due to the need, after the first visit of the preliminary study, for more information than had been obtained during the first visit, several items were revised and procedures were refined. Consequently, stability was not able to be adequately measured. Further studies would need to examine whether QOU scores are affected by repeated measures of the same classrooms. These studies would need to have preliminary studies to verify that wording of scale items was appropriate; and, if any wording was changed, repeated use of the instrument completed before main studies were done.

Although inter-rater reliability was examined during the preliminary study by use of an offsite materials reviewer during the preliminary study, the second reviewer was not on-site. The only reviewed materials were copied and orally presented. To address this limitation, a stronger measure of inter-rater reliability, that of having two or more observers on-site, is suggested for future studies.

In this study, the identification of which student did receive Speech services and/or had language proficiency issues could not be identified from the data collected. Therefore the relationship of Speech services and/or language proficiency issues to the effects of QOU could not be determined. Future studies should collect data that would link these services to individual student scores so these effects could be measured and it could be determine if these were a significant influence on changes seen in the dependent variable.

Another possible determinant of change in student results, not ruled out, was the presence of other reading programs. This could not be examined as SAISD has put in place a policy that requires at least one other reading program be included in all Special Education reading classes; therefore, all classes examined had at least one other program present and no comparison with classrooms without a secondary program was possible. It is suggested that future studies attempt to include similar classroom situation that do not include a secondary reading program in order to rule out the possible effects from other programs.

Finally, this QOU study was completed using a very small sample from one school district in Texas. Further studies would need to have larger numbers, drawn from a larger pool, to determine if the relationships noted do exist or if these findings represent something unique to this particular sample. Also, the pool should include students from general education to rule out the possibility of an effect only seen due to the use of a limited sample of the population drawn from special education students.

Conclusions

A Quality of Use (QOU) summative scale was developed to answer research questions about the QOU of teachers using a specific program in order to determine if the results of the program were depended upon the quality of the fidelity of teacher use or upon some other variable. The scale was applied to the classroom teaching of 20 Special Education teachers who were using the curriculum, Integrated Skills Method (ISM) Linguistic Pattern Series (LPS), in 13 schools in the San Antonio Independent

School District (SAISD). The final QOU scale was composed of five items (#1- Initial Placement, #3- Freq - Direct Instruction, #5- Materials – LPS, #6- Materials – Lit/Test Prep, and #7- Scheduling) that possessed low-moderate cohesiveness. In examining the correlation between QOU summary scores and residualized Score Grade Equivalent (RSGE) gains, the QOU scale predictions were strong for a non-student external measure in education. This QOU effect was shown to be independent of the possible determinants of change in the student results in the areas of: ISM Teacher Experience, SES and Ethnicity.

In agreement with these findings, Biggs, et al. (2008) found that intervention use is related to outcomes indicating the teacher's implementation of a program is important for it to be successful. More adherence to the use of a program, i.e. high QOU, by teachers also could see an increase in effectiveness. Decreased intervention integrity generally showed poorer student response (Noell, Gresham & Gansle, 2002). If a program is implemented with undocumented variations from the original design, it would be difficult to interpret results. If results are negative, this could be seen as a problem with the program itself rather than a problem with program delivery (Dane & Schneider, 1998). Many intervention failures could be attributed to lack of appropriate implementation (Gresham, 1989).

Various methods have been used to determine how well interventions are implemented. Two systems of measurement reviewed were Levels of Use (LoU) and Fidelity of Implementation (FOI). Hall, Loucks, Rutherford, & Newlove (1975) present a Levels of Use (LoU) chart that defines use levels. Studies and observations by Hall and Hord (2001) showed that there were several different patterns of behavior for users

and non-users of an intervention. Measurement of LoU has presented problems. Hall & Hord (2001) note that researchers have attempted to measure LoU through a paper and pencil self-report. They state that this questionnaire method will not work (Hall & Hord). Observations were considered to measure LoU, but was rejected due not only to the need for long observation times, but also to the problem of the act of observing the intervention implementer could cause a change in the implementation just by the observer's presence. Use of interviews is recommended as the best, most efficient way to collect LoU data (Loucks, Newlove, and Hall, 1975). However, one problem with both interviews and questionnaires, is the tendency for respondents to provide responses to researchers or interviewers that they think will be most desirable (Alreck & Settle, 1995).

An associated concept that is related to how a program is used is Fidelity of Implementation (FOI). FOI can be defined as content and instructional strategies delivered both accurately and consistently as they were intended (Fullan & Pomfret, 1977; Gresham et al., 2000; National Research Center on Learning Disabilities, 2006, 2007). Five aspects that studies have measured to determine FOI are: adherence – extent of delivery of program components as prescribed, exposure – quantity of program content given to participants, quality of program delivery – qualitative aspects of delivering program content, participant responsiveness – engagement of participants, and program differentiation – determination that subjects receive only planned, unique interventions (Dane & Schneider, 1998; Ruiz-Primo, 2006; Dusenbury, Brannigan, Falco & Hansen, 2003). Besides these five factors, there are other treatment integrity related

factors presented by Gresham (1989). These factors are: “(a) complexity of treatments, (b) time required to implement treatments, (c) materials/resources required for treatments, (d) number of treatment agents required, (e) perceived and actual effectiveness of treatments, and (f) motivation of treatment agents” (Gresham, 1989, p. 38).

Methods suggested, by Gresham et al. (2000), for assessing FOI include both direct and indirect assessment. Direct assessment involves direct observation or videotaping for later evaluation. Indirect assessment includes such instruments as: interviews, self-reports, rating scales and permanent products. Each method has potential issues. Observer reactivity is a problem with using observation methods as the intervention may only be correctly implemented when observers, live or videotaping, are present. As with LoU, the survey type instruments are susceptible to response bias (Alreck & Settle, 1995). Permanent products, which are produced as artifacts during use of intervention components, are less likely to be affected by problems such as social desirability, reactivity and they require less time to assess (Gresham et al., 2000). Records alone, however, may lack details which only those using the program could provide (King, Morris & Fitz-Gibbon, 1987).

Quality of Use (QOU) of a program incorporates aspects of both these intervention use measures to reduce the limitations of each. The QOU method involves carefully planned, documented, short-term observations of interventions artifacts, with follow up questions, using a summative scale giving QOU levels in multiple areas that are combined to produce an average QOU for a particular teacher or other

interventionist. A demographic questionnaire is used to determine if there were other threats to internal validity present. QOU observations are short and do not involve observing the intervention being implemented, only the artifacts produced, would not be susceptible to observer reactivity. The questionnaire is completed while the observer is present; but not verbally, to reduce the tendency to give desirable responses. Details missing from artifact, or permanent product, observation are overcome by the observer being able to ask questions of the implementers to clarify aspects being observed.

It is possible that the five item QOU scale could be used with a variety of programs. Some areas, such as the Frequency of Direct Instruction, could be used with very little change. Other portions would need to be amended to directly address similar components of other programs. The QOU scale areas of Initial Placement, as designed for the LSP program, refers to the use of the DST and/or mastery tests to initially place students. If a different test is used to place students, it should be referred to instead of the DST and mastery tests. It would not be a good scale to be used with programs that did not produce permanent products, or artifacts, as this a critical aspect of QOU.

Witt, Noell, LaFluer and Mortenson (1997) found, in their study of teacher intervention use, although the teachers started with complete treatment adherence, the adherence decreased after training. Findings that intervention use is related to outcomes indicate the teacher's implementation of a program is important for it to be successful. More adherence to the use of a program, i.e. high QOU, by teachers also could see an increase in effectiveness (Biggs, Vernberg, Twemlow, Fonagy, & Dill, 2008). Student results from this small study are in agreement with the literature and do vary

significantly depending on the quality of teacher use of the program. The QOU scale did predict $.771^2 = 50\%$ of score variance. This is a strong prediction for a non-student external measure in education.

REFERENCES

- Alreck, P. L. & Settle, R. B. (1995). *The survey research handbook (2nd ed.)*. Boston, MA: Irwin/McGraw-Hill.
- Anderson, S. E. (1997). Understanding teacher change: Revisiting the concerns based adoption model [Electronic version]. *Curriculum Inquiry*, 27(3), 331-367.
- Babbie, E. R. (1990). *Survey research methods (2nd ed.)*. Belmont, CA: Wadsworth, Inc.
- Baker, B. D., Keller-wolff, C. & Wolf-wendel, L. (2000). Two steps forward, one step back: Race/ethnicity and student achievement in education policy research [Electronic version]. *Educational Policy*, 14(4), 511-529.
- Baron, R. M. & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations [Electronic version]. *Journal of Personality and Social Psychology*, 51(6), 1173-1182.
- Battle, J. & Pastrana, A. (2007). The relative importance of race and socioeconomic status among hispanic and white students [Electronic version]. *Hispanic Journal of Behavioral Sciences*, 29(1), 35-49.
- Biggs, B. K., Vernberg, E. M., Twemlow, S. W., Fonagy, P., & Dill, E. J. (2008). Teacher adherence and its relation to teacher attitudes and student outcomes in an elementary school-based violence prevention program [Electronic version]. *School Psychology Review*, 37(4), 533-549.

- Brinegar, K. (2010). "I feel like I'm safe again:" A discussion of middle grades organizational structures from the perspective of immigrant youth and their teachers [Electronic version]. *Research in Middle Level Education Online*, 33(9), 1-14.
- Carter, E. W. & Pesko, M. J. (2008). Social validity of peer interaction intervention strategies in high school classrooms: Effectiveness, feasibility, and actual use [Electronic version]. *Exceptionality*, 16, 156-173.
- Cobb, J. B. (2001). The effects of an early intervention program with preservice teachers as tutors on the reading achievement of primary grade at risk children [Electronic version]. *Reading Horizons*, 41(3), 155-73.
- Cohen, J. (1992). A power primer [Electronic version]. *Psychological Bulletin*, 112(1), 0033-2909.
- Dane, A. V. & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control [Electronic version]. *Clinical Psychology Review*, 18(1), 23-45.
- de Leeuw, E. D. (1992). Data quality in mail, telephone and face to face surveys [Electronic version]. *Netherlands Organization for Scientific Research*, ISBN-90-801073-1-x(NUGI-659).
- DeVellis, R. F. (2003). *Scale development: Theory and applications* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Dictionary.com Unabridged (v 1.1). (n.d.). *Bayesian*. Retrieved September 15, 2007, from Dictionary.com website: <http://dictionary.reference.com/browse/bayesian>

- Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. B. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings [Electronic version]. *Health Education Research, 18*(2), 237-256.
- Frudden, S. J. (1984). Lesson plans can make a difference in evaluating teachers [Electronic version]. *Education, 104*(4), 351-353.
- Fullam, M. & Pomfret, A. (1977). Research on curriculum and instruction implementation [Electronic version]. *Review of Educational Research, 47*(1), 335-397.
- Gall, M. D., Borg, W. R. & Gall, J. P. (1996). *Educational research: An introduction (6th ed.)*. White Plains, NY: Longman Publishers USA.
- Gardner, R. C. & Neufeld, R. W. J. (1987). Use of the simple change score in correlational analyses [Electronic version]. *Educational and Psychological Measurement, 47*(4), 849-864.
- Gray, P. J. (1997). Viewing assessment as an innovation: Leadership and the change process [Electronic version]. *New Directions for Higher Education, 100*, 5-15.
- Gresham, F. M. (1989). Assessment of treatment integrity in school consultation and prereferral intervention [Electronic version]. *School Psychology Review, 18*(1), 37-50.
- Gresham, F. M., MacMillan, D. L., Beebe-Frankenberger, M. E. & Bocian, K. M. (2000). Treatment integrity in learning disabilities intervention research: Do we really know how treatments are implemented [Electronic version]. *Learning Disabilities Research & Practice, 15*(4), 198-205.

- Grow, L. L., Carr, J. E., Gunby, K. V., Charania, S. M., Gonslaves, L., Ktaech, I. A., et al. (2009). Deviations from prescribed prompting procedures: Implications for treatment integrity [Electronic version]. *Journal of Behavioral Education, 18*(2), 142-156.
- Hall, G. E. & Hord, S. M. (2001). *Implementing change: Patterns, principles, and potholes*. Needham Heights, MA: Allyn and Bacon.
- Hall, G. E., Loucks, S. F., Rutherford, W. L., & Newlove, B. W. (1975). Levels of use of the innovation: A framework for analyzing innovation adoption [Electronic version]. *Journal of Teacher Education 26*(1), 52-56.
- Hancock, R. & Knezek, G. (2007). Relationships between course management system (CMS) use and teacher technology adoption [Electronic version]. *International Journal of Instructional Technology & Distance Learning 4*(8), 1-10.
- Hinkin, T. R. (1995). A review of scale development practices in the study of organizations [Electronic version]. *Journal of Management 21*(5), 967-988.
- Hough, L. & Piper, M. (1982). The relationship between attitudes toward science and science achievement [Electronic version]. *Journal of Research in Science Teaching, 19*(1), 33-38.
- Hull, G. & Zacher, J. (2007). Enacting identities: An ethnography of a job training program [Electronic version]. *Identity: An International Journal of Theory and Research, 7*(1), 71-102.

Hunt Jr., L. C. (1996). The effect of self-selection, interest, and motivation upon independent, instruction, and frustrational levels [Electronic version]. *Reading Teacher, 40*(4), 278-282.

Individuals with Disabilities Education Improvement Act of 2004 (IDEA), Pub. L. No. 108-446 [H.R. 1350] (2004) Section 614(b)(6)(B). Retrieved September 12, 2007 from http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=108_cong_public_laws&docid=f:publ446.108

ISM Teaching Systems, Inc. (2003). *The ISM reading program an information paper* [Brochure]. El Paso, TX: Author.

ISM Teaching Systems, Inc. (2004-a). *San Antonio ISD annual dyslexia program report 2003-04 school year*. El Paso, TX: Author.

ISM Teaching Systems, Inc. (2004-b). *San Antonio ISD the ISM in resource & BAC annual project report 2003-04 school year*. El Paso, TX: Author.

ISM Teaching Systems, Inc. (2008). *Excerpts from the ISM in the SAISD dyslexia program annual program report 2007-2008*. El Paso, TX: Author.

King, J. A., Morris, L. L., & Fitz-Gibbon, C. T. (1987). *How to access program implementation*. Newbury Park, CA: Sage Publications, Inc.

Lane, K. L., Kalberg, J. R., Bruhn, A. L., Mahoney, M. E., & Driscoll, S. A. (2008). Primary prevention programs at the elementary level: Issues of treatment integrity, systematic screening, and reinforcement [Electronic version]. *Education and Treatment of Children, 31*(4), 465-494.

- Loucks, S. F., Newlove, B. W., & Hall, G. E. (1975). *Measuring levels of use of the innovation: A manual for trainers, interviewers, and raters*. Austin, TX: The University of Texas.
- McIntyre, L. L., Gresham, F. M., DiGennaro, F. D. & Reed, D. D. (2007). Treatment integrity of school-based interventions with children in the journal of applied behavior analysis 1991-2005 [Electronic version]. *Journal of Applied Behavioral Analysis, 40(4), 659-672*.
- Mills, S. C. & Ragan, T.J. (2000). A tool for analyzing implementation fidelity of an integrated learning system [Electronic version]. *Educational Technology Research and Development, 48(4), 21-41*.
- NASDSE & CASE White Paper on RtI. (2006, May). *Response to Intervention*. Retrieved September 10, 2007, from <http://www.nasdse.org/documents/RtIANAdministratorsPerspective1-06.pdf>
- Nathan, L., Stackhouse, J., Goulandris, N., & Snowling, M. (2004). Educational consequences of developmental speech disorder: Key stage I national curriculum assessment results in English and mathematics [Electronic version]. *British Journal of Educational Psychology, 74, 173-186*.
- National Research Center on Learning Disabilities (2006). *Fidelity of implementation*. Retrieved September 3, 2007, from http://www.nrclld.org/rti_manual/pages/RTIManualSection4.pdf

- National Research Center on Learning Disabilities (2007). *RTI: Responsiveness to intervention*. Retrieved September 3, 2007, from http://www.nrcl.org/RTI_Practices/glossary.shtml
- No Child Left Behind Act of 2001, Pub. L. No. 107-110 [H.R. 1] (2002) Sections 1111(b)(2)(F) & 1111(b)(3)(C)(v)(v). Retrieved September 7, 2007, from http://web.lexis-nexis.com.ezproxy.tamu.edu:2048/congcomp/document?_m=b30edeac756cef9d83435fc135febbb8&_docnum=8&wchp=dGLbVlb-zSkSA&_md5=9e0a947318af4da6fec3cd53ae2f5764
- Noell, G. H., Gresham, F. M. & Gansle, K. A. (2002). Does treatment integrity matter? A preliminary investigation of instructional implementation and mathematics performance [Electronic version]. *Journal of Behavioral Education, 11(1)*, 51-67.
- Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill, Inc.
- O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research [Electronic version]. *Review of Educational Research, 78(1)*, 33-84.
- Putnam, L. R. (Ed.). (1996). *How to become a better reading teacher strategies: For assessment and intervention*. Englewood Cliffs, NJ: Prentice-Hall, Inc.
- Richardson, E. & DiBenedetto, B. (1999). *Integrated skills method a resource guide for teaching reading*. El Paso, TX: ISM Teaching Systems, Inc.

- Richardson, K. (1987). Test review: Decoding skills test (DST) [Electronic version]. *The Reading Teacher, 41*, 220-223.
- Roberts, J., Jurgens, J. & Burchinal, M. (2005). The role of home literacy practices in preschool children's language and emergent literacy skills [Electronic version]. *Journal of Speech, Language & Hearing Research, 48*(2), 345-359.
- Rogosa, D. R. & Willett, J. B. (1983). Demonstrating the reliability of the difference score in the measurement of change [Electronic version]. *Journal of Educational Measurement, 20*, 335-343.
- Ruiz-Primo, M. A. (2006). *A multi-method and multi-source approach for studying fidelity of implementation*. Los Angeles, CA: Stanford University. (CSE Report 677)
- SAISD. (2006). *SAISD facts & figures – history and community*. Retrieved April 12, 2006, from <http://www.saisd.net/factfig/history.shtm>
- Spector, Paul (1981). *Research designs*. Sage University Papers Series on Quantitative Applications in the Social Sciences series no. 23. Newbury Park, CA: Sage.
- Spector, Paul (1992). *Summated rating scale construction*. Sage University Papers Series on Quantitative Applications in the Social Sciences series no. 82. Newbury Park, CA: Sage.
- Statistics Solutions. (2009). *Moderator variable*. Retrieved August 5, 2009, from <http://www.statisticssolutions.com/moderator-variable>

- StatSoft, Inc. (2010). Principal components and factor analysis. *In electronic statistics textbook*. Retrieve October 02, 2010, from <http://www.statsoft.com/textbook/principal-components-factor-analysis/>
- Stewart, M. T. (2004, May). Early literacy instruction in the climate of no child left behind [Electronic version]. *Reading Teacher*, 57(8), 732-743.
- Stockard, J. (2010). An analysis of the fidelity implementation policies of the what works clearinghouse. *Current Issues in Education*, 13(4). Retrieved December 17, 2010, from <http://cie.asu.edu/ojs/index.php/cieatasu/article/view/398/43>
- Texas Education Agency. (2002). *No child left behind public law 107-110*. (BR03 312 01 Fall, 2002) Retrieved December 30, 2005, from http://www.tea.state.tx.us/nclb/hottopics/Brochure_English_6pg.pdf
- Texas Education Agency. (2005-a). *2005 adequate yearly progress (AYP) guide for Texas public school district and campus*. (GE05 602 03 June, 2005) Retrieved December 30, 2005, from <http://www.tea.state.tx.us/ayp/2005/guide.pdf>
- Texas Education Agency. (2005-b). *Office of accountability and data quality final 2005 AYP results by district name*. Retrieved April 12, 2006, from <http://www.tea.state.tx.us/ayp/2005/distcampfinal05.pdf>
- Texas Education Agency. (2005-c). *Office of accountability and data quality school year 2005-2006 school improvement status*. Retrieved April 16, 2006, from <http://www.tea.state.tx.us/nclb/PDF/SIPCampus.pdf>

- Texas Education Agency. (2006-a). *2005-06 academic excellence indicator system campus reports*. Retrieved March 31, 2007, from <http://www.tea.state.tx.us/perfreport/aeis/2006/campus.srch.html>
- Texas Education Agency. (2006-b). *2005-06 academic excellence indicator system district reports*. Retrieved September 20, 2009, from <http://ritter.tea.state.tx.us/perfreport/aeis/2006/district.srch.html>
- Texas Education Agency. (2007-a). *Special education in Texas*. Retrieved September 9, 2007, from <http://www.tea.state.tx.us/special.ed/rti/>
- Texas Education Agency. (2007-b). *Assessments for students served by special education for the 2007-2008 school year-administrator letter*. Retrieved September 21, 2007, from http://www.tea.state.tx.us/student.assessment/resources/letters/2008/070919_Special_Ed_Assessments_Letter_with_AttachmentA.pdf
- Trcka, Z. L. (1994). *Teacher career ladder, teacher experience and student achievement growth*. Ph.D. dissertation, The University of Texas at Austin. Retrieved September 16, 2007, from *ProQuest Digital Dissertations database*. (Publication No. AAT 9506112).
- U. S. Department of Education. (2002). *Key policy letters signed by the education secretary or deputy secretary*. Retrieved September 06, 2007, from <http://www.ed.gov/policy/elsec/guid/secletter/020724.html>

- U. S. Department of Education. (2006). *Improving basic programs operated by local education agencies (Title I, Part A)*. Retrieved September 08, 2007, from <http://www.ed.gov/programs/titleiparta/index.html>
- U. S. Department of Education. (2009). *Reading first*. Retrieved September 11, 2010, from <http://www2.ed.gov/programs/readingfirst/index.html>
- Valentine, J. C. & Cooper, H. (2003). *Effect size substantive interpretation guidelines: Issues in the interpretation of effect sizes*. Washington, DC: What Works Clearinghouse. Retrieved August 08, 2009, from <http://ies.ed.gov/ncee/wwc/pdf/essig.pdf>
- Webster 1913 Dictionary (n.d.). *Monotonic*. Retrieved September 14, 2007, from Answers.com Web site: <http://www.answers.com/topic/monotonic-2>
- What Works Clearinghouse (2008). *WWC procedures and standards handbook version 2.0 – December 2008*. Retrieved December 17, 2010, from <http://ies.ed.gov/ncee/wwc/references/idocviewer/Doc.aspx?docId=19&tocId=2>
- What Works Clearinghouse (2010). *About us*. Retrieved December 18, 2010, from <http://ies.ed.gov/ncee/wwc/aboutus/>
- Williams, R., Zimmerman, D., Rich, J., & Steed, J. (1984). An empirical study of the relative error magnitude in three measures of change [Electronic version]. *The Journal of Experimental Education*, 53(1), 55-57.

- Witt, J. C., Noell, G. H., LaFleur, L. H., & Mortenson, B. P. (1997). Teacher use of interventions in general education settings: Measurement and analysis of the independent variable [Electronic version]. *Journal of Applied Behavior Analysis, 30(4)*, 693-696.
- Ysseldyke, J. E., Christenson, S., Pianta, B., & Algozzine, B. (1983). An analysis of teachers' reasons and desired outcomes for students referred for psychoeducational assessment [Electronic version]. *Journal of Psychoeducational Assessment, 1(1)*, 73-83.
- Zvoch, K. (2009). Treatment fidelity in multisite evaluation: A multilevel longitudinal examination of provider adherence status and change [Electronic version]. *American Journal of Evaluation, 30(1)*, 44-61.

APPENDIX A
SCHOOL POOL

School	Average ISM Level	Reading TAKS %	School #s	Program #s
A*	High	93%	340	9
B*	High	80%	385	20
C	High	77%	740	15
D*	High	81%	443	11
E*	High	82%	663	10
F*	High	76%	299	9
G*	High	70%	671	12
H*	High	93%	404	11
I*	Low	90%	697	12
J	Low	87%	404	9
K*	Low	85%	496	18
L	Low	92%	631	11
M*	Low	88%	323	9
N	Low	70%	507	24
O*	Low	81%	556	27
P*	Low	85%	436	9
Q*	Low	75%	563	15
R	Low	86%	247	10

(The High and Low school pool is based on 05/06 ILGE gains. High schools included those schools with nine or more students that showed gains greater than or = to 0.7. Low schools included those schools with nine or more students that showed gains less than or = to 0.3. Seven High and six Low schools, marked with an *, were chosen for inclusion in the study. Ten teachers at High schools and ten teachers at Low schools were observed for this study.)

APPENDIX B

QUALITY OF USE INNOVATION CONFIGURATION MAP WITH

SUMMATIVE SCALE – PRELIMINARY STUDY

INITIAL PLACEMENT

(A) +4	(B) +3	(C) +2	(D) +1	(E) +0
Use of DST ILGE/RSGE to initially place students. Students not grouped above recommended level.	Use of DST ILGE/RSGE to initially place students. If grouped above recommended level, verified with appropriate mastery test.	Use of non-DST scores/availability to group students. If grouped above recommended level, verified with appropriate mastery test.	Use of DST ILGE/RSGE to initially place students. No verification if grouped above recommended level.	Use of non-DST scores/availability to group students. No verification if grouped above recommended level.

DIRECT INSTRUCTION

(A) +4	(B) +3	(C) +2	(D) +1	(E) +0
Direct instruction by teacher in 1 to 3 small groups of one to six students with approximately same reading level.	Direct instruction by teacher in 1 to 3 small groups of more than six students up to ten students with approximately same reading level.	Direct instruction by teacher in more than 3 small groups of one to six students with approximately same reading level.	Direct instruction by teacher in more than 3 small groups of more than six students up to ten students with approximately same reading level or more than ten students in a single group with any number of total groups.	No direct instruction. Students grouped with approximately same reading level reading to each other.

FREQUENCY OF DIRECT INSTRUCTION

(A) +4	(B) +3	(C) +2	(D) +1	(E) +0
Direct Instruction of all groups 3 or more times per wk.	Direct Instruction of all groups at least 2 times per wk. & beginning readers 3 or more times per wk.	Direct Instruction of all groups at least 2 times per wk.	Direct Instruction of all groups at least 1 time per wk. & beginning readers at least 2 times per wk.	Direct Instruction of all groups 0-1 times per wk.

FREQUENCY OF INDEPENDENT/ INTERDEPENDENT RELATED ACTIVITIES

(A) +4	(B) +3	(C) +2	(D) +1	(E) +0
Use of Independent/ Interdependent Literature and/or Test Prep. Activities by all groups 2 or more times per wk.	Use of Independent/ Interdependent Literature and/or Test Prep. Activities by all groups at least 1 time per wk. and some groups 2 or more times per wk.	Use of Independent/ Interdependent Literature and/or Test Prep. Activities by all groups at least 1 time per wk.	Use of Independent/ Interdependent Literature and/or Test Prep. Activities by some groups at least 1 time per wk. and all groups at least 1 time every 2 wks.	Use of Independent/ Interdependent Literature and/or Test Prep. Activities by all groups 1 time or less every 2 wks.

MATERIALS – ISM (LINGUISTIC PATTERN SERIES)

(A) +4	(B) +3	(C) +2	(D) +1	(E) +0
LPS current assigned book & Mastery test available for all program students and showing some student work in the majority of them. Some LPS games/strategy materials present and put together for use for each LPS level assigned.	LPS current assigned book & Mastery test available for all program students and showing some student work in the majority of them. Some LPS games/strategy materials present and put together for use.	LPS current assigned book & Mastery test available for all program students and showing some student work in the majority of them. Some LPS games/strategy materials present.	LPS current assigned book & Mastery test available for all program students and showing some student work in the majority of them.	LPS current assigned book &/or Mastery test not available or showing no student work in the majority of them.

MATERIALS – LITERATURE, TEST PREP

(A) +4	(B) +3	(C) +2	(D) +1	(E) +0
Literature <u>and</u> Test Prep materials at all LPS assigned levels present and evidence of use of both types present.	Literature <u>and</u> Test Prep materials at all LPS assigned levels present and evidence of use of one type present.	Literature <u>or</u> Test Prep materials at all LPS assigned levels present and evidence of use present.	Literature <u>and/or</u> Test Prep materials at some LPS assigned levels present and evidence of use present.	No Literature or Test Prep materials LPS at assigned levels present or no evidence of use present.

SCHEDULING

(A) +4	(B) +3	(C) +2	(D) +1	(E) +0
ISM classes scheduled at least 45 mins 5 times a week.	ISM classes scheduled at least 30 mins (but less than 45 mins) 5 times a week or at least 45 mins 3 or 4 times per week.	ISM classes scheduled at least 30 mins (but less than 45 mins) 3 or 4 times a week or at least 45 mins 2 times per week.	ISM classes scheduled at least 30 mins (but less than 45 mins) 2 times a week.	ISM classes scheduled less than 30 mins 2 times a week or less than 60 mins total per week.

LESSON PLANS

(A) +4	(B) +3	(C) +2	(D) +1	(E) +0
Individual lesson plans present for all students for ISM (LPS), Literature & Test Prep.	Lesson plans present for all groups for ISM (LPS), Literature & Test Prep.	Lesson plans present for all groups for ISM (LPS), and either Literature or Test Prep.	Lesson plans present for all groups for ISM (LPS).	No consistent lesson plans present.

APPENDIX C

QUALITY OF USE INNOVATION CONFIGURATION MAP WITH

SUMMATIVE SCALE – FULL STUDY

INITIAL PLACEMENT

(A) +4	(B) +3	(C) +2	(D) +1	(E) +0
Use of initial/annual DST ILGE to initially place students. Students not grouped above recommended level.	Use of initial/annual DST ILGE to initially place students. If grouped above recommended level, verified with appropriate mastery test.	Use of non-DST scores/availability to group students. If grouped above recommended level, verified with appropriate mastery test.	Use of initial/annual DST ILGE to initially place students. No verification if grouped above recommended level.	Use of non-DST scores/availability to group students. No verification if grouped above recommended level.

DIRECT INSTRUCTION

(A) +4	(B) +3	(C) +2	(D) +1	(E) +0
Direct instruction by teacher/aide in 1 to 3 small groups of one to six students with approximately same reading level.	Direct instruction by teacher/aide in 1 to 3 small groups of more than six students up to ten students with approximately same reading level.	Direct instruction by teacher/aide in more than 3 small groups of one to six students with approximately same reading level.	Direct instruction by teacher/aide in more than 3 small groups of more than six students up to ten students with approximately same reading level or more than ten students in a single group with any number of total groups.	No direct instruction. Students grouped with approximately same reading level reading to each other or direct instruction with groups of students, of any size, on various reading levels reading orally at the same time from different books.

FREQUENCY OF DIRECT INSTRUCTION

(A) +4	(B) +3	(C) +2	(D) +1	(E) +0
Direct Instruction of all groups 3 or more times per wk.	Direct Instruction of all groups at least 2 times per wk. & beginning readers 3 or more times per wk.	Direct Instruction of all groups at least 2 times per wk.	Direct Instruction of all groups at least 1 time per wk. & beginning readers at least 2 times per wk.	Direct Instruction of all groups 0-1 times per wk.

FREQUENCY OF INDEPENDENT/ INTERDEPENDENT RELATED ACTIVITIES

(A) +4	(B) +3	(C) +2	(D) +1	(E) +0
Use of Independent/ Interdependent Literature and/or Test Prep. Activities by all groups 2 or more times per wk.	Use of Independent/ Interdependent Literature and/or Test Prep. Activities by all groups at least 1 time per wk. and some groups 2 or more times per wk.	Use of Independent/ Interdependent Literature and/or Test Prep. Activities by all groups at least 1 time per wk.	Use of Independent/ Interdependent Literature and/or Test Prep. Activities by some groups at least 1 time per wk. and all groups at least 1 time every 2 wks.	Use of Independent/ Interdependent Literature and/or Test Prep. Activities by all groups 1 time or less every 2 wks.

MATERIALS – ISM (LINGUISTIC PATTERN SERIES)

(A) +4	(B) +3	(C) +2	(D) +1	(E) +0
LPS current assigned book & Mastery test available for all program students and showing some student work in the majority of them. Some LPS games/strategy materials present and put together for use for each LPS level assigned.	LPS current assigned book & Mastery test available for all program students and showing some student work in the majority of them. Some LPS games/strategy materials present and put together for use.	LPS current assigned book & Mastery test available for all program students and showing some student work in the majority of them. Some LPS games/strategy materials present.	LPS current assigned book & Mastery test available for all program students and showing some student work in the majority of them.	LPS current assigned book &/or Mastery test not available or showing no student work in the majority of them.

MATERIALS – LITERATURE, TEST PREP

(A) +4	(B) +3	(C) +2	(D) +1	(E) +0
Literature <u>and</u> Test Prep materials at all LPS assigned levels present and evidence of use at assigned grade levels of both types present.	Literature <u>and</u> Test Prep materials at all LPS assigned levels present and evidence of use at assigned grade levels of one type present.	Literature <u>or</u> Test Prep materials at all LPS assigned levels present and evidence of use at assigned grade levels present.	Literature <u>and/or</u> Test Prep materials at some LPS assigned levels present and evidence of use at some assigned grade levels present.	No Literature or Test Prep materials LPS at assigned levels present or no evidence of use at assigned grade levels present.

SCHEDULING

(A) +4	(B) +3	(C) +2	(D) +1	(E) +0
ISM classes scheduled at least 45 mins 5 times a week.	ISM classes scheduled at least 30 mins (but less than 45 mins) 5 times a week or at least 45 mins 3 or 4 times per week.	ISM classes scheduled at least 30 mins (but less than 45 mins) 3 or 4 times a week or at least 45 mins 2 times per week.	ISM classes scheduled at least 30 mins (but less than 45 mins) 2 times a week.	ISM classes scheduled less than 30 mins 2 times a week or less than 60 mins total per week.

LESSON PLANS

(A) +4	(B) +3	(C) +2	(D) +1	(E) +0
Individual lesson plans present for all students for ISM (LPS), Literature & Test Prep.	Lesson plans present for all groups for ISM (LPS), Literature & Test Prep.	Lesson plans present for all groups for ISM (LPS), and either Literature or Test Prep.	Lesson plans present for all groups for ISM (LPS).	No consistent lesson plans present.

APPENDIX D
DEMOGRAPHIC QUESTIONNAIRE
PRELIMINARY STUDY

1. In what school do you teach?

2. What grade level(s) do you teach?

3. What subject(s) do you teach?

4. How many months/years have you been teaching? _____

5. How many months/years have you been teaching the ISM reading program?

6. Were you trained by ISM personnel to use the program? _____ If not,
what training did you receive and who provided the training?

7. Have you taught other reading programs? _____ If so, what programs have
you taught? _____

How long were these taught? _____ Are you currently teaching any of these

programs? _____ If so, is it in combination with the ISM program or is it taught to
different students than you teach the ISM program? _____

8. How many students do you teach in the ISM program? _____ How many girls and how many boys are included? Girls _____ Boys _____
9. How many of your special education ISM students are: AI? ___ AU? ___ ED? ___ MR? ___ NEC? ___ OI? ___ OHI? ___ (S)LD? ___ SI? ___ TBI? ___ VI? ___ Multiple Disabilities? ___ Bi-lingual/Dual Lang? ___
10. How many of your special education students are in Resource? ___ Content Mastery? ___ Both? ___
11. Are any of your ISM students also served in Speech? _____ How many? _____
12. Do you teach ISM to students in groups? _____ If so, how many are in a group? _____ Are these groups composed of multiple grade level students? _____
13. If you teach ISM in groups, how do you determine which group in which to place students?
- _____
- _____
- _____
- _____
- _____
- _____
- _____
14. How many times per week do the students receive the ISM program? ___ How long is each session? _____ Are missed sessions made-up? _____

15. What is the ethnic makeup of your ISM program students?

APPENDIX E
DEMOGRAPHIC QUESTIONNAIRE
FULL STUDY

1. In what school do you teach?

2. What grade level(s) do you teach?

3. What subject(s) do you teach?

4. How many months/years have you been teaching? _____

5. How many months/years have you been teaching the ISM reading program?

6. Were you trained by ISM personnel to use the program? _____ If not,
what training did you receive and who provided the training?

7. Have you taught other reading programs? _____ If so, what programs have
you taught? _____

How long were these taught? _____ Are you currently teaching any of these

programs? _____ If so, is it in combination with the ISM program or is it taught to
different students than you teach the ISM program? _____

8. How many students do you teach in the ISM program? _____ How many girls and how many boys are included? Girls _____ Boys _____
9. How many of your special education ISM students are: AI? _____ AU? _____ ED? _____ MR? _____ NEC? _____ OI? _____ OHI? _____ (S)LD? _____ SI? _____ TBI? _____ VI? _____ Multiple Disabilities? _____ Bi-lingual/Dual Lang? _____ Bi-lingual/Dual Lang Denials? _____
10. How many of your special education students are in Resource? _____ Content Mastery? _____ Both? _____
11. Are any of your ISM students also served in Speech? _____ How many? _____
12. Do you teach ISM to students in groups? _____ If so, how many are in a group? _____ Are these groups composed of multiple grade level students? _____
13. If you teach ISM in groups, how do you determine which group in which to place students? (Please be specific concerning measures used such as: current grade level, class scheduling, DST scores, mastery tests, standardized assessments, teacher made tests, etc.)
- _____
- _____
- _____
- _____
14. How many times per week do the students receive the ISM program? _____ How long is each session? _____ Are missed sessions made-up? _____

15. What is the ethnic makeup of your ISM program students? (Please state approximate numbers per ethnic group.)

APPENDIX G

SAMPLES OF MATERIAL GIVEN OFFSITE REVIEWER

Campus _____ Teacher _____

ISM Progress Form
February 2008

Record the daily attendance for each student.

Student Name	ID	Grade	February 1	February 4	February 5	February 6	February 7	February 8	February 11	February 12	February 13	February 14	February 15	February 18	February 20	February 21	February 22	February 25	February 26	February 27	February 28	February 29
Royler		4	A1	A1-1-1-1-2	A1-1-1-1-2	A1-1-1-1-2	A1-1-2	A1-1-2	A1-2	A1-2	A1-2	A1-2	A1-2	A1-2	A1-2	A1-2	A1-2	A1-2	A1-2	A1-2	A1-2	A1-2
Enic		4	A3	A3	A3	A3	A3	A3	A3	A3	A3	A3	A3	A3	A3	A3	A3	A3	A3	A3	A3	A3
Matthew		2	A5	A5	A5	A5	A5	A5	A5	A5	A5	A5	A5	A5	A5	A5	A5	A5	A5	A5	A5	A5
Ernie	NO MT	4	A5	A5	A5	A5	A5	A5	A5	A5	A5	A5	A5	A5	A5	A5	A5	A5	A5	A5	A5	A5
Gelela		2	A4	A4	A4	A4	A4	A4	A4	A4	A4	A4	A4	A4	A4	A4	A4	A4	A4	A4	A4	A4
Paul	No Book	5																				
Francisco	No Book	5																				
Jonathan		2	A5	A5	A5	A5	A5	A5	A5	A5	A5	A5	A5	A5	A5	A5	A5	A5	A5	A5	A5	A5
Levie		5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5
Michael		5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5	4.5
Isaac	No Book	5																				
Spencer		5																				
Marina	No Book	1																				
Andrew		1	1-4	1-4	1-4	1-4	1-4	1-4	1-4	1-4	1-4	1-4	1-4	1-4	1-4	1-4	1-4	1-4	1-4	1-4	1-4	1-4
Andrea		5																				
Jatira		3																				

A1: Student absent from school A2: Student present, but did not attend intervention T: Teacher absent I: Intervention not offered
 Record the LPS Book/Lesson progress SDK - Star Base Kelly
 3 or 2-4 Record the Book: Lesson on the date started OMT: Record on the date the oral mastery test is administered
 THESE FORMS ARE FOR DOCUMENTATION OF SERVICES PROVIDED AND PROGRAM PROGRESS

Lesson P1-1

Matching Sounds

Handwritten notes: F B O T / O O

				
				
				
				
				
				
				

Part 1: Sentence Completion

Directions: Fill in the blanks with words from the following list to complete the sentences.

cheater speaker dreaming sneakers steamer
 neater screaming sweeping reading leaning

1. When I am at meetings, I always pay heed to the Speaker.
2. You can't always get what you want by Screaming.
3. If you put your things away, your room will be neater.
4. No one likes to see a cheater at school.
5. It is not a good idea to be leaning back in your seat.
6. You may be able to run faster if you wear sneakers.
7. When I am sleeping, I like to be dreaming.
8. I really like it when the teacher teaches reading.
9. A big ship may sometimes be called a steamer.
10. If your mother is cleaning, you can help with the Sweeping.

APPENDIX H

PEARSON CORRELATIONS SECTION

Pearson Correlations Section (Pair-Wise Deletion)

	5_item	ILGEGain	resILGE Gain	RSGE_Gain	resRSGE Gain	Pre_RSGE
5_item_StzSc	1.000	0.329	0.345	0.774	0.770	-0.116
ILGE_Gain	0.329	1.000	0.983	0.507	0.526	0.482
resILGE Gain	0.345	0.983	1.000	0.513	0.526	0.321
RSGE_Gain	0.774	0.507	0.513	1.000	0.999	-0.038
resRSGE Gain	0.771	0.526	0.526	0.999	1.000	0.000
Pre_RSGE	-0.116	0.482	0.321	-0.038	0.000	1.000
Post_RSGE	0.119	0.617	0.463	0.261	0.298	0.955
Pre_ILGE	-0.056	0.184	0.000	0.017	0.051	0.907
Post_ILGE	0.110	0.617	0.462	0.252	0.288	0.953

Cronbach's Alpha = 0.830 Standardized Cronbach's Alpha = 0.880

Pearson Correlations Section (Pair-Wise Deletion)

	Post_RSGE	Pre_ILGE	Post_ILGE
_5_item_StzSc	0.119	-0.056	0.110
ILGE_Gain	0.617	0.184	0.617
resILGgain	0.463	0.000	0.462
RSGE_Gain	0.261	0.017	0.252
resRSGEgain	0.298	0.051	0.288
Pre_RSGE	0.955	0.907	0.953
Post_RSGE	1.000	0.881	0.995
Pre_ILGE	0.881	1.000	0.887
Post_ILGE	0.995	0.887	1.000

Cronbach's Alpha = 0.830 Standardized Cronbach's Alpha = 0.880

VITA

Zelma Jane Gragg
Coordinator/Educational Diagnostician
Department of Educational Psychology
Texas A&M University

Texas A&M University
College of Education
Department of Educational Psychology
College Station, Texas 77843-4225
(979) 845-1831

Burleson-Milam Special Services
607 W. Alligator Street
Caldwell, TX 77836
(979) 567-9811
e-mail: zjsurfer@yahoo.com

Educational Background:

B.A., Sociology/Psychology, The University of Texas at El Paso, August 1976.

Alternative Certification Generic Special Education, The University of Texas at El Paso, August 1991.

M.Ed., Educational Diagnostician, The University of Texas at El Paso, December 1998.

Principal Certification, Texas A&M University, May 2005.

Ph.D., Educational Psychology, Texas A&M University, May 2011.