

THE EFFECTS OF MULTIPLE-STEP AND SINGLE-STEP DIRECTIONS
ON FOURTH AND FIFTH GRADE STUDENTS' GRAMMAR
ASSESSMENT PERFORMANCE

MATTHEW B. MAZERIK

MAY, 2006

THE EFFECTS OF MULTIPLE-STEP AND SINGLE-STEP DIRECTIONS
ON FOURTH AND FIFTH GRADE STUDENTS'
GRAMMAR ASSESSMENT PERFORMANCE
PRESENTED TO
THE FACULTY OF THE TEACHER EDUCATION PROGRAM
OF BETHANY UNIVERSITY OF THE ASSEMBLIES OF GOD
IN PARTIAL FULFILLMENT OF THE MASTER OF ARTS IN EDUCATION
BY
MATTHEW BRIAN MAZERIK
NOVEMBER, 2006

Acknowledgements

I would like to thank the Landmark Elementary school community for their encouragement and participation in my thesis project. I would also like to acknowledge the dedication of members of the faculty of the Teacher Education Program at Bethany University whose encouragement greatly aided me in the process of writing this thesis. Finally, I would like to thank my wife, Karie, for her confidence in, and support of, my work and her enduring patience during this lengthy process.

Table of Contents

	Page
Thesis Committee Acceptance.....	ii
Acknowledgements.....	iii
Table of Contents.....	iv
List of Tables.....	vi
Abstract.....	vii
Chapter I: Introduction and Literature Review.....	1
Introduction.....	1
Background of the Problem.....	1
Purpose of the Study.....	3
Research Question.....	3
Research Hypothesis/Null Hypothesis.....	4
Assumptions.....	4
Delimitations/Limitations.....	5
Operational Definitions.....	6
Literature Review.....	7
Cognitive Development Theory.....	8
Cognitive Development and Memory.....	14
Memory and Following Directions.....	17
Psychological Development Theory.....	19

Assessments and English Language Learners	19
Test Language and English Language Learners	20
Measurement of Academic Language Complexity.....	21
Test Accommodations and English Language Learners.....	23
English Language Proficiency Tests.....	25
California English Language Development Test (CELDT)	27
Chapter II: Method.....	30
Participants.....	30
Instrumentation	31
Procedures.....	32
Analyses	34
Chapter III: Results	36
Chapter IV: Discussion	41
References.....	45
Appendix A: CELDT Levels Matrix	51
Appendix B: Student Assessment Booklet	53
Appendix C: Assessment Administration Instructions	58
Appendix D: Student Testing Permission Slip	60

List of Tables

Table	Page
1. Correlation Coefficients of Mean Scores on Tests	37
2. Comparison of Mean Scores for Males and Females	37
3. Comparison of Mean Scores for 4 th and 5 th Graders.....	38
4. Comparison of Mean Scores for EO and ELL Participants.....	38
5. Comparison of Mean Scores for Advanced to Redesignated ELL and EO Participants	39
6. Correlation Coefficients of Participants' CELDT Levels and Test Scores	40

Abstract

The mean scores of English Language Learners (ELL) and English Only (EO) students in 4th and 5th grade ($N = 110$), across the teacher-administered Grammar Skills Test, were examined for differences in participants' scores on assessments containing single-step directions and assessments containing multiple-step directions. The results indicated no significant differences between participants' mean scores on the pre, mid, and posttest. Differences in the mean scores of 4th and 5th grade participants yielded statistically significant results ($p < .05$). ELLs with high levels of English language proficiency had higher mean scores than EOs. ELLs' language proficiency levels and mean scores revealed moderately strong correlations ($p < .01$). This study is relevant to teachers and researchers interested in direction complexity, memory, and ELL issues.

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Introduction

Background of the Study

Since the inception of high-stakes testing and the public school system's reliance on the results of standardized tests to measure student achievement, researchers have given much attention to the reliability, validity, and fairness of such assessments (Rahn, Stecher, Goodman, & Alt, 1997). However, these types of tests are not the only important forms of assessment. Bass and Glaser (2004) write about "informative assessments that improve teaching and learning by communicating learning goals, interpreting student performance, tracking progress over time, and suggesting appropriate corrective actions" (p. 2). Such assessments have a different purpose than high-stakes, standardized tests; a purpose more closely tied to the cycles of classroom instruction and assessment that constitute the academic year. It would behoove researchers and educators to devote more energy to creating formative and summative classroom tests and evaluating currently used tests for clarity, accuracy, and reliability (Fitt, Rafferty, Presner, & Heverly, 1999).

Teachers depend on daily or weekly assessments to gather evidence of student learning and to make decisions regarding a classroom's next instructional steps (Brown, 1998). The majority of research articles on classroom tests examined for this current study focused on multiple choice test items. Perhaps this is because the majority of

classroom curricula use these types of tests as measures of student learning, and most classroom teachers rely on textbook tests (Fitt, Rafferty, Presner, & Heverly, 1999). Numerous studies can be found on all aspects of the multiple choice assessment: the number of items in an answer, the number of items on a test, conventional and non-restrictive questions, test organization according to difficulty, and others (Aamodt & McShane, 1992; Costin, 1972; Newman, Kundert, Lane, & Bull, 1987; Rosemarie Kolstad & Robert Kolstad, 1994). However, teachers should be aware of the important issues surrounding all types of tests to be able to accurately and fairly evaluate their students. A problem occurs when teachers over-rely on one type of assessment, to the detriment of using other types of effective, applicable measures (Fitt, Rafferty, Presner, & Heverly, 2004).

One important issue concerning classroom assessments is the effect that test directions have on students' results. Directions are a basic, but key portion of an assessment. One qualitative study on mathematics questions showed that small differences in the wording and format of questions can mean the difference between error and accuracy for certain students (Wilson, 2004). These effects are especially pronounced among English Language Learner (ELL) populations who can be more sensitive to the syntactical complexity of test questions (Abedi, 2000).

This current study grew out of anecdotal observations of increasing and decreasing English Language Learner (ELL) and English Only (EO) students' scores across assessments. These assessments covered similar content areas but contained directions with varying levels of complexity. Assessment data meetings with other teachers revealed that increasing and decreasing scores were a trend across three classes

within at least one grade level, which prompted several discussions about the possibility of the assessment directions confounding the students' results. Upon closer observation, assessment scores at the 4th and 5th grade level in the content area of grammar, tended to be higher when directions were less complex e.g., 1 to 2 steps, than when they were more complex e.g., 3 or more steps.

Purpose of the Study

The purpose of this study was to investigate the effects of written directions on the results of a series of grammar assessments taken by 4th and 5th grade students. The study evaluated the participants' results from a series of grammar assessments with similar content but alternating single-step and multiple-step directions to determine whether the directions had a statistically significant effect on students' performance on the assessment.

Research Questions

This study poses questions regarding the effects that directions have on the results of assessments performed by elementary students and whether the results vary according to students' levels of English language proficiency, grade level, or gender. What differences, if any, exist between English Language Learners' mean scores from grammar assessment items with single-step directions and grammar assessment items with multiple-step directions? What differences, if any, exist between the mean scores of ELLs at various levels of English proficiency on grammar assessment items with single-step directions and grammar assessment items with multiple-step directions? What differences, if any, exist between English Only students' mean scores from grammar assessment items with single-step directions and grammar assessment items with

multiple-step directions? What differences, if any, exist between the mean scores of male and female students, 4th and 5th graders, and ELL and EO students on grammar assessment items with single-step directions and grammar assessment items with multiple-step directions?

Research Hypothesis/Null Hypothesis

This study hypothesizes that there will be a statistically significant difference between students' grammar assessment results on items with single-step directions and items with multiple-step directions. The null hypothesis states that there will be no statistically significant difference between students' assessment results on single-step and multiple-step directions.

Assumptions

The researcher of this study assumed that three rounds of testing would be sufficient to test the accuracy of the hypothesis. It was also assumed that the results of the study would be consistent across all groups in the study: ELL, EO, grade level, and gender. The directions and sentences in this study's assessment were closely modeled after the directions and sentences used in the 2nd and 3rd Editions of the *6-8 Week Skills Assessments Developed for Districts Using Houghton Mifflin Reading* test series developed by the Reading Lions Center at the Sacramento County Office of Education. This series is used throughout the state of California in all districts with *Houghton Mifflin Reading* as their English Language Arts curriculum and is assumed to be valid and reliable. The researcher assumed that a single pilot test with no more than eight to ten students would be sufficient in determining the validity and reliability of the assessment used in this study. Furthermore, it was assumed that a teacher board of three veteran

teachers, each with ten or more years of experience, would also serve as sufficient examiners of the assessment's accuracy and grade level appropriateness. It was assumed that the tests would be administered according to written procedures, on the same day and same time across classrooms, so that test administration will not confound the results. It was assumed that students from the 4th and 5th grades who have not received direct, explicit instruction during the course of the year in the grammatical content areas tested in the assessment could confound the data. Therefore, the results from students who had participated in reading intervention classes and who had not received regular classroom instruction in this academic content area were not included as participants in the study.

Delimitations/Limitations

The delimitations established by this study apply to the participants and assessment used for the research. This study focused on 4th and 5th grade students in Sheltered English Instruction (SEI) classrooms at a public elementary school, grades Kindergarten through 5th, in the Central Coast region of California. The participants' test results were based upon three rounds of assessments that alternated between simple, one-step directions and complex, three-step directions. The assessments tested the grammatical areas of complete and simple subjects and predicates. The assessments were taken in a pre-mid-post test format, one week apart, during the last trimester of the school year, with no intervention in between testing cycles. The 4th and 5th grade students performed the same assessments.

As a result of the above delimitations, results of this study cannot be generalized to other grade levels, to students in non SEI classrooms, or to other schools in California.

Nor, can they be generalized to other types of assessments or grammatical topics other than those assessed in this study.

Operational Definitions

California English Language Development Test (CELDT): The CELDT is the language test used by schools in the state of California to assess an English Language Learner's level of English language proficiency.

Directions: As used in this study, directions refer to the written instructions that students follow to complete the assessment.

English Language Learner (ELL): An English Language Learner is a student whose primary language is anything other than English. It includes the various language proficiency designations, English Learner, Limited English Proficient, Fluent English Proficient, and Redesignated student.

English Learner (EL): A student whose primary language is anything other than English is also referred to as an English Learner. This term is used synonymously with ELL.

English Only (EO): A students whose primary language is English.

Fluent English Proficient (FEP): Fluent English Proficient is a term referring to a student whose primary language is anything other than English and who is advanced in the knowledge and use of the English language to the point of being considered fluent upon first entering the California public school system.

Limited English Proficient (LEP): Limited English Proficient refers to a student, whose primary language is anything other than English, who is not advanced in the knowledge and use of the English language to the point of being considered fluent upon first entering the California public school system.

Multiple-step directions: Assessment directions requiring three steps which students must follow to complete the assessment successfully were referred to as multiple-step directions in this study. An example of multiple-step directions would be: Draw a line separating the complete subject and complete predicate. Circle the simple subject. Underline the simple predicate.

Redesignated student: A Redesignated student is a student whose primary language is anything other than English and who has reached a point in the knowledge and use of the English language that it is no longer considered necessary to test her for English language progress using assessments specifically designed for ELLs, such as the CELDT exam.

Single-step directions: Assessment directions requiring only one step which students must follow to complete the assessment successfully were referred to as single-step directions in this study. An example of single-step directions would be: Underline the simple predicate.

Literature Review

A number of areas should be considered when undertaking a study that utilizes assessments and student participants. An understanding of cognitive development theory

and the cognitive levels of the participants in the study will aid in creating and administering academically and developmentally appropriate assessments. Findings on memory and the ability to follow directions are important, considering the varying complexity of the directions used in this study's assessment. Psychological development levels of the participants will be examined to gain a holistic view of their cognitive development, especially as it applies to processing written material. The large number of ELL participants in this study requires a close look at research pertaining to English Language Learners and assessments, and more specifically, the effects of test language complexity and test accommodations on ELLs' assessment results. Finally, examining the leveling of ELLs, based on the results of English language proficiency tests, will clarify the relationship between language proficiency and the results of this study's assessment.

Cognitive Development Theory

Cognitive development theory, finding its rise in the 20th century, has become the criteria with which most psychologists and educators describe intellectual growth in humans from infancy to adulthood. Jean Piaget is the individual most responsible for shaping the language, research, and beliefs surrounding childhood cognitive-developmental theory (Siegler & Ellis, 1996).

Siegler and Ellis (1996) describe three overarching contributions that Piaget made to the theory of childhood cognitive development. First, children are constructivist by nature. Part of cognitive development includes an innate desire to find new problem-solving strategies even when adequate ones are in place. Constructivism also theorizes that "construction of new understandings involves an integration of prior understandings

and new experience” (Siegler & Ellis, 1996, p. 212). Piaget’s second contribution is that there are certain essentials or characteristics that occur within children’s cognitive development at different age levels. Piaget defined these as operational stages and gave theorists the most widely used and succinct terms by which to describe the stages. Lastly, there is dynamism or the possibility of change within and across cognitive stages of development. Piaget allowed for and expected cognitive variability and competition of ideas in individual learners as means to cognitive transformations. Children may or may not exhibit a prescribed pattern of cognitive development depending on individual circumstances (Siegler & Ellis, 1996).

Piaget’s stages of cognitive development, with their corresponding ages are as follows: the sensorimotor period, birth to age 2; preoperational thought, age 2 to age 6 or 7; concrete operations, ages 6 or 7 to ages 11 or 12; and formal operations ages 11 or 12 to adult (as cited in Mooney, 2000). Piaget (1969/1970) delineates some of the characteristics of each developmental level. During the first stage of cognitive development, children gain understandings through actions and intelligence directly linked to their senses. They also begin to build important cognitive schemas like conservation of objects, reversibility, and causal relationships which they will add onto in later stages. In preoperational thought, children begin to be able to function on a symbolic level. They create mental images, engage in symbolic play, and, most importantly, expand their use of language, a symbolic tool. At the concrete operations stage, children start forming mental operations. They are able to classify objects, deal with more complex numerical problems, look at relationships between objects, and understand operational reversibility. The final stage, formal operations, allows children

to begin to operate in the realm of the hypothetical. They can look at situations from multiple viewpoints, employ logic, and use reasoning to reach hypothetical solutions.

Piaget (1969/1970) also discusses the necessary elements that help move children through the various stages of cognitive development. The first catalyst for cognitive development is the natural, physical maturation of the individual. Regardless of acceleration due to education or possible delays, cognitive development occurs in conjunction with physical development. A second catalyst comes in two forms of what Piaget terms “acquired experience” (p. 37). Physical experience entails physically interacting with objects, i.e. experimentation, and making discoveries about their properties through abstraction. Logico-mathematical experience also involves physical interaction with objects. However, discoveries are made, not about the objects themselves, but about the processes surrounding the interactions or manipulations of the objects. The third catalyst for cognitive development is the presence of “educational or social communications” (p. 39). Though most linguistic communication take place verbally, Piaget continues to emphasize children’s physical experiences and experimentations as central for developing the necessary logic to interpret verbal communications.

The participants in this current study ranged in age from 9 to 11 years old. Most children at this age range function in the concrete operations stage of cognitive development with the exception of some 11 year olds who could be functioning in the formal operation stage (Mooney, 2000; Piaget, 1969/1970). As stated earlier, children in the concrete operation stage are forming and using mental operations that allow them to classify objects, to examine relationships and reversibility between items, and to think

logically about number operations. Their mental abilities are still limited to dealing with the concrete world and, typically, they have not moved into the realm of the hypothetical. The process of transformation from this stage to the next is not instantaneous, but occurs gradually, during transition periods, where children may be using concrete mental operations which they cannot adequately explain (Piaget & Inhelder, 1964). One study found that concrete-operational children in the 9 to 11 age range, who comprehend metaphorical phrases, are starting to approach formal operational thought with the aid of the mental operation, intersection. Intersectional thought allows that a common attribute can be shared between two classes or groups that are generally mutually exclusive. When coupled with symbolic understanding of language, intersectional thought allowed the cognitive-operational participants to adequately paraphrase and explain metaphorical statements with nearly the same frequency as the formal operational group of participants (Cometa & Eson, 1978). Based on the presence of operational intersection in the 9 to 11-year-old age range and Piaget's views on children's progress from one stage to another, it would be appropriate to consider the participants in this current study as somewhere in the transitional continuum from the concrete operations stage to the formal operations stage.

Vygotsky, another cognitive stage theorist, considers the onset of physical puberty in the child as the specific point where true concept formation begins. He states, "No new elementary function, essentially different from those already present, appears at this age [puberty], but all the existing functions are incorporated into a new complex structure, form a new synthesis, become parts of a new complex whole" (Vygostky, 1962, p. 59). Prepubescent children display cognitive developments that allow them to

perform operations similar to those found in genuine concept formation, but these are simplistic, individual forms that fully develop and synthesize through puberty. Vygotsky observes three basic phases of cognitive development: syncretism, complex thought, and conceptual formation. He also emphasizes the importance of social interactions through play and language as an essential means for cognitive development (Vygotsky, 1962).

Participants in this current research study, nearing or beginning puberty, could be characterized as beginning the final stage of cognitive development, conceptual formation. They are beginning their cognitive transformation by forming “potential concepts” that can be changed or built upon as a means of practice for abstraction and pure conceptualization (Vygotsky, 1962, p. 81). Vygotsky also addresses children’s cognitive development as it relates to the areas of writing and grammar, two content areas pertinent to this current study. It is often observed that there is significant developmental lag in a students’ writing when compared to oral communication, throughout their schooling. This discrepancy can be attributed to several factors related to the complexity of written language. Written language is an abstract form of linguistics in much the same way that algebra relates to basic arithmetic. It requires students to detach from their natural, oral inclination toward speech and operate in the deeply symbolic territory of language. Writing also demands more language analysis and conscious work than oral language (Vygotsky, 1962).

The same situation is found in the area of grammar. Children come into school already possessing a basic, oral command of grammar that is used unconsciously. It is then, after explicit instruction and conscious effort, that children become aware of the grammar and structure of the language they already use (Vygotsky, 1962). Therein lays

the potential difficulty for many English Language Learners. They are working on analytical, abstract, and structural levels with a language that they have yet to unconsciously internalize.

Some current trends in cognitive development theory have moved away from the traditional stage theories of Piaget and Vygotsky. One such approach is termed process-oriented because of its concern with the process of how cognitive development occurs as opposed to the stages in which it occurs (Granott, 1998). Seigler's wave theory, a type of process-oriented developmental theory, has changed the way researchers view variability, and he has advocated the use of the microgenetic method as a more accurate means of research (as cited in Granott, 1998). Wave theorists postulate that cognitive development occurs as a series of overlapping waves that build and shift according to an individual's thought patterns and cognitive skills at any point in his development. At one point in development, an individual may be displaying a high frequency of seriation thought processes, and then, the weight of that particular wave may push him toward the development and frequent use of conservation thought. This theory accounts for variabilities, inconsistencies, and contradictions in children's cognitive development, all issues normally explained away or ignored by traditional stage theory (Granott, 1998). The microgenetic research method is characterized by "observations of individual children throughout the period of change, a high density of observations relative to the rate of change in that period, and intensive trial-by-trial analyses" (Seigler & Crowley, 1991, p. 606). It appears to incorporate instances of quantitative and qualitative research in the hopes of yielding more accurate data regarding individual subjects.

The integration of social and emotional behavior into the study of cognitive development is also becoming an important trend in current research (Guavain, 2005). This is an important integration that focuses on the holistic development of the child as opposed to extracting and studying one area in the hopes of understanding that area as a separate function. This approach also strives to take into account the social interactions and influences that help shape a child's cognitive development (Guavain, 2005).

Cognitive Development and Memory

The study of memory and its development is crucial to the field of education and educational assessments because students' comprehension and academic achievement are most often measured by assessments that require students to recall, from memory, information that they have learned over the course of instruction. Assessments that require the recall of information begin to have higher stakes as students' progress through their schooling. Eventually, students' ability to remember information for assessments determines the types of academic courses they can take, whether or not they graduate high school, and the type of college or university they can enter.

Piaget and Inhelder were the first researchers to discuss the relationship between cognition and memory development (as cited in Kail, 1979). They focused their research on memory as it relates to "reactions associated with recognition (in the presence of the object) and recall (in the absence of the object)" (Piaget & Inhelder, 1973, p. 4-5). Memory is closely tied to a subject's level of understanding and cognitive development. Its function is best described as assimilating new information that a subject deems relevant into her existing schema, and then, accommodating the schema to allow for the

new information (Piaget & Inhelder, 1973). Once the information has been assimilated, it is then ready for recognition or recall.

Kail (1979) described a number of mnemonic strategies that children develop to assist in recognition and recall. Rehearsal is a strategy in which an individual will audibly or silently mouth the material which is to be recalled at a later point. One study's results indicated that only 10% of 5 year olds used this strategy, while 60% of 7 year olds and 85% of 10 year olds relied on rehearsal (Flavell, Beach, & Chinsky, 1966). The researchers discussed the possibility of cognitive development and linguistic maturation coinciding with the increased use of this mnemonic device. Rehearsal is most often a spontaneous strategy utilized by individuals in higher stages of cognitive development (Kail, 1979). It can be expected that participants in this current study may use rehearsal as a strategy for recalling the directions, even though they have access to them throughout the assessment.

Other mnemonic strategies can be separated into categories of storage and retrieval (Kail, 1979). Children's development of information organizing strategies, such as outlining, mirrors their use of rehearsal strategies. In one study, children were given sets of picture cards to study for later recall and were observed for categorization activity. Fifth graders showed a sharp increase in spontaneous categorization of the pictures, while kindergartners, 1st graders, and 3rd graders all displayed either ineffective or little-to-no categorizing activities (Moely, Olson, Halwes, & Flavell, 1969). In a study of the use of cue cards for retrieval of picture items, Kobasigawa (1974) found that 33% of 6 year olds, 75% of 8 year olds, and over 90% of 11 year olds used the cue cards as a retrieval strategy. The effective use of the strategy for greater item recall also increased

significantly as age levels increased. The use and refinement of the various memory strategies appears to coincide with the cognitive development level of the individual.

In addition to the increased use of mnemonic strategies as children develop, Piaget and Inhelder (1973) found that memory may actually increase over periods of cognitive development. They found that recall of information in a group of 4 to 7 year olds improved over a period of eight months. The subjects were given a simple seriation test with a set of lines ordered from shortest to longest and asked to redraw them after a short, time period. When the subjects were asked, eight months later, to redraw what they had done earlier, all but two showed some level of improvement over their first drawing. Their increasing cognitive development in the area of seriation had aided their memory and helped them show improved skills. Fourth and 5th graders in this current study, who have properly learned and integrated the grammatical content on this assessment, could show improved ability in finding subjects and predicates in sentences than when it was first taught and tested, eight months ago.

There is considerable debate as to why short-term memory improves with children's cognitive development. Pascual-Leone, using Piaget's various memory studies, postulated that short-term memory capacity increases significantly with every two years of cognitive development, while Case, Kurland, and Goldberg's study attributed increased memory recall to the development of faster, more efficient encoding and retrieval of information (as cited in Engle, Carullo, & Collins, 1991). Chi and Dempster, in separate studies, both argued that memory improves with cognitive development due to the increased use of mnemonic strategies like the ones previously mentioned (as cited in Engle, Carullo, & Collins, 1991). Whatever the determining

factor, it can be expected that the 4th and 5th graders in this current study may show differences in their performance of the assessment due in part to differences in their cognitive development levels and short-term memory ability.

Memory and following directions.

Memory as it relates to following directions has particular relevance to this current study and even greater relevance in the classroom given the frequent number of directions students are given throughout the day. However, there has been little research conducted concerning short-term memory as it relates to following directions (Kaplan & White, 1980). Furthermore, most studies surrounding the topic have focused on simple actions given to subjects for the purpose of recall (Engle, Carullo, & Collins, 1991). Early 20th century researchers, Binet and Simon, and Thorndike, examined memory and following directions as it relates to general intelligence. They placed items assessing students' ability to follow directions within intelligent quotient tests (as cited in Engle, Carullo, & Collins, 1991).

It was not until more recently, that researchers have examined the complexity of directions and students' ability to follow them. Kaplan and White (1980) assessed directions for complexity by counting the number of behaviors and qualifiers contained in a set of directions. They define a behavior as a single response that satisfies the condition of the direction. A qualifier is a particular condition or qualification under which the behavior must be carried out. For example, the set of directions, "Please be seated and open your textbook to page 2," would contain two behaviors and one qualifier. As the number of behaviors and qualifiers grows, the direction becomes more complex. Their study also recorded the frequency and complexity of directions given by a number of

teachers in Kindergarten through 5th grade classrooms. Teachers administered an average of one direction every 40 seconds, and 71% of directions were considered simple, containing one behavior and one qualifier. Kaplan and White (1980) found that there was a significant increase in the ability of students to follow increasingly complex directions in grades Kindergarten through 2nd, but there were no significant differences found in 3rd through 5th grades, possibly due to a ceiling effect on the complexity of the directions. The upper grades were able to follow more complex directions with greater accuracy. However, their accuracy began to decline when more than one qualifier was present in a direction or when one qualifier was introduced in a direction with multiple behaviors.

Kaplan and White's (1980) method of measuring direction complexity was used in another study to examine the relationship between short-term memory capacity and cognitive skills. Engle, Carullo, and Collins (1991) determined elementary students' short-term memory capacity using a series of word-span tests and compared those results with students' scores in reading comprehension and following directions. They found strong, significant correlations between 6th graders' memory-span scores, their performance on reading comprehension tests, and their ability to follow directions. As the complexity of the directions increased, 6th graders in the highest quartile of memory-span were able to maintain a higher percentage of accuracy, while students in the lowest quartile made greater declines in accuracy. However, at the highest level of complexity, both sets of 6th graders showed similarly decreased accuracy. Given the interrelatedness of short-term memory, reading comprehension, and following directions, it would be

interesting to examine the effects that an additional language barrier might have on the test results of English Language Learners, such as the ones in this current study.

Psychological Development Theory

To gain a better understanding of the psychological development of this study's participants, one could explore the work of Erikson. Erikson created the psychosocial development theory of the Eight Ages of Man. Participants in this current study would be placed in the stage of Industry vs. Inferiority (Anselmo & Franz, 1995; Mooney, 2000). This is the last stage of childhood before the onset of adolescence and transformation into adulthood. Industry vs. Inferiority is framed in terms of school life and systematic instruction. Erikson (1950) states, “[The child] now learns to win recognition by producing things...To bring a productive situation to completion is an aim which gradually supersedes the whims and wishes of play” (p. 259). Children in this stage learn to utilize the technology, e.g. literacy, of adults. They also learn to work “beside and with others” (p. 260) as preparation for adult work. The dangers of this stage include feelings of inadequacy or inferiority related to learning and productivity, and also, the view of work as purely obligation and conformity as opposed to fulfillment (Erikson, 1950).

Assessments and English Language Learners

Recent state and federal legislation requires that increasing numbers of English Learners participate in national and state-wide standardized assessments (Abedi, 2004). There are also an increasing number of states using high school exit exams which have the potential to drastically affect the number of ELL students graduating from high school (Adam, 2004). The growing use of standardized assessments for measuring

achievement led some researchers to examine the effects of the language used in such tests on English Learners' results and possible accommodations that could be made to allow for more equitable test-taking. The effects of language used in standardized test items and directions are important issues to this current study, in which language comprehension is an integral part of both the instructions and the sentence items used in this study's assessment.

Test Language and English Language Learners

Abedi, Leon, and Mirocha (2003) performed an analysis of extant testing data from multiple grade levels on the Stanford Achievement Tests (SAT 9) and the Iowa Test of Basic Skills (ITBS) in four sites across the country. They found that there was a relationship between English proficiency levels and performance on content-based assessments, and that ELLs' performance decreased as the assessments' language load increased. Further analysis revealed that assessment items with high language complexity may contain measurement error for ELLs, and the results of content-based assessments could have been confounded by language proficiency levels. These findings indicate that English Learners may have difficulty showing proficiency in a content area due to the language complexity of the test and not simply due to the lack of content knowledge of the subject.

Stevens, Butler, and Castellon-Wellington (2000) examined test language used in the ITBS and an English language proficiency test and found that the ITBS contained more complex language structures across all language areas. They also found low internal reliability, 0.56 on the ITBS Social Studies test, for the group of English Learners. Despite the language complexity and reliability issues, item-analyses of results

for ELLs with higher language proficiency revealed that test language may be less troublesome than the test content itself, thus showing variability within the ELL population. A study by Cunningham and Moore showed that when academic vocabulary, specifically test jargon, was reduced on comprehension questions in a standardized reading assessment, English-speaking students in the 4th through 6th grades performed significantly better on the comprehension items (as cited in Stevens, Butler, & Castellon-Wellington, 2000). Bailey (2000) analyzed a standardized assessment for language complexity and difficulty in its different content areas. Language demands were lower in the mathematics and science subsections and much higher in the reading comprehension subsection both in terms of the test content and test questions. A second finding indicated an increased difference in the scores of ELL and non ELL students on test items with language requiring a high level of processing skills.

Measurement of academic language complexity.

To determine the level of language difficulty in assessment directions or items, one must first have a scale by which to measure the complexity therein. Bailey (2000) distinguished academic language from “specialized content-specific language” and “everyday informal speech” (p. 82). Cunningham and Moore contrasted formal terms like “examine” and “cause” with their less formal counterparts, “look at” and “make” as a demonstration of the use of academic language (as cited in Bailey, 2000, p.82). Bailey created a qualitative scale of 0 to 3 for rating language demand and difficulty of test items based on three determiners. The first is where the difficulty is located in the item, i.e. a stimulus passage, stem, or response item. The second determiner is the area of language affected by the item, i.e. vocabulary, syntax, or discourse. The final determiner is the

type of language difficulty, i.e. “uncommon vocabulary, atypical parts of speech, non-literal use of language” (p. 86). The study examined reliability of the scale by comparing two coders’ rating of items across content areas and the language areas. Reliability ratings were based on exact agreements in rating and fluctuated from 60% to 100% with most scores falling in the 75 to 85 percentiles.

Solomon and Rhodes (1995) examined literature surrounding academic language and found two dominant definitions and a third, emerging view. The first definition characterizes academic language as language functions and structures that pose difficulty for ELLs. The second definition is based on the divergence of social language and academic language, which contains less contextual assistance. This view, developed by Cummings, appears to have come to the forefront in recent years, with the widely used terms of Basic Interpersonal Communication Skills (BICS) for social language and Cognitive Academic Language Proficiency (CALP) for academic language (as cited in Bailey, 2000). The final view of academic language is that it is a register of language uses and devices that fluctuate from subject to subject, depending on their applicability (Solomon & Rhodes, 1995).

Stevens, Butler, and Castellon-Wellington (2000) analyzed and compared the language complexity of two standardized tests. Language demand was determined by comparing each assessment’s topics, discourse, test language, syntactic complexity, vocabulary, and function. The area of syntactic complexity, as it applies to assessment language demand, is especially relevant to this current study, considering that all of the test items consisted of sentences with varying degrees of syntactical complexity, in which the subjects had to find the same grammatical features.

Test Accommodations and English Language Learners

Due to the language issues confronting English Learners on standardized tests, various studies have been conducted to examine potential accommodations that might address the equity problems arising from standardized tests. Test accommodations for ELs can be divided into two main categories: modifications to the test and modifications to the test's procedures (Butler & Stevens, 1997). Butler and Stevens (1997) indicated that there are numerous factors of eligibility and appropriateness when considering accommodations for students. Furthermore, they emphasized the lack of a body of empirical research concerning the validity of test accommodations. Abedi, Courtney, and Lord (2003) analyzed accommodation strategies while considering four criteria: the accommodation's effectiveness in closing the ELL/nonELL achievement gap, its effect on the validity of the assessment, its usefulness for students' with different background variables, and the accommodation's ease of implementation.

Current research investigating accommodations focused on modifications that aid ELLs in accessing the language structures of standardized test items. Abedi (2000) took a sample pool of word problems from the mathematics section of the National Assessment of Educational Progress (NAEP) and modified the linguistic structures of the problems. The mathematical content remained in tact, but the surrounding language was simplified, where possible. Though there was small improvement in ELL students' performance on the linguistically modified items, no statistically significant differences were found between the results on regular items and modified items. These findings contradicted an earlier study by Abedi, Lord, and Hofstetter (1998), which found marked improvement in ELLs' performance in 49% of linguistically modified test items from the

math section of the NAEP. Students in this study showed better performance on the linguistically modified items than both the original items and items translated into Spanish. Wilson (2004) demonstrated how changing the format and/or the wording on math test items can play to students' strengths or weaknesses, either aiding them in an accurate display of their knowledge or inhibiting them from showing a correct answer.

Several studies attempted to determine what specific types of test accommodations were most effective for the subjects. In examining 4th and 8th grade EL students' performance on science test items released from the NAEP, Abedi et al. found that the use of an English dictionary among 4th graders and linguistic modification of test items among 8th graders were the most effective accommodation strategies (Abedi, Courtney, Leon, Mirocha, & Goldberg, 2005). English Only students were given the accommodations also, but there was no significant increase in their test results. This suggests that the accommodations did not harm the assessment's validity. An earlier study by Abedi, Courtney, and Lord (2003) tested three different accommodations on science items for 4th and 8th graders. Students were given standard test items with no accommodations, linguistically modified items, items with the use of a customized English dictionary, and items with a bilingual glossary. The 4th grade ELL students showed no improvements with the use of accommodations. The 8th grade ELL students showed significant improvement when using the linguistically modified test items. The researchers suspected the differences in accommodation effectiveness between grade levels are due, in part, to the complexity of the subject matter of the assessments. Science textbooks and test items at an 8th grade level contain more complex linguistic structures than 4th grade material. Therefore, when some of the linguistic complexity is

removed or simplified through accommodation, 8th grade English Learners are able to show improvement on test performance. As with the previous study, none of the accommodation strategies appeared to harm the internal validity of the unmodified test. Castellon-Wellington (2000) allowed ELLs to choose which accommodations they preferred, either extra assessment time or having the directions and test items read aloud to them. Neither accommodation had a significant effect on students' test results. Based on Castellon-Wellington's findings, it could be assumed that the reading aloud of test directions should not have a significant impact on the participants' assessment results in this current study.

English Language Proficiency Tests

Over a decade ago, 83% of school districts nationwide were using various English language proficiency tests to aid in determining a student's status as Limited English Proficient (LEP) or nonLEP. Sixty-four percent of districts were using the tests to determine what kind of classroom placement would best suit the student, and 74% used the tests to help reclassify LEP students as proficient (Hopstock, Bucaro, Fleischman, Zehler, & Eu, 1993). Legislation, like the Improving America's School Act (1994), California's Proposition 227 (1998), and the No Child Left Behind Act (2001) have reinforced the necessity of tests used to measure ELL students' gains in English proficiency. These legislative acts also rely on English proficiency tests to monitor the number of EL students being redesignated to FEP and the number of LEP students making growth towards English proficiency (Jepsen & de Alth, 2005).

A number of English proficiency tests measure EL students' growth, and various studies measure the reliability and validity of these tests. Zehler, Hopstock, Fleischman,

and Greniuk compared the six major English proficiency tests being used at that time and found that they all tested different language skills, different tasks within skills, and that tabulation discrepancies prevented consistent proficiency leveling across the tests (as cited in Abedi, 2004). Pray (2005) administered three different English proficiency tests to nonHispanic and Hispanic native English-speakers, assuming that a reliable test would score the participants as fluent or proficient. The Woodcock-Munoz Language Survey scored none of the participants as fluent, 87% of the participants were proficient on the Idea Proficiency Tests (IPT), and 100% were proficient on the Language Assessment Scales (LAS).

Current studies are also concerned with the link between language proficiency tests, academic standards, and standards achievement tests. One such study states that, “Under Title III of the [No Child Left Behind] NCLB (2001b) every state needs to show linkage between state content standards and state ELD standards as input to the development of state English proficiency tests” (Bailey, Butler, & Sato, 2005, p. 1). The stronger the links between the language tests, academic content standards, and achievement tests, the more prepared the students will be to succeed across all tasks (Bailey, Butler, & Sato, 2005). Bailey, Stevens, Butler, Huang, and Miyoshi (2005) reported on creating language proficiency test items that are academic and “standards-informed,” and are linked to classroom texts and instruction (p. 4). These test items utilize the academic language that EL students must comprehend in order to show proficiency in academic English and in a particular academic content area like mathematics or social studies (Bailey et al., 2005). However, these items have not yet been tested on students. Abedi (2004) computed correlation coefficients to rate the

relationships between students' LEP classification codes, their scores on the Language Assessment Scales (LAS), and their standardized achievement test scores from the Stanford 9 (SAT-9) or Iowa Test of Basic Skills (ITBS). Findings showed a very weak relationship between the students' LEP classification and their scores on the LAS and the standardized achievement tests (Abedi, 2004). A study comparing the language used on the LAS and ITBS found that "the language of the LAS is less complex, more discreet and decontextualized, and more limited in its range of grammatical constructions than the language of the ITBS" (Stevens, Butler, & Castellon-Wellington, 2000, p. 22).

The necessity of properly identifying ELLs in relation to achievement tests and current studies showing serious flaws in English proficiency tests, have led to the continued development of new exams. The goal of these tests are to align language proficiency standards with academic content standards and achievement tests, thereby gaining a more accurate understanding of a student's ability according to his language proficiency test results (Abedi, 2004).

California English Language Development Test (CELDT)

This current study relies on the CELDT to group ELL participants according to their levels of English language proficiency. The CELDT was developed in 2000 as a response to California State legislation.

As stated in California Assembly Bill 748 (Statutes of 1997), the Superintendent of Public Instruction was required to select or develop a test that assesses the English language development of pupils whose primary language is a language other than English. Subsequently, California Senate Bill 638 (Statutes of 1999) required school districts to assess the English language development of all English Learners. The

California English Language Development Test (CELDT) was the test designed to fulfill these requirements (Technical Report for the CELDT 2001-2002, 2003, p. 1).

Since its development, it has become California's primary vehicle for classifying English Learners' levels of language proficiency and measuring their yearly progress towards fluency (Jepsen & de Alth, 2005). During the 2003-2004 school year, 1,795,101 California students were administered the CELDT Form C (Technical Report for the CELDT 2003-2004 Form C, 2005). The CELDT contains four sections that focus on the different areas of language: listening, speaking, reading, and writing. The listening and speaking sections are administered individually, using the provided protocol, with responses scored by the administrator. The reading and writing portions are given in a whole group setting. The student's scaled score is turned into a proficiency level of 1 through 5, beginning through advanced (Jepsen & de Alth, 2005). A more thorough description of the proficiency levels of the CELDT can be found in Appendix A (Technical Report for the CELDT 2003-2004 Form C, 2005, pg. 7). The Technical Report for the CELDT 2003-2004 Form C (2005) released by CTB/McGraw-Hill found strong reliability coefficients of 0.85 to 0.90 across all grades and areas of the test. The report also investigated the standard error of measurement, a measure of the margin of error should the student's score be compared to a completely reliable test. The CELDT's range of standard error is between 16 and 27 in scale score points. This would mean 1 to 2 points error in terms of a raw score (Technical Report for the CELDT 2003-2004 Form C, 2005).

Some researchers have been less concerned with the internal validity of the CELDT as much as its relationship to other academic areas. Murphy, Bailey, and Butler (2006) examined the alignment between the items from the language areas tested in the

CELDT and California's English Language Development (ELD) content standards. There was an average alignment of 55% between the ELD standards and all of the areas tested in the CELDT. Alignment varied between weak and moderate across grade spans of the test and across the dimensions of complexity, modality, and language demands (Murphy, Bailey, & Butler, 2006). The researchers recommend the development of "CELDT items that reflect more academic language functions and higher levels of complexity" (p. 59) as a way to improve alignment with ELD standards and improve the accuracy of students' proficiency classifications.

CHAPTER II

METHOD

Participants

This study relied on student participants to generate the results of the grammar skills assessment. The student participants performed the grammar skills assessment over three testing periods. The participants attended a public elementary school, grades Kindergarten through 5th, in the Central Coast area of California. Student participants ($N = 110$) from the 4th ($n = 60$) and 5th ($n = 50$) grades, who had received parental consent, took the grammar skills assessment. There were 55 male and 55 female participants in the study. The study contained 75% ELL students ($n = 81$) and 25% EO students ($n = 25$). The majority of ELL students were Hispanic, with Spanish as their first language; only one ELL student had a different ethnic background and first language. The EO students were predominantly Hispanic and Caucasian, and included several Filipino students.

Some students who participated in the assessment did not have their scores recorded in the database. The group excluded from the study consisted of students who had been removed from their normal classroom setting for English language arts instruction. Students who were removed from the normal classroom setting during English language arts could have missed instruction in the areas tested by the grammar skills assessments in this study, and therefore, could have potentially confounded the data.

Instrumentation

The instrument used to gather data for this study was adapted, and in the case of some of the directions, directly reproduced from the 2nd and 3rd editions of the 5th grade *6-8 Week Skills Assessments Developed for Districts Using Houghton Mifflin Reading*. This assessment series was created and edited by the assessment development team at the Reading Lions Center, a department in the Sacramento County Office of Education, Sacramento, California.

There was one test booklet used in this current study to gather data on the effects of assessment directions on student performance (Appendix B). The test booklet contained a cover sheet and three pages of assessments. The multiple-step directions on page one of the test booklet were reproduced from the checking skills sections of the Reading Lions Center assessments. The test items in the booklet were adapted from the items used in the checking skills section of the Reading Lions Center assessments. The sentences used in this study's assessment were similar in format to the sentences written in the Reading Lion Center assessments, with attention given to the number of sentences with action, helping, linking, and being verbs, and the types of subjects, articles, adjectives, and direct objects therein. These grammatical components can be constructed to form overly simplistic or overly complex sentences that hinder accurate evidence of English Language Learners' content knowledge (Abedi, 2000).

A second instrument created for this study was the form containing the teacher directions for administering the grammar skills assessments (Appendix C). This form was reviewed, along with the student test booklet, by three veteran teachers not involved

in this study, for clarity, accuracy, and appropriateness before the teachers administered the first assessment.

Procedures

The study was conducted beginning April 24, 2006. The researcher and the data scorer/recorder formed a list using the 4th and 5th grade class rosters of all the students whose parents had consented to their participation in three grammar skills assessments and who were not excluded from the study for the reason listed in the participants section of this study (see Appendix D for the parent consent form). This list was used to create a confidential master list that contained the student participants' names and the corresponding code used to label the students' results. The data scorer/recorder kept the only copy of this list, and the researcher had access to it upon request. The data scorer/recorder used the student participants' codes to create a data base for the group, in which the assessment results were to be recorded. The data scorer/recorder was not affiliated with the school district in which the study was being performed, and had no personal knowledge of the student participants.

The researcher conducted a meeting with the teacher participants on Tuesday, April 25, 2006. The classroom teacher participants were all multiple-subject credentialed teachers with one to sixteen years teaching experience. There were three female teachers, two 4th grade and one 5th grade, and three male teachers, one 4th grade and two 5th grade, including the researcher. The researcher explained the purpose of the study to the teachers, reviewed the procedures for administering the assessments, discussed the study's confidentiality policies, and established three consistent days and times on which to administer the assessment. The teachers also reviewed the directions and items in the

test booklet. However, they were not given teacher or student copies of the booklet at this time.

A single pilot test was also administered during this week using 4th and 5th grade students ($N = 10$) from the same school where the study was conducted. The students performed one or two pages from the test booklet, but did not perform all three cycles of the assessment. The pilot assessments were administered by classroom teachers and the students' results were not accessible to the teachers or participants. Immediately following the pilot test, the researcher conducted a brief discussion with the pilot group pertaining to the clarity of the directions, test items, and test administration procedures. Also during this time, three teachers, not involved in administering the assessments, reviewed the test administration procedures, directions, and contents of the assessments for clarity, accuracy and grade-level appropriateness. After reviewing the procedures and test booklets, the teachers returned the materials to the researcher with necessary comments or questions. Some slight revisions were made to clarify the steps in the test administration procedures.

The teachers administered the pretest, a multiple-step directions measure, on Wednesday, May 10th at 9:45 a.m. The mid test, a single-step directions measure, was administered on May 17th and the posttest, returning to multiple-step directions, on May 24th, both at 9:45 a.m. The researcher gave the teachers the student test booklets and a copy of the procedures for administering the test the mornings of the assessments. The directions for the assessment were read aloud once to the student participants before they were allowed to begin the assessment. The researcher and teacher participants followed the same procedure for the three assessment cycles.

The assessments were returned to the researcher immediately following the assessments, and they were secured until they were taken to the data scorer/recorder. The assessments were scored and the results recorded in a coded data base. Once the results for a student were recorded, the data scorer/recorder labeled the page that had been scored with the student's code and removed it from the test booklet. The data scorer/recorder kept the original test document of the results and the researcher had access to them upon request. The data scorer/recorder followed the same procedure for all three cycles of assessments.

Analyses

This study used paired samples *t*-tests, independent samples *t*-tests, bivariate correlation coefficients, and a one-way ANOVA test to analyze the statistical significance ($p < .05$) of participants' results. Paired samples *t*-tests were used to determine if there were significant differences between the means on the pretest, mid test, and posttest for both ELL and EO participants. The means of various subgroups of participants, male and female, 4th and 5th grade, and ELL and EO were all compared using independent samples *t*-tests. Bivariate correlation coefficients were used to determine the strength of relationships between the mean scores of all the participants on the three tests and also between CELDT levels of ELL participants and their mean scores on the three tests. A one-way ANOVA was calculated to determine differences in the mean scores of ELL participants leveled according to their CELDT designations.

For the purpose of this study, Fluent English Proficient (FEP) participants were included with ELL participants in ELL test results. However, they were excluded from test results that used CELDT designations as a factor because their language proficiency

level was not categorized by the CELDT. Redesignated participants were included with ELL participants in ELL test results and were also included in tests using CELDT designations as a factor because their language proficiency level was categorized by the CELDT. To simplify analysis in the one-way ANOVA test and to accommodate low numbers of participants in certain levels, CELDT designations were grouped into three categories: Early to Early Intermediate, Intermediate to Early Advanced, and Advanced to Redesignated.

CHAPTER III

RESULTS

The results of the ELLs' paired samples *t*-test ($n = 81$) indicated that there were no statistically significant differences ($p < .05$) between the mean scores on the pretest ($M = 3.64, SD = 3.49$), mid test ($M = 4.02, SD = 3.36$), and posttest ($M = 4.30, SD = 4.11$). There were strong, significant correlations in the relationships between the mean scores on the pretest and mid test ($r = .53, p < .01$). The relationship between the mean scores on the mid test and posttest was moderately strong ($r = .41, p < .01$).

A paired samples *t*-test that was run with EOs' assessment scores ($n = 25$) yielded similar results to the ELLs' scores. There were no significant differences between mean scores of the three tests: pretest ($M = 4.04, SD = .747$), mid test ($M = 4.84, SD = .697$), and posttest ($M = 4.64, SD = .764$). The correlation coefficients were very strong between the mean scores on the pretest and mid test ($r = .64, p < .01$) and the mean scores on the mid test and posttest ($r = .74, p < .01$).

To further examine the relationship of the assessment results, the correlation coefficients of the mean scores of all participants, ELL and EO, were studied across the three tests. Results showed moderate to strong relationships between the mean scores of the three assessments (Table 1).

Table 1

Correlation Coefficients of Mean Scores on Tests

		Pretest	Mid-test	Posttest
Pretest	<i>r</i>	1	.567**	.379**
	Sig. (2-tailed)	.	.000	.000
	<i>N</i>	110	110	110
Mid test	<i>r</i>	.567**	1	.472**
	Sig. (2-tailed)	.000	.	.000
	<i>N</i>	110	110	110
Posttest	<i>r</i>	.379**	.472**	1
	Sig. (2-tailed)	.000	.000	.
	<i>N</i>	110	110	110

** Correlation is significant at the 0.01 level (2-tailed).

A series of independent samples *t*-tests were used to determine differences in the mean scores between gender groups, grade levels, and the English language proficiency levels of the subjects in the study. There were no significant differences in the mean scores of males and females on any of the tests (Table 2).

Table 2

Comparison of Mean Scores for Males and Females

	Gender	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Std. Error</i>	
					<i>Mean</i>	<i>t</i>
Pretest	Male	55	3.45	3.214	.433	-.694
	Female	55	3.89	3.814	.514	
Mid test	Male	55	4.45	3.511	.473	1.015
	Female	55	3.80	3.246	.438	
Posttest	Male	55	4.60	3.690	.498	.716
	Female	55	4.05	4.275	.576	

Note. Scores are based on a 15-point scale

Fifth grade subjects showed higher mean scores on each of the tests than did 4th graders, with small, but statistically significant differences (Table 3).

Table 3

Comparison of Mean Scores for 4th and 5th Graders

	Grade	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Std. Error</i>	
					<i>Mean</i>	<i>t</i>
Pretest	5 th Grade	50	4.30	4.249	.601	1.723*
	4 th Grade	60	3.15	2.692	.348	
Mid test	5 th Grade	50	5.04	3.675	.520	2.655*
	4 th Grade	60	3.37	2.934	.379	
Posttest	5 th Grade	50	5.94	4.225	.598	4.155*
	4 th Grade	60	2.98	3.234	.417	

* $p < .05$, one-tailed.

EO students and ELL students' mean scores showed no significant differences on any of the tests (Table 4).

Table 4

Comparison of Mean Scores for EO and ELL Participants

	Language Proficiency	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Std. Error</i>	
					<i>Mean</i>	<i>t</i>
Pretest	ELL	81	3.64	3.490	.388	-.490
	EO	25	4.04	3.736	.747	
Mid test	ELL	81	4.02	3.361	.373	-1.051
	EO	25	4.84	3.484	.697	
Posttest	ELL	81	4.30	4.112	.457	-.371
	EO	25	4.64	3.818	.764	

An independent samples *t*-test comparing English Learners in the Advanced to Redesignated category of the CELDT and English Only students showed that the ELL

participants had higher mean scores than the EO participants in each assessment, although the differences were not statistically significant (Table 5).

Table 5

Comparison of Mean Scores for Advanced to Redesignated ELL and EO Participants

	Language Proficiency	<i>n</i>	<i>M</i>	<i>SD</i>	<i>Std. Error Mean</i>	<i>t</i>
Pretest	ELL	28	5.64	4.183	.791	1.464
	EO	25	4.04	3.736	.747	
Mid test	ELL	28	6.25	3.439	.650	1.481
	EO	25	4.84	3.484	.697	
Posttest	ELL	28	6.71	4.713	.891	1.747
	EO	25	4.64	3.818	.764	

A comparison was done to determine significant differences between mean scores on three ELL groups leveled according to the CELDT variable. On the pretest measure, the Intermediate to Early Advanced group's mean score was lower ($M = 2.32$, $SD = 2.440$, $n = 38$) than the Advanced to Redesignated group ($M = 5.64$, $SD = 4.183$, $n = 28$), and the difference was significant, $F(2, 71) = 8.744$, $p < .001$. The difference between the Early to Early Intermediate group's mean score ($M = 2.75$, $SD = 2.915$, $n = 8$) and the other two ELL groups was not significant.

On the mid test measure, the Intermediate to Early Advanced group's mean score ($M = 2.58$, $SD = 2.657$, $n = 38$) was lower than the Advanced to Redesignated group ($M = 6.25$, $SD = 3.439$, $n = 28$), and the difference was significant, $F(2, 71) = 11.871$, $p < .001$. The difference between the Early to Early Intermediate group's mean score ($M = 4.25$, $SD = 3.151$, $n = 8$) and the other two ELL groups was not significant.

On the posttest measure, the Advanced to Redesignated group's mean score ($M = 6.71$, $SD = 4.713$, $n = 28$) was higher, at a statistically significant level, $F(2, 71) = 10.993$, $p < .001$, than both the Intermediate to Early Advanced group ($M = 2.53$, $SD = 2.948$, $n = 38$) and Early to Early Intermediate group ($M = 2.88$, $SD = 2.357$, $n = 8$).

Correlation coefficients were performed to determine the strength of the relationship between ELL participants' CELDT levels and results from the three assessments. The results indicated moderately strong relationships between the two factors (Table 6).

Table 6

Correlation Coefficients of Participants' CELDT Levels and Test Scores

		Pretest	Mid test	Posttest
CELDT level	Pearson Correlation	.394**	.370**	.454**
	Sig. (2-tailed)	.001	.001	.000
	<i>n</i>	74	74	74

** Correlation is significant at the 0.01 level (2-tailed).

CHAPTER IV

DISCUSSION

The results of this study did not support the original research hypothesis that there would be statistically significant differences in students' assessment results between single-step and multiple-step direction grammar assessments across the pre, mid, and posttests. Therefore, it cannot be concluded, based on this study, that single-step and multiple-step directions have an effect on 4th and 5th grade students' grammar test results. Results indicated that students' scored low across all three assessments. Out of 15 possible points, participants' averages were 3.67 on the pretest, 4.13 on the mid test, and 4.33 on the posttest. The low scores indicate that participants' lack of test content knowledge or recall was a potential confound to the results of the study. To eliminate this confound, future tests should be administered using content that was recently taught by instructors or reviewed by the participants, ensuring adequate content knowledge to answer the test items.

The assessment was administered over a three-week period with one week in between each test. The assessment was administered toward the end of the school year, and, given normal conditions, should be administered with longer periods of time in between each cycle. Despite the short periods between assessments, test effect does not appear to have been a confound, given the lack of variance in the participants' average scores over the three tests. Further research could include lengthening the time frames to several weeks between the pre, mid, and posttest and including one instructional session of content review between each assessment. This would mitigate test effect and reduce

the potential confound that participants will be unable to recall the content as they move further into the testing cycle.

The results from this study showed no statistically significant differences in the mean scores of EO and ELL students. California Standardized Testing and Reporting (STAR) results from 2006 and previous years indicate that English Only students scored significantly higher than English Language Learners on most standardized test areas. The differences are especially exaggerated on assessments that contain English language skills items (see also Abedi, Leon, & Mirocha, 2003). Despite previous research findings and STAR score indications, a comparison of mean scores in this current study between EO students and ELL students in the Advanced and Redesignated category of the CELDT continuum revealed that the ELL students scored from 1.5 to 2 points higher than EO students on each assessment. The Advanced and Redesignated ELL students scored higher than all the other subgroups, including their ELL counterparts, whose mean scores were lower at a statistically significant level. However, the Advanced and Redesignated ELL students displayed average scores that were only slightly higher than all participants' averages and still well below what would be considered proficient for these assessments. The lack of any important differences and very few proficient scores in the subgroups' test results may indicate that, despite EO students' linguistic advantages, both groups do equally poorly recalling specific grammatical skills that require explicit instruction, in this case, locating subjects and predicates. Exploring the possible sources for the two groups' low performance in this grammatical area could be a basis for further research.

Fifth graders in this study had slight, but statistically significant, higher mean averages than 4th graders on all three assessments. This difference could be attributed to

the 5th grade participants having been exposed to the test content area for two consecutive years, whereas the 4th graders would have only learned about the content area during one period of instruction. Further research that includes equal amounts of content area exposure and review for both grades before testing would help clarify if, and for what reasons, the differences in grade level scores exist.

Various adjustments could be made to this current study to further investigate the effects of assessment instructions on ELL and EO students. In addressing ELL participants and their CELDT levels, this current study combined CELDT levels into three groupings due to insufficient numbers of participants in some levels (Early, $n = 2$; Early Intermediate, $n = 6$; Intermediate, $n = 17$; Early Advanced, $n = 21$; Advanced, $n = 5$; Redesignated, $n = 23$). Increasing the number of participants within each CELDT level and analyzing the results of the grammar assessments at each CELDT level could yield a clearer picture of how participants' English language proficiency levels affect their scores. Additional research could also include assessments containing other content areas such as mathematics or science to further investigate participants' performance of single-step and multiple-step directions across subject areas. This design would also allow researchers to examine how the assessment's content areas potentially confound participants' test results. Further studies could also be adapted to include a wider range of grade levels such as 1st through 5th grade, or 4th through 8th grade, as either range has the potential of spanning at least two important cognitive developmental levels according to Piaget (1969).

The use of Kaplan and White's (1980) method for determining the complexity of directions would have yielded more accurate descriptions of the assessment instructions than the terms "single-step and multiple-step directions" used for this study (see *Memory*

and following directions in this study's literature review). Employing their method on this current study's assessment directions, it was discovered that the single-step directions contained two behaviors and one qualifier (2 x 1), and the multiple-step directions contained four behaviors and four qualifiers (4 x 4). The implications of this rating system are important for this study, especially as it concerns the more complex directions. Kaplan and White (1980) found that only 65% of 3rd through 5th graders were able to follow nonacademic directions with four behaviors and four qualifiers. The level of complexity in the multiple-step directions coupled with the required academic content knowledge for the assessment may have proved too difficult to yield reliable results, regardless of subjects' levels of test content recall.

Finally, a more uniform use of test items containing similar levels of syntactical complexity would be beneficial when constructing an assessment such as the one used in this study. Anecdotal observation of students' assessments revealed that many participants who were able to distinguish between the complete subject and predicate and locate the simple subject and predicate in relatively simple sentences were unable to perform the same task in sentences containing prepositional phrases, an excessive number of modifiers, verb phrases, being verbs, and linking verbs. These additional levels of language complexity in the sentences most likely led to participants' lower scores on those items. These observations are congruent with the findings of other studies examining language complexity in standardized tests (Abedi, 2000; Bailey, 2000; Stevens, Butler, & Castellon-Wellington, 2000; Wilson, 2004).

In conclusion, modifying the test items' language complexity and testing procedures and adjusting language level groups would provide more insight into direction complexity and students' test results, as well as add pertinent information to this study.

References

*References marked with an asterisk indicate studies included in the meta-analysis.

*Aamodt, M. G., & McShane, T. (1992). A meta-analytical investigation of the effect of various test item characteristics on test scores and test completion times. *Public Personnel Management*, 21(2), 151.

Abedi, J. (2000). *Confounding of students' performance and their language background variables*. (Research Report 143, UD 033 934.). Author.

Abedi, J. (2004). Inclusion of students with limited English proficiency in NAEP: Classification and measurement issues (Tech. Rep. No. 629.). Los Angeles, CA: University of California, Los Angeles, Center for the Study of Evaluation (CSE).

Abedi, J., Courtney, M., & Leon, S. (2003). Effectiveness and validity of accommodations for English language learners in large-scale assessments (Tech. Rep. No. 608.). Los Angeles, CA: University of California, Los Angeles, Center for the Study of Evaluation (CSE).

Abedi, J., Courtney, M., Mirocha, J., Leon, S., & Goldberg, J. (2005). Language accommodations for English language learners in large-scale assessments: Bilingual dictionaries and linguistic modification (Tech. Rep. No. 666.). Los Angeles, CA: University of California, Los Angeles, Center for the Study of Evaluation (CSE).

- Abedi, J., Leon, S., & Mirocha, J. (2003). Impact of student language background on content-based performance: Analyses of extant data (Tech. Rep. No. 603.). Los Angeles, CA: University of California, Los Angeles, Center for the Study of Evaluation (CSE).
- Abedi, J., Lord, C., & Hofstetter, C. (1998). Impact of selected background variables on students' NAEP math performance (Tech. Rep. No. 478.). Los Angeles, CA: University of California, Los Angeles, Center for the Study of Evaluation (CSE).
- Adam, M. (2004). High school exit exams and Hispanics. *The Hispanic Outlook in Higher Education*, 15, 12-15.
- Anselmo, S., & Franz, W. (1995). *Early childhood development: Prenatal through age eight*. Second Edition. Englewood Cliffs, NJ: Prentice Hall.
- Bailey, A. (2000). Language analysis of standardized achievement tests: Considerations in the assessment of English language learners. In *The validity of administering large-scale content assessments to English language learners: An investigation from three perspectives* (Tech. Rep. No. 663.). Los Angeles, CA: University of California, Los Angeles, Center for the Study of Evaluation (CSE).
- Bailey, A. L., Butler, F. A., & Sato, E. (2005). Standards-to-standards linkage under title III: Exploring common language demands in ELD and science standards (Tech. Rep. No. 667.). Los Angeles, CA: University of California, Los Angeles, Center for the Study of Evaluation (CSE).
- Bailey, A. L., Stevens, R., Butler, F. A., Huang, B., & Miyoshi, J. N. (2005). Using standards and empirical evidence to develop academic English proficiency test items in reading (Tech. Rep. No. 664.). Los Angeles, CA: University of California, Los Angeles, Center for the Study of Evaluation (CSE).

- Bass, K. M., & Glaser, R. (2004). Developing assessment to inform teaching and learning (Tech. Rep. No. 628.). Los Angeles, CA: University of California, Los Angeles, Center for the Study of Evaluation (CSE).
- Brown, W. R. (1998). Test mapping: Planning the classroom test. *The Clearing House*, 71, 153.
- Butler, F. A., & Stevens, R. (1997). Accommodation strategies for English language learners on large-scale assessments: Student characteristics and other considerations (Tech. Rep. No. 448.). Los Angeles, CA: University of California, Los Angeles, Center for the Study of Evaluation (CSE).
- Castellon-Wellington, M. (2000). The impact of preference for accommodations: The performance of English language learners on large-scale academic achievement tests (Tech. Rep. No. 524.). Los Angeles, CA: University of California, Los Angeles, Center for the Study of Evaluation (CSE).
- Cometa, M. S., & Eson, M. E. (1978). Logical operations and metaphor interpretation. *Child Development*, 49, 649-659.
- *Costin, F. (1972). Three-choice versus four-choice items: Implications for reliability and validity of objective achievement tests. *Educational and Psychological Measurement*, 32, 1035-1038.
- CTB/McGraw-Hill. (2003, January). *Technical report for the California English Language Development Test (CELDT) (2000-2001)*. Monterey, CA: Author.
- CTB/McGraw-Hill. (2005, January). *Technical report for the California English Language Development Test (CELDT) (2003-2004 Form C)*. Monterey, CA: Author.

- Engle, R. W., Carullo, J. J., & Collins, K. W. (1991). Individual differences in working memory for comprehension and following directions. *Journal of Educational Research, 84*(5), 253-262.
- Erikson, E. (1950). *Childhood and society*. New York: W. W. Norton & Company.
- Fitt, D. X., Rafferty, K., Presner, M. T., & Heverly, M. V. (1999). Improving the quality of teachers' classroom tests. *Education, 119*, 643.
- Flavell, J. H., Beach, D. R., & Chinsky, J. M. (1966). Spontaneous verbal rehearsal in a memory task as a function of age. *Child Development, 37*, 283-299.
- Granott, N. (1998). A paradigm shift in the study of development [Review of the book *Emerging Minds*]. *Human Development, 41*, 360-365.
- Guavain, M. (2005). With eyes to the future: A brief history of cognitive development. *New Directions for Child and Adolescent Development, 109*, 119-126.
- Hopstock, P. J., Bucaro, B. J., Fleishchman, H. L., Zehler, A. M., & Eu, H. (1993). Descriptive study of services to limited English proficient students. Volume 2: Survey results (Contract No. LC91003001, Development Associates, Inc.). Washington, DC: Office of Policy and Planning U.S. Department of Education.
- Jepsen, C., & de Alth, S. (2005). *English learners in California schools*. San Francisco: Public Policy Institute of California.
- Kail, R. (1979). *The development of memory in children*. New York: W. H. Freeman and Company.
- Kaplan, C. H., & White, M. A. (1980). Children's direction-following behavior in grades k-5. *Journal of Educational Research, 74*(1), 43-48.
- Kobasigawa, A. (1974). Utilization of retrieval cues by children in recall. *Child Development, 45*, 127-134.

- Kolstad, R. K., & Kolstad, R. A. (1994). Applications of conventional and non-restrictive multiple-choice examination items. *The Clearing House*, 67, 317.
- Moely, B. E., Olson, F. A., Halwes, T. G., & Flavell, J. H. (1969). Production deficiency in young children's clustered recall. *Developmental Psychology*, 1(1), 26-34.
- Mooney, C. G. (2000). *Theories of childhood: an introduction to Dewey, Montessori, Erickson, Piaget & Vygotsky*. St. Paul, MN: Redleaf Press.
- Newman, D. L., Kundert, D. K., Lane, D. S. Jr., & Bull, K. S. (1988). Effect of varying item order on multiple-choice test scores: Importance of statistical and cognitive difficulty. *Applied Measurement In Education*, 1(1), 89-97.
- Murphy, A., Bailey, A., & Butler, F. (2006). *California English language development standards & assessment: Evaluating linkage & alignment*. Monterey, CA: CTB/McGraw-Hill.
- Piaget, J. (1970). *Science of education and the psychology of the child*. (D. Coltman, Trans.). New York: The Viking Press. (Original work published in 1969).
- Piaget, J., & Inhelder, B. (1973). *Memory and intelligence*. (A. J. Pomerans, Trans.). New York: Basic Books, Inc. (Original work published in 1968).
- Pray, L. (2005). How well do commonly used language instrument measure English oral-language proficiency? *Bilingual Research Journal*, 29(2), 387-410.
- Rahn, M. L., Stecher, B. M., Goodman, H., & Alt, M. N. (1997). Making decisions on assessment methods: Weighing the tradeoffs. *Preventing School Failure*, 41, 85.
- Siegler, R. S., & Crowley, K. (1991). The microgenetic method: A direct means for studying cognitive development. *American Psychologist*, 46(6), 606-620.
- Siegler, R. S., & Ellis, S. (1996). Piaget on childhood. *American Psychological Society*, 7(4), 211-215.

- Solomon, J., & Rhodes N. C. (1995). Conceptualizing academic language: Research report 15 (Contract No. R117G10022, National Center for Research on Cultural Diversity and Second Language Learning). Washington, DC: Office of Educational Research and Improvement.
- Stevens, R. A., Butler, F. A., & Castellon-Wellington, M. (2000). Academic language and content assessment: Measuring the progress of English language learners (Tech. Rep. No. 552.). Los Angeles, CA: University of California, Los Angeles, Center for the Study of Evaluation (CSE).
- Vygotsky, L. S. (1962). *Thought and language*. (E. Hanfmann, G. Vakar, Trans.). Cambridge, Massachusetts: The M.I.T. Press (Original work published in 1934).
- Wilson, L. D. (2004). On tests, small changes make a big difference. *Teaching Children Mathematics*, 11(3), 134.

APPENDIX A
CELDT PROFICIENCY LEVEL DESCRIPTIONS

CELDT 2003-2004 Form C Technical Report

Table 4 CELDT Proficiency Level Descriptions

Proficiency Level	Description
Advanced	Students performing at this level of English language proficiency communicate effectively with various audiences on a wide range of familiar and new topics to meet social and academic demands. In order to attain the English proficiency level of their native English-speaking peers, further linguistic enhancement and refinement are necessary.
Early Advanced	Students performing at this level of English language proficiency begin to combine the elements of the English language in complex, cognitively demanding situations and are able to use English as a means for learning in other academic areas.
Intermediate	Students performing at this level of English language proficiency begin to tailor the English language skills they have been taught to meet their immediate communication and learning needs.
Early Intermediate	Students performing at this level of English language proficiency start to respond with increasing ease to more varied communication tasks.
Beginning	Students performing at this level of English language proficiency may demonstrate little or no receptive or productive English skills. They may be able to respond to some communication tasks.

APPENDIX B
STUDENT ASSESSMENT BOOKLET

Test Booklet #1

Student Name _____

Grammar Skills Tests
Test Booklet #1

Read each sentence. Draw a slash mark (/) between the complete subject and the complete predicate. Then circle the simple subject and underline the simple predicate.

1. The scary story describes what happens during a tornado.
2. Strong winds knocked over the tall trees.
3. The two boys were caught in the rain storm.
4. Each thunderbolt in the sky made a loud boom.
5. The children had left the bedroom.

Read each sentence. Draw a line (|) between the complete subject and complete predicate.

1. My friend's cousin practices with the swim team.
2. The new girl likes to play in the water too.
3. Swimming is a challenging sport.
4. All the people in the water look happy.
5. That boy will jump off of the diving board.

Read each sentence. Underline only the simple subject.

1. My friend's cousin practices with the swim team.
2. The new girl likes to play in the water too.
3. Swimming can be a challenging sport.
4. All the people in the water look happy.
5. That boy will jump off of the diving board.

Read each sentence. Circle only the simple predicate.

1. My friend's cousin practices with the swim team.
2. The new girl likes to play in the water too.
3. Swimming can be a challenging sport.
4. All the people in the water look happy.
5. That boy will jump off of the diving board.

STOP

Read each sentence. Draw a slash mark (/) between the complete subject and the complete predicate. Circle the simple subject and double underline the simple predicate.

1. A strong basement is a safe place to hide during a storm.
2. The family's dog hid during the thunder.
3. Rumbling sounds filled the house.
4. Large tornadoes can cause major damage to a town.
5. The storm was scary and exciting.

APPENDIX C

ASSESSMENT ADMINISTRATION INSTRUCTIONS

Administering the Grammar Skills Assessments

Teacher Directions:

1. Have students put up privacy folders if their desk is connected to a classmate's desk or whatever method you usually employ to discourage copying.
2. Tell the students that they will have 15 minutes after they receive the directions to complete the assessment.
3. Pass out labeled test booklets to the students. Do not allow students to open the test booklets until you give them instructions to do so.
4. Have students open to the first page of the booklet. Read the directions from **page 1** of the *Test Booklet 1*, one time, as students follow along. Then, indicate the word STOP at the bottom of the page. When students reach the word stop at the bottom of the page they are working on, they will close the test and return it to the teacher. Students will only be completing one page of the test booklet during each testing period.
5. Start the clock and begin the testing period.
6. After all the tests have been returned or time has elapsed, collect the remaining tests, the teacher's copy of *Test Booklet 1*, and this paper and have two students bring the tests to room #5.
7. Follow the same administration instructions for the 2nd and 3rd round of tests. Student directions for the 2nd and 3rd round of tests are found on page 2 and 3 of the teacher copy of *Test Booklet 1*.
8. You may want to tell the students that you are aware that the sentences repeat themselves on the assessment.

APPENDIX D
STUDENT TESTING PERMISSION SLIP

Name _____

Student Testing Permission Slip

Parents,

This letter is to inform you that there will be three 20 minute grammar tests given to the 4th and 5th grades over the last trimester of the school year for the purpose of a special study conducted by a teacher here at the school. Your child's name and scores will be kept confidential, the results will not be seen by any person that is not involved in the study (including their classroom teacher), and your child's grade report will not be affected. Your child's participation in these tests is voluntary; they do not have to participate if you do not wish them to. However, we would appreciate full participation from all the students as it will help us get more accurate information for the study. Thank you for your help.

If you **do not** wish your child to participate in these tests, please indicate by signing below.

Parent's Signature

Student's Name

Nombre _____

PERMISO PARA ADMINISTRAR PRUEBA A ESTUDIANTE

Estimados Padres,

Esta carta es para informarles que habra tres pruebas de 20 minutos en gramatica los cuales seran administrados a los estudiantes del cuarto y quinto grado a lo largo del ultimo trimestre del ano escolar con el proposito de un estudio especial que sera conducido por un maestro aqui en la escuela. El nombre de su nino(a) y su puntuacion sera mantenido confidencialmente, los resultados no seran vistos por ninguna otra persona que no este enbolucrada en el estudio(incluyendo al maestro del salon), ni la nota academica de su hijo(a) sera afectada. La participacion de su hijo(a) en estas pruebas es totalmente voluntaria; no tienen que participar si usted no lo desea. No obstante, nos gustaria que todos participaran en el estudio porque eso nos daria informacion mas precisa del estudio. Muchas gracias por su ayuda.

Si usted no desea que su hijo(a) participe en estas pruebas, por favor indíquelo firmando en las líneas de abajo.

Firma del Padre/Madre

Nombre del Estudiante