THE TEST OF LANGUAGE COMPETENCE: A VALIDITY STUDY WITH LANGUAGE DISABLED AND NORMAL CHILDREN

Bу

CHERYL ANNE AINSWORTH

B.A., The University of British Columbia, 1974

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF ARTS

in

THE FACULTY OF EDUCATION

(Department of Educational Psychology and Special Education)

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

October 1989

(c) Cheryl Anne Ainsworth, 1989

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Educational Rychologic and Special Education

The University of British Columbia 1956 Main Mall Vancouver, Canada V6T 1Y3

Date October 13, 1989

Abstract

This study investigated the validity and related psychometric characteristics of the Test of Language Competence (TLC), published in 1985 by Wiig and Secord. The TLC was developed as a measure of higher order language functioning in children and adolescents between the ages of nine and eighteen years. Evidence concerning the psychometric characteristics of the TLC is reported in the test manual; however, to date, no studies addressed primarily to the subject of TLC validity have been reported in the literature. Moreover, no information is available concerning the effectiveness of its use with local school children. This study endeavored to examine the technical characteristics of the TLC using data obtained from 23 language disordered (LLD) and 23 control subjects sampled from the local school population. At the same time, the criterion-related validity of an informal language sample analysis was investigated.

Item analysis statistics, including indices of item difficulty, item discrimination, internal consistency, and interrater reliability were prepared for the TLC. Discriminant function analyses were used to assess criterion validity of the TLC, with and without corrections in TLC scores for Verbal IQ. Because of the multiethnic nature of the sample, English as a second language (ESL) and English as a first language (EFL) group means were tested for significant differences on six variables. LLD and control group performance on the language sample analysis were tested for significant differences, using Wilcoxon rank sum tests.

Results of the item analyses indicated support for the internal consistency of the TLC subtests and the test composite, with the exception of Subtest Two (Making Inferences), which obtained an internal consistency coefficient below the designated .8 criterion. Subtest Two and Subtest Three (Recreating Sentences) were found to contain items of questionable validity, and all four subtests contained items that were misordered in terms of difficulty. Subtests Two and Three exhibited satisfactory criterion validity;

however, Subtest One (Understanding Ambiguous Sentences) and Four (Understanding Metaphoric Expressions) failed to discriminate between LLD and Control groups in a stepwise analysis. The language sample analysis discriminated between the two groups. Possible explanations for the findings, along with implications for clinical practice and recommendations for further research, are discussed.

TABLE OF CONTENTS

Abstract	ii
Table of Contents	iv
List of Tables	vii
Acknowledgements	viii
Dedication	ix

Contents

<u>Chapter</u>

<u>Page</u>

Ι	Introduction	1
	Background of the Problem Purpose of the Study Signficance of the Study Summary	1 3 5 5
II	Literature Review	7
	Introduction The Assessment Process Test Validity Theoretical Background Validation Research Assessing ESL Populations The Test of Language Competence Theoretical Background Standardization Content Validity Criterion-Related Validity Construct Validity Internal Consistency Interrater Reliability Informal Assessment Summary Research Questions	7 7 10 13 15 18 19 19 20 20 20 20 20 21 22 23 24 25 27 27
III	Methodology	29
	Sampling Procedures Measuring Instruments Test of Language Competence Subtest One Subtest Two Subtest Three Subtest Four WISC-R Informal Language Sample	29 30 30 33 33 33 34 35 35

TABLE OF CONTENTS CONT'D

Chapter

IV

59

70

Method of Data Analysis	36
Item and Test Analysis	36
Correlational Analyses	37
Interrater Reliability	37
Regression Analysis	37
Stepwise Discriminant Function Analysis	38
Adjusted Discriminant Function Analysis	39
Hotelling's T-Square	39
LARSP	39
Wilcoxon Rank Sum Tests	41
Method	42
Results	43
Descriptive Statistics	43
Item and Test Analysis	43
Item Difficulty	47
Item Discrimination	48
Internal Consistency	49
Interrater Reliability	49
Correlational Analysis	49
Discriminant Function Analysis	53
Simple Regression Analysis	
Adjusted Discriminant Function Analysis	55
Hotelling's T-Square	56
Wilcoxon Rank Sum Tests	56

V Discussion

Purpose of the Study 59 Summary and Discussion of Results 60 Internal Characteristics of the TLC 60 Item Difficulty 60 Item Discrimination 61 Internal Consistency (TLC Subtests) 62 Internal Consistency (TLC Composite) 63 Intercorrelations 63 **Interrater Reliability** 64 Relationship Between TLC and VIQ 64 Language Disabled and Control Group Discrimination (TLC) 65 Test Differences in ESL/EFL Group Performance 68 Language Disabled and Control Group Discrimination (Language Sample Analysis) 68 68

Conclusions Limitations of the Study

TABLE OF CONTENTS CONT'D

<u>Chapter</u>		<u>Page</u>
·	Recommendations for Clinical Practice Recommendations for Further Research	71 73
Reference	s	74
Appendix	A Parents/Student Information	80
Appendix	B Item Analysis Data	90
Appendix	C Stimulus Items Used to Elicit Language Samples	97
Appendix	D Language Sample Transcripts	99
Appendix	E LARSP Summary Sheets	120
Appendix	F Computational Procedure for Wilcoxon Rank Sum Tests	131

LIST OF TABLES

		Page
1.	Sample Characteristics: Age, Sex, English Second Language (ESL), English First Language (EFL)	31
2.	Cultural Background of Sample	32
3.	WISC-R Means, Standard Deviations (SD) and Standard Error of Measurement	44
4.	TLC Means, Standard Deviations (SD) and Standard Errors of Measurement (SEM)	44
5.	Scaled Scores and Standard Scores for the Mean Age Group (13 years) on the TLC Subtests and Composite	45
6.	TLC Subtest Internal Consistency Reliabilities and Descriptive Statistics: Combined Sample	46
7.	Correlations Between TLC and WISC-R for the Combined Sample	51
8.	Intercorrelations Among TLC Subtests for the Combined Sample	52
9.	Summary of Discriminant Function Analysis: Stepwise Selection of TLC Subtests (Total Scores)	54
10.	Summary of Discriminant Function Analysis: Classification of Language Disabled (LLD) and Control Groups	54
11.	Summary of Simple Regression Analyses: TLC subtests regressed on VIQ	55
12.	Hotelling's T ² : Significance of Multivariate and Univariate Differences Between ESL and EFL Group Means	57
13.	Wilcoxon Rank Sum Tests: Sum of Ranks (R ₁) by Stage of LARSP Analysis	58

Acknowledgements

I would like to thank my faculty advisor and committee chairperson, Dr. Julianne Conry, for her assistance in planning and completing this study. Thanks also to Dr. Robert Conry, second member of the committee, for contributing his time and expertise to the data analysis and interpretation. Further acknowledgements are extended to the staff and students of School District #39 (Vancouver) who participated in the data collection; to Joanna Neilson for re-scoring the TLC protocols; to Cathy Margetts who donated her secretarial skills to preparing the final draft; and to the many friends, relatives and colleagues whose interest encouraged me to persevere. Finally, but not least in importance, I would like to thank my husband, John, without whose patience and support this thesis could not have been finished; and my children, Christopher and Kimberly, who have waited many hours for mommy to be finished working. For my father

Ŗ

Introduction

Wiig and Semel (1984) describe language assessment as a hierarchical process. The first level of language assessment involves screening for language disability. This is followed by more in-depth diagnostic assessment to identify specific areas of weakness, and goals for intervention. Finally, assessment is concerned with monitoring and evaluating progress. The choice of methods used within the assessment process depends upon the level or purpose of the assessment, and the theoretical orientation of the examiner. Although theoretical perspectives may vary, current practice generally includes a combination of formal and informal assessment techniques.

Formal assessment, by definition, involves the use of standardized evaluation procedures. Among the suggested advantages of standardized language tests are their objectivity, their replicability, and controlled administration procedures which help to eliminate unwanted sources of variance. These features, together with the fact that standardized tests yield quantitative data, make them useful for diagnostic and research purposes. Despite the appeal of standardized tests, a number of issues surround their use. These concern the technical and practical merits of many commonly-used instruments.

Background of the Problem

Sommers, Erdige and Peterson (1978) predicted that the impact of PL 94-142 (U.S. Office of Education, 1975) would be to stimulate the development and widespread use of formal language tests; a forecast which has come true. The years since PL 94-142 was enacted have given rise to a plethora of standardized language tests. Standardized test results are now required by law in the United States for placement and funding purposes. As Stephens and Montgomery (1985) have pointed out, speech-language clinicians do not have a choice as to whether they will use standardized tests, but only which tests they will use.

Despite their widespread use, standardized language tests have become the subject of increasing criticism. Particular dissatisfaction has gathered around "the proliferation of published tests and materials being marketed with insufficient information concerning their effectiveness and psychometric characteristics" (American Speech-Language-Hearing Association [ASHA] 1988, p. 75). Theoretical and technical information reported in test manuals is frequently inadequate (Lieberman, Heffron, West, Hutchinson & Swem, 1987; McCauley & Swisher, 1984a; Stephens & Montgomery, 1985). Moreover, a review of the professional literature reveals a paucity of research concerning the psychometric characteristics of even the most commonly used instruments. Finally, critics maintain that the validity of standardized language tests used to identify language disorders among those who speak English as a second language (ESL) is highly questionable (Damico; in press, Evard & Sabers, 1979; Vaughn-Cooke, 1983).

The increased dissatisfaction with standardized language tests has had two effects. First, there has been a movement away from formal assessment procedures toward the use of informal, descriptive techniques, including language sampling and analysis. Proponents of this view argue that in addition to their technical inadequacies, standardized tests offer no practical utility. Apart from determining the existence of a language disorder, standardized tests do little to describe the nature of the problem or how to fix it. Descriptive procedures, on the other hand, take no more time to administer than a battery of standardized tests, and ultimately yield information which is useful at all levels of the assessment process (Damico, in press; 1988; Muma, Lubinski, Pierce, 1982; Simon, 1984).

A second movement resulting from the dissatisfaction with standardized tests has been toward improving standards for test use and test construction. A recent example of this effort appeared in the form of the ASHA Guidelines on Instrument Evaluation (ASHA, 1988). The ASHA Guidelines were modeled after the American Psychological Association (APA) Standards For Educational and Psychological Tests (American Psychological Association, 1985) and were intended "as general criteria for judging the adequacy of measurement and intervention instruments or procedures" (p. 76). The Guidelines raise issues concerning the significance of theory in test development, standardization procedures, and test reliability; however, no direct reference is made to test validity, which according APA standards "is the most important consideration in test evaluation" (APA, 1985).

Regardless of divergent professional opinion concerning the relative merits of formal assessment, standardized tests continue to be used extensively by speech-language pathologists for assessment purposes. Lieberman and Michael (1986) and McCauley and Swisher (1984a; 1984b) have drawn attention to the fact that serious errors in diagnosis and remediation can occur if the tests used by clinicians fail to meet adequate technical standards. They maintain that in order to ensure accuracy within the assessment process, and to influence the quality of future instruments, clinicians have a responsibility to scrutinize the technical characteristics of the instruments they use. This view is consistent with APA (1985) recommendations for test use, and implies that clinicians must have access to technical information concerning the various instruments available, and that the information provided should include evidence of test validity.

Purpose of the Study

The first purpose of the present study was to investigate the validity and related psychometric characteristics of the Test of Language Competence (TLC) (Wiig and Secord, 1985). A second purpose of the study was to determine if significant differences in the performance of language disabled (LD) and control subjects would be observed on an informal language sample analysis. Each of these objectives is discussed separately below.

The Test of Language Competence (TLC) is an individually administered measure of language competence developed for use with older children and adolescents between

the ages of nine and eighteen years. Language competence is defined as "the appropriate understanding and/or expression of language content and a responsiveness to the communicative demands of a specific situation" (Wiig & Secord, 1985, p. 1). The TLC is intended to complement other formal and informal methods of language assessment, and is recommended for use with measures of receptive vocabulary and language sample analyses.

The TLC Technical Manual presents information concerning the theoretical background and development of the test. Evidence supporting TLC validity and reliability is reported on the basis of data obtained during standardization. Separate investigations using LD and control subjects are also reported; however, sampling procedures employed in these latter investigations are not well-described, and intelligence test data are available for the language disabled subjects only. Santos (1987) included the TLC in a study of variance in reading comprehension; however, no studies addressed primarily to the subject of TLC validity or reliability have as yet been reported in the literature.

The primary objective of this research was to examine the technical characteristics of the TLC, and in particular, to evaluate the validity of the instrument. This was accomplished first by examining the internal characteristics of the TLC, including item difficulty, item discrimination, internal consistency, and interrater reliability. Second, two discriminant function analyses were calculated to determine the capacity of the TLC to discriminate between language-learning disabled (LLD) and control subjects both before and after the effects of verbal intelligence had been removed. Verbal intelligence was operationally defined as the Verbal Intelligence Quotient (VIQ) on the Wechsler Intelligence Scales for Children-Revised (WISC-R) (Wechsler, 1974). Third, in order to determine if subjects who spoke English as a second language (ESL) performed differently than subjects who spoke English as their

first native language (EFL), group means on the TLC and WISC-R (VIQ, PIQ, FSIQ) were tested for significant differences.

A second objective of this research was to compare the performance of LD and control subjects on an informal language sample analysis. Language samples were obtained from five LD and five control subjects. These were analyzed using the Language Assessment, Remediation, Screening Procedure (LARSP) (Crystal, Fletcher & Garman, 1976). Results of the LARSP analysis for both groups were then tested for significant differences.

Significance of the Study

Data for this research was collected in School District 39 (Vancouver). Vancouver is a broadly multicultural district with a high proportion (approximately 46%) of students who speak English as a second language (ESL). Although the TLC is used by speech-language pathologists within the district, little is known about the effectiveness of its use with Vancouver students. The present study investigated the psychometric characteristics of the TLC using data obtained from within this culturally diverse population. In so doing, it contributes to existing evidence of TLC validity and reliability, and is of interest to speech-language pathologists in culturally diverse educational jurisdictions who are using or who may consider including the TLC in their assessment batteries. In the same way, this study contributes to evidence concerning the criterion-related validity of an informal language sampling procedure for the local population.

Summary

Language assessment is a hierarchical process that includes formal and informal methods of assessment. Although standardized tests are widely used, there is growing dissatisfaction concerning their technical and practical merits. The Test of Language Competence (TLC) is a recently developed measure of language suitable for use with older children and adolescents between the ages of nine and eighteen years. Information concerning the theoretical background and development of the TLC is presented in the test manual. Research evidence of TLC validity and reliability is limited to that reported in the test manual, and one other unpublished investigation. The present study investigated the validity and related technical characteristics of the TLC using data obtained from a local and culturally diverse population. At the same time, the criterion-related validity of an informal language sample was investigated.

Chapter Two

Review of the Literature

Introduction

This chapter will review the literature concerning the validity of standardized language tests, and in particular, those developed for use with older children and adolescents. Three issues relevant to this discussion were identified in the previous chapter, and will be discussed here in the following order: theoretical considerations during test construction, the lack of test validation research, and the validity of standardized language tests applied in multicultural settings. The TLC will be reviewed within the context of this review, followed by a discussion of informal language sampling as an adjunct to formal assessment. The chapter opens with an overview of the assessment process, and a definition of test validity.

The Assessment Process

Language assessment is a hierarchical process that involves screening, diagnosis, program planning, and evaluation (Wiig & Semel, 1984). The first level of assessment concerns screening for possible language disorder. Screening data may be obtained from a variety of sources, including observation, student records, informal, clinician-prepared tasks, and standardized tests (Larson & McKinley, 1987; Tibbits, 1982; Wiig & Semel, 1984). Few screening tests suitable for use with adolescents have been developed. Those available include the Screening Test of Adolescent Language (STAL; Prather, Beecher, Stafford & Wallace, 1980), and the Clinical Evaluation of Language Functions -Screening Tests (CELF; Semel & Wiig, 1980), currently under revision. A short form of The Test of Language Competence (TLC), which is the subject of this study, is intended for use as a screening instrument.

The second level of language assessment concerns diagnosis. The purpose of assessment at this level is to confirm the existence of a language disorder. Typically, this objective is met by the administration of standardized tests (Larson and McKinley, 1987; Tibbits, 1982; Wiig and Semel, 1984). A number of standardized language tests have been developed for use at this level with older children and adolescents. The Test of Adolescent Language, now the TOAL-2 (Hammill, Brown, Larsen & Wiederholt, 1987) may be used with subjects between the ages of 12 and 18. The Test of Language Development-Intermediate, revised as the TOLD-2-I (Hammill & Newcomer, 1988) was developed for older children between the ages of 8 and 12 years. Other instruments include The Fullerton Language Test for Adolescents (FLTA) (Thorum, 1980), which is intended for use with subjects between the ages of 11 and 18 years. The Clinical Evaluation of Language Fundamentals-Revised (CELF-R; Semel, Wiig, & Secord, 1987) includes norms for subjects between the ages of 5 and 16 years of age. Other measures of specific language skills and abilities that have not been developed exclusively for use with adolescents but may be suitable for use with this population are described by Larson and McKinley (1987), Tibbits (1982), and Wiig and Semel (1984).

Data obtained from standardized tests may be used at the following two levels of assessment for describing the nature of the language problem and planning intervention goals. To accomplish these objectives, clinicians may examine differences among subtest scores to determine areas of strength or weakness. Other methods include conducting error analyses on the basis of item responses, and altering task formats to observe where student performance breaks down.

McCauley and Swisher (1984b) have cautioned against use of the above methods for planning therapy objectives, claiming that they may lead to "a mistaken understanding of a client's problem, to inappropriate and fruitless therapy programs, or to inaccurate conclusions regarding the efficacy of therapy" (p. 338). Errors associated with response analysis stem from the fact that no single test covers an exhaustive range of skills; therefore some skills that need to be addressed in therapy may be overlooked in the assessment. A second problem is that an incorrect response to a specific test item may not represent a true deficit in the skill represented. Third, subject responses obtained under standardized conditions may be unrepresentative of the individual's language in other contexts. Finally, altering test items or teaching to specific items invalidates a test for future purposes. Profile analysis is considered an acceptable method for determining strengths and weaknesses if appropriate statistical procedures governing the interpretation of significant differences are observed; however, McCauley and Swisher point out that the information required to do this is frequently omitted from test manuals.

Although standardized tests may contribute to the diagnostic profile, information is often obtained at this level through the use of informal assessment procedures. The term refers to interviews, observations, questionnaires, and other non-standardized procedures, including language sample analyses. Informal procedures are, for some, the preferred method of language assessment. Proponents argue that standardized tests collapse what is essentially a complex process (language) into a few meaningless test scores, whereas informal procedures yield descriptive information that may be translated into instructional objectives (Damico, in press; Muma et al., 1982; Leonard, Perozzi, Prutting & Berkley, 1978). Others maintain that informal procedures enable the clinician to sample behavior in a variety of contexts, leading to a more accurate portrayal of language functioning (Bloom & Lahey, 1978; Larson & McKinley, 1987; Lund & Duchan, 1983). Moreover, informal procedures permit the direct application of theory "thereby bridging the gap between some less timely standardized tests and what is currently understood about the nature of receptive and expressive competence" (Simon, 1984, p. 84). Finally, standardized tests are viewed as suffering from many technical inadequacies; this, it has been suggested, "forces the clinician and teacher to use clinical judgement in the diagnostic process" (Cupples & Lewis, 1984, p. 131).

Stark, Tallal and Mellits (1982) have enumerated the pitfalls of relying on clinical judgement in language assessment. Clinical judgement, they argue, is based on inexplicit criteria; therefore it cannot be replicated, it is subject to bias, and is not suitable for

research purposes. Clinical judgement cannot be used independently to assess the language of children from other cultures, and it cannot be used to distinguish between a language disorder and a more global intellectual impairment. McCauley and Swisher (1984b) advise that the reliability and validity of informal procedures in general require further clinical and research attention.

The final level of language assessment concerns progress evaluation. Evaluation is an ongoing part of the assessment process, and may include the use of formal or informal procedures. Hammill et al. (1987) comment that "the use of criterionreferenced enroute objectives does not obviate the need to be sure that the enroute objectives do in fact lead to the desired, general integrated language goals" (p. 3), and recommend retesting students with the same, or similar instruments, that were used to identify them for special programmes in the first place. McCauley & Swisher (1984b) cite three reasons against the use of standardized tests for monitoring progress. First, standardized tests are designed to compare individuals, and may be insensitive to intraindividual changes over time. Second, changes in test scores may be related to the unreliability of the instrument. Finally, repeated administration of standardized tests may result in practice effects which invalidate test results.

While there are extremes of opinion concerning formal and informal assessment procedures, the more broadly-held view is that the two approaches are complementary (Blau, Lahey, Oleksiuk-Velez, 1984; Kelly & Rice, 1986; Launer & Lahey, 1981; McCauley & Swisher, 1984b; Stephens & Montgomery, 1985). Larson and McKinley (1987) and Tibbits (1982) consider a combined approach essential for assessing the language of older children and adolescents.

Test Validity

Test validity concerns the appropriateness of inferences made from test scores. Traditional definitions of test validity have distinguished between content, concurrent and construct validity. More recently, validity has been defined as a unitary concept that includes all three types of evidence (APA, 1985). For the sake of clarity, each will be defined separately at this point.

A preliminary step in the process of test development is the selection of a theoretical model which defines the trait or behavior to be measured, and provides a rationale for item selection (APA, 1985; Cronbach, 1971). Content validity concerns the adequacy of content sampling from within this theoretical framework, or the extent to which test items represent the behavior of interest in its proper proportion. Nunnally (1978) maintains that content validity "rests mainly on appeals to reason"; however, in some situations empirical methods may be employed to enhance content validity. These include, for example, using item analysis procedures during test development, or obtaining correlations between the test of interest and measures of the same trait or behavior. Detailed discussions of item development and content validity are located in Anastasi (1982), Cronbach (1971), Henrysson (1971) and Nunnally (1978).

Criterion-related validity is of primary interest when the test under investigation is intended for classification or decision-making purposes (Anastasi, 1982; Cronbach, 1971). The extent to which test scores predict performance on one or more outcome criteria is a measure of criterion-related validity. Anastasi (1982) suggests "a test may be validated against as many criteria as there are specific uses for it" (p. 138); typical criteria are academic achievement, group membership, diagnostic classification, and so forth. Criterion-related evidence may be concurrent or predictive. Concurrent validity is examined when data are obtained for the test of interest and the outcome criteria at the same point in time. Predictive validity is examined when data pertaining to the outcome criteria are obtained at some point in the future; however, the term may be used to refer to prediction at any time (Anastasi, 1982).

Although test validation should employ all three types of evidence (APA, 1985), construct validity is of critical significance when the test of interest is a proposed measure of some unobservable trait, or construct. Construct validity concerns how well a test measures the construct it is intended to measure, and "any data throwing light on the nature of the trait under consideration and the conditions affecting its development and manifestations are grist for this validity mill" (Anastasi, 1982, p. 144). Thus, content and criterion-related evidence may contribute to evaluations of construct validity. It is on this point that Guion (1977) has argued "all validity is at its base some form of construct validity" (p. 410). Similarly, Messick (1980) defines construct validity as "the unifying concept of validity that integrates criterion and content considerations into a common framework for testing rational hypotheses about theoretically relevant relationships " (p. 1015).

Messick's point can be traced to earlier discussions concerning the significance of the "nomological net" (Cronbach & Meehl, 1955) or theoretical framework within which constructs are defined in relation to other constructs and observable behaviors. Construct validation is a process of testing hypotheses or predictions made on the basis of test performance. The extent to which an hypothesized relationship is supported is evidence of construct validity. Examples of construct validation studies include, for example, correlating the test of interest with measures of the same trait, or measures of different traits, examining item and subtest intercorrelations, and factor analysis. In situations where the proposed relationship is not supported, one may assume that the test, the research methodology, or the theoretical framework is unsound. (Cronbach & Meehl, 1955).

To summarize, test validity is a unitary concept comprised of content, criterionrelated and construct validity. All three types of evidence are bound by a common theoretical framework which provides a rationale for test score interpretation. This definition assumes several points. First, in order to demonstrate validity, a test must be based on a theoretical framework or model that provides a rationale for item selection and test interpretation. Second, evidence of construct validity supports not only the validity of the test under investigation, but the theory on which the test is based. Finally, test validation is a ongoing process of accumulating evidence which may ultimately be used in an evaluation of construct validity.

Theoretical Background

It may be concluded from the foregoing discussion that validity is built into a test from the outset through the articulation of a sound theoretical model. As stated, the function of a test model is to provide a rationale for item selection and the interpretation of test scores; a factor which bears upon content and construct validity.

Despite the significance of theory in test construction, many language tests are not theory-driven (Muma, 1985; McCauley & Swisher, 1984a). For example, Stephens and Montgomery (1985) reviewed six tests of adolescent language (STAL, WORD Test, TOLD-I, FLTA, CELF, TOAL) and concluded only the TOAL and the TOLD-I were constructed with reference to any theoretical model. Lieberman & Michael (1986) evaluated the content relevance and content coverage of three standardized language tests (CELF, CELI, TOLD); two of which are suitable for use with older children. Content relevance was evaluated according to five criteria, including the existence of a theoretical model. Only one test (TOLD) was judged to be adequate in this area.

Content coverage in the same study was evaluated by analyzing the grammatical requirements of each test item using the Language Assessment Remediation Procedure, or LARSP (Crystal et al., 1976), which is based on a developmental model of grammar. Results of the LARSP analysis led the researchers to conclude that for all three tests, content coverage was incomplete and unrepresentative of the grammatical domain. Each instrument, for example, was found to overrepresent the earlier stages of grammatical development, indicating that it might be too easy to identify language problems in older children.

In a later study, Lieberman et al. (1987) compared the performance of 30 randomly-selected sixth-graders (11.6 to 12.5 years) on four adolescent language tests (FLTA, TOAL, CELF, STAL). The findings indicated that 21 subjects obtained scores

below the designated cutoff point on the FLTA, compared to 22 on the TOAL, 18 on the CELF, and 6 on the STAL (a screening test). The researchers attributed the observed differences in group performance in part to the atheoretical nature of the tests, adding: "it is possible that neither the content nor the procedures of these tests may represent the essential forms, features, and systems of adolescent language in their proper proportion and balance" (p. 260-61).

Some language tests have been constructed according to models that are not supported in research. For example, Muma (1984) was critical of the CELF (Semel & Wiig, 1980) because rather than proposing a broadly-based theoretical model, the test authors cited "various domains of presumed deficits that have been reported in the special education literature" (p. 101). Noting the methodological flaws inherent in much of the learning disabilities research, Muma concluded that the CELF authors had "managed to stack together several strawmen in the components of the CELF" (p. 102). Elsewhere, Lieberman et al. (1987) argued that the results of existing research into adolescent language development are "incomplete and fragmentary", adding that "until researchers broaden this language base and authors use it in test construction, the development of adolescent language tests seems premature and especially susceptible to problems of test inadequacy" (p. 263).

To summarize, content validity depends upon the existence of a theoretical framework. Without a well-defined theoretical domain, "assessment becomes a circular endeavor of merely claiming a domain, attaching a label, and constructing presumed tasks with their attendant responses, scores, norms, and results " (Muma, 1984, p. 102). Furthermore, by definition, construct validity assumes the existence of a theoretical, or conceptual framework. The evidence presented would suggest that many standardized language tests may be inadequate with regard to content and construct validity because they are weak on the level of theory.

Validation Research

Test developers have a responsibility to provide evidence of test validity in test manuals (APA, 1985). Reviewers often criticize the adequacy of technical information reported in language test manuals. For example, reviewers of the WORD test, a measure suitable for use with older children up to the age of 11 years, concur that the test authors offer minimal evidence of validity and reliability (Stephens & Montgomery, 1985; Donahue, 1985; Raju, 1985). In their review of the CELF, Stephens & Montgomery (1985) referred to the reported evidence of validity and reliability as "singularly unimpressive" (p. 36). Sommers (1985) criticized the evidence of criterionrelated evidence in the STAL as "inappropriate" because it used the DTLA as a criterion measure, and "there is no reason to believe that the four subtests from the DTLA measure language processing either" (p. 1332). Following a review of 30 language and articulation test manuals, McCauley & Swisher (1984a) concluded "those criteria that require the application of considerable psychometric expertise, time, and money--criteria related to empirical evidence of validity and reliability--were met least often" (p. 40).

Test validation refers to the process of gathering evidence to support specific inferences made from test scores (APA, 1985). This definition implies that test validation continues beyond the initial research reported in test manuals. Nevertheless, few test validation studies are reported in the literature. A number of studies have been reported which raise questions concerning the criterion-related validity of several adolescent language tests. These are touched upon briefly below.

Stephens and Montgomery (1985) reported that clinicians surveyed found the STAL, a screening test for adolescent language disorder, "too easy", adding that the STAL manual reported a false negative rate of 32% in students passing the STAL but falling below the designated cutoff score on the DTLA. Lieberman et al. (1987) observed that the STAL identified 6 students out of 30 as being at risk. Three other measures in the same study (TOAL, CELF, and FLTA) identified no fewer that 18

students as being language disordered. Considering the latter three as criterion measures, the criterion-related validity of the STAL seems questionable.

This low failure rate on the STAL might be attributable to the fact that it has a lower cutoff score (10th percentile) than the other measures. Other researchers have concluded, however, that the TOAL and FLTA may, in fact, overidentify individuals as being language disordered. Caskey and Franklin (1986) and Aram, Ekelman and Nation (1984) concur that the TOAL appears to be too difficult for evaluating the lower range of language functioning in adolescents. As an example, Caskey & Franklin found that in a sample of 20 "gifted" students (WISC-R IQ of 128 or higher), 10 obtained adolescent language quotients on the TOAL in excess of 15 standard score points (+1 standard deviations) below their IQ scores, thus qualifying them for services as learning disabled (Caskey & Franklin, 1986). These results might be explained in part by item-ordering on the TOAL. Caskey & Franklin observed that when all TOAL items were administered, some individuals achieved several basals after reaching a ceiling. The researchers concluded that items on the TOAL are not well-ordered with respect to difficulty.

Lieberman et al. (1987) addressed the question of construct validity in their study of four adolescent language tests. Differences in group performance were observed among the four measures; however these were not significant in three out of four instances, and intertest correlations were moderately high. These results were thought to support the theory of a general language construct underlying many language tests, including those claiming to measure distinct skills and abilities. Studies reported by Damico and Damico (Damico, personal communication, August, 1989), Schery (1985) and Sommers et al. (1978) have resulted in similar conclusions. Although the data suggested some redundancy among the different measures, Lieberman et al. concluded that further research into the factor structure of adolescent language tests is necessary, and that the substitution of one adolescent language test for another is inadvisable at the present time.

The fact that few validation studies are reported in the speech-language literature is disconcerting for two reasons. First, there is insufficient evidence to support the use of any test as a criterion against which to measure the construct validity of new instruments. Second, there is limited information on which to base the revision of existing tests. This point is particularly relevant in view of the fact that a number of standardized language tests have recently been revised. A case in point is the TOAL, which served as a criterion instrument in TLC validation research, and has now been revised as the TOAL-2 (Hammill et al., 1987).

Tests should be revised "when new research data, significant changes in the domain represented, or new conditions of test use and interpretation make the test inappropriate for its intended use" (APA, 1985). An inspection of the TOAL-2 Manual revealed no explicit purpose or rationale for test revision. Validity and reliability data are reported on the basis of research using either the TOAL or the TOAL-2, because, the test authors explain, "The two versions of the test are essentially the same" (p. 47). The main difference between the two tests appears to be in the range of item difficulty. On the basis of user "comments" that the TOAL was too difficult, a number of "easy" items have been added to seven of the eight subtests. No reference is made in the TOAL-2 manual to independent research investigations of TOAL item difficulty, and it appears that no attempt has been made to correct for suggested problems of item ordering. Thus, although the TOAL-2 may represent an improvement over the previous edition, it appears that the authors have been less than thorough in their efforts toward improving the test.

In summary, validation evidence presented in language test manuals is frequently inadequate, and few studies investigating the validity of standardized language tests are reported in the professional literature. Consequently, insufficient evidence exists to support the use of any one test as a criterion instrument in validation research.

Moreover, there is little empirical justification for the revision of existing instruments. Assessing ESL Populations

In its position paper on social dialects, the Committee on the Status of Racial Minorities (ASHA, 1983) maintained that "no dialectal variety of English is a disorder or a pathological form of speech or language" (p. 23). Bernstein (1989) allows that distinguishing between a communication difference and a communication disorder in language assessment "is not an easy task". Two kinds of errors are possible. The first is to misclassify children with language differences as language disordered. A second type of error is to overlook children with language disorders because of an assumption that they have had insufficient opportunity to learn the language.

Few standardized tests have been developed to identify language disorders among the ESL population. As a result, clinicians rely on tests that have been developed for use with populations whose first language is Standard English. Among the threats to validity associated with using tests under these circumstances are the unrepresentativeness of test norms, the possibility of culturally-biased test items, lack of test-taking skills among children from minority backgrounds, examiner effects on the test behavior of culturally different children, and motivational factors, all of which may lead to errors in assessment (Evard and Sabers, 1979; Sattler, 1988; Vaughn-Cooke, 1983).

To address these issues, alternatives to existing instruments and procedures have been proposed. Examples include developing norms for distinct linguistic groups, including a percentage of minority groups in standardization samples, modifying test items, administering tests in the subjects' native language(s), and developing new tests (Evard and Sabers, 1979; Sattler, 1988; Vaughn-Cooke, 1983). Others have advocated the use of informal and criterion-referenced procedures to improve the quality of language assessment (Bernstein, 1989; Damico, in press; Holland & Forbes, 1986). Regarding the assessment of ESL adolescents, Larson & McKinley (1987) support the development of new instruments, but maintain that informal language samples should be included in the assessment process.

The Test of Language Competence

A relatively recent publication, the TLC has arrived on the heels of considerable criticism concerning the technical adequacy of standardized language tests. It would appear that TLC authors Wiig and Secord (1985) have been sensitive to this criticism. The test manual reports extensively on the theoretical background and technical characteristics of the test. These are summarized below in relation to the foregoing discussion.

Theoretical Background. The TLC authors define language competence as a single construct requiring both the understanding of language content, and a responsiveness to the context in which communication occurs. Each of the four TLC subtests is intended to measure a unique aspect of language competence. The subtests are categorized into a model which features semantics (word meaning), syntax (grammar) and pragmatics (rules governing social/verbal communication). Within these three levels, the content of each subtest is further divided into (a) propositions in narrow contexts and (b) propositions in communicative contexts. In the former context, subtest content deals primarily with semantic or syntactic meaning, while in the latter context, subtest content includes pragmatic considerations.

This model of communicative competence represents what the test authors view to be a shift away from the assessment of specific language skills such as phonology, vocabulary or syntax, to the assessment of linguistic processes or strategies. The model was derived from an extensive review of the literature concerning linguistic strategy development. Four general areas were selected on the basis of that review and are represented by each of the TLC subtests. A second literature review led to the selection of specific models within each area from which the subtests were developed. The TLC is then an integrated model based on not one, but several theories of language

processing. Each subtest is proposed as a measure of the broader construct, language competence, and is based on existing theory or research.

The same criticism raised by Muma (1984), regarding the patchwork of theory and research that went into the making of the CELF, may also apply to the design of the TLC. Little contemporary research evidence is reported to support the theoretical design of each subtest. Moreover, relationships among these various bodies of theory and research have not been established. Further examination of the technical aspects of the TLC should contribute to judgements concerning the adequacy of the test model.

<u>Standardization</u>. The TLC was standardized on 1,796 students from three geographic regions in the United States. The cultural characteristics of the sample are described as 86.2% "white", 8.6% "black", and 3.6% "other", while the proportion of distinct linguistic groups, other than Spanish, is not described.

Given these considerations, the appropriateness of TLC norms for use with students in the Vancouver school district, where data for this research were collected, seems questionable. In 1982, for example, 24,524 Vancouver students were found to speak English as a second language (ESL). This figure represented approximately 46% of the total district population. Of these, Chinese, East Indian and Italian were the most common language groups (La Torre, 1983). More recently, 27% of students surveyed in Vancouver reported that they had learned at least two different languages simultaneously as native languages (Watson-Russell, 1986).

In their discussion of TLC development and standardization, authors Wiig and Secord explain that separate norms for separate races or ethnic groups were not considered because test validity is unrelated to the representativeness of a norming sample; however, the authors advise caution in interpreting the test scores of minority subjects and have provided information to assist in the development of local norms.

<u>Content Validity</u>: Evidence of TLC content validity is claimed on the basis of four criteria outlined by Kretschmer and Kretschmer (1978). These criteria focus on the

role of theory in test construction, specifying that tests must be based on a theoretical definition, that there should be contemporary research support for this theoretical framework, and sufficient information concerning the development of test items should be provided in test manuals to permit the generation of new test items. In response to these criteria it might be argued that there is insufficient contemporary research evidence to support the theoretical scaffolding on which the TLC is based.

In addition to the theoretical evidence offered in support of content validity, item analysis procedures were employed during TLC development. These are not discussed in detail in the Technical Manual, but are reported to have included internal consistency coefficients using Cronbach's Alpha to increase the homogeneity of the subtests, and studies of item difficulty (Secord, personal communication, August, 1989).

<u>Criterion-Related Validity</u>. Evidence of criterion-related validity is reported in the Technical Manual on the basis of correlations obtained between the TLC and three criterion measures: the TOAL, WISC-R, and the Educational Abilities Series (EAS; Thurstone, 1978) for a sample of 28 LLD and 28 controls. LLD subjects were so identified on the basis of school referral procedures which are not described. Controls were described as normally achieving; again, how this determination was made is unclear. Correlations between the TLC, the TOAL and the EAS were calculated separately for both groups. TLC and WISC-R correlations were calculated for the LLD group only. WISC-R data were not reported for the control group.

Results of a discriminant function analysis using TLC and TOAL scores for the same 56 subjects indicated that 96% were correctly classified as language-disabled, while 93% of controls were correctly classified. A subsequent stepwise discriminant function procedure indicated that Subtest Four (Understanding Metaphoric Expressions) contributed most to group discrimination, followed by Subtest Three (Recreating Sentences), and Subtest Two (Making Inferences). Subtest One (Understanding Ambiguous Sentences) did not account for a significant proportion of variance and was not entered into the discriminant function.

In a later study, Santos (1987) investigated the variance in reading comprehension among a combined sample of 20 reading disabled and 20 control subjects (ages 15-17 years). A significant relationship was hypothesized between each of the TLC subtests and the Durrell Analysis of Reading Difficulty. Results indicated that 16 of the 20 reading disabled subjects obtained TLC composite scores >1 SD below the mean. In contrast, 19 of 20 control subjects scored at or above the 50th percentile. Accepting group membership as a criterion, these results might be claimed to support the concurrent validity of the instrument.

<u>Construct Validity</u>. Evidence of TLC construct validity is offered on the basis of correlations between the TLC and WISC-R scores for the same 28 LLD subjects described above. Higher (convergent) correlations were observed between scores on the TLC subtests and VIQ (.48 to .78), while lower (divergent) correlations were observed between scores on the TLC subtests and PIQ (.18 to .53). While this is a reasonable interpretation, it may once gain be pointed out that these correlations are available for the language disabled group only.

Further evidence of construct validity is reported on the basis of intersubtest correlations obtained from the standardization sample at various age intervals. These range from .17 to .50. Moderate correlations are explained by the fact that each subtest represents a different content domain. This is reasonable; however, more support for this interpretation could have been demonstrated if subtest-to-total-test correlations had been reported. Assuming that the total test score is an overall measure of language competence, and assuming that each subtest measures some aspect of language competence, subtest-to-total-test correlations should be higher (Anastasi, 1982).

In a separate investigation, subtest intercorrelations were calculated using data obtained from the same group of 28 LLD subjects and 28 controls described above.

Correlations ranged from .24 to .57 for the LLD's and from -.1 to .39 for the controls. In contrast, Santos (1987) reported a considerably higher range of subtest intercorrelations (.47 to .75) for a mixed sample of reading disabled/control subjects. The higher range of correlations observed in Santos' investigation are explained by the heterogeneity of the subject sample involved.

Factor analysis results are reported on the basis of subtest intercorrelations obtained from the standardization sample. The percentage of variance explained by the first unrotated factor at all but two age levels was greater than 90%. A subsequent oblique rotation factor analysis using TLC item intercorrelations and yielded four factors at the 9-11 and 12-17 year age groups. These results are claimed to support both the existence of an underlying language factor, as well as the specificity of the TLC subtests. Thus, "the TLC emerges as an attractive blend of both worlds--a strong general factor supported by four specific subgroups of items." (Wiig and Secord, 1985, p. 47).

Internal Consistency: The validity of a test is a function of its reliability. One method of estimating test reliability that is useful in the evaluation of test validity is to examine the degree of association among test items, or the internal consistency of the test. Internal consistency coefficients (Cronbach's Alpha) are reported for the four TLC subtests and the total test at each age level in the standardization sample. Because the TLC was intended to measure language competence across different content areas, greater homogeneity was expected within subtests than across items (Wiig and Secord, 1985, pp. 2, 48). In fact, the range of coefficients reported for the TLC subtests is from .52 to .79, while the range of coefficient observed for the TLC subtests is explained by the test authors as the effect of test length on estimates of reliability. That is, the four TLC subtests "were designed to be as short as possible" in order to reduce administration time (Wiig and Secord, 1985, p. 48). In limiting the length of each subtest, internal

consistency estimates have likewise been reduced. The combined length of the total test is considered to have resulted in a higher range of internal consistency coefficients than observed for the TLC subtests.

Interrater Reliability: Another consideration in the evaluation of test reliability is the extent of interrater agreement on subjectively scored items. Interrater reliability figures are reported in the Technical Manual for Subtests Three and Four, which require some judgement in scoring. Interrater agreement was defined as the percentage of agreements observed between sixteen raters and the test authors on one protocol. The final estimates were 97% for Subtest Three and 98% for Subtest Four.

The above figures were obtained following a three-step procedure during which all sixteen raters were trained in the application of scoring criteria by author Secord. Intermediate measures of agreement were obtained, followed by additional instruction. Not surprisingly, an increase in the percentage of agreements was observed between the intermediate and final calculations. This method of estimating interrater agreement illustrates the effect of direct training on scorer reliability. Sample protocols and scoring criteria identical to those used in the training procedure are included in the Administration Manual. The magnitude of the interrater reliability coefficients obtained in this study will give some indication as to the adequacy of the scoring guidelines.

To summarize, the TLC represents a significant improvement over many other standardized language tests currently available for use with older children and adolescents in several respects. First, the test is based on a clearly stated theoretical model. Second, the test manual includes extensive information concerning the technical characteristics of the test. The information as reported, however, raises a number of questions. In particular, there has as yet been no attempt to examine the relationship between TLC performance and VIQ for control subjects. Second, there is some evidence to suggest that at least one of the subtests does not contribute to LLD/control group

discrimination; however; the test authors have not addressed this issue. Finally, the validity of using the TLC with ESL subjects has not been explored.

Informal Assessment

As the general dissatisfaction with standardized language tests continues to increase, more extensive use is being made of informal assessment procedures. Informal language assessment involves the use of interviews, questionnaires, observational techniques, and other non-standardized procedures, including informal language sampling and analysis. Language sampling is the process of eliciting, recording and transcribing a sample of spontaneous language, and then analyzing it to determine areas of strength and weakness. The focus of discussion here will be on the informal language sample.

Elicitation procedures used in the collection of language samples vary. These include picture stimuli, where the subject is shown a picture and is asked to describe it. Other methods include prompting statements, such as "tell me about..", or direct questions. Unstructured conversation between the clinician and subject is the preferred method of obtaining a language sample (Larson & McKinley, 1987; Atkins & Cartwright, 1982); however this is not always practical due to situational and time constraints (Simon, 1984).

Opinion varies as to the length of the sample required to ensure adequate coverage of the subject's language. The minimum sample required to calculate Mean Length of Utterance (MLU) (Nice, 1925), for example, is 50 utterances. Crystal et al. (1976) recommend continuous sampling of 15 - 30 minutes. Muma et al. (1982) suggest sampling over time in a variety of situations. Little research evidence exists to help resolve the issue. The consensus appears to be that 50 - 100 utterances is an acceptable minimum (Darley & Spriestersbach, 1978; Wiig & Semel, 1984)

The advantages and disadvantages of language sampling have been widely discussed (Bloom & Lahey, 1978; Larson & McKinley, 1987; Muma et al., 1982; Wiig

and Semel, 1984). Often cited among the advantages are the flexible administration procedures, and the opportunity for observing language behaviors in various contexts. Moreover, because language sampling is a descriptive procedure, proponents argue that it provides valuable information for planning intervention strategies.

Among the disadvantages of language samples is that they do not yield standardized scores and therefore are not useful for making classification decisions. They lack objectivity, and because administration procedures are unstandardized, they are not replicable, and results may vary from sample to sample. More practical disadvantages include the time required to obtain and transcribe language samples, and the expertise required to interpret them.

Larson & McKinley (1987) have suggested that the greatest difficulty with language samples is knowing how to analyze them. This is particularly the case in analyzing the spontaneous language of older children and adolescents. Numerous language sampling and analysis procedures have been developed (Crystal et al., 1976; Lund & Duchan, 1983; Miller & Chapman, 1983; Muma, 1981; Tyack & Gottsleben, 1974). Many of these are referenced to a developmental framework of early language development, and are considered inappropriate for use with older children and adolescents. There is at present little empirical evidence concerning the critical stages of language development within this age group, or the sequence in which these occur (Hammill et al., 1987; Larson & McKinley, 1987; Lieberman et al., 1987); however, this does not preclude the use of language samples with these groups. Despite its limitations, language sampling yields valuable descriptive information concerning social-verbal skills and expressive language characteristics at any age level (Larson & McKinley, 1987; Tibbits, 1982).

The expressive language characteristics of language-disabled youth have been described in some detail by Larson and McKinley (1987), Tibbits (1982) and Wiig and Semel (1984). These include word-finding difficulties such as overuse of fillers (um,
uh), pronouns, and circumlocutions. In addition, language disabled youth demonstrate patterns of deficit syntactic development. As a result, they have difficulty mastering the rules which govern the construction of complex sentences, and as a group show lower MLU than their non-disabled peers. It might be hypothesized that language disabled adolescents as a group would demonstrate fewer complex sentence structures in natural conversation than nonhandicapped adolescents, and that this difference would be demonstrated by an informal language sample analysis.

Summary

Language assessment is a hierarchical process that involves the use of formal and informal assessment procedures. Issues surround the use of both; however, they are generally viewed as complementary procedures. The Test of Language Competence (TLC) is a relatively new test designed for use with children and adolescents between the ages of 9 and 18 years. It represents a significant improvement over many tests currently available for use with this age group; however, evidence of TLC validity is limited to that reported in the Technical Manual, and one other unpublished investigation. The purpose of the present study was be to further investigate the technical characteristics of the TLC. A second purpose was be to compare the performance of language disabled and control subjects using an informal language sample.

Research Questions

This study compared data obtained by a mixed sample of LLD and control group subjects on the TLC and an informal language sample. The following research questions were addressed:

1. What are the internal characteristics of the TLC with respect to:

- a. item difficulty indices
- b. item discrimination indices
- c. internal consistency of subtests and composite

27

- 2. What is the interrater reliability of the subjectively-scored sections of the TLC?
- 3. What is the relationship between the TLC and VIQ?
- 4. How effectively does the TLC discriminate between language-disabled and control groups? To what extent does this discrimination reflect individual differences in the IQ?
- 5. Will scores obtained by ESL subjects differ significantly from scores obtained by EFL subjects on six variables (VIQ, PIQ, TLC Subtests One through Four)?
- 6. Do language disabled and control subjects differ in characteristics of their informal language?

Chapter III

Methodology

This chapter presents a description of research methodology. Sampling procedures and sample characteristics are described, followed by a discussion of data collection procedures and data analysis.

Sampling Procedures

Data were collected in School District 39 (Vancouver). Members of the language disabled group (LLD) were selected from classrooms for the communicatively disordered located in two schools, one elementary and one secondary, situated on the city's east side. Children are declared eligible for placement in these classes on the basis of psychoeducational and speech-language assessment data. Eligibility criteria allow a significant discrepancy between WISC-R VIQ and PIQ on the WISC-R in favour of Performance; although this may not always occur. Children must demonstrate language delays of two or more years in the primary grades and three years in the intermediate/secondary grades, and their language deficits must not be the result of physical, intellectual, or emotional impairments. Although children in these classes may speak English as a second language, this is not considered to be the primary cause of their language difficulties (Education Services Group, 1985). For the purpose of this study, children were required to have no greater than average verbal ability as measured on the WISC-R Verbal scale (VIQ<115), and have at least average nonverbal ability as measured on the WISC-R Performance Scale (PIQ >85).

Each student in both communications classes received an information package to be taken home and read by parents. The package contained a letter explaining the purpose of the study, parent/student consent forms, and a brief questionnaire (Appendix A) concerning family linguistic background, area of residence and educational history. A total of 25 students agreed, with parent permission, to participate in the study. Existing WISC-R scores were then obtained from student records. Students whose WISC-R data was more than two years old were retested. Those who obtained PIQ's below 85 were eliminated. No subjects were eliminated on the basis of VIQ. The final number of students in the LLD group was 23.

Control group members (controls) were selected from two regular education classes, one in the same elementary school as LLD subjects, and one in a neighboring high school. Teachers and counsellors were provided with lists of matching criteria; these included the age, sex and linguistic background of each subject in the LLD group. Staff were asked to nominate students who met these criteria, and who were judged to be of average classroom achievement on the basis of school grades. Each nominated student received an information package addressed to parents. Students who, with parent permission, agreed to participate were administered the WISC-R. Students of average verbal ability (VIQ 85-115) and at least average nonverbal ability (PIQ \geq 85) were retained. Students receiving learning assistance, or instruction in English as a second language, were excluded, as were those whom school counsellors considered to be of above average achievement according to school records. A total of 23 subjects were retained.

The final sample consisted of 46 subjects ranging in age from 11 to 15 years. Subjects were matched for age, sex and linguistic background. A summary of sample characteristics is provided in Tables 1 - 2.

Measuring Instruments

The Test of Language Competence (Wiig and Secord, 1985)

The Test of Language Competence (TLC) was developed for use with older children and adolescents between the ages of 9.0 and 18.11 years. It is intended to assess delays in the development of language competence. "Language competence" is defined by the test authors as a unitary construct consisting of (a) the understanding and expression of language and (b) sensitivity to the social context within which the

Sample Characteristics: Age, Sex, English Second Language (ESL), English First

Language (EFL)

	Number in Each Group							
Characteristic	Language Disabled	Controls	Combined					
Sex								
Male	13	12	25					
Female	<u>10</u>	11	<u>21</u>					
Total	23	23	46					
Language								
ESL	9	11	20					
EFL	13	12	25					
Unknown	1	<u>0</u>	1					
Total	23	23	46					
Age Statistics (in years/mon	ths)							
Range	11-3 to 15-5	11-3 to 15-1	_					
Mean	13.4	13.5						

Cultural Background of Sample

	No. Language Disordered	No. Controls	No. Combined
Croatian	1		1
Japanese	1	-	1
Portuguese	-	1	1
Greek	1	1	2
Native Indian	1	1	2
Philipino	1	1	2
Punjabi	1	1	2
Vietnamese	-	2	2
Hindi/Fijian	2	2	4
Chinese	7	6	13
English	8	8	16
Total	23	23	46

communication occurs. The TLC consists of four subtests, each intended to measure a unique aspect of language competence. A brief description of each subtest follows:

Subtest One (Understanding Ambiguous Sentences)

The subject is presented with one sentence that may be interpreted in two ways (e.g. "I saw the girl take his picture"). The task is to verbalize both interpretations. The subtest includes thirteen scoreable items, and two training items. The entire test is given; no basal or ceiling levels are established. The subtest may be discontinued if the subject fails to respond three times in a row; however this is not a requirement. The subject is allowed a maximum response time of "10-20" seconds. Scoring is objective, and weighted as follows: two correct responses = 3 points; one correct response = 1 point; no correct responses = 0 points. The range of raw scores possible is from 0 to 39.

Subtest Two (Making Inferences)

The subject is presented with two sentences that describe the beginning and ending of a situation (e.g. "Jack went to a Mexican restaurant"; "He left without giving a tip"). The task to verbalize two inferences that describe what might have transpired between these two instances. The subtest includes twelve scoreable items, and one training item. No basal or ceiling rules apply; the subtest may be discontinued if the subject fails to respond three consecutive times. The response time allowed per item is 60 seconds. Scoring is objective and weighted as in Subtest One. Raw scores possibly range from 0 to 36.

Subtest Three (Recreating Sentences)

The subject is presented with three stimulus words and a picture depicting a social situation (e.g. A picture of people hiking is accompanied with the words 'fall', 'leg', 'and'). The words are read aloud to the subject, whose task is to combine the three stimulus words in a sentence that might have been spoken by someone in the picture. The subtest contains 13 scoreable items and 2 trial items. The maximum allowable response time is 60 seconds. Items are scored twice; once for correctness on the basis of

scoring criteria provided in the test manual. This system entails some subjectivity on the part of the examiner, and rates responses on a scale of 0, 1, or 3. Items are scored a second time for the number of stimulus words included in each sentence (three stimulus words = 3 points, two stimulus words = 1 point, and 1 or 0 stimulus words = 0 points). These two sets of scores are then combined to obtain the subtest raw score. Raw scores possibly range from 0 to 4, or from 0 to 6. The range of scores possible for the subtest total is from 0 to 78.

Subtest Four (Understanding Metaphoric Expressions)

The subject is presented with a common metaphor (e.g. "She sure casts a spell over me"). Each item is presented verbally, then in print. The task is first to explain what the metaphor means. Second, the subject is choose the correct interpretation from among four distractors. The subtest includes 12 items plus two trials. Fifteen seconds is allowed for the first part of the task, and 45 seconds for the second. No basal or ceiling rules apply; the subtest may be discontinued after three consecutive failures to respond. Points per item are awarded as follows: 3 points for two correct responses, 1 point for one correct, or 0. Possible total raw scores range from 0 to 36.

The TLC package includes a Technical Manual, an Administration Manual, and a Stimulus Manual. Subtest One and Two items are presented verbally by the examiner, and then in print using the Stimulus Manual. In this way, the subject first hears and then reads the items. Subtest Three items are read aloud by the examiner while the subject looks on. For Subtest Four, the first part of each item is presented verbally, and then in print using the Stimulus Manual. The second part of each item is presented visually while the examiner reads the four options. All responses are verbal, and recorded on the test protocol by the examiner.

The TLC yields scaled scores for each of the four subtests ($\overline{X} = 10$; SD = 3), and a standard score for the TLC Composite ($\overline{X} = 100$; SD = 15). A Partial test score, consisting of Subtests Three and Four, may be calculated. The Partial composite is recommended for screening purposes. Scaled scores and standard scores may be converted to age equivalent or percentile scores. Confidence intervals are provided. A detailed discussion of TLC validity and reliability is located in Chapter Two. Wechsler Intelligence Scale for Children-Revised (WISC-R) (Wechsler, 1974)

The WISC-R is an individually-administered test of general intelligence suitable for use with children between the ages of six and sixteen years. The test is comprised of twelve subtests organized into two scales. The Verbal Scale contains six subtests and provides a measure of verbal reasoning ability. The Performance Scale is made up of the remaining six subtests and provides a measure of nonverbal reasoning ability. Subtests yield scaled scores with a mean of ten and standard deviation of three. Ten of the subtest scaled scores are combined to form the Verbal, Performance and Full Scale IQ's, each with a mean of 100 and a standard deviation of 15. The validity and reliability of the WISC-R is well-supported in research. Sattler (1988) and Kaufman (1979) provide detailed discussions of the WISC-R and its psychometric properties. Informal Language Sample

Language samples were obtained by tape-recording a fifteen-minute conversation between the examiner and each subject. An imperative format (eg: "Tell me about...") was used to elicit the language samples. The imperative format was chosen for three reasons. First, this format is replicable (Atkins and Cartwright, 1982). Second, there is evidence to suggest that imperatives elicit more fluent and more complex language than other procedures (Wiig and Semel, 1984). Third, this method was appealing for use with language disabled adolescents whose spontaneous conversation might be limited.

Stimulus items used to elicit language samples are presented in Appendix C. Five language samples were randomly selected from each group of subjects. Fifty utterances from the middle of each sample were transcribed for analysis using the LARSP procedure (Crystal et al., 1976), described below. Utterances were excluded from the analysis if they were partially or completely unintelligible, if they were repetitions of earlier responses, or if they were unfinished (e.g." I went to the..."). Single word utterances, or starters and fillers (like, um, you know) were not included. Repetitions due to dysfluency were treated as one utterance (eg: "I went to the...to the store).

Methods of Data Analysis

Item and Test Analysis. The reliability and validity of a test is largely a function of its item characteristics. Item analysis is a term applied to the examination of item characteristics. Typically it includes measures of item difficulty, item discrimination, and item homogeneity.

Item difficulty is defined as the proportion of persons passing or failing a test item. It is related to the total distribution of test scores, and to test reliability. For dichotomously scored items, item difficulty is expressed in terms of p values, or the percentage of persons passing each item. An index of item difficulty for items scored on a continuous scale, such as is the case with the TLC, is the mean score for each item (Nunnally, 1978).

Measures of item discrimination indicate the extent to which a test item differentiates among individuals on the behavior being measured (Anastasi, 1982). Item discrimination may be evaluated on the basis correlations between each item and an external criterion, or between each item and the total test score. Point biserial correlations are appropriate for use when test items are scored dichotomously. When items are scored on a multipoint scale, as in the case of the TLC, product moment correlations are appropriate. An item demonstrates adequate discrimination when it reaches a level of .3 or better when correlated with the total test score (Nunnally, 1978).

Estimates of internal consistency describe the degree association among test items. The Kuder-Richardson formulas (K-R 20 and K-R 21) are among the most commonly used methods of calculating internal consistency. Cronbach's Alpha is a generalization of the K-R 20 formula suitable for use with items scored on a multipoint scale. Hoyt's Analysis of Variance is a less frequently used procedure that produces the same results as K-R 20. Internal consistency estimates of .8 or better are considered to indicate acceptable test reliability (Nunnally, 1978).

In this study, TLC item characteristics were analyzed using LERTAP (Nelson, 1974), an extensive test analysis program. In order to study the effectiveness of the subjective scoring criteria during this analysis, Subtest Three (Recreating Sentences) was coded as three subtests: Subtest Three (H), Subtest Three (W), and Subtest Three (T). These correspond to the holistic, word count, and total subtest scores respectively. TLC items are scored on a multipoint scale; therefore, the mean and standard deviation of each item were inspected for relative degrees of difficulty. Likewise, product moment correlations were obtained between items and subtest/total test scores to evaluate item discrimination. In addition, correlations were calculated between each TLC item and an external criterion (VIQ). Hoyt coefficients were obtained for each of the TLC subtests and the composite. Cronbach's Stratified Alpha was calculated for the test composite only.

<u>Correlational Analyses</u>. Pearson product-moment correlation coefficients (Pearson r) were obtained between TLC subtests and the test composite. TLC and WISC-R scores were correlated to study the relationship between the TLC and VIQ. Correlations were obtained using the computer program SPSSX (Lai, 1986).

Interrater Reliability. In order to measure the extent of interrater agreement on the subjectively scored sections of the TLC, reliability coefficients (Pearson r) were calculated between scores obtained from two independent raters. Rater One was the graduate student researcher, and rater two was a speech-language pathologist who uses the TLC in her employment with the Vancouver school district.

<u>Regression Analysis</u>. Regression analysis is a method by which scores on a dependent or criterion variable are predicted from scores on an independent variable (Pedhazur, 1982). The extent to which an observed criterion score deviates from its

predicted score is the residual, or error of estimate for that individual. Residual scores represent that proportion of variance which is unique to criterion and unaccounted for by variance in the independent variable. In this study, a series of simple regression analyses were calculated using Verbal IQ as the independent variable, and each of the four TLC subtests as dependent variables. The purpose was to obtain standardized residual scores representing that proportion of variance unaccounted for by Verbal IQ, and unique to the TLC, for each individual. Residual scores were standardized to have a mean of 0 and standard deviation of 1. These analyses were conducted using the computer program SPSSX.

Stepwise Discriminant Function Analysis. Discriminant function analysis is an extension of regression analysis suitable for use with multiple variables when the criterion is group membership (Pedhazur, 1982). In his discussion, Klecka (1980) divides discriminant analysis into two levels of activity: interpretation and classification. At the level of interpretation, one is concerned with obtaining the canonical discriminant functions. A discriminant function is a composite of variables that has maximum potential for discriminating between groups. The number of functions possible is one minus the number of groups, or the number of discriminating variables, whichever is smaller. Discriminant functions are applied at the second level of activity to predict group membership.

In stepwise discriminant function analysis, the variable contributing most to group discrimination is entered first into the equation. That variable is then paired with each of the remaining variables, and the most discriminating of the remaining variables is entered. This stepwise selection of variables continues until all variables in the function have been entered, or until the remaining variables provide no significant contribution to group discrimination. Wilk's lambda (U) is a measure of residual discrimination employed in stepwise procedures. Values of U range from 0 to 1, with 0 indicting maximum group discrimination, and 1 indicating negative discrimination. The

38

U statistic may be converted to an F statistic and tested for significance. By inspecting the U and F statistics, it is possible to distinguish the discriminating variables from those which do not contribute substantially to group discrimination (Klecka, 1980).

In this study, the computer program BMDP-P7M (Dixon, 1988) was used to calculate a forward stepwise discriminant function analysis. Each of the four TLC subtests was employed as a discriminating variable, and the criterion was group membership. The purpose of the analysis was to observe first the relative contribution of each subtest to group discrimination, and second, the capacity of the test to predict group membership. This was followed by a second discriminant function analysis using Subtest Three (Holistic Scoring) in place of the Subtest Three total, together with the remaining three subtests as discriminating variables. The purpose was to determine if removal of the Word Count scoring would significantly alter the results.

Adjusted Discriminant Function Analysis. The standardized residual scores obtained by the regression analyses represented that proportion of variance unique to the TLC and unaccounted for by VIQ. In order to observe the relative contribution of each TLC subtest to group discrimination after the effects of Verbal IQ had been removed, these standardized residual scores were entered into a second discriminant function analysis using the computer program BMDP-P7M.

Hotelling's T-Square. Hotelling's T-square (T^2) is a multivariate technique for measuring the distance between group means on two or more variables simultaneously. It can be converted to an F statistic and evaluated for significance. In this study, a total of 22 subjects were known to speak English as a second language (ESL). In order to rule out the effect of ESL on test performance, individuals were assigned to one of two groups (ESL or non-ESL). Group means were then tested for equality on six variables: WISC-R Verbal IQ, WISC-R Performance IQ, and each of the four TLC subtests. The computer program used for this analysis was BMDP-3D.

LARSP. The Grammatical Analysis of Language Disability Crystal et al. (1976)

(LARSP) is a syntactic analysis procedure. A computerized version of LARSP (Bishop, 1985) was used in this study. The program analyses sentences at three levels. First, sentences are broken into words, and each word is classified as a part of speech. At the next level, sentences are divided into noun, verb or adverbial phrases. Relationships between clauses are analyzed at the final stage of the program. Structures at each level are assigned to one of seven developmental stages as frequency counts. Mean length of utterance (MLU) is also calculated. Results are printed out on the LARSP summary sheet. The seven developmental stages represented in the LARSP analysis are as follows:

Stage I: One Word Sentences

Stage II: Two Word Sentences

Stage III: Three Word Sentences

Stage IV: Sentences of Four Words or More

Stage V: Recursion

Stage VI: System Completion

Stage VII: Discourse structure, syntactic comprehension and style

Stage one (single word) utterances were not included in this analysis. Stages six and seven are not handled by the Bishop version of LARSP and were not dealt with in this study. The analysis concentrated on stages two, three, four, and five. Stage five was of principal interest in this study because it represents that point at which children begin to use complex patterns of sentence structure. Stage Five is defined as follows:

Essentially what the child has to learn here is a set of connecting devices which can be used to interrelate clauses, and the transformational processes whereby one can be used within ('embedded within') another. Once these devices have been learned, of course, the process can continue indefinitely, longer and more complex sentences being built up as a result. It is this feature of language, to take a basic structure and use it repeatedly to produce extensive sequences, which is the primary characteristic of the creativity of language...It is accordingly a stage of great significance in normal development, as at this point the range of expression available to the child is enormously increased. (Crystal et al, 1976, p. 76).

If LLD adolescents do, in fact, use fewer complex sentence structures than non-LD youngsters, some divergence in the performance of LLD and control group subjects might be expected at the Stage Five level.

The LARSP system was selected over other available language analyses procedures for several reasons. First, it is suitable for use with both children and adults, and with speakers of nonstandard English (Crystal et al, 1976; Holland & Forbes, 1986; Tibbits, 1982). Second, it is based on a descriptive framework of English grammar (Quirk, Greenbaum, Leech, Svartvik, 1972), and thus has adequate content validity (Lieberman et al., 1987). Finally, the LARSP procedure has been found to discriminate among groups of language disabled and control subjects at different age levels (Hawkins & Spencer, 1985; Kearns & Simmons, 1983; Penn, 1983; Penn & Behrmann, 1986).

<u>Wilcoxon Rank Sum Test</u>. The Wilcoxon rank sum test is a nonparametric test of significance of group differences, appropriate for use with two independent samples when data are in the form of frequency counts. Data from two groups are combined and assigned ranks. The ranks are then summed, and a value (R_1) is obtained for each group. The probability of this value is then tested in relation to the theoretical distribution of R_1 .

In this study, rank sum tests were calculated using data obtained from the LARSP procedure, described above. The purpose was to determine whether LARSP would discriminate between LLD and Control group members, particularly at the Stage Five level. Frequency data were converted to percentages in order to adjust for fluency. Rank sum tests were calculated using procedures outlined by Ferguson (1976).

<u>Method</u>

Subjects were administered the Test of Language Competence (TLC), the Wechsler Intelligence Scale for Children Revised (WISC-R), and an informal language sample. Testing was conducted over a six month period by two graduate students trained in the administration, scoring and interpretation of standardized tests. The WISC-R was not re-administered to LLD group students who had been tested within the previous two years. In these instances, WISC-R testing had been conducted by school psychologists employed in School District 39.

Chapter IV

Results

This chapter presents the results of the data analysis. Descriptive and item analysis statistics are presented first, followed by results of the correlational analyses. Results of the discriminant function are described next. Finally, results of Hotelling's T-Test and Wilcoxon Rank Sum Tests for equality of group means will be summarized. Descriptive Statistics

Group descriptive statistics summarized in Tables 3 - 4. These include the mean, standard deviation and standard error of measurement for each group (LLD and Control) as well as the combined sample on each of the TLC and the WISC-R Verbal and Performance Scales. TLC results are reported here in raw scores. The mean subtest and composite score obtained by each group and the combined sample were converted to scaled scores and standard scores respectively using TLC norms. The mean age for each group (13 years) was used to make this conversion. Results are located in Table 5. These indicate that LLD and Control subjects obtained generally lower than average scores on all but two subtests

Item and Test Analysis

Table 6 presents results of the LERTAP analysis. These include the means, standard deviations, and standard errors of measurement for each of the TLC subtests, and the total test using combined sample. In this instance, Subtest Three (Recreating Sentences) was analyzed as three subtests (Holistic, Word Count, Total); however, only the subtest composite (Subtest Three-Total) was included in the total test statistics. Individual item statistics (mean, standard deviation, item correlations) for each subtest are summarized in Appendix B. Appendix B includes the percentage distribution of subjects on each item.

Table 3

.

WISC-R Means,	Standard	Deviations	(SD)	and St	tandard	Error	of_N	Measureme	ent

	WISC-R Intelligence								
	<u>Verbal</u>			Performance			<u>F</u>	Full Scale	
Group	Mean	SD	SEM	Mean	SD	SEM	Mean	SD	SEM
Language Disabled	74.8	8.9	1.9	101.0	10.7	2.2	86	8.3	1.7
Control	103.0	9.1	1.9	111.0	11.0	2.3	107.3	10.4	2.1

Table 4

TLC Means, Standard Deviations (SD) and Standard Errors of Measurement (SEM)

			LLD			Controls			
Subtest	No. of Items	Mean	SD	SEM	Mean	SD	SEM		
Ambiguous	13	7.4	6.8	1.4	19	8.0	1.7		
Inferences	12	20.9	5.0	1.0	29.7	3.3	.7		
Recreate	13	46.8	10.1	2.1	63.3	6.9	1.4		
Metaphors	12	7.3	5.5	1.1	21.8	8.5	1.8		
TLC Total	50	82.4	20.8	4.3	133.7	19.8	4.1		

Scaled Scores and Standard Scores for the Mean Age Group (13 years) on the TLC

Subtests and Composite

	TLC Scaled Scores					
TLC Subtests	LLD	Controls	Combined			
Ambiguous Sentences	03	06	04	•		
Making Inferences	04	08	06			
Recreating Sentences	03	08	06			
Metaphoric Expressions	03	07	05			
		<u>TLC Standard Sc</u>	ores			
Composite	65	83	69			

<u>Note</u>. Scaled score mean = 10; SD = 3

Standard Score mean = 100; SD = 15

TLC Subtest Internal Consistency Reliabilities and Descriptive Statistics: Combined

<u>Sample</u>

Subtest	Mean	SD	SEM	Hoyt	r
Ambiguous	13.17	9.37	2.97	.89	·
Inferences	25.30	6.11	3.08	.72	
Recreate(H)	23.76	7.29	3.27	.78	.92
Recreate(W)	31.35	7.85	3.10	.83	
Recreate(T)	55.04	11.98	4.94	.82	
Metaphors	14.54	10.16	3.15	.89	.99
Composite	108.07	32.76	9.27	.95	

<u>Note</u>. n = 46 for Tables 6 through 11

r = Interrater reliability

H = Holistic Scoring; W = Word Count; T = Subtest Total

Coefficient Alpha for the test composite = .87

<u>Item Difficulty</u>. Item means for the combined sample were compared for evidence of item difficulty. The range of scores possible for each item was 0, 1 or 3, thus item means falling near 1.5 (or 3 for Subtest Three-Total) represented the midrange of difficulty.

Results indicated that, for this sample, Subtest One (Understanding Ambiguous Sentences) was the most difficult of the TLC subtests. Item means ranged from .56 to 1.6, and although items were arranged toward increasing levels of difficulty, some variation within this pattern was observed. For example, 19 subjects (41%) obtained scores of 0 on item 13 ($\overline{X} = .8$), as compared to 28 subjects (60%) who scored 0 on item 10 ($\overline{X} = .56$), indicating item 13 to be an easier item.

All item means for Subtest Two (Making Inferences) fell above the mid-range of difficulty (1.7 to 2.5), indicating this to be the least difficult subtest, with the exception of Subtest Three (Holistic), described below. A general trend toward increasing item difficulty was noted; however, as in Subtest One, there was some variation within this pattern.

Item means for Subtest Three (Holistic) ranged from 1.3 to 2.5. The range of item means for Subtest Three (Word Count) was 1.8 to 2.7. This was indicated to be the least difficult of the TLC subtests. The higher range of item means observed score for the Word Count subtest stems from the fact that most subjects received credit for attempting to include all three stimulus words in their responses. Item means for Subtest Three (Total) ranged from 3.5 to 5.2. Items did not appear to be arranged in order of difficulty on this subtest, nor on either of the two separate scoring systems.

Subtest Four (Understanding Metaphoric Expressions) was the second most difficult subtest. The range of item means observed for this subtest was .72 to 1.8. Items 1 to 9 were not ordered in terms of difficulty. Items 10, 11 and 12 obtained

lower mean scores relative to the other items ($\overline{X} < 1.0$); thus more difficult items were located at the end of the subtest.

Because some discrepancies in item ordering were noted among the different subtests, item order difficulty correlations were calculated using Spearman's rank order correlation coefficient (Rho). The range of coefficients was .53 (Subtest Three-Holistic), .65 (Subtest Three-Total), .66 (Subtest Three-Word Count), .76 (Subtest One-Understanding Ambiguous Sentences), .77 (Subtest Two-Making Inferences), and .80 (Subtest Four-Understanding Metaphoric Expressions). These results indicate that some TLC items are not well-ordered in terms of difficulty.

<u>Item Discrimination</u>. Each item in the TLC was correlated with its respective subtest, the TLC composite, and an external criterion (VIQ). Item to subtest correlations of .3 or greater indicated that the item was discriminating adequately among subjects.

All items in Subtest One (Understanding Ambiguous Sentences) and Subtest Four (Understanding Metaphoric Expressions) met the .3 criterion for adequate discrimination when correlated with their respective subtests.

Four items in Subtest Two (Making Inferences) failed to meet the .3 criterion when correlated with the subtest total. These included the following:

Item two: Tim stopped on his way to school to play a video game. At the locker, he realized he had to hurry back in order to be in class on time. Tim had to go back because...

Item three: The sun was shining, when the Robertson's started out for the picnic. Unfortunately they had the picnic in the livingroom. They had the picnic in the livingroom because...

Item five: Bob and Ray rode on a crowded bus to the shopping mall. They told the story of Bob's bad luck to a policeman. They talked to a policeman because... Item nine: Lori took the bus downtown because it was her mother's birthday.

She left the fashionable stores with tears in her eyes. Lori cried because... The results suggest that these four items discriminate poorly among individuals on the specific behavior of interest (inferential thinking). Moreover, item two failed to discriminate adequately among individuals on a broader verbal construct represented by the TLC composite.

Three items (two, eight, twelve) in Subtest Three (Recreating Sentences Holistic), and six items (one, two, three, five, eight) in Subtest Three (Recreating Sentences - Word Count) did not meet the .3 criterion when correlated with their respective subtests. The more discriminating set of items was produced by Subtest Three (Total). Only two items were shown to be unacceptable. These included item eight (without, difficult, again) and twelve (fresh, nor, here). Item eight failed to reach the level of .3 when correlated with either the TLC composite or VIQ, suggesting that this item discriminates poorly among individuals on both the general and specific factors it was intended to measure.

Internal Consistency. Hoyt estimates of internal consistency for the TLC subtests are located in Table 6. These range from .72 to .89, indicating a strong degree of association among items within each subtest. Hoyt and Alpha coefficients obtained for the TLC composite were .95 and .87 respectively. Alpha coefficients reported in the Technical Manual across eight age intervals for the TLC composite ranged from .77 to .82.

Interrater Reliability. Interrater reliability coefficients for the subjectively scored sections of the TLC are presented in Table 6. Coefficients ranged from .92 to .99, indicating close agreement between raters.

Correlational Analyses

Correlations between the TLC and the WISC-R based on the combined sample are presented in Table 7. Coefficients above .34 are significant at $\propto = .01$. A correlation of .82 was observed between the TLC composite and FSIQ. This compares to a correlation of .75 reported in the TLC manual for a group of 28 language disabled youngsters. Correlations of .90 and .50 were observed between the TLC composite and the WISC-R Verbal and Performance IQ's respectively. This is compared to correlations of .78 and .53 reported in the Technical Manual.

Correlations between the TLC composite and the WISC-R verbal subtests ranged from .76 to .86. The range of correlations between the TLC composite and the WISC-R Performance subtests was .13 to .40. Corresponding values were not reported in the TLC manual.

Correlations between each of the four TLC subtests and VIQ ranged from .74 to .83. The TLC manual reports a corresponding range of correlations from .40 to .79 for a group of 28 language disabled subjects. The range of correlations between the TLC subtests and PIQ was .34 to .59 This compares to a range of .18 to .45 reported in the test manual.

The range of correlations observed between the four TLC subtests and five WISC-R verbal subtests ranged from .53 to 78. Correlations between the TLC subtests and five WISC-R performance subtests were predictably lower, from .00 to .47. Corresponding values were not reported in the Technical Manual.

TLC subtest intercorrelations are reported in Table 8. These ranged from .56 to .77. Subtest intercorrelations reported in the TLC manual ranged from .24 to .57 for a group of 28 language disabled subjects, and from -.11 to .39 for 28 nonhandicapped individuals. Subtest intercorrelations obtained by the standardization sample are reported in one year intervals between the ages of nine to fourteen; and two year intervals above age fourteen. These ranged from .17 to .50.

Correlations Between TLC and WISC-R for the Combined Sample

			TLC		
WISC-R	AMBS	INFS	RECR	METS	TOTAL
Information	68	68	61	78	79
Similarities	69	70	71	7.6	82
Arithmetic	53	73	65	76	76
Vocabulary	76	76	76	71	86
Comprehension	67	64	74	66	79
P. Completion	17	20	15	37	26
P. Arrangement	25	10	-00	12	13
Block Design	34	38	34	46	44
Obj. Assembly	27	24	12	47	31
Coding	32	27	40	33	40
Verbal IQ	74	7 9	78	84	90
Performance IQ	43	39	34	59	50
Full Scale IQ	69	69	66	81	82

Note. Coefficients have been rounded to two significant figures and decimals omitted;

 $r \ge .34$ is significant at $\alpha = .01$.

						_
AMBS	INFS	REC(H)	REC(W)	REC(T)	METS	
100	(-05)			(24)	(08)	_
57	100			(08)	(32)	
64	54	100				
45	47	25	100			
68	64	77	81	100	(16)	
65	77	66	46	70	100	
84	82	77	65	89	89	
(53)	(57)				(65)	
	AMBS 100 57 64 45 68 65 84 (53)	AMBS INFS 100 (-05) 57 100 64 54 45 47 68 64 65 77 84 82 (53) (57)	AMBS INFS REC(H) 100 (-05) 57 100 64 54 100 45 47 25 68 64 77 65 77 66 84 82 77 (53) (57)	AMBS INFS REC(H) REC(W) 100 (-05) 57 100 64 54 100 64 54 100 45 47 25 100 68 64 77 81 65 77 66 46 84 82 77 65 (53) (57)	AMBS INFS REC(H) REC(W) REC(T) 100 (-05) (24) 57 100 (08) 64 54 100 45 47 25 100 68 64 77 81 100 65 77 66 46 70 84 82 77 65 89 (53) (57)	AMBS INFS REC(H) REC(W) REC(T) METS 100 (-05) (24) (08) 57 100 (08) (32) 64 54 100 45 47 25 100 68 64 77 81 100 (16) 65 77 66 46 70 100 84 82 77 65 89 89 (53) (57) (65)

Intercorrelations Among TLC Subtests for the Combined Sample

Note. Correlations rounded to two figures and decimals omitted; "Total" variables not corrected for overlap.

H = Holistic Scoring, W = Word Count, T = Subtest Total.

Values in parentheses represent intercorrelations among VIQ - corrected residuals

Results of a forward stepwise discriminant function analysis are presented in Table 9. Subtest 2 (Making Inferences) and Subtest 3 (Recreating Sentences) were selected in the first two steps of the analysis. No other subtest made further significant contribution to the discriminant function for differentiating between LLD and Control groups.

Table 10 presents the group classifications produced by the discriminant analysis. Of the 46 subjects, 19/23 (83%) were correctly classified as language disabled, while 21/23 subjects (91%) were correctly classified as controls. A total of four language disordered students (9%) were misclassified as controls (false negatives), and 2 controls (4%) were misclassified as language disordered (false positives). These results did not change when the analysis used Subtest Three (Holistic Scoring) in place of the total subtest score.

Results of the discriminant function analysis reported in the TLC manual indicated that Subtest Four (Understanding Metaphoric Expressions) followed by Subtest

Three (Recreating Sentences) were the most discriminating subtests. Subtest Two accounted for sufficient variance to be entered into the analysis, but its contribution to group discrimination was considered minimal, as indicated by a relatively small increase in the squared canonical correlation which resulted when Subtest Two was entered into the equation (Wiig & Secord, 1985). Subtest One did not contribute significantly to group discrimination. Results of the classification function indicated that the TLC correctly identified 27/28 language disabled students (96%) and 26/28 controls (92%).

Summary of Discriminant Function Analysis: Stepwise Selection of TLC Subtests (Total

Scores))

Step	Variable Entered	U	F	p
1.	Making Inference	0.47	49.36	<.01
2.	Recreating Sentences (Total)	0.38	34.83	<.01

<u>Note</u>. U = Wilk's lambda

Table 10

Summary of Discriminant Function Analysis: Classification of Language Disabled (LLD)

and Control Groups

	Number of Cases Classified								
Group	LLD	Controls	% Correct	% Incorrect					
LD	19	04	83	17 ^a					
Controls	02	21	91	09 ^b					
Total	21	25	87	13					

Note. ^apercentage of false negatives

^bpercentage of false positives

Simple Regression Analysis

The purpose of this analysis was to obtain a set of standardized residual scores representing that proportion of variance unique to the TLC and unaccounted for by VIQ. These adjusted scores were entered into a second discriminant function analysis (discussed below) to determine the capacity of the TLC to discriminate between LLD and Control groups after the effects of VIQ had been removed. Table 11 presents the results of four simple regression analyses using Verbal IQ as the independent, or predictor variable, and each of the TLC subtests as dependent, or criterion variables. It may be seen that Verbal IQ accounted for 55% of the variance in Subtest One (Understanding Ambiguous Sentences), 62% of the variance in Subtest Two (Making Inferences), 60% of the variance in Subtest Three (Recreating Sentences-Total) and 70% of the variance in Subtest Four (Understanding Metaphoric Expressions). These results confirm a substantial relationship between VIQ and each of the TLC subtests.

Table 11

Subtest	r	r-Square	F	р
Ambiguous	.74	.55	54.32	<.01
Inferences	.79	.62	72.47	<.01
Recreate	.78	.60	66.89	<.01
Metaphors	.84	.70	101.55	<.01

<u>Summary of</u>	f Simple	Regression	Analyses:	TLC su	btests regressed	1 on VIC

Adjusted Discriminant Function Analysis

A second discriminant function analysis was undertaken using the standardized residual scores obtained in the regression analyses, discussed above. Not one of the TLC

subtests was entered into the analysis, indicating group differences are almost wholly explained by VIQ. Intercorrelations among the VIQ-corrected residuals for each subtests are located in Table 8. These ranged from -.05 to .32, and none reached significance. <u>Hotelling's T-Square</u>

The purpose of this analysis was to determine if significant differences between ESL and EFL group performance existed on six variables: VIQ, PIQ, TLC Subtest One (Understanding Ambiguous Sentences), Two (Making Inferences), Three (Recreating Sentences), or Subtest Four (Understanding Metaphoric Expressions). Results are located in Table 12. No significant differences were indicated on the multivariate test ($T^2 =$ 12.83), or on any one variable.

Wilcoxon Rank Sum Test

Language samples were analyzed using the LARSP procedure. Language Sample transcripts are located in Appendix D. Summary sheets of the LARSP analysis are located in Appendix E. Rank sum tests were calculated using results of the LARSP analyses at phrase and clause level, stages two, three, four and five. The purpose of this analysis was to determine if significant differences would be observed between the expressive language characteristics of LLD and Control group members at one or more levels. Results are presented in Table 13. Significant differences were observed between groups at both phrase and clause level, stages three and five. LLD group members demonstrated a significantly higher percentage of phrase and clause usage at stage three than control group members. Conversely, control group members used significantly fewer stage three utterances than LLD's at clause level, and a higher percentage of stage five level phrases and clauses than members of the LLD group.

Hotelling's T²: Significance of Multivariate and Univariate Differences Between ESL and EFL Group Means

	ES	<u>L</u>	<u>EF</u>	<u>L</u>		
Variable	Mean	SD	Mean	SD	T ²	р
Multivariate					12.83	>.01
VIQ	88.3	16.4	88.67	17.6	.26	>.01
PIQ	107.4	13.5	105.1	10.5	66	>.01
TLCI	11.8	8.8	14.3	9.9	.90	>.01
TLC2	23.9	6.1	26.5	5.9	1.46	>.01
TLC3	54.1	10.7	56.0	13.7	.52	>.01
TLC4	11.9	7.2	16.8	11.8	1.62	>.01

<u>Note</u>. ESL (n = 20); EFL (n = 25).

Wilcoxon Rank Sum Tests: Sum of Ranks (R1) by Stage of LARSP Analysis

	Phrase		Clause	
LARSP Stage	LLD	Control	LLD (Control
п	27	28	30	25
III	36*	21*	37*	18*
IV	22	33	24.5	30.5
V	17*	38*	15**	40**

<u>Note</u>. * <u>p</u> < .05 ** <u>p</u> < .01

Chapter V

Discussion

This chapter reviews the purpose of the study, research methodology, and results of data analysis. Implications of the research findings are discussed, and limitations of the study considered. Finally, clinical applications of the TLC and suggestions for future research will be proposed.

<u>Purpose of the Study</u>

The purpose of this study was first to investigate the validity and related psychometric characteristics of the Test of Language Competence (TLC) for use with language learning disabled and control subjects. Five research questions were addressed concerning the internal structure of the TLC, the concurrent validity and reliability of the instrument, and the viability of its use within a multilingual/multicultural population. A second purpose of the study was to investigate the criterion-related validity of an informal language sample.

The principal methods of analysis used to investigate the internal structure of the TLC were measures of item difficulty, item discrimination, and internal consistency. Correlational analyses yielded additional information concerning internal consistency and interrater reliability. TLC criterion-related validity was studied on the basis of a discriminant function analysis to determine the capacity of the TLC to predict LLD and control group membership. A second discriminant function analysis examined the criterion-related validity of the TLC with the effects of Verbal IQ removed. An important consideration during this investigation was the possible influence of ESL on TLC or WISC-R performance. Hotelling's T^2 was used to test the equality of group means on the TLC subtests, VIQ and PIQ.

In order to determine if an informal language sample would differentiate among LLD and control subjects, language samples were obtained from five members of each group. These were analyzed using the LARSP procedure, and the results tested for significant differences using Wilcoxon Rank Sum tests.

Summary and Discussion of Results

Internal Characteristics of the TLC

Item Difficulty. Present results suggest some variation in item difficulty among the four TLC subtests for the combined sample. Subtest One (Understanding Ambiguous Sentences) appears to be the most difficult subtest, followed by Subtest Four (Understanding Metaphoric Expressions), Subtest Three (Recreating Sentences-Total), and Subtest Two (Making Inferences). Further variation in item difficulty was observed between the two scoring systems which are included in Subtest Three (Total), with Subtest Three (Word Count) obtaining the highest mean raw score relative to any other subtest.

Conversion of the mean TLC subtest and composite scores to scaled scores using norms provided in the test manual for the mean age group (thirteen years) indicated that LLD subjects obtained scaled scores below average (>1 SD) on each of the four subtests. Controls obtained scaled scores below average on Subtests One (Making Inferences) and Four (Understanding Metaphoric Expressions). Scaled scores for the combined group were below average for all four subtests. Although it might be argued that control subjects in this study obtained lower than average scaled scores because, unlike the TLC standardization sample, the current sample included a high proportion of ESL subjects, this argument is not upheld by results of Hotellings T^2 , which indicated no significant differences between ESL/EFL subjects on the TLC subtests. A second argument to explain the lower scaled scores obtained by the current sample might be that the sample included a disproportionate number of language disabled subjects. Given the high correlation observed between the TLC and VIQ, however, combined with the fact that control group subjects were known to be of average verbal intelligence, this explanation is rejected. It is therefore reasonable to conclude that TLC norms may not accurately represent the performance of local children, and consequently may overidentify individuals as language disordered. Moreover, these results would suggest caution in the use of profile analysis based on subtest scaled score comparisons.

Results of Spearman's rank order correlation coefficient (Rho) indicated that TLC items are generally not well-ordered in terms of difficulty. This observation would contraindicate the practice of discontinuing a subtest after three consecutive failures to respond, as subjects may respond correctly to subsequent items that are less difficult.

Item Discrimination. The data indicate that all items in Subtest One (Understanding Ambiguous Sentences) and Four (Making Inferences) discriminated effectively on the specific behavior they were intended to measure. A total of four items (33%) in Subtest Two (Making Inferences) and two items (15%) in Subtest Three (Recreating Sentences-Total) failed to meet the required .3 criterion when correlated with their respective subtests. It was noted that fewer items in Subtest Three met the .3 criterion when correlated with either the Holistic or Word Count totals; therefore the most discriminating set of items was produced by the subtest composite.

The fact that some TLC items fail to discriminate adequately among individuals may be related to several factors, including item difficulty. Extreme values of item difficulty tend to reduce item discrimination (Nunnally, 1978). All items in Subtest Two (Making Inferences) were observed to fall above the mid-range of difficulty. It might be argued that items which fail to discriminate among individuals in this subtest do so because they are too easy. Items in Subtest Three (Recreating Sentences-Total) likewise fell above the mid-range of difficulty; therefore the same argument may apply.

Inadequate item discrimination may also be the result of poor content sampling. For example, of the six TLC items which did not discriminate among individuals on the specific behaviors they were intended to measure, five showed inadequate discrimination on a broad verbal factor represented by the TLC composite. The results suggest a

61

minimal relationship between these items and either a general or specific verbal construct.

Internal Consistency (TLC Subtests). Internal consistency coefficients were calculated using Hoyt's Analysis of Variance procedure and Cronbach's Alpha. Coefficients of .8 or better were considered acceptable. Hoyt's estimate of reliability was .89 for Subtest One (Understanding Ambiguous Sentences), .83 for Subtest Three (Recreating Sentences-Word Count), .82 for Subtest Three (Total) and .89 for Subtest Four (Understanding Metaphoric Expressions). Subtest Two (Making Inferences) and Subtest Three (Recreating Sentences-Holistic) fell below .8 (.72 and .78 respectively). Several explanations may account for these results.

First, test reliability is a function of item difficulty and item discrimination. Items which are far-removed from the mid-range of difficulty may reduce the size of the reliability coefficient. Low item to subtest correlations have the same effect. The lower estimate of internal consistency observed for Subtest Two (Making Inferences) is consonant with the observation that this subtest contains the least acceptable combination of items in terms of item difficulty and item discrimination.

A second explanation for the low internal consistency estimates obtained by Subtest Two, and by Subtest Three (Holistic) may be that at least some items in each subtest measure dissimilar constructs, a point raised earlier. The effect of this would be to reduce the inter-item correlations, and estimates of internal consistency as a result.

One final explanation for these results may be related to the length of the TLC subtests. Subtest Two (Making Inferences) is one of the shorter subtests, consisting of only twelve items. Subtest Three includes the maximum thirteen items. It might be pointed out, however, that Subtests One and Four obtained reliability coefficients above .8, despite the fact that each contains twelve and thirteen items respectively. Thus, although subtest length is a possible explanation for lower estimates of reliability, it is an unlikely one.
Internal Consistency (TLC Composite). Given that each of the TLC subtests is intended to measure a specific skill or ability, it might be expected that higher estimates of internal consistency would be observed for individual subtests than for the total test. Hoyt's estimate of reliability for the TLC composite contradicts this interpretation. A higher estimate was observed for the TLC composite than for any of the subtests (.95). Cronbach's Alpha for the test composite was likewise substantial (.87).

These results are similar to internal consistency data reported in the TLC manual for the standardization sample. TLC authors Wiig and Secord observed a higher range of internal consistency coefficients (Cronbach's Alpha) for the TLC composite across ages than for individual subtests. The test authors attributed this difference to the effect of increased test length on estimates of reliability when all the items were combined. This interpretation is reasonable, and receives some support here in view of the fact that Cronbach's Alpha, which is lowered when subtests are not highly correlated, is somewhat lower than Hoyt's estimate for the composite. This difference is small, however, and both results suggest a high degree of association among items in the total test.

Intercorrelations. A pattern of moderate, positive intercorrelations was observed among TLC subtests (.56 to .77). Correlations between Subtest Three (Holistic) and Subtest Three (Word Count) did not reach significance (.34), suggesting that these two scoring systems yield different results.

A higher range of subtest to total test correlations was observed (.82 to .89), lending support to the internal consistency of the subtests; however, higher correlations might also be explained by the effects of increased test length, and the fact that correlations were not corrected for overlap.

The range of correlations between each subtest and the TLC composite was relatively higher than that observed between each subtest and the external criterion (VIQ). These ranged from .74 to .83. This difference might be interpreted as supporting the uniqueness of the measured construct; although higher subtest to total test

correlations could again be explained by the effects of overlap. It might be argued that correcting for overlap would result in comparable correlations between each subtest and the test composite or VIQ.

Interrater Reliability. Interrater reliability coefficients for Subtest Three (Recreating Sentences-Holistic) and Subtest Four (Understanding Metaphoric Expressions) were .92 and .99 respectively. These results indicate very close agreement between raters, and support the adequacy of the subjective scoring criteria.

Relationship Between TLC and VIO

The range of correlations observed between the TLC subtests and subtests on the WISC-R Verbal Scale for the combined sample was .53 to .78 (correlations above .34 are significant at $\alpha = .01$). A more pronounced relationship, which may be explained by the effect of increased test length on correlations, was observed between VIQ and the TLC composite (.90). This value is considerably higher than that reported in the Technical Manual for a sample of 28 language-disabled subjects (.78). Likewise, the range of values observed in the present study between the four TLC subtests and VIQ (.74 to .83) is considerably higher than that reported in the Technical

The lower range of values reported in the TLC manual might be explained by effect of restricted range on correlation. That is, homogeneous samples produce lower correlation coefficients than heterogeneous samples (Anastasi, 1982). Correlations reported in the Technical Manual were obtained using a homogeneous sample of language disabled youngsters; whereas the subject sample in the present study included handicapped and nonhandicapped individuals. The greater variability in the present sample has likely resulted in higher correlations.

Results of the simple regression analyses indicated that VIQ accounted for 55% of the variance in Subtest One (Understanding Ambiguous Sentences), 62% of the variance in Subtest Two (Making Inferences), 60% of the variance in Subtest Three (Recreating Sentences-Total) and 70% of the variance in Subtest Four (Understanding

Metaphoric Expressions). Intercorrelations among the subtest residuals failed to reach significance; thus, the proportion of variance remaining that could be considered unique to the TLC was minimal.

In total, these results raise the issue of what Sommers (1985) referred to as "the troublesome distinction between a language disorder and cognitive abilities" (p. 1087). The same issue has been addressed elsewhere by Oller (1978) and Gunnersson (1978) who challenge the view that language and intelligence tests measure different constructs. Likewise, present results do not support a distinction between the proposed construct (language competence) and VIQ.

Language Disabled and Control Group Discrimination (TLC)

Results of the discriminant function analysis indicated that Subtest Two (Making Inferences) was the most discriminating subtest, followed by Subtest Three (Recreating Sentences-Total). The remaining two subtests, Subtest One (Understanding Ambiguous Sentences) and Subtest Four (Understanding Metaphoric Expressions) did not contribute significantly to group discrimination and were not entered into the equation. Present results are inconsistent with those originally reported by the test authors, who observed that Subtest Four (Understanding Metaphoric Expressions) accounted for the major proportion of variance in group membership, followed by Subtests Three and then Two.

In terms of classification, present results are somewhat inconsistent with those reported by the test authors. In this study, the TLC correctly classified 83% of the LLD's, as opposed to 93% originally reported. 91% were correctly classified as controls, compared to 93% reported in the test manual. The data would suggest that the TLC does not discriminate as effectively among local students at the lower end of language functioning, thus resulting in increased numbers of false negatives.

Although current results do not agree with the original findings reported by the test authors, the two studies are not directly comparable. Criteria for group membership in the present study were defined as average or better nonverbal ability as measured on the WISC-R, and delays of two or more years as determined by speech-language pathologists on the basis of objective test data, which were not the result of other handicapping conditions or ESL. Moreover, this research employed a multiethnic subject sample matched for age, sex, and linguistic background. In contrast group selection criteria used in the original study were not reported in the TLC Technical Manual, intelligence test data were not available for the control group, and no information was provided regarding the cultural makeup of the sample involved. It is possible that disparate findings observed between the two studies are the result of different grouping criteria.

Several explanations may account for the observation that Subtest One (Understanding Ambiguous Sentences) and Subtest Four (Understanding Metaphoric Expressions) were not included in the discriminant function reported here, despite the fact that each demonstrated adequate item characteristics and internal consistency. First, the results may have been distorted by sampling error. For example, LLD group members in this study were selected from within a population of language disabled students already identified as such by qualified speech-language pathologists. It is possible that of those individuals retained in the final sample, some were originally misclassified.

A second explanation may be that neither subtest measures a unique aspect of language competence which is not already accounted for by the other two subtests. This conclusion would challenge the specificity of the TLC subtests, each of which is intended to measure a unique aspect of a broad verbal factor. The test authors claimed support for subtest specificity on the basis of an oblique rotation factor analysis (Wiig & Secord, 1985, pp. 42-47). This method has been found to demonstrate the existence of a general language factor common to different language tests which may be divided into subcomponents, each possessing its own share of reliable variance (Oller &

Damico, in press). Current results, however, offer limited support for TLC subtests specificity.

One further point regarding the discriminant function analysis concerns the number of false positives and false negatives observed in the classification function. It was noted in earlier discussion that Subtest Two (Making Inferences), which explained the largest proportion of variance between groups, was found to be one of the least difficult subtests, to contain several items which fail to discriminate adequately among individuals on either a general or specific language factor, and to demonstrate inadequate reliability. The number of false negatives observed in the present analysis may be a function of the poor internal characteristics associated with this subtest.

The results described above for the present study did not change when the discriminant function analysis was re-calculated using Subtest Three (Recreating Sentences-Holistic) in place of the subtest total, which uses the combined scoring system. This means that Subtest Three (Word Count) does not contribute substantially to group discrimination; however, its exclusion from the battery for classification purposes is not advised. Several items in Subtest Three (Holistic) were found to discriminate poorly among individuals. Moreover, the Holistic scoring obtained an internal consistency coefficient below the acceptable .8 minimum. Both item discrimination and internal consistency were greatest for the combined scoring system. Exclusive use of the Holistic scoring system for classification purposes might result in increased numbers of false positives or false negatives due to the poor internal qualities of that subtest.

A third discriminant function analysis was intended to determine how well the TLC would discriminate between groups after the effects of VIQ had been removed. In fact, insufficient variance was remaining to calculate the additional analysis. Allowing for sampling error, the remaining variance which could be considered to account for a unique construct (i.e. language competence) is insignificant; thus, VIQ explains the major proportion of variance between LLD and control groups.

Test for Differences in ESL/EFL Group Performance

Significant differences were not observed between groups on the six variables of interest (VIQ, PIQ, TLC Subtests One, Two, Three, Four). These results indicate that ESL was not a factor in LLD or control group performance in this study. Language Disabled and Control Group Discrimination (Language Sample Analysis)

Results of the Wilcoxon Rank Sum Tests indicated that control subjects demonstrated a lower frequency of stage three clause level utterances, and a higher frequency of stage five phrase and clause level utterances than LLD's on the LARSP analysis. Conversely, language-disabled students used a higher proportion of stage three level utterances and a lower percentage of stage five level utterances (phrase and clause). These results support the capacity of the LARSP analysis to discriminate between LLD and control groups on the basis of language complexity, and further, demonstrate that language disabled adolescents as a group tend to produce less complex sentence structures in spontaneous speech than their nonhandicapped peers.

Conclusions

The results obtained in this study suggest a number of conclusions concerning the validity and internal characteristics of the TLC, as well as the criterion-related validity of the LARSP analysis. These have implications for the diagnostic and practical utility of both measures.

Conclusions regarding the technical characteristics of the TLC are based on the results of item analyses, estimates of internal consistency and interrater reliability coefficients. The results indicate that Subtests One (Understanding Ambiguous Sentences), Three (Recreating Sentences) and Four (Understanding Metaphoric Expressions) demonstrate adequate item discrimination and internal consistency. Interrater reliability for Subtests Three and Four is very high, supporting the adequacy of the subjective scoring criteria. Subtest Two (Making Inferences) demonstrates a relatively high percentage of items (33%) which fail to discriminate adequately among

individuals, and the internal consistency estimate for this subtest is below the desired .8. These results are consistent with the observation that this is the least difficult of the subtests. All four subtests contain items which are not well-ordered in terms of difficulty.

TLC criterion-related validity was judged on the basis of the discriminant function analysis which indicated that Subtest Two (Making Inferences) and Subtest Three (Recreating Sentences) discriminate between LLD and control groups. These results do not imply that Subtest One and Four lack the capacity to discriminate between groups; an inspection of LLD and control group means for each subtest indicated sizeable differences in the performance of both groups on all four subtests. From this it may be concluded that the variance in Subtests One and Four was accounted for by Subtests Two and Three in the discriminant function analysis. The criterion-related validity of Subtests One and Four, independent of the other two subtests, might be the subject of further investigation.

TLC content and construct validity were evaluated on the basis of the above, together with results of item and subtest intercorrelations, correlations between the TLC and VIQ, the discriminant function analyses; and, the extent to which these results were consistent with the stated theoretical design of the test. The TLC is based on a model of language competence which assumes a broad verbal factor subdivided into four specific content areas represented by each of the TLC subtests. A pattern of moderate, positive intercorrelations was observed among the TLC subtests, indicating some support for subtest specificity. This interpretation was not upheld by the high internal consistency coefficients, which indicated that TLC items measure a common construct; nor by the results of the discriminant function analysis, which suggested that most of the variance in Subtests One and Four was accounted for by Subtests Two and Three. In total, limited support has been demonstrated for the view that each subtest measures a unique aspect of language competence.

Correlations between the TLC and VIQ support the TLC as a measure of a verbal construct; however, the magnitude of these correlations suggests that the TLC and VIQ are measuring a common factor. This interpretation is supported by the results of the adjusted discriminant function analysis, which indicated that LLD and Control group discrimination was explained by VIQ. These findings are consistent with research conducted by Damico (Damico, personal communication, August, 1989), Schery, (1985), and Sommers et al. (1978), which support the view that many language tests claimed to measure a unique language factor are, in fact, measuring a common construct. These results would further suggest that this common factor may be VIQ.

Results of the Wilcoxon Rank Sum tests support the validity of the LARSP procedure for distinguishing between LLD and control groups; however, the analysis does not allow for comparisons between individuals. Moreover, the descriptive power of the LARSP analysis is lost when individual performance is reduced to a set of numbers. Clearly the TLC and LARSP might be viewed as complementary procedures. The TLC has been demonstrated to possess adequate criterion-related validity; however, the proposed interpretive rationale for TLC results is not strongly supported. Conversely, the LARSP system may not discriminate among individuals, but it does provide extensive descriptive information. These results would lead to the conclusion that the two instruments may be most effective if used together for the identification of language disorders, and the identification of intervention goals.

Limitations of the Study

The present study was intended to investigate the technical characteristics of the TLC using a locally selected sample of language disabled and control subjects. The generalizability of current findings is limited by several factors. First, current findings were obtained from within a distinctly multiethnic populations, and are not considered representative of populations in other regions. Second, the age range of the present sample was limited to 11 through 15, thus the representativeness of these results is

confined to that age range. Third, the selection of average achieving students for the control group was limited to the opinion of school personnel and school records, either of which may not have been objective. It might be argued, however, that as achievement is highly dependent upon verbal ability, and as all control group subjects were found to be of average verbal ability, it is unlikely that control group subjects were of above average achievement.

Another limitation of the present study is related to the collection, transcription and coding of the language sample analyses. First, the language sample was obtained under structured conditions (imperatives), and not obtained across a number of settings. It might be suggested that these factors limited the representativeness of the samples collected. Secondly, language sample transcription and analysis is a complex procedure, and the limited experience of the researcher in conducting these analyses may have increased the possibility for error. Although Bishop's computerized version of LARSP helped to direct the analyses somewhat, it is an interactive program requiring user judgement as the analysis proceeds. Moreover, the program was found to have limited capacity for analyzing the longer and more complex sentences generated by control subjects. For example, the program will not analyze sentences beyond 25 words. Clearly the program is effective for use with young children, or subjects in the lower range of language functioning.

Recommendations for Clinical Practice

1. The observation that subtest items are not well-ordered in terms of difficulty would contraindicate the procedure of discontinuing a subtest after three consecutive failures to respond. If a subject fails to respond because three consecutive items are too difficult, he/she may respond to subsequent items which are less difficult.

- 2. The TLC short form, which consists of Subtest Three (Recreating Sentences) and Subtest Four (Understanding Metaphoric Expressions) is not recommended for screening purposes at this time because of these two, only Subtest Three has been shown to discriminate adequately between LLD and control subjects. Moreover, the content of Subtest Four may be too specific to North American culture to discriminate among LLD and non-LLD subjects from multicultural backgrounds.
- 3. Although Subtest One (Understanding Ambiguous Sentences) and Subtest Four (Understanding Metaphoric Expressions) did not contribute to group discrimination in this study, their exclusion from the battery for the purpose of classifying LLD subjects is not recommended without further research.
- 4. Because VIQ explains a significant proportion of variance in TLC performance, the administration of both instruments for the purpose of classifying language disabled subjects seems time-consuming and redundant; however substitution of one instrument or the other is not recommended without further research.
- 5. Cautious interpretation of TLC results based on the test norms is recommended, as these may tend to overidentify individuals as language disabled.
- 6. The TLC authors recommend the use of profile analysis for determining individual strengths and weakness; however, this practice is not recommended, first because TLC norms may not be representative of the performance of local children, and second, because limited support has been demonstrated for TLC subtest specifity.
- 7. Detailed suggestions for developing individual education plans (IEP's) on the basis of TLC data are provided in the test manual. Current results, however, lend little support to this practice. First, there is no evidence offered by the test authors to suggest that remediation based on TLC results has an effect on actual performance over time. Second, IEP's are based on the assumption that each

subtest measures a unique aspect of language competence; however, current results fail to support the specificity of the TLC subtests.

8. It is suggested that the TLC be used only in conjunction with other language measures which provide reliable diagnostic information. These might include an informal language sample, such as the LARSP analysis.

Recommendations for Further Research

- Suggestions for further research would include additional criterion-related validity studies at different age levels to observe test characteristics and possible developmental patterns in TLC performance.
- 2. Given that current results were obtained from a multilingual/multicultural sample, future studies might focus on the performance of distinct ethnic groups, including EFL.
- 3. Further investigation of the criterion-related validity of the TLC short form (Subtests Three and Four) for use with local subjects is suggested. The validity of using Subtests Two and Three as a short form for local children might be investigated.
- 4. Further investigation of the relationship between VIQ and the TLC is suggested to determine the relative efficacy of each instrument for classifying LLD subjects, and whether one instrument might be substituted for another to avoid redundancy.
- 5. Further investigations of TLC subtest specificity using factor analysis are suggested.
- 6. Experimental research to determine the effectiveness over time of instructional objectives based on TLC performance might be considered.
- 7. Further examination of the appropriateness of TLC norms for local children is suggested. The development of local norms might be considered.

- American Psycholgical Association (1985). <u>Standards for educational and</u> <u>psychological testing</u>. Washington, DC: American Psychological Association.
- American Speech-Language Hearing Association (1988). Report of the ad hoc committee on instrument evaluation. <u>ASHA</u>, <u>23</u>, 75-76.
- Anastasi, A. (1982). Psychological testing (5th ed.). New York, NY: MacMillan.
- Aram, D., Ekelman, B., & Nation, J. (1984). Preschoolers with language disorders: 10 years later. <u>Journal of Speech and Hearing Research</u>, <u>27</u>, 232-244.
- Atkins, C.P., & Cartwright, L.R. (1982). Preferred language elicitation procedures used in five age categories. <u>ASHA</u>, 24, 321-323.
- Bernstein, D.K. (1989). Assessing children with limited English proficiency: Current perspectives. <u>Topics in Language Disorders</u>, 9, 15-20.
- Bishop, D. (1985). Language assessment, remediation and screening procedure (LARSP) by David Chrystal, Paul Fletcher & Mike Garman: Computerized version for Apple II and IIe. (Available from Department of Speech, University of Newcastle upon Tyne, Great Britian)
- Blau, A., Lahey, M., & Oleksiuk-Velez, A. (1984). Planning goals for intervention: Can a language test serve as an alternative to a language sample? <u>Journal of</u> <u>Childhood Communication Disorders</u>, 7, 27-37.
- Bloom, L., & Lahey, M. (1978). <u>Language development and language disorders.</u> New York, NY: John Wiley & Sons.
- Caskey, W., & Franklin, L. (1986). The Test of Adolescent Language (TOAL) and WISC-R scores: A caveat. Language. Speech and Hearing Services in Schools, 17, 307-311.
- Committee on the Status of Racial Minorities (1983). Social dialects (position paper). ASHA, 25, 23-25.
- Cronbach, L. (1971). Test validation. In R. Thorndike (Ed.), <u>Educational</u> <u>measurement</u>. Washington, DC: American Council on Education.
- Cronbach, L., & Meehl, P. (1955). Construct validity in psychological tests. <u>Psychological Bulletin</u>, 52, 281-302.
- Crystal, D., Fletcher, P., & Garman, M. (1976). <u>The grammatical analysis of</u> <u>language disability</u>. London, Great Britain: Edward Arnold.
- Cupples, W.P., & Lewis, M.E.B. (1984). Language-based learning disabilities: differential diagnosis and implementation. <u>Learning Disabilities</u>, <u>3</u>, 129-140.
- Damico, J.S. (1988). The lack of efficacy in language therapy: A case study. Language, Speech and Hearing Services in Schools, 19, 51-66.

- Damico, J.S. (in press). Descriptive assessment of communicative ability in limited English proficient students. In E. Hamayan, & J.S. Damico (Eds.), <u>Nonbiased</u> <u>assessment of limited English proficient special education children</u>. San Diego, CA: College Hill Press.
- Darley, F.L., & Spriestersbach, D.C. (1978). <u>Diagnostic methods in speech pathology</u> (2nd ed.). Reading, MA: Addison-Wesley.
- Dixon, W.J. (Ed.) (1988). <u>BMDP statistical software manual</u> (Vol. 1). Berkely, CA: University of California Press.
- Donahue, M. (1985). Review of the WORD Test. In J.V. Mitchell (Ed.). <u>Ninth</u> <u>mental measurements yearbook</u>. Lincoln, NE: Buros Institute of Mental Measurement, University of Nebraska.
- Education Services Group (1985). <u>Student services handbook</u>. Vancouver, B.C.: Vancouver School Board.
- Evard, B., & Sabers, D. (1979). Speech and language testing with distinct ethnic -racial groups: A survey of procedures for improving validity. Journal of Speech and Hearing Disorders, 44, 271-281.
- Ferguson, G.A. (1976). <u>Statistical analysis in psychology and education</u>. Toronto, Canada: McGraw-Hill.
- Guion, R.M. (1977). Content validity: Three years of talk-What's the action? <u>Public</u> <u>Personnel Management</u>, 6, 407-414.
- Gunnarsson, B. (1978). A look at the content similarities between intelligence, achievement, personality, and language tests. In J.W. Oller Jr. (Ed.), Language in education: Testing the tests. Rowley, MA: Newbury.
- Hammill, D., & Newcomer, P. (1988). <u>Test of Language Development 2 -</u> <u>Intermediate</u> (TOLD-2-I). Austin, TX: Pro-Ed.
- Hammill, D., Brown, V., Larsen, S., Wiederholt, J. (1987). <u>Test of Adolescent</u> <u>Language - 2</u> (TOAL-2). Austin, TX: Pro-Ed.
- Hawkins, P., & Spencer, H. (1985). Imitative versus spontaneous language assessment: A comparison of CELI and LARSP. <u>British Journal of Disorders</u> of Communication, 20, 191-200.
- Henrysson, S. (1971). Gathering, analyzing, and using data on test items. In R. Thorndike (Ed.) <u>Educational measurement</u>. Washington, DC: American Council on Education.
- Holland, A., & Forbes, M. (1986). Nonstandardized approaches to speech and language assessment. In O.L. Taylor (Ed.), <u>Communication disorders in</u> <u>linguistically diverse populations</u> (pp. 49-66). San Diego, CA: College-Hill Press.

Kaufman, A.S. (1979). Intelligent testing with the WISC-R. New York, NY: Wiley.

- Kearns, K., & Simmons, N. (1983). A practical procedure for the grammatical analysis of aphasic language impairments: The LARSP. In R. Brookshire (Ed.), clinical aphasiology conference proceedings. Minneapolis, MN: BRK.
- Kelly, D.J., & Rice, M.L. (1986). A strategy for language assessment of young children: A combination of two approaches. <u>Language. Speech and Hearing Services in Schools</u>, 17, 83-92.
- Klecka, W.R. (1980). Discriminant analysis. Beverly Hills, CA: Sage
- Kretschmer, R.R., & Kretschmer, L.W. (1978). <u>Language development and</u> <u>intervention for the hearing impaired.</u> Baltimore, MD: University Park Press.
- Lai, C. (1986). <u>U.B.C. SPSSX Statistical Package for Social Sciences extended</u> version. Vancouver, Canada: University of British Columbia.
- Larson, V., & McKinley, N. (1987). <u>Communication assessment and intervention</u> <u>strategies for adolescents</u>. Eau Claire, WI: Thinking Publications.
- Launer, P.B., & Lahey, M. (1981). Passages: From the fifties to the eighties in language assessment. <u>Topics in Language Disorders</u>, <u>1</u>, 11-29.
- LaTorre, R.A. (1983). <u>The 1982 survey of pupils in Vancouver schools for whom</u> <u>English is a second language</u> (Report No.83-02). Vancouver: Board of School Trustees, Evaluation and Research Services, Program Resources.
- Leonard, L.B., Perozzi, J.A., Prutting, C.A., & Berkley, R.K. (1978). Nonstandardized approaches to the assessment of language behaviors. <u>ASHA</u>, <u>20</u>, 371-379.
- Lieberman, R., Heffron, A., West, S., Hutchinson, E., & Swem, T. (1987). A comparison of four adolescent language tests. Language, Speech, and Hearing Services in Schools, 18, 250-266.
- Lieberman, R.J., & Michael, A. (1986). Content relevance and content coverage in tests of grammatical ability. Journal of Speech and Hearing Disorders, 51, 71-81.
- Lund, N., & Duchan, J. (1983). <u>Assessing children's language in naturalistic</u> <u>contexts</u>. Englewood Cliffs, NJ: Prentice Hall.
- McCauley, R.J., & Swisher, L. (1984a). Psychometric review of language and articulation tests for preschool children. <u>Journal of Speech and Hearing</u> <u>Disorders, 49</u>, 34-42.
- McCauley, R.J., & Swisher, L. (1984b). Use and misuse of norm-referenced tests in clinical assessment: A hypothetical case. Journal of Speech and Hearinng Disorders, 49, 338-348.
- Messick, S. (1980). Test validity and the ethics of assessment. <u>American</u> <u>Psychologist</u>, <u>35</u>, 1012-1027.

- Miller, J.F., & Chapman, R. (1983). <u>SALT: Systematic analysis of language</u> <u>transcripts.</u> Madison, WI: University of Wisconsin, Language Analysis Laboratory, Waisman Center.
- Muma, J. (1981). <u>Language primer for the clinical fields.</u> Lubbock, TX: Natural Child Publishing.
- Muma, J.R. (1984). Semel and Wiig's CELF: Construct validity? (Letter to the editor). Journal of Speech and Hearing Disorders, 49, 101-104.
- Muma, J.R. (1985). "No news is bad news": A response to McCauley and Swisher (letter to the editor). Journal of Speech and Hearing Disorders, 50, 290-292.
- Muma, J.R., Lubinski, R., & Pierce, S. (1982). A new era in language assessment: Data or evidence. In N.J. Lass (Ed.), <u>Speech and language: Advances in basic</u> research and practice. Toronto, Canada: Academic.
- Nelson, L.R. (1974). <u>Guide to LERTAP use and interpretation</u>. Dunedin, New Zealand: University of Otago.
- Nice, M.M. (1925). Length of sentences as a criterion of a child's progress in speech. Journal of Educational Psychology, 16, 370-179.

Nunnally, J.C. (1978). Psychometric theory. Toronto, Canada: McGraw-Hill.

- Oller, J.W., Jr. (1978). How important is language proficiency to IQ and other educational tests? In J.W. Oller Jr. (Ed.), <u>Language in education: Testing the</u> tests. Rowley, MA: Newbury.
- Oller, J.W., Jr., & Damico, J. (in press). Theoretical considerations in the assessment of LEP students. In E. Hamayan, & J.S. Damico (Eds.), <u>Nonbiased</u> <u>assessmentof limited English proficient special education children</u>. San Diego, CA: College Hill Press.
- Pedhazur, E.J. (1982). <u>Multiple regression in behavioral research</u>. New York, NY: Holt, Rinehart and Winston.
- Penn, M.A.C. (1983). <u>Syntactic and pragmatic aspects of aphasic language.</u> Unpublished doctoral dissertation, University of Witwatersrand, South Africa.
- Penn, C., & Behrmann, M. (1986). Towards a classification scheme for aphasic syntax. <u>British Journal of Disorders of Communication</u>, 21, 21-38.
- Prather, E., Beecher, S., Stafford, M., & Wallace, E. (1980). <u>Screening Test of</u> <u>Adolescent Language.</u> Seattle, WA: University of Washington Press.
- Public Law 94 142. <u>Education for all Handicapped Children Act.</u> Federal Register, 1975, p. 42478.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1972). <u>A grammar of contemporary English</u>, London, Great Britain: Longman.

- Raju, N.S. (1985). Review of the WORD Test. In J.V. Mitchell (Ed.). <u>Ninth mental</u> <u>measurements vearbook</u>. Lincoln, NE: Buros Institute of Mental Measurement, University of Nebraska.
- Santos, O. B. (1987). The variance in reading comprehension in terms of language skills and cognitive processes (Doctoral dissertation, Boston University, 1987). Dissertation Abstracts International, 48, 8716098.
- Sattler, J. (1988). <u>Assessment of children</u> (3rd ed.). San Diego, CA: Jerome M. Sattler.
- Schery, T.K. (1985). Correlates of language development in language-disordered children. Journal of Speech and Hearing Disorders, 50, 73-83.
- Semel, E., & Wiig, E. (1980). <u>Clinical Evaluation of Language Functions</u>. Columbus, OH: Merrill.
- Semel, E., Wiig, E.H., & Secord, W. (1987). <u>Clinical Evaluation of Language</u> <u>Fundamentals - Revised (CELF-R)</u>. San Antonio, TX: Psychological Corporation.
- Simon, C.S. (1984). Functional-pragmatic evaluation of communication skills in school-aged children. Language, Speech, and Hearing Services in Schools, 15, 83-98.
- Sommers, R.K. (1985). A review of the Screening Test of Adolescent Language. In J.V. Mitchell (Ed.), <u>Ninth mental measurements yearbook</u>. Lincoln, NE: Buros Institute of Mental Measurement, University of Nebraska.
- Sommers, R.K., Erdige, S., & Peterson, M.K. (1978). How valid are children's language tests? <u>The Journal of Special Education</u>, 12, 393-407.
- Stark, R.E., Tallal, P., Mellits, E.D. (1982). Quantification of language abilities in children. In N.J. Lass (Ed.), <u>Speech and language: Advances in basic</u> research and practice. Toronto, Canada: Academic Press.
- Stephens, M., & Montgomery, A. (1985). A critique of recent relevant standardized tests. <u>Topics in Language Disorders</u>, 5, 21-45.
- Thorndike, R. (Ed.) (1971). <u>Educational measurement.</u> Washington, DC: American Council on Education.
- Thorum, A. (1980). <u>Fullerton Language Test for Adolescents</u>. Palo Alto, CA: Consulting Psychologists Press.
- Thurstone, T.G. (1978). <u>Educational Abilities Series</u>. Palo Alto, CA: Science Research Associates.
- Tibbits, D.F. (1982). Language disorders in adolescents. Lincoln, NE: Cliffs Notes.
- Tyack, D., & Gottsleben, R. (1974). <u>Language sampling, analysis and training</u>. Palo Alto, CA: Consulting Psychologists Press.

Vaughn-Cooke, F. (1983). Improving language assessment in minority children. ASHA, 25, 29-34.

- Watson-Russel, A. (1986). <u>Student ethnic survey</u> (Research Report No. 86-07). Vancouver: Board of School Trustees, Student Assessment and Research, Education Services Group.
- Wechsler, D. (1974). <u>Wechsler Intelligence Scale for Children Revised</u>. New York, NY: Psychological Corporation.
- Wiig, E.H., & Secord, W. (1985). <u>Test of Language Competence</u>. Toronto, Canada: Merrill.
- Wiig, E., & Semel, E. (1984). Language assessment and intervention for the learning disabled. Toronto, Canada: Merrill.

Appendix A

Parent/Student Information

Letter to Parents of Control Subjects

Dear Parents:

's school has agreed to participate in a research project: "Validity of the Test of Language Competence". The Test of Language Competence (TLC) is a testing instrument used by speech/language specialists in the Vancouver School District to appraise children's use of language. The purpose of this study is to investigate how well the TLC does the job for which it is intended with secondary school students.

For this project to be successful, 60 students in the Vancouver School District will take 3 tests: the TLC, an ability test, and an informal language sample. The language sample is gathered by audiotaping an informal conversation between each student and a graduate student examiner. These conversations centre around nonthreatening real-life topics such as "What would you do if you won a million dollars?". The sole purpose of these conversations is to obtain a representative sample of each student's usual language. In addition, parents of participating students will be asked to complete a brief questionnaire.

The researcher seeks to determine if the TLC is an effective measure of language functioning, and also how it compares to anothter commonly used procedure, the language sample analysis.

The research project is being undertaken as a master's thesis in the Department of Educational Psychology at the University of British Columbia. it has been approved by the Vancouver School Board's Student Assessment and Research offfice, by the Principal of your son's/daughter's school, and by Faculty research specialists at the University. 's name has been chosen as a possible participant in this research. If you and your son/daughter agree to participate, he/she will be asked to take part in two individual testing sessions of approximately 60 minutes and 2 hours respectively. A trained graduate student will do the testing in the school. Such tests are common in schools. Students usually find them interesting and enjoyable.

Parents will also be asked to complete a brief questionnaire on family background.

The results of the tests will be strictly confidential; your child's name will not appear on the test forms. No individual test results will be released. The purpose is not to test any one child's performance, but rather to evaluate the usefulness of the Test of Language Competence. Parents interested in receiving a copy of the group results should request this on the consent form.

I wish to emphasize that participation is voluntary. Participation in or withdrawl from the project at any time will not in any way influence your son's/daughter's class standing. I would, however, greatly appreciate your cooperation in this research. Please complete the Parent Consent Form and questionnaire and return it in the envelope provided as soon as possible.

Thank you. Feel free to contact me for any further information at

_____, or _____ for messages.

Sincerely,

C. Ainsworth

- 2 -

Letter to Parents of LLD Subjects

Dear Parents:

's school has agreed to participate in a research project: "Validity of the Test of Language Competence". The Test of Languge Competence (TLC) is a testing instrument used by speech/language specialists in the Vancouver School District to appraise children's use of language. The purpose of this study is to investigate how well the TLC does the job for which it is intended with secondary school students.

For this project to be successful, 60 students in the Vancouver School District will take 3 tests: the TLC, an ability test, and an informal language sample. The language sample is gathered by audiotaping an informal conversation between each student and a graduate student examiner. These conversations centre around nonthreatening real-life topics such as "What would you do if you won a million dollars?". The sole purpose of these conversations is to obtain a representative sample of each student's usual language. In addition, parents of participating students will be asked to complete a brief questionnaire.

The researcher seeks to determine if the TLC is an effective measure of language functioning, and also how it compares to anothter commonly used procedure, the language sample analysis.

The research project is being undertaken as a master's thesis in the Department of Educational Psychology at the University of British Columbia. it has been approved by the Vancouver School Board's Student Assessment and Research offfice, by the Principal of your son's/daughter's school, and by Faculty research specialists at the University. 's name has been chosen as a possible participant in this research. If you and your son/daughter agree to participate, he/she will be asked to take part in one individual testing session of approximately 60 minutes. A trained graduate student will do the testing in the school.

Such tests are common in schools. Students usually find them interesting and enjoyable.

In order to avoid re-testing children for whom test data is already available, your permission for the investigator to obtain such existing information is requested. Parents will also be asked to complete a brief questionnaire on family background. The results of the tests will be strictly confidential; your child's name will not appear on the test forms. No individual test results will be released. The purpose is not to test any one chid's performance, but rather to evaluate the usefulness of the Test of Language Competence.

Parents interested in receiving a copy of the group results should request this on the consent form.

I wish to emphasize that participation is voluntary. Participation in or withdrawl from the project at any time will not in any way influence your son's/daughter's class standing. I would, however, greatly appreciate your cooperation in this research. Please complete the Parent Consent Form and the questionnaire and return it to the school as soon as possible.

Thank you. Feel free to contact me for any further information at ______, or _____ for messages.

Sincerely,

C. Ainsworth

- 2 -

Parent Consent Form (Controls)

VALIDITY STUDYOF THE TEST OF LANGUAGE COMPETECE PARENT CONSENT FORM

I am willing / not willing to give my consent for ______'s participation in the research study at _______school. I am aware that this will involve testing sesions totalling approximately three hours duration. In understand that confidentiality of test results will be maintained and that no individual scores will be released. I also understand that participation in this project is voluntary and may be terminated at any time.

Name

Signature

Date

I would / would not like a copy of the group results to be mailed to:

VALIDITY STUDY OF THE TEST OF LANGUAGE COMPETENCE PARENT CONSENT FORM

I am willing / not willing to give my consent for _____'s participation in the research study at ______school. I am aware that this will involve testing sessions totalling approximately 60 minutes duration. I further consent to the release of test data which has been obtained for my son/daughter to the principal investigator in this research. In understand that confidentiality of test results will be maintained and that no individual scores will be released. I also understand that participation in this project is voluntary and may be terminated at any time.

Name

Signature

Date

I would / would not like a copy of the group results to be mailed to:

VALIDITY STUDYOF THE TEST OF LANGUAGE COMPETENCE STUDENT CONSENT FORM

I am willing / not willingto participate in the research study at

school. I am aware that this will involve testing sessions totalling about three hourse in length. I understand that my test results will be kept confidential. My name will not appear on any of the test papers or in the final report. I also understand that my participation in this project is voluntary, and that I can quit at any time without affecting my school grades.

Name

Signature

Date

VALIDITY STUDY OF THE TEST OF LANGUAGE COMPETENCE STUDENT CONSENT FORM

I am willing / not willing to participate in the research study at

Name

Signature

Date

QUESTIONNAIRE

Your assistance in providing the following information would be very helpful in making this a meaningful study:

- 1. What language do adults speak in the home?
- 2. What language do children speak in the home?
- 3. How often do adults speak English in the home?

2.	rion orien de dadas speak English in the nome,
	always
	3/4 of the time
	1/2 of the time
	1/4 of the time
	never
4.	How often do children speak English in the home?
	always
	3/4 of the time
	1/2 of the time
	1/4 of the time
	never
5.	In which area of the city do you live?
	Downtown (west-end)
	Vancouver west of Main Street
	Vancouver east of Main Street and north of 41st Avenue
	Vancouver east of Main Street and sound of 41st Avenue
6.	What is your son's/daughter's birthdate and age?
	Age:
	year month day
7.	How would you describe your son's/daughter's school achievement?
	Below Average Average Above Average
	Reading
	Writing
	Spelling
	Arithmetic
^	••• · · · · · · · · · · · · · · · · · ·
8.	Has your son/daughter ever received special assistance for learning
	difficulties?(yes)(no)
0	OUESTIONS ADDRESSED TO THE MOTHER :
9.	QUESTIONS ADDRESSED TO THE MOTHER
	9a. Which estages helew hart describes way as what the stages
	90. Which category below best describes your completed level of
	education?
	Less than High School Completion
	High School Completion
	Post-Secondary, no degree
	University or college degree

10. QUESTIONS ADDRESSED TO THE FATHER

10a. What is your occupation?_____

10b. Which category below best describes your completed level of education?

Less than High School Completion	on
 High School Completion	
 Post-Secondary, no degree	
 University or college degree	

Thank you for your cooperation.

Appendix B

.

Item Analysis Data

				Correlations		<u>% Dis</u>	tributio	<u>n</u>
Item	Mean	SD	Subtest	Total Test	VIQ	0	1	3
1.	1.63	1.21	63	53	44	19.6	39.1	41.3
2.	1.23	1.30	78	70	63	41.3	26.1	32.6
3.	1.02	1.20	47	35	25	45.7	30.4	23.9
4.	1.06	1.23	58	64	60	21.7	37.0	41.3
5.	1.15	1.13	76	63	58	32.6	43.5	23.9
6.	1.00	1.09	66	62	55	39.1	41.3	19.6
7.	1.02	.95	66	52	50	28.3	56.5	15.2
8.	.78	1.05	61	51	48	52.2	32.6	15.2
9.	.89	1.14	60	57	47	50.0	30.4	19.6
10.	.56	.88	51	42	42	60.9	30.4	8.7
11.	.73	1.06	52	56	40	56.5	28.3	15.2
12.	.71	.93	47	63	60	50.0	39.1	10.9
13.	.80	.91	35	54	46	41.3	47.8	10.9

Summary Item Statistics: Subtest One: Understanding Ambiguous Sentences

<u>Note</u>. n = 46

Item order difficulty correlation: Rho = 76

			. <u></u>	Correlations		% Dis	tributio	<u>n</u>
Item	Mean	SD	Subtest	Total Test	VIQ	0	1	3
1.	2.61	.80	32	22	22	0	.6	80.4
2.	2.19	1.09	17	25	32	6.5	30.4	63.
3.	2.37	1.04	15	35	30	6.5	21.7	71.7
4.	2.58	.86	54	49	46	2.2	17.4	80.4
5.	2.13	1.00	28	35	39	0.0	43.5	56.5
6.	2.04	1.13	58	72	72	8.7	34.8	56.5
7.	1.84	1.03	55	47	54	2.2	54.3	43.5
8.	1.69	1.03	43	40	40	4.3	58.7	37
9.	2.17	.99	16	32	22	0	41.3	58.7
10.	1.78	1.05	47	46	32	4.3	54.3	41.3
11.	1.93	1.10	36	49	37	6.5	43.5	50.
12.	1.93	1.10	34	34	34			

Summary Item Statistics: Subtest Two: Making Inferences

<u>Note</u>. n = 46

5

Item order difficulty correlation: Rho = .77

.

			· .	· · ·				
	· · · · · · · · · · · · · · · · · · ·			Correlations	•	<u>% Dis</u>	tributio	n
Item	Mean	SD	Subtest	Total Test	VIQ	0	1	3
1.	2.50	.91	38	42	40	2.2	21.7	76.1
2.	1.63	.97	21	25	23	2.2	65.2	32.6
3.	2.13	1.07	40	37	35	4.3	37.	58.7
4.	2.09	1.13	43	55	54	8.7	32.6	58.7
5.	2.10	1.04	33	34	42	2.2	41.3	56.5
6.	1.83	1.12	64	69	63	8.7	45.7	45.7
7.	1.52	1.15	59	54	55	17.4	47.8	34.8
8.	1.56	1.05	18	06	-01	8.7	58.7	32.6
9.	2.17	1.06	33	44	38	4.3	34.8	60.9
10.	1.30	.96	56	50	45	13.	65.2	21.7
11.	1.61	1.06	39	33	29	8.7	56.5	34.8
12.	1.41	1.13	28	24	19	19.6	50	30.4
13.	1.89	1.16	60	51	40	10.9	39.1	50

Subtest Three: Recreating Sentences (Holistic)

<u>Note</u>. n = 46

Item order difficulty correlation: Rho = .53

Subtest Three: Recreating Sentences (Word Count)

	·····		·	· · · ·				
				Correlations		<u>% Distribution</u>		
Item	Mean	SD	Subtest	Total Test	VIQ	0	1	3
1.	2.74	.68	31	28	29	0	13	87
2.	2.54	.89	20	19	15	2.2	19.6	78.3
3.	2.69	.81	34	12	03	4.3	8.7	87
4.	2.24	1.14	37	36	41	10.9	21.7	67.4
5.	2.67	.87	47	28	27	6.5	6.5	87
6.	2.52	1.00	51	37	30	8.7	10.9	80.4
7.	2.52	1.07	71	49	35	13.0	4.3	82.6
8.	2.28	1.13	34	17	08	10.9	19.6	69.6
9.	1.80	1.20	58	63	62	15.2	37	47.8
10.	2.30	1.20	55	41	32	17.4	8.7	73.9
11.	2.48	1.09	66	54	42	13	6.5	80.4
12.	2.13	1.29	38	44	31	21.7	10.9	67.4
13.	2.41	1.08	76	50	32	10.9	13	76.1

<u>Note</u>. n = 46

Item order difficulty correlation: Rho = .664

Subtest Three: Recreating Sentences (Total Test)

<u>Correlations</u> <u>Distribution</u>												
Item	Mean	SD	Subtest	Total Test	VIQ	0	1	2	3	4	5	6
1.	5.24	1.21	34	47	47	0	2.2	2.2	0	28.3	0	67.4
2.	4.17	1.39	33	30	26	2.2	0	13	0	58.7	0	26.1
3.	4.83	1.44	35	34	27	4.3	0	0	0	45.7	0	50
4.	4.33	1.94	41	53	56	6.5	0	17.4	65	19.6	0	80
5.	4.78	1.41	33	43	48	2.2	0	4.3	4.3	39.1	0	50.0
6.	4.35	1.69	57	68	59	6.5	0	65	4.3	43.5	0	39.1
7.	4.08	1.81	66	63	54	8.7	4.3	2.2	4.3	47.8	0	32.6
8.	3.85	1.55	13	16	05	8.7	0	6.5	2.2	65.2	0	17.4
9.	3.87	1.69	58	70	63	4.3	2.2	17.4	10.9	37	0	28.3
10.	3.61	1.68	63	58	49	13	2.2	2.2	2.2	67.4	0	13
11.	4.04	1.67	55	60	50	8.7	2.2	2.2	2.2	58.7		26.1
12.	3.53	1.96	28	42	32	13	10.9	0	4.3	50		21.7
13.	4.35	1.83	72	62	46	6.5	4.3	4.3	4.3	37		43.5

<u>Note</u>: n - 46

Item order difficulty correlation: Rho = .65

				Correlations		<u>% Dis</u>	tributio	<u>n</u>
Item	Mean	SD	Subtest	Total Test	VIQ	0	1	3
1.	1.80	1.20	37	53	50	15.2	37.0	47.8
2.	1.22	1.26	69	73	64	39.1	30.4	30.4
3.	1.26	1.24	67	61	62	34.8	34.8	30.4
4.	1.67	1.33	50	49	45	28.3	23.9	47.8
5.	1.37	1.27	49	43	35	32.6	32.6	34.8
6.	1.15	1.25	74	69	63	41.3	30.4	28.3
7.	1.37	1.27	63	68	69	32.6	32.6	34.8
8.	1.22	1.26	48	52	55	39.1	30.4	30.4
9.	1.15	1.44	76	68	64	58.7	4.3	37.0
10.	.76	1.23	69	65	58	67.4	10.9	21.7
11.	.72	1.07	56	57	50	58.7	26.1	15.2
12.	.85	1.03	76	80	72	45.7	39.1	15.2

Subtest Four: Understanding Metaphoric Expressions

Note: n = 46

Item order difficulty correlation: Rho = .80

x

Appendix C

Stimulus Items Used to Elicit Language Samples

Stimulus Items Used to Elicit Language Samples

- 1. What was the last movie you saw? Tell me about it.
- 2. Describe what you usually do on Saturdays from the time you get up until the time you go to bed.
- 3. Tell me about the funniest thing that has ever happened at your house.
- 4. Tell me about your favourite singer/actor.
- 5. What is your favourite TV show? What happened on that show the last time you saw it?
- 6. What's the best book/story you've read. Tell me about it.
- 7. Tell me about your favourite teacher.
- 8. Tell me how you would spend a million dollars.
- 9. Tell me about the best time you can remember having with your family or friends.
- 10. What's the funniest/most embarrasing/exciting thing that's happened to you? Tell me about it.
- 11. Where do you live? Tell me how to get there from here.
- 12. Describe your home.
- 13. Tell me about your hobbies.
- 14. Describe how your family celebrates Christmas.
- 15. How will you spend your Easter vacation?
Appendix D

Language Sample Transcripts

Language Sample Transcript: Control Case One

There was this guy and he went around looking for all these jobs.

He wouldn't fit into any of them.

Finally he went to a department store.

Some accident happened.

A sign came falling.

There was a lady under it.

He grabbed on to it and started swinging up in the air.

He liked working with manikans.

The lady asked him what kind of job he wanted.

He explained to her what he did.

She gave him a job.

There was these people that did not agree and disliked the person with the job.

He finally made this manikan.

One day he turned arond and she was alive.

All these happened.

Everyone starts noticing this change in him.

They're wondering why he's talking to a manikan.

At the end she turns into real life.

She stays like that.

They showed these scenes right at the start.

It was in Asia.

She was in a tomb.

She wanted a different life but she didn't know how she was gonna get it.

That's how it ended out.

It was good.

I liked it alot.

There was this family.

Mack is old.

Someone said that she'd died.

All of a sudden he found out that he had a daughter.

She came there.

A whole bunch of scenes were happening.

Life was changing in the house.

Everybody was getting into arguments.

Then Ann came along.

He thought that she was dead.

Now she moved into the house next door and his wife doesn't agree with it.

They always argue.

At the end he went over there and they started dancing.

They were having good times.

Karen's starting to get thoughts about what's going on between them again.

She's married to a man named Ben now.

She had twins.

She found out they were Gary's.

They weren't Ben's.

Ben doesn't approve of it.

Val and Gary are good friends.

While she was being a teacher she was being a friend too.

Girls could go and talk to her.

Language Sample Transcript: Control Case Two

It would be my teacher because she had a farm and we got to go to her farm and feed her goats.

It is about teenagers.

They went to find a dead body.

They went along train tracks.

When they found it there was this older group that wanted to find it.

They stook up for it.

They never did say who got it.

They just brought it.

His brother had died.

I guess when he saw this dead one he might have thought of him.

Actually it was really good.

It was about friendship.

They always stood by each other.

They were trying to figure out what girls were like.

He went through puberty or whatever.

He saw his babysitter and I guess he liked her.

He was trying to get to know her.

Her sister came in.

He said that he was a college man.

He lied.

She found out.

They got mad.

They asked their dad what it was about girls.

That is the plot.

My dad says I don't need braces.

So he won't get them.

I guess I would get them.

It's got grey carpet.

My room has a balcony.

We have different rooms.

In the middle there is a bathroom.

We just walke through doors and we go there.

I guess it was Sunday.

I went down to Metrotown with Joy.

We just had fun.

We went to those picture booths and we took a couple of pictures.

It was great.

Usually we just sit around.

My dad makes popcorn.

We just watch TV.

Yesterday Lesley came over.

She said my stairs remind her of Psycho.

Me and my sister are in the upstairs.

We are the only ones there.

My sister was babysitting.

Then she left.

I was remembering Psycho when I was laying there.

It was really scary.

Language Sample Transcript: Control Case Three

He's working at this store.

He makes this manikan.

Then he gets fired.

He sees the manikan in a store.

One day he's working on something and she comes alive.

Near the end somewhow she gets into this machine where she's going to be chopped up.

He saves her.

At the end everyone else can see her.

The son Theo wanted to take flying lessons.

The little girl brought home a boyfriend and she was ordering him around and telling him what to do.

He used to stay after school and help you with problems.

Spend alot of time with you.

First put it in the bank.

When I get older buy a car.

Probably move to Hawaii.

We went there for my dad's convention.

He was mostly at the convention.

My mom used to go shopping and me and my sister just go to the beach.

Walk around.

I went to the Polynesian Cultural Centre, Pearl Harbour, the zoo.

We walked around quite a bit.

Met some new people there.

It was pretty sad when they showed the movie and all the people dying.

It's kind of just like a normal boat.

Then you go on a memorial.

There's a big sign of all the people who died.

Some people throw flowers into the water.

Some of them are sticking out from the water.

In the evenings we's all go out for dinner and then go shopping or see a movie.

This mother takes her two kids to live with her parents but her parents don't know that she has the kids.

She leaves them locked away in an attic.

She leaves them there for a really long time.

She used to come and visit.

She got married to someone else.

The kids would get really bored.

One of the kids died because it got sick.

They took her to the hospital but there wasn't enough time.

At the end all the kids sneak out of the window and run away.

It was just lying there.

One day its mom finally came in.

The brother and sister told the mom.

Once a long time ago I put my pants on backwards.

I had the holes at the back.

I came out and we had some guests over.

I like their songs and their drummer is really good.

I like Sylvester Stalone and Harrison Ford.

Language Sample Transcript: Control Case Four

He was being killed.

He asked a friend to help.

He must have a revenge on that friend.

Now he started to have his revenge but it's not the end yet.

It's only one chapter a day.

It's a Chinese way.

I don't know which one to pick.

It's about a hero that helps people.

Ther's lots of girls liked him.

Girls are following him around.

He knows Kung Fu.

He knows some other friends that knows Kung Fu too.

They're all heroes.

It was about people finding out what's happening.

There's a water flooding.

People would go over there and check what's happening.

They would go and help them.

She's a nice teacher.

She gives out candies every week on Friday.

He's real funny.

He jokes a lot.

He doesn't give us hard work.

I would buy a new house.

Go on field trips with my parents.

Buy lots of things.

We went to a restaurant to eat dinner.

We just go on special days like mother's day or when something comes to Canada.

Then we go to a restaurant to eat.

Some live in Hong Kong.

I don't know.

I read all different kinds of books.

It's about the mixed-up twins.

It's about a twin that gets mixed-up.

They're the same.

Policeman and friends get all mixed-up with those two.

One time when they were lost policemen were trying to find them.

When the policeman find one of them the other one ran away.

When the policeman find that one again the policeman got all mixed-up.

Sometime they do something wrong and the it's real funny.

Sometimes they make mistakes.

There are Chinese.

They're singers and actors.

They sing.

Their songs are excellent.

I go shopping with my mother sometimes.

Go to my grandma's house.

Do my homework.

Play with my friend.

Language Sample Transcript: Control Case Five

The last movie I saw was Star Trek Four.

I don't know if you're into science fiction.

It was really weird.

It was different from all the other ones because all the other ones were really science fiction.

This one is more comical because instead of the group being in their spaceship they were coming back in the twenty-fifth century.

There was a great big probe or a spaceship or something.

It was terrorizing the earth and was planning on destroying it.

It was sending off these messages that only whales could hear

They had to go back to our century to find these whales and bring them back.

It was really comical.

It was good.

It left you in suspense for a fifth part coming out.

The last time I saw it was probably last Thursday.

He was telling everybody how nobody could fool him with all these practical jokes they were trying to play on him.

The whole family made up this really elabourate joke to play on him.

He overheard them talking about it.

It backfired on them and he got them instead.

He was really too smart for them to actually play practical jokes on.

My favourite teacher was probably my grade seven teacher.

He wasn't old.

He was in his forties.

He knew where his students stood.

He wasn't all oldfashioned.

He knew all the terms we used and everything they meant.

We could talk to him as if he was just another student.

We didn't have to talk to him as if he was a teacher.

He did alot of things with us.

When we went places he let us make suggestions of where to go and then he would pick the best places.

We went camping two or three times with him.

We went one time for a whole week.

We missed a whole week of school.

He took the whole class on a camping trip.

If I won a million dollars I'd probably put it in the bank for a year and let the interest grow.

My parents were talking about this just the other day.

They were saying that they'd put it in the bank for a year.

Then they'd take half of it out and use it for a downpayment on a house.

They wouldn't take the whole thing out and use it at one time.

Every day there was all sorts of things to do because I met a whole bunch of new friends.

We did everything together.

Language Sample Transcript: LLD Case One I would put it in my bank. Going vacation. I forget. Buy car. Visit my aunt and cousin. I did not watch any movies. We have a part at my house. We celebrate. Everybody came to our house. Put all the dishes in the other side. Wash the other side. Put my clothes together and opened the suitcase. It is big. There is a livingroom there. They have a kitchen livingroom and one bedroom. I would say you got the wrong number. I will not give it to them. I will not open the door. I might call the police. I will mail it back. Give it to the teacher. Put a bandaid. Keep looking for the library book. I would give it to the police. I would phone the fire department. Go to the neighborhood.

Ask them to phone the police.

Who did it.

Give it back.

Tell them to stop.

They are strong.

It is cold.

Thy rob something.

They kill someone.

It is too hard.

It is smaller.

It is rough.

Do not tell anyone.

She teach me new things.

She help us math.

Do you like school?

It is fun.

It is small.

There is a blue creature.

Do you want to be a teacher?

Do you want to go to college?

I ran out of questions.

How old is your sister?

It is not a doll.

It is a stuffed animal.

I call them my cute cub.

I got dog.

Language Sample Transcript: LLD Case Two

She's kind and helpful. She's mean. Get angry easily. Go around places. Buy a new house. I go to Chinese school. Learn Chinese. Came back at 3:30 and help my dad. We get memorize the words and have dictation. It's a Chinese movie. There's a twin prince. Got mixed-up. They fight the bad guys. Long time I know but forget. Some people going out and found a moon. The moon was dead. No live there. There's a animal. It was an elephant. It was dead. Then went into another moon. Somebody was deep sleep. The master try to scaped last. The god kill him. A bird flew in the house. My dad go open the door.

Clean the tanks.

He's a famous singer.

Jave a Christmas tree and dinner.

I always help my father the most.

My youngers have all sorts of spare time.

Doesn't help.

My father's a carpenter.

I help him in the roof.

We have somewhere chop down trees.

My sister fights.

They fight with the other small ones.

They keep trying to hit the little ones.

Get a nice job.

I like to be myself.

Wouldn't talk to them.

Just hang up.

Call the police.

There's a skytrain to Main Street.

Then walk down.

Pay for it.

Tell somebody.

I don't know.

It's too far to walk.

My feet get tired.

Language Sample Transcript: LLD Case Three

A ninja was fighting this boy.

There is a man in the show.

The man is a police officer and the kid is a karate guy.

The man always catch robbers and the little kid always helps him.

The other school doesn't even let us go to the washroom.

She does not teach us not to mark it.

When you ask her some question she would say go back to your desk.

She is nice.

She helps people alot.

She does not scream at us.

She let us go to the washroom.

I will keep it for college.

Help the family pay their insurance.

My friend invite me to his house and then he ask me to stay over for a night.

We had a time.

He showed me all his videos.

We were playing computer games.

We have a tree.

We decorate our house.

We have lots of friends.

We have turkey dinners.

We go out for dinner.

We invite lots of friends over.

They are nice.

Bird came inside my house.

We had the screen door open.

We had a barbecue out. This bird came in. We did not know. Then we went back to the house. We closed the door. The bird and the cat got trapped. The cat was under my mom's room. The bird was in the flowers. Me and my father were finding something. I was finding something under my mom's room. We saw this cat. Dad was finding something around the plants. Then he found a bird. We opened the door and then we chased it around the house. We did the same thing. They are good. They have good songs. I can't name them all. I forgot.

It's pretty interesting.

Language Sample Transcript: LLD Case Four

I'd buy a lambourgine.

I didn't really get it.

It had Michael J. Fox in it.

Guy has to go back in the future.

He has to stop this guy from shooting him.

Then he has to go in the future to change his kids.

Bring em back.

He explains stuff good.

There was a snake in their house.

Bill Cosby was scared of it.

They had a string.

They were trying to catch it but I forget what happens.

It's football.

You have all the equipment.

It's tackle football.

It's small people playing.

There's coaches.

They're a rock group.

The best song I like is Walk this Way.

They want kids to go to school.

They don't want gangs around.

ACDC is like Highway to Hell.

I watch cartoons till 12:00.

Watch wrestling till 4:00.

I go out and play football.

All my weeks are different.

I have different things.

I call it an arcade cause that's the closes real arcade we have.

They get different ones each week.

I play 1942.

It's a wargame.

You shoot down airplanes.

I like football more.

I don't read books.

I read my own books.

I make books up.

There's this boy.

It's the night before Halloween.

He's on this island.

There's a skeleton.

The skull is all set to kill him.

An axe is coming towards him.

You wake up.

You're floating down the river.

You go back on this island.

This happened now.

It's two pages long.

I wrote a 37 page book.

It's not a book.

It's paper.

Language Sample Transcript: LLD Case Five

It's a situation comedy.

It's about that family.

It's a family.

It has a housekeeper.

The housekeeper has a daughter.

They all live together.

The housekeeper is a man.

He seems to be always solving the problems.

I like comedy shows.

His name's Tony.

Her daughter is Samantha.

She's good in basketball.

He wanted her daughter to join basketball.

She didn't join it.

He was pretty upset.

Later she joined the team.

She had a boyfriend named Todd.

I think they broke up.

Last part was she had a new boyfriend.

He happened to be around a eighteen-year old boy and she was only about thirteen or fourteen.

That was pretty funny.

I would give half to my parents.

Buy some things I would like.

Keep it for my future or maybe for college.

She's really nice.

I guess we were kinda close together.

She was easy to talk to.

She was really nice and helpful.

I liked her.

I go and visit her.

I added everything wrong.

I thought I was going to get four hundred dollars.

My dad goes you added all wrong.

He put it on the wall so everybody culd see.

It was really funny and I was so embarassed.

I was ading one of my cheques.

Everybody was laughing at me.

They've been my favourite group since grade five.

I still like them.

They play rock and roll.

It's not hard and it's not soft.

It's just in the middle.

They broke up but now they're back together again.

They lost one person.

I'm glad they're back together because i like them.

She seems to be getting songs that are kind of normal.

She seems to be dressing up normal.

Most of the times my mom works on the weekends.

She's a nurse.

I end up cleaning the house.

Appendix E

LARSP Summary Sheets

Control Case One

	COMM V	Q	v	N	OTHER I		
	сомм ух	QX	SV 8	AX 4	D N 15	v v	ING 17
			so	V0 3	ADJ N 5	V PART 11	
			SC	vc	NN	INT X	PL 13
			NEG X	OTHER II	PRN5	OTHER II 10	ED 53
	X+S(NP) 3	X+V(VP) 4	X+C(NP)	X+O(NP) 3	X+A(AP) 4		
	COMM VXY	QXY	SVC 10	VCA	D ADJ N 2	PRON-P 49	EN
	LET XY	VS?	SVO 17	VOA 1	ADJ ADJ N	PRON-O 9	
	DO XY		SVA 15	VOI	PRDNB	AUX-M 3	3S 26
			NEG XY	OTHER III 1	COP 15	AUX-O 14 OTHER III 3	GEN
	XY+S(NP) 4	XY+V(VP) 19	XY+C(NP) 4	XY+O(NP) 13	XY+A(AP) 12		
	COMM +S	QVS(+)	SVOA 5	AAXY 3	NP PR NP 4	NEG V 6	N'T 5
		QXYZ	SVCA 2	OTHER IV 3	PR D ADJ N	NEG X	'COP 1
	VXY+	VS+?	SVOI 1		cx	2 AUX 2	'AUX 6
•		TAG	svoc		X C X 1	OTHER IV 3	
ND 7			COORD(1) 7	COORD(1+)	POSTM CL1 2	POSTM CL1+ 1	EST
I LNO			SUBA(1) 3	SUBA (1+)	POSTM PHR1+		ER
UB 11 THER CONN			CL S CL C 1	CL O 14 COMPARATIVE			LY 2
			PASSIVE	HOWI	NP INIT 3	CMPLX VP 2	_ <u> i</u> _
			COMPLEMENT 1	WHAT!	NP COORD		
CONNA						<u></u>	
OMMNT CI	II THERE 4	1					
MPH ORDER	INGUE 4						
	ND 7 ONJ 1 UB 11 THER CONN CONN 3 COMMT CL MPH ORDER	COMM VX X+S(NP) 3 COMM VXY LET XY DO XY XY+S(NP) 4 COMM +S VXY+ ND 7 ONJ 1 UB 11 THER CONN CONN 3 IT THERE 4 MPH ORDER	COMM VX QX X+S(NP) 3 X+V(VP) 4 COMM VXY Q XY LET XY VS? DO XY XY+S(NP) 4 XY+V(VP) 19 COMM +S QVS(+) QXYZ VXY+ VS+? TAG ND 7 ONJ 1 UB 11 ITHER CONN CONN 3 IT SOMMNT CL THERE 4	COMM VX QX SV 8 SO SC NEG X X+S(NP) 3 X+V(VP) 4 X+C(NP) COMM VXY Q XY SVC 10 LET XY VS? SVO 17 DO XY SVA 15 NEG XY SVA 15 NEG XY SVA 15 VY SVA 15 NEG XY SVA 2 VXY+ VS? SVA 2 VXY+ VS? SVOC ND 7 COORD(1) 7 SUBA(1) 3 UB 11 CL 5 CL 5 VTHER CONN CL 5 COMPLEMENT 1 CONN 3 IT SUBANT CL MPH ORDER IT SUBANT CL	COMM VX QX SV 8 AX 4 SO Y0 3 SC VC NEG X OTHER II VC10 VC2 X+S(NP) 3 X+V(VP) 4 X+C(NP) X+O(NP) 3 COMM VXY Q XY SVC 10 VCA LET XY VS? SVO 17 VOA 1 DO XY SVA 15 VOI NEG XY OTHER III 1 XY+S(NP) 4 XY+V(VP) 19 XY+C(NP) 4 XY+O(NP) 13 COMM +S QVS(+) SVOA 5 AAXY 3 QXYZ SVCA 2 OTHER IV 3 VXY+ VS+? SVOI 1 TAG VXY+ VS+? SVOI 1 SUBA(1) 3 SUBA(1+) UB 11 CL 5 CL 0 14 COMPLEMENT 1 WHAT! CONN 3 IT THERE 4 MPH ORDER HOW1	COMM VX QX SV 8 AX 4 D N 15 SG Y0 3 ADJ N 5 SC NN NEG X OTHER II PR N 5 X+S(NP) 3 X+V(VP) 4 X+C(NP) X+O(NP) 3 X+A(AP) 4 COMM VXY Q XY SVC 10 VCA D ADJ N 2 LET XY VS7 SV0 17 VOA 1 ADJ ADJ N DO XY SY SV0 17 VOA 1 ADJ ADJ N XY+S(NP) 4 XY+V(VP) 19 XY+C(NP) 4 XY+O(NP) 13 XY+A(AP) 12 COMM +S QV5(+) SVOA 5 AAXY 3 NP PR NP 4 QXYZ SVCA 2 OTHER III 1 COP 15 VXY+ VS+? SVOI 1 CX CX X VXY+ VS+? SVOI 1 CX X X X X X X X X X X X X X X X X X X X X	COMM VX QX SV 8 AX 4 D N 15 V V SO V03 ADJ N 5 V PART 11 SC VC N N INT X NEG X OTHER II PR N 5 OTHER II 10 X+S(NP) 3 X+V(VP) 4 X+C(NP) X+O(NP) 3 X+A(AP) 4 COMM VXY Q XY SVC 10 VCA D ADJ N 2 PRON-P 49 LET XY VS1 SVO 17 VOA 1 ADJ ADJ N PRON-0 9 DO XY SV1 SVO 17 VOA 1 ADJ ADJ N PRON-0 9 MEG XY OTHER III 1 COP 15 AUX-014 OTHER III 3 XY+S(NP) 4 XY+V(VP) 19 XY+C(NP) 4 XY+0(NP) 13 XY+A(AP) 12 COMM +S QVS(+) SVOA 5 AAXY 3 NP PR NP 4 NEG V 6 QXYZ SVOA 5 CAXY 3 NP PR NP 4 NEG V 6 XX + 3 VXY+ VS+7 SVOC X C X 1 OTHER IV 3 ND 7 COORD(1) 7 COORD(1+) POSTM CL 1 2 <

52 ANALYSED SENTENCES 0 UNINTELLIGIBLE 0 SYMBOLIC NOISE 0 DEVIANT 0 INCOMPLETE 0 AMBIGUOUS 0 MINOR SOCIAL 0 MINOR STEREOTYPES

.

121

.

Control Case Two

0 DEVIANT 0 INCOMPLETE 0 AMBIGUOUS 0 MINOR SOCIAL 0 MINOR STEREOTYPES

.

	CONNCTVY	COMMAND	QUESTION					
STAGE 1		сомм у	Q	v	И	OTHER I		
STAGE II		COMM VX	QX	SV 5	 AX	D N 14	V V 6	ING 5
				SO	VO 3	ADJ N 2	V PART 8	
				SC	vc	NN	INT X 8	PL 11
				NEG X	other 11	PRN7	OTHER II 6	ED 43
		X+S(NP) 3	X+V(VP) 5	X+C(NP)	X+0(NP) 1	X+A(AP)		
STAGE III		COMM VXY	Q XY	SVC 6	VCA	D ADJ N 2	PRON-P 61	EN 3
		LET XY	VS?	SVO 28	VOA	ADJ ADJ N	PRON-O 4	
		DO XY		SVA 19	VOI	PRDN4	AUX-M 3	35 21
				NEG XY	other III	COP 15	AUX-0 9 OTHER III 5	GEN
<u></u>		XY+S(NP) 5	XY+V(VP) 18	XY+C(NP) 4	XY+O(NP) 16	XY+A(AP) 12		
STAGE IV		COMM +S	QVS(+)	SVOA 3	AAXY	NP PR NP 2	NEG V 2	דיא
			QXYZ	SVCA 4	OTHER IV	PR D ADJ N	NEG X	COP
		¥XY+	VS+?	SVOI		сх	2 AUX 1	'AUX
			TAG	SVOC		X C X 2	OTHER IV	
STAGE V	AND 5			COORD(1) 5	COORD(1+) 1	POSTM CL1 3	POSTM CL1+	EST
	CONJ 1			SUBA(1) 5	SUBA (1+)	POSTM PHR1+		ERI
	SUB 9			CL S	CL O 8		•	LY
	OTHER CONN			CL C	COMPARATIVE			
STAGE VI		· · · · · · · · · · · · · · · · · · ·		PASSIVE	HOW	NP INIT	CMPLX VP 2	
				COMPLEMENT	WAAT	M COOLD		
STAGE VII	A CONN 1	IT						
	COMMNT CL EMPH ORDER	THERE 2						
MLU (IN MOR	PHEMES) = 8.18	- <u></u>		······································				
48 ANALYSED 0 UNINTELLIC	SENTENCES SIBLE							

N

122

Control Case Three

	CONNCTVY	COMMAND	QUESTION					•
STAGE I		COMM V	.Q	v	N	OTHER 1		
STAGE II		сомм ух	QX	SV 4	AX 2	D N 22	V V 11	ING 11
				so	V0 1	ADJ N 5	V PART 12	
				SC	VC 1	NN	INT X 3	PL 19
				NEG X	OTHER II	PR N 3	OTHER II 6	ED 32
		X+S(NP) 1	X+V(VP) 4	X+C(NP)	X+0(NP) 1	X+A(AP) 1		· .
	· · · · · · · · · · · · · · · · · · ·							`
STAGE III		COMM VXY	Q XY	SVC 6	VCA	D ADJ N 5	PRON-P 47	EN
		LET XY	VS?	SVO 11	VOA 4	ADJ ADJ N 1	PRON-O 5	
		DO XY		SVA 9	VOI	PR D N 19	AUX-M 3	35 21
			а. 1	NEG XY	OTHER III 1	COP 6	AUX-0 8 OTHER III 5	GEN 1
	· · · · · · · · · · · · · · · · · · ·	XY+S(NP) 8	XY+V(VP) 14	XY+C(NP) 5	XY+O(NP) 12	XY+A(AP) 9	<u> </u>	
STAGE IV	··· · · · · · · · · · · · · · · · · ·	COMM +S	QVS(+)	SVOA 8	AAXY 11	NP PR NP 1	NEG V 3	N'T 2
			QXYZ	SVCA 1	OTHER IV 4	PR D ADJ N	NEG X	'COP 2
		VXY+	VS+?	SVOI 1		C X	2 AUX	'AUX 4
			TAG	SVOC		X C X 6	OTHER IV 4	
STAGE V	AND 10			COORD(1) 13	COORD(1+) 1	POSTM CL1 3	POSTM CL1+	EST
	CONJ 4			SUBA(1) 6	SUBA (1+)	POSTM PHR1+		ER 1
	SUB 6			CL S 1	CLOI			LY
	OTHER CONN			CL C	COMPARATIVE			
STAGE VI			······································	PASSIVE 1	HOW!	NP INIT 6	CMPLX VP 4	
				COMPLEMENT 4	WHAT!	NP COORD 1		
STAGE VII	A CONN 2						<u></u>	- <u></u>
	COMMNT CL	THERE 2						
	EMPH ORDER							

50 ANALYSED SENTENCES 0 UNINTELLIGIBLE 0 SYMBOLIC NOISE 0 DEVIANT 0 INCOMPLETE 0 AMBIGUOUS 0 MINOR SOCIAL 0 MINOR STEREOTYPES

.

Control Case Four

	CONNCTVY	COMMAND	QUESTION	N				
STAGE I		COMM V	Q	v	N	OTHER I		
STAGE II	· · · · ·	сомм ух	QX	SV 6	AX 3	D N 11	VV 1	ING 8
				so	V0 8	ADJ N 3	V PART 6	
				SC	vc	N N	INT X	PL 22
				NEG X	OTHER II	PRN6	OTHER II 8	ED 15
٢		X+S(NP) 1	X+V(VP) 9	X+C(NP)	X+O(NP) 5	X+A(AP) 3		
STAGE III		COMM VXY	Q XY	SVC 10	VCA	D ADJ N 4	PRON-P 40	EN
		LET XY	VS?	SVO 9	VOA 1	ADJ ADJ N	PRON-O 9	
		DO XY		SVA 11	VOI	PRDN8	AUX-M 4	3S 25
		· *		NEG XY .	OTHER III 2	COP 17	AUX-O 8 OTHER III 5	GEN
		XY+S(NP) 3	XY+V(VP) 6	XY+C(NP) 8	XY+O(NP) 6	XY+A(AP) 11	· · · · · · · · · · · · · · · · · · ·	
STAGE IV		COMM +S	QVS(+)	SVOA 5	AAXY 3	NP PR NP 4	NEG V 4	אית 3
			QXYZ	SVCA 3	OTHER IV 4	PR D ADJ N 2	NEG X	'COP 14
		VXY+	VS+?	SVOI 1		сх	2 AUX 1	'AUX 3
			TAG	SVOC		X C X 3	OTHER IV	
STAGE V	AND 2			COORD(1) 4	COORD(1+)	POSTM CL1 7	POSTM CL1+ 2	EST
	CONJ 2			SUBA(1) 6	SUBA (1+)	POSTM PHR1+		ER
	SUB 10			CLS	CL 0 4			LY
	OTHER CONN			CL C 1	COMPARATIVE			
STAGE VI				PASSIVE 1	HOW	NP INIT 4	CMPLX VP 1	
	•			COMPLEMENT 3	WHAT!	NP COORD		
			., 		·			
STAGE VII	A CONN 3	IT						
	COMMNT CL EMPH ORDER	THERE 3						
MLU (IN MOR	(PHEMES) = 7.76							
	,							
52 ANALYSED	SENTENCES			•				

0 UNINTELLIGIBLE 0 SYMBOLIC NOISE 0 DEVLANT 0 INCOMPLETE 0 AMBIGUOUS 0 MINOR SOCIAL

0 MINOR STEREOTYPES

.

124

Control Case Five

	CONNCTVY	COMMAND	QUESTION	I				
STAGE I		COMM V	Q	v	N	OTHER I		
STAGE II		COMM VX	ox	SV 6	AX 1	D N 9	V V 3	ING 14
				so	V0 5	ADJN 6	V PART 14	
				SC	vc	NN	INT X	PL 18
				NEG X	OTHER II	PRN4	OTHER II 10	ED 48
	· · ·	X+S(NP) 3	X+V(VP) 3	X+C(NP)	X+O(NP) 3	X+A(AP) 1	· · · · · · · · · · · · · · · · · · ·	m 7 . <u>m</u>
STAGE III		COMM VXY	Q XY	SVC 11	VCA	D ADJ N 12	PRON-P 63	EN
		LET XY	VS?	SVO 6	VOA 4	ADJ ADJ N	PRON-O 8	
		DO XY		SVA 5	VOI	PRDN9	AUX-M 8	3S 21
				NEG XY	OTHER III	COP 19	AUX-O 10 OTHER III 5	GEN
		XY+S(NP) 6	XY+V(VP) 8	XY+C(NP) 8	XY+O(NP) 2	XY+A(AP) 7		
STAGE IV		COMM +S	QVS(+) 1	SVOA 10	AAXY 3	NP PR NP 12	NEG V 5	N'T 5
			QXYZ	SVCA 4	OTHER IV 5	PR D ADJ N 3	NEG X	COP 1
		VXY+	VS+?	SVOI		сх	2 AUX	'AUX 3
			TAG	SVOC		X C X 3	OTHER IV 6	
STAGE V	AND 8			COORD(1) 9	COORD(1+)	POSTM CL1 12	POSTM CL1+ 3	EST 1
	CONJ	-		SUBA(1) 9	SUBA (1+)	POSTM PHR1+		ER 1
	SUB 11			CL S	CL O 3			LY 4
	OTHER CONN			CLC	COMPARATIVE			
STAGE VI				PASSIVE	HOWI	NP INIT 8	CMPLX VP 9	· ·
				COMPLEMENT	WHAT!	NP COORD		
	· · · · · · · · · · · · · · · · · · ·				·			
STAGE VII	A CONN 2	IT					-	
	COMMNT CL EMPH ORDER	THERE 2						
MLU (IN MOR	PHEMES) = 13.21				<u> </u>	<u> </u>	<u></u>	
41 ANALYSED	SENTENCES							
O UNINTELLIO	GIBLE							
0 SYMBOLIC	NOISE							
0 DEVIANT								
0 INCOMPLET	CE .							

0 MINOR SOCIAL

0 AMBIGUOUS

.

0 MINOR STEREOTYPES

LD Case One

0 DEVIANT

.

0 INCOMPLETE

0 AMBIGUOUS0 MINOR SOCIAL 0 MINOR STEREOTYPES

.

	CONNCTVY	COMMAND	QUESTION					
STAGE I		сомм у	Q	ν	N	OTHER I		
STAGE II		COMM VX	QX	SV 3	AX 2	D N 14	V V 2	ING 2
				so	V0 10	ADJ N 1	V PART 1	
				SC	vc	NN	INT X	PL 3
				NEG X	OTHER II	PRN2	OTHER II 2	ED 7
		X+S(NP)	X+V(VP) 3	X+C(NP)	X+O(NP) 7	X+A(AP) 2		
STAGE III		COMM VXY	Q XY 1	SVC 10	VCA	D ADJ N 5	PRON-P 47	EN
		LET XY	VS?	SVO 13	VOA'3	ADJ ADJ N	PRON-0 7	
		DO XY		SVA 2	VOI	PRDN6	AUX-M 8	35 12
				NEG XY	OTHER III	COP 14	AUX-0 6 OTHER III 2	GEN
		XY+S(NP)	XY+V(VP) 12	XY+C(NP) 4	XY+O(NP) 10	XY+A(AP) 4		
STAGE IV	· · · · · · · · · · · · · · · · · · ·	COMM +S	QVS(+) 1	SVOA 4	AAXY	NP PR NP 2	NEG V S	N'T
	•		QXYZ	SVCA 2	OTHER IV	PR D ADJ N 1	NEG X	'COP
· .		VXY+	VS+? 3	SVOI 3		сх	2 AUX	'AUX
			TAG	SVOC		XCX2	OTHER IV 1	
STAGE V	AND 1		~	COORD(1) 1	COORD(1+)	POSTM CL1	POSTM CL1+	EST
	CONJ			SUBA(1)	SUBA (1+)	POSTM PHR1+		ER 1
	SUB			CLS	CL O 3		•	LY
	OTHER CONN			CLC	COMPARATIVE			
STAGE VI				PASSIVE	HOWI	NP INIT 1	CMPLX VP 1	
				COMPLEMENT	WHAT!	NP COORD 1		
STAGE VII	A CONN	IT						
	COMMNT CL EMPH ORDER	THERE 2				~		
MLU (IN MOR	PHEMES) = 5.07							
54 ANALYSED 0 UNINTELLIC 0 SYMBOLIC I	SENTENCES SIBLE NOISE				· ·			

126

LD Case Two

	CONNCTVY	COMMAND	QUESTION					
STAGE I		COMM V	Q	v	N	OTHER I		
STAGE II		COMM VX	QX	SV 3	AX 5	D N 18	V V 3	ING 2
			-	so	V0 12	ADJ N 3	V PART 7	
				SC	vc	NN	INT X 2	PL 9
				NEG X	other 11	PR N 3	OTHER II 7	ED 12
		X+S(NP) 3	X+V(VP) 6	X+C(NP)	X+O(NP) 7	X+A(AP) 3		
STAGE III		COMM VXY	Q XY	SVC 13	VCA 1	D ADJ N 6	PRON-P 21	EN
		LET XY	VS?	SVO B	VOA .	ADJ ADJ N	PRON-0 7	
		DO XY		SVA 4	VOI	PRDN3	AUX-M 4	35 15
				NEG XY	OTHER III 2	COP 13	AUX-O 4 OTHER III 2	GEN
		XY+S(NP) 8	XY+V(VP) 4	XY+C(NP) 9	XY+0(NP) 6	XY+A(AP) 5		
STAGE IV		COMM +S	QVS(+)	SVOA 2	AAXY 1	NP PR NP 1	NEG V 3	N'T 3
			QXYZ	SVCA	OTHER IV 1	PR D ADJ N	NEG X 1	'COP 9
		VXY+	VS+?	SVOI		cx	2 AUX	-AUX
			TAG	SVOC		XCX2	OTHER IV I	
STAGE V	AND 3			COORD(1) 4	COORD(1+)	POSTM CLI	POSTM CL1+	EST 1
	CONJ 1			SUBA(1) 2	SUBA (1+)	POSTM PHR1+		ER IVI
	SUB			CLS	CL O 1		¥	
	OTHER CONN			CLC	COMPARATIVE			
STAGE VI		· · · · · · · ·		PASSIVE COMPLEMENT 2	HOW! WHAT!	NP INIT 1 NP COORD	CMPLX VP 3	
			·					
STAGE VII	A CONN 1 COMMNT CL EMPH ORDER	IT THERE 3					· .	
MLU (IN MOR	1PHEMES) = 5.36							
52 ANALYSED	SENTENCES							
0 UNINTELLIGIBLE		*						
0 SYMBOLIC	NOISE							
0 DEVIANT								
0 INCOMPLET	re							
8 AMBIGUOU	S							
U MINOR SOC								
U MINOR STE	REOTYPES							

١

127

LD Case Three

	CONNCTVY	COMMAND	QUESTION					
STAGE I		сомм у	Q	v .	N	OTHER I		
STAGE II		COMM VX	QX	SV 4	AX	D N 24	V V	ING 5
				50	VU 2	ADJ N 1	V PART 2	D7 10
				SC NDC Y		N N	INT X Z	PL 10
			·	NEG X	01HER 11	PRN 1		ED 24
		X+S(NP) 2	X+V(VP) 4	X+C(NP)	X+O(NP) 1	X+A(AP)		
STAGE III		COMM VXY 1	QXY	SVC 6	VCA	D ADJ N 3	PRON-P 46	EN
		LET XY	VS?	SVO 19	VOA	ADJ ADJ N	PRON-O 4	
		DO XY		SVA 8	VOI	PR D N 10	AUX-M 4	3S 15
<u></u>				NEG XY	OTHER III	COP 9	*UX-0 9	GEN
							OTHER III S	
		XY+S(NP) 8	XY+V(VP) 8	XY+C(NP) 3	XY+0(NP) 17	XY+A(AP) 9		
STAGE IV		COMM +S	QVS(+)	SVOA 11	AAXY	NP PR NP	NEG V 5	דיא 1
			QXYZ	SVCA 1	OTHER IV 3	PR D ADJ N 2	NEG X	'COP
		VXY+	VS+?	SVOI 2		сx	1 AUX	'AUX
			TAG	SVOC		XCX2	OTHER IV 1	
STAGE V	AND 4	<u> </u>		COORD(1) 4	COORD(1+)	POSTM CL1	POSTM CL1+	EST
	CONJ			SUBA(1)	SUBA (1+)	POSTM PHR1+		ER
	SUB 1			CL S	CL O 7			LY
	OTHER CONN			CLC	COMPARATIVE			
STAGE VI				PASSIVE	HOWI	NP INIT 4	CMPLX VP	
				COMPLEMENT	WHAT!	NP COORD		
STAGE VII	A CONN 4	IT						<u></u>
	COMMNT CL	THERE 1						
	EMPH ORDER							
MLU (IN MOR	rnemes) = 7.36							

47 ANALYSED SENTENCES 0 UNINTELLIGIBLE 0 SYMBOLIC NOISE 0 DEVIANT 0 INCOMPLETE 0 AMBIGUOUS 0 MINOR SOCIAL

0 MINOR STEREOTYPES

LD Case Four

		COMMAND	QUESTION					
STAGE I		сомм у	Q	ν	N	OTHER I	·	
STAGE II		COMM VX	QX	SV 4	AX	D N 13	V V 5	ING 5
				so	V0 2	ADJ N 4	V PART 4	
				SC	vc	N N	INT X 1	PL 13
				NEG X	OTHER II	PRN4	OTHER II 4	ED 9
		X+S(NP)	X+V(VP) 3	X+C(NP)	X+O(NP) 1	X+A(AP)	<u> </u>	, ,, ,, ,, ,, ,, ,, ,, <u>,, ,, ,, ,, ,, ,</u>
STACE III			0.77	SVC 16	NCA	D ADIN 3	PRON-P 45	
JINGS III		LET VV	4.4.1	SVC 16	VCA		PRON-O 8	211
		DOXY	¥5!	34011	VOA Z			26.32
		DUXI		SVA 6		PR D N B	AUX-M 2	33 43 OTN
				NEG XY	OTHER III	COPB	OTHER III 2	GEN
		XY+S(NP) 3	XY+V(VP) 13	XY+C(NP) 9	XY+O(NP) 10	XY+A(AP) 6		
STAGE IV		COMM +S	QVS(+)	SVOA 6	AAXY 3	NP PR NP 1	NEG V 4	N'T 3
			QXYZ	SVCA 1	OTHER IV	PR D ADJ N	NEG X	COP 14
		VXY+	VS+?	SVOI		сх	2 AUX	AUX 2
			TAG	SVOC 1		XCX1	OTHER IV 2	
STAGE V	AND			COORD(1) 2	COORD(1+)	POSTM CL1 2	POSTM CL1+	EST 2
	CONJ 2			SUBA(1)	SUBA (1+)	POSTM PHR1+		ER 1
	SUB			CL S	CL O 3			LY
	OTHER CONN			CL C	COMPARATIVE			
STAGE VI				PASSIVE COMPLEMENT 1	HOW! WHAT!	NP INIT 2 NP COORD 1	CMPLX VP 2	
					·			<u> </u>
STAGE VII	COMMNT CL EMPH ORDER	THERE 4						
MLU (IN MOR	RPHEMES) = 6.23		·	<u> </u>				· · ·
52 ANALYSEI	SENTENCES							
O UNINTELLI	GIBLE							
0 SYMBOLIC	NOISE							
0 DEVIANT								
6 INCOMPLET	TE					i		

0 AMBIGUOUS

0 MINOR SOCIAL

.

0 MINOR STEREOTYPES

LD Case Five

	CONNCTVY	COMMAND	QUESTION	I				
STAGE I		сомм у	Q	v	N	OTHER I		
STAGE II	. '	COMM VX	QX	SV 4	AX	D N 14	V V 2	ING 7
				SO	V0 1	ADJ N 4	V PART 5	
				SC	vc	NN	INT X 6	PL 9
	- ,			NEG X	OTHER II	PRN1	OTHER II 2	ED 26
		X+S(NP)	X+V(VP) 4	X+C(NP)	X+0(NP) 1	X+A(AP)		
STAGE III		COMM VXY	Q XY	SVC 16	VCA	D ADJ N 3	PRON-P 60	EN 1
		LET XY	VS?	SVO 23	VOA 1	ADJ ADJ N	PRON-0 5	
		DO XY		SVA 9	VOI	PRDN7	AUX-M 5	3S 31
				NEG XY	OTHER III 1	COP 25	AUX-0 7 OTHER III 3	GEN
		XY+S(NP) 8	XY+V(VP) 9	XY+C(NP) 11	XY+O(NP) 14	XY+A(AP) 7		
STAGE IV		COMM +S	QVS(+)	SVOA 6	AAXY 4	NP PR NP 2	NEG V 3	N'T 1
			QXYZ	SVCA 2	OTHER IV	PR D ADJ N 1	NEG X	'COP 12
		VXY+	VS+?	SVOI		сх	2 AUX 1	'AUX 2
			TAG	SVOC		XCX4	OTHER IV	
STAGE V	AND 3	· ·	· · · ·	COORD(1) 6	COORD(1+)	POSTM CLI 3	POSTM CL1+	EST
	CONJ 2			SUBA(1) 2	SUBA (1+)	POSTM PHR1+		ER .
	SUB 2			CLS	.CL 0 7			LY
	OTHER CONN			CLC	COMPARATIVE			
STAGE VI		· · · · · · · · · · · · · · · · · · ·		PASSIVE COMPLEMENT 2	HOW! WHAT!	NP INIT 1 NP COORD	CMPLX VP 2	
STAGE VII	A CONN 1 COMMNT CL EMPH ORDER	IT There						
MLU (IN MOR	PHEMES) = 7.3	<u>, in an a</u>					Ammen -	
53 ANALYSED	SENTENCES							
0 UNINTELLIC	JIBLE							
0 SYMBOLIC	NOISE							
0 DEVIANT								
0 INCOMPLET	г г .							
0 AMBIGHOU	5							
0 MINOR SOC				. ·				
	** ***							

Appendix F

Computational Procedure for Wilcoxon Rank Sum Tests

Computational Procedure for Wilcoxon Rank Sum Tests

- 1. Proceeding from smallest to largest, ranks were assigned to each case in both groups.
- 2. When ties occurred, each case was assigned the average of the ranks it would occupy if no ties had occurred.
- 3. The sum of ranks (R₁) was calculated for each group at phrase and clause level for Stages 2, 3, 4, and 5.
- 4. $\overline{\mathbf{R}}$ was calculated for both groups:

$$Mean = \overline{R} = \underline{N(N + N + 1)}_2$$

- 5. R_1 for each group was compared to \overline{R} . If less than \overline{R} , R_1 was compared to the critical values required for significance. The critical lower tail values of R_1 for 5 and 5 cases are 19 ($\alpha = .05$) and 16 ($\alpha = .01$).
- 6. If R_1 exceeded \overline{R} , the corresponding lower tail value was obtained as follows: $2\overline{R} R_1$. This result was then compared to the critical lower tail values indicated above.