Assessing Language Progress in Children with Autism:
A Comparison of Two Approaches

Susan Margaret Fawcett
B.Sc., University of Victoria, 1997

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

in
THE FACULTY OF GRADUATE STUDIES
(School of Audiology and Speech Sciences)

We accept this thesis as conforming to the required standard

UNIVERSITY OF BRITISH COLUMBIA
May, 2004
© Susan M. Fawcett, 2004
Library Authorization

In presenting this thesis in partial fulfillment 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.

Susan Fawcett
Name of Author (please print) 16/04/04
Date (dd/mm/yyyy)

Title of Thesis: Assessing Language Progress in children with Autism: A Comparison of Two Approaches

Degree: M.Sc. (Speech-Language Pathology) Year: 2004
Department of Audiology & Speech Sciences
The University of British Columbia
Vancouver, BC Canada
In both the research and clinical arenas, the measurement of children’s progress in language therapy is commonly completed by the administration of formal, norm-referenced tests. This is despite the fact that more naturalistic or informal measures, such as language sample analyses, may be better suited to this assessment task. This issue may be particularly relevant in the assessment of young children with autism due to their unique set of communication challenges. The question addressed in the present study, then, is whether the data resulting from both types of measures, formal and informal, provides convergent information about the progress made in several areas of expressive communication by seven children with autism. Seven children with autism were assessed using both types of tools at two successive data points, and the information gleaned from each type of assessment was then compared. It was found that, in the majority of cases, the two data sets did not converge on the type of information they provided about how the children progressed as a result of therapy. It appeared as though this was due to two factors: child factors, such as not having the prerequisite test-taking abilities to earn scores on the formal measures, and assessment tool factors, such as some language domains not being addressed adequately by formal tests.
# TABLE OF CONTENTS

**ABSTRACT**

**TABLE OF CONTENTS**

**LIST OF TABLES**

**ACKNOWLEDGEMENTS**

## CHAPTER 1: INTRODUCTION

1.1 Overview

1.2 Types of Expressive Language Assessment Used in the Present Study

1.2.1 Formal assessment

1.2.2 Informal assessment

1.2.3 Empirical Measurement of Progress Using Formal and Informal Assessments

1.3 Assessment Challenges Specific to Autism

1.3.1 Brief Summary of Typical Language Characteristics in Children with Autism

1.3.2 Behavioural Challenges in Formal Testing of Children with Autism

1.4 Current Research Question

## CHAPTER 2: METHODS

2.1 Overview

2.1.1 General Description of Current Study

2.1.2 Description of the Early Intensive Behavioural Intervention Evaluation Project

2.2 Participants

2.2.1 Description of Participants

2.2.2 Communication Characteristics of Participants

2.3 Procedures - Communication Assessment

2.3.1 Overview of Procedures

2.3.2 Formal Measures Used in the Current Study

2.3.3 Informal Measures Used in the Current Study

2.3.3.1 Dependent Variables for All Children: Summary of Formal and Informal Measures

2.3.3.2 Dependent Variables for Preverbal Children: Communication Sample Analyses

2.3.3.3 Dependent Variables for Verbal Children: Language Sample Analyses

2.4 Procedures - Analysis Strategy

2.5 Summary of Methods
CHAPTER 3: RESULTS

3.1 Overview 69
3.2 Preverbal Children 69
  3.2.1 EMMETT 69
    3.2.1.1 Time 1 results 69
    3.2.1.2 Time 2 results 72
    3.2.1.3 Comparison summary: Formal and informal measures 74
  3.2.2 JACOB 78
    3.2.2.1 Time 1 results 78
    3.2.2.2 Time 2 results 79
    3.2.2.3 Comparison summary: Formal and informal measures 82
  3.2.3 WESLEY 85
    3.2.3.1 Time 1 results 85
    3.2.3.2 Time 2 results 87
    3.2.3.3 Comparison summary: Formal and informal measures 90
  3.2.4 Brief Summary: Preverbal Children 92
3.3 Verbal Children 93
  3.3.1 GRAHAM 93
    3.3.1.1 Time 1 results 93
    3.3.1.2 Time 2 results 95
    3.3.1.3 Comparison summary: Formal and informal measures 97
  3.3.2 CHRIS 99
    3.3.2.1 Time 1 results 99
    3.3.2.2 Time 2 results 101
    3.3.2.3 Comparison summary: Formal and informal measures 102
  3.3.3 BRANDON 104
    3.3.3.1 Time 1 results 104
    3.3.3.2 Time 2 results 106
    3.3.3.3 Comparison summary: Formal and informal measures 109
  3.3.4 ALEX 110
    3.3.4.1 Time 1 results 110
    3.3.4.2 Time 2 results 112
    3.3.4.3 Comparison summary: Formal and informal measures 113
  3.3.5 Brief Summary: Verbal Children 115
3.4 Therapy Goal Results 116
3.5 Summary of Results 117

CHAPTER 4: DISCUSSION

4.1 Summary of Results 119
  4.1.1 Brief Summary of Findings 119
  4.1.2 Comparison with Other Studies 120
  4.1.3 Discussion Overview 121
4.2 Methodological Issues 121
  4.2.1 Range and Selection of Variables 122
  4.2.2 Coding Reliability 123
LIST OF TABLES

Table 2.1: Dates for EIBI testing and language sample collection expressed as cumulative months-in-program 35
Table 2.2: Contexts of each Child’s Videotaped Communication Sample – T1 41
Table 2.3: Contexts of each Child’s Videotaped Communication Sample – T2 42
Table 2.4: Dependent Variables for All Children – Formal and Informal Measures 44
Table 2.5: Dependent Variables for each Preverbal Child 45
Table 2.6: Dependent Variables for each Verbal Child 56
Table 3.1a: Standard Scores on Formal Measures for Emmett 76
Table 3.1b: Communication Sample Analysis Results for Emmett 77
Table 3.2a: Standard Scores on Formal Measures for Jacob 83
Table 3.2b: Communication Sample Analysis Results for Jacob 84
Table 3.3a: Standard Scores on Formal Measures for Wesley 91
Table 3.3b: Communication Sample Analysis Results for Wesley 92
Table 3.4a: Standard Scores on Formal Measures for Graham 98
Table 3.4b: Language Sample Analysis Results for Graham 99
Table 3.5a: Standard Scores on Formal Measures for Chris 103
Table 3.5b: Language Sample Analysis Results for Chris 104
Table 3.6a: Standard Scores on Formal Measures for Brandon 109
Table 3.6b: Language Sample Analysis Results for Brandon 110
Table 3.7a: Standard Scores on Formal Measures for Alex 114
Table 3.7b: Language Sample Analysis Results for Alex 115
Table 3.8a: Summary of Results – Do formal and informal expressive language measures converge on the presence / direction of progress? 118
I would like to express my sincere gratitude to everyone who supported me throughout this project, including my expert committee members, family and friends.

Many thanks to all the hard-working research assistants involved with this project: Alison Louie, Deb Vigh, Lynette Winsor, and Kathleen Jackson. Special thanks to Sherri Millham for all her coding expertise and support. I also extend thanks to Donna Seedorf-Harmuth and her staff at Queen Alexandra Hospital for their enthusiastic support of this endeavor.

Thank you to Barb Bernhardt for her support, knowledge, insightfulness and editorial expertise.

I am very grateful for Pat Mirenda, who generously shared not only her data, but also her passion for this project with me.

A very special thank you to Judith Johnston for everything she generously gave me to make completion of this project possible. I am so grateful for her caring support, her infectious enthusiasm for research and this project in particular, and for all that I have learned from her.

Finally, to my partner, Nitsa, who supported me in so many ways through this process, I extend my gratitude for her unwavering patience and love, and for being such a great source of inspiration. I could not have done this without you.
1. INTRODUCTION

A recent survey revealed that many speech-language pathologists working with children use standardized tests as the method of choice for a variety of assessment purposes, including diagnosis of language impairment, establishment of treatment goals, and the measurement of progress in therapy (Kerr et al., 2003). This is despite the fact that such tests may not be equally well suited to all of these language assessment tasks.

Assessment of progress in therapy is one of the tasks that is questionably addressed by standardized testing. Lund and Duchan (1993), among others (McCauley & Swisher, 1984a; Paul, 1995), argue that more informal types of procedures are better able to tackle this assessment question. The current study explores this issue further by comparing two approaches to the assessment of therapeutic progress in the expressive language and communication of young children with autism.

1.1 Overview

In this chapter, the two types of assessment methodologies, “formal” and “informal,” employed in this study will first be defined and described. Positive aspects to the usage of both specifically in the measurement of progress in therapy will be outlined, as will the problems inherent in their usage. Empirical examples of the usage of both formal and informal procedures in the assessment tasks of documenting change in individual therapy, as well as in program evaluation, will be provided. Finally, the language characteristics of children with autism will be summarized, highlighting the aspects of their communication that are most problematic, and therefore most important to include in a thorough assessment of these children.
1.2 Types of Expressive Language Assessment Used in the Present Study

The present study focuses on language assessment procedures of two sorts: "formal" and "informal." "Formal" assessment here applies to the family of assessments that are characterized by both the use of standardized elicitation tasks and the use of normative scores. These assessments will be contrasted with the set of "informal" procedures that make use of naturalistic behaviour samples and descriptive analyses. "Informal" in this context is not used in the sense of "casual," since both types of assessment and their analysis phases are pre-planned.

These two approaches, thus defined, represent the poles of a continuum of assessment procedures. It must be acknowledged from the outset that there are other assessment approaches between these two extremes. In this middle ground, one finds methods for comparing naturalistic behaviours to norms and for eliciting behaviours that are then used descriptively. However, the contrast between these extremes will bring the key characteristics of major assessment approaches into better focus.

1.2.1 Formal assessment

The prototypical formal language assessment measure is a standardized test. Tests, while varying widely in the aspects of language they measure, have several characteristics in common. They typically measure a range of language skills with a few items that are administered with the same instructions and materials across all children, and they provide the examiner with explicit rules for scoring the resultant behaviour (Tomblin, 2000). Tests are typically administered to a large sample of children in order
to derive the norms used for comparison for individuals who have been given the test, hence the alternate term *norm-referenced test* (Lund & Duchan, 1993).

Once administered to an individual, her or his raw scores are converted via data from the norming sample into standard scores, percentiles and/or age-equivalence scores. This comparison of a particular child to her or his peers supports judgments about the presence or absence of developmental delays, and is the most widely accepted use for tests (McCauley and Swisher, 1984a & 1984b). Indeed, identification of language "impairment" or "normalcy" is the task that standardized tests are designed to perform (McCauley and Swisher, 1984a & 1984b).

As documented in the Kerr study cited earlier, standardized tests tend also to be used for assessment purposes other than the one they are best suited to address. Speech-language pathologists working with elementary school-aged children designated standardized tests as their most important decision-making tool for diagnosis, description of language systems, and establishment of therapy goals. Tests were rated as the second most important decision-making tool for measuring progress and for screening, being rated only somewhat less important than anecdotal evidence. Fey et al. (1993), as well as Nye et al. (1987) document similar usage patterns and additionally the use of tests in program evaluation. Other uses include measuring progress in therapy or program evaluation as well as profiling of language skills.

Several probable reasons exist for the widespread usage of norm-referenced tests. First, perhaps due to their quantifiable nature, these tests are widely assumed to be an objective, reliable, and valid means of measuring language ability (Danwitz, 1981; Ukrainetz McFadden, 1996). While standardized administration procedures do result in
well-constructed tests having high repeatability or reliability, validity is more questionable. Validity, or the ability of a test to measure what it claims to measure, is difficult to ascertain and is a somewhat subjective concept (Paul, 1995). This is particularly true for the subscales of language tests, which often require broader competencies than their names imply. For example, the grammatical subtest of the 4th edition of the Utah Test of Language Development correlates most highly with the vocabulary portions of the Test of Language Development (Johnston, in press).

A related characteristic of tests is that they typically sample a wide variety of language skills but devote only a few items to each. As a result, scores derived from tests may not indicate much more than a general reflection of language functioning (Ukrainetz McFadden, 1996). For example, on the Preschool Language Scale, a single number represents the entire area of expressive language skills, including lexical semantics, morphology and syntax (Zimmerman, Steiner, & Pond, 1992). A score of 80 on this test does not, in and of itself, give any indication of the child’s language production strengths and needs. Mecham (2003) estimates that “Even on the best of norm-referenced tests, any single item contains a minimum of 34% of the content or trait being measured by the entire test, and about 64% of some content or trait in which the test is not interested in measuring” (p.2). The author also reminds us that “norm-referenced tests automatically omit items that are passed by an entire normative group, yet these items … should be of interest … when planning intervention goals” (p.2).

A second reason for the ubiquitous use of tests stems from their ease of administration, or the amount of time or preparation required for clinicians to complete them. It has been found that the most commonly used norm-referenced language tests
among clinicians are ones that are reportedly relatively simple to administer, for example, the Peabody Picture Vocabulary Test and the Expressive One Word Picture Vocabulary Test (Huang, Hopkins, & Nippold, 1997).

To summarize, the recent studies make it clear that clinicians use tests for a wide range of purposes, on the grounds that they are objective and efficient. While they certainly do provide a convenient and efficient means of assessment, their objectivity may be over-rated by clinicians.

Clinicians are also surprisingly tolerant of several well-documented problems with tests. The discussion here of these problems will be limited to those that are directly relevant to the present study, although other limitations of standardized testing have been discussed in the literature (see Danwitz, 1981; Kerr et al., 2003; McCauley & Swisher, 1984a, b; Paul, 1995 for further information). The two main limitations that will be addressed here are the use of standardized testing to measure children’s progress in therapy and the use of tests to assess the language behaviour of children with autism.

McCauley and Swisher make a convincing argument about the shortcomings of norm-referenced tests when used for the purpose of measuring progress in therapy (1984b). They state that the assessment of progress as a result of a treatment program must be sensitive enough to target the specific language behaviours that the therapy was designed to improve. However, clinicians may assume that tests or other assessment tools that are appropriate for comparison of individuals and subsequent identification of language impairment are equally appropriate for the assessment of change produced by a therapeutic program. The assessment tool used for measuring progress should provide a finely detailed, descriptive portrait of the child’s communicative skills (Paul, 1995). This
is the same descriptive approach to assessment that is necessary when setting or revising goals for therapy.

Using norm-referenced tests for the purpose of assessment of behavioural change may result in either underestimation or overestimation of a child’s abilities (McCauley & Swisher, 1984b). This practice may result in underestimation of skills due to the nature of the items on a test. As noted earlier, the items are designed to compare individuals within a developmental level, and so must examine broad, stable language characteristics. An example of such broad language production characteristics is given by the following item taken from the Preschool Language Scale - 3: a child “produces basic sentences” consisting of 4 or 5 words when prompted to “Tell me about your pets” (Zimmerman, et al., 1992, p.4 of test manual). From a developmental point of view, a wide range of sentence types could meet this criterion (e.g., “My dog has fleas” vs. “Dogs try to fool people”). With such a broad characterization, a child could make gains in language abilities that are not captured. A final example of possible underestimation of language abilities as portrayed by test scores stems from the fact that a child must accelerate her or his learning as she or he increases in chronological age in order to just maintain the same standard score. If a child is language delayed yet learning at a consistent, steady pace, his standard test score will still decrease over time.

Standardized testing to assess progress over time may also inflate a child’s performance. Gain scores, small or large, may be due to chance or measurement error. However, clinicians may interpret these gain scores as the child having made progress, even when this “progress” is not evident in the child’s daily activities or therapy sessions. Secondly, a child may learn the test items with repeated administrations, or the clinician
may "teach to the test," both resulting in an overestimation of the child’s capabilities (McCauley & Swisher, 1984b).

As well as their potential for leading clinicians to over or underestimate a child’s language skills, the fact that tests measure children only within a particular age range presents further barriers to their use in estimating progress. For example, the Preschool Language Scale-3 (Zimmerman et al., 1992) measures only children between birth and age 6 years, 11 months. A clinician would not then be able to test a child repeatedly with this measure beyond the age of 7 years, and could not directly compare the scores from this test with the scores from a different norm-referenced test. Related to this problem is that children who are functioning at a prelinguistic level of development will achieve scores at floor level on many standardized language tests (i.e. the child is not able to perform even the first item on a test, or is not able to achieve a baseline score). Furthermore, some of these children, particularly those with developmental disabilities, will still score at floor levels even months later if they remain prelinguistic. A child who achieves floor scores on successive testing occasions may or may not have made gains in communication and/or language abilities during therapy.

An equivalent error in judgment occurs when a clinician or researcher compares test scores at floor level of a child beginning a therapy program to the same child who later is able to achieve higher scores on the same test. This is because it is impossible to determine the language level of the child at the baseline if they could not be assigned a score at that time. An example of this may be seen in children whose behaviour, rather than their language level, prevents them from being able to complete a testing session.
1.2.2 Informal assessment

The key characteristic that distinguishes informal types of assessment is their use of naturalistic rather than elicited behaviour. While standardized tests are often completely decontextualized, informal assessments allow the context in which a child's spontaneous language occurs to play a central role in the interpretation of the child's abilities. Informal assessment methods include behavioural observations and language or communication sample analysis (Paul, 1995). Unlike formal tests, which answer the question of "How does this child compare to her or his peers?," more naturalistic measures answer the question, "What are this child's language abilities?." In other words, while formal test scores answer a normative question, informal measure results answer a descriptive question (Johnston, 2001; Paul, 1995).

For the present study, communication/language sample analysis was the informal assessment procedure of choice. This type of assessment involves collecting a representative sample of a child's communicative behaviour in a naturalistic setting, and then examining the resulting sample for patterns of abilities and weaknesses in the child's language (Lund & Duchan, 1993). The sample is ideally collected in a setting where the child is engaged in play with a familiar other. When obtaining a sample, the clinician or researcher must attempt to ensure that the sample is highly representative of the child's typical communicative behaviour. Assessing the child in a familiar environment with a person she or he is used to interacting with (such as a mother or a clinician) are ways to ensure this representativeness.

All areas of precommunicative behaviours, communication and language can be examined using this method, and the ways of doing so are limited only by the clinician's
imagination and her or his knowledge of descriptive frameworks in various domains. Numerous coding schemes can be found in the literature and can be used as guidelines (Coggins & Carpenter, 1981; McCathren et al., 1996; Miller, 1981; Paul, 1995; Sing, 2000). Language sample analysis allows the clinician to take a close look at language behaviours that are not examined via traditional standardized tests, at least not easily, as well as those that are. Areas that are best looked at in an interactional, spontaneous context include the functions of language and conversational behaviours (Leonard & Weiss, 1983). Frequencies and regularities of the child’s language forms can be summarized. For example, counts of morphosyntactic forms, as well as a syntactic complexity measure (mean length of utterance) can be calculated. The child’s production of language content can also be explored by looking at the vocabulary and semantic categories used. Finally, the child’s patterns of the use of language and communication within a spontaneous context can be analyzed (in other words, the area of pragmatics), often through examining such phenomena as turn-taking behaviours and topic maintenance (Lund & Duchan, 1993; Stockman, 1996).

This approach to assessment is used both clinically and in the research literature, and can be seen to have strengths and limitations. The nature of language sampling techniques makes them particularly suitable for addressing specific assessment goals, such as detailed description of language abilities and how these abilities change over time. In the same way that standardized tests are designed to answer normative questions, informal measures, including those involving language and communication sample analysis, are designed to answer descriptive questions.
This was shown in a study by Wittman (1998), who compared formal tests to informal measures on their ability to detect the degree, range and complexity of a young child’s language impairment. Traditional standardized language tests such as the Clinical Evaluation of Language Fundamentals – Preschool (Wiig et al., 1992) did not detect the extent of the child’s language production problems in either the areas of form or content. The tests overall showed that the child was either normal or had only mild difficulties, and were therefore concluded to have a lack of sensitivity to both detect and describe a child who exhibited significant difficulties in her everyday communicative interactions. Informal assessments in the form of observational checklists, however, not only described her difficulties well, but also identified for the author those environments in which the child was having the most difficulty. This aspect of the results gave much more useful information for the purposes of treatment planning and progress tracking. Echoing these results is the case of a different child, whose numerous expressive morphosyntactic difficulties were not detected by formal measures, such as the Test of Language Development – Primary, the Clinical Evaluation of Language Fundamentals – Revised, and the Expressive One Word Picture Vocabulary Test (Semel, Wiig, & Secord, 1987; Hamill & Newcomer, 1990; Gardner, 1990). However, these difficulties were revealed in an analysis of the child’s spontaneous language sample (Schuele, 1992). Bopp likewise captured progress in intervention in a group of 9 children with phonological disorders using spontaneous speech samples, although her study did not make use of standardized measures (1995).

Finally, a large sample study is reported by Aram, Morris, and Hall (1993), involving data from 252 children with specific language impairment. Although all of
these children were enrolled in treatment programs and were judged by parents and/or clinicians to have significant language deficits, these problems were not captured by the standardized tests. The authors identify pragmatic abilities and text formulation as two of the areas particularly overlooked by tests. Ultimately it was again language sample data that proved most effective in identifying children with language difficulty (Dunn et al., 1996). Thus, even in the arena in which tests are usually thought of as effective, the descriptive capabilities of informal measures have provided an even better picture of language competencies. It is exactly the descriptive strengths of language sample analysis that are the key to measuring progress in therapy.

Since the process of language sample analysis allows the clinician to examine behaviours in which the child is engaged, regardless of whether they are precommunicative or fully linguistic in nature, there is no problem of scores at floor level with this type of assessment. That is, the clinician can glean some useful information about the child's level of communicative functioning with this methodology, rather than summarizing the child as "untestable." While there are some newer standardized measures that attempt to elicit nonverbal communication (such as the Communication and Symbolic Behavioral Scales; Prizant & Wetherby, 2002), these remain limited by the elicited nature of the behaviour. Some communicative acts are inherently spontaneous in function, and therefore cannot be elicited. It would seem, then, that communication sample analysis measures are a better choice of assessment tool for nonverbal children than norm-referenced tests.

McCauley and Swisher (1984b) also raised concerns about the possibility of overestimation of language skills with formal tests. Again, this is a non-issue with
informal measures, as there are no items for a clinician to teach to the child, or for the child to learn over repeated assessments. Another advantage of language sample analysis is that it does not impose age-range constraints. A clinician or researcher could presumably assess a child via this method from a very young age through to adulthood, keeping in mind that the setting of the analysis might change from free-play on the floor to a natural conversation with a second adult communicator. The fundamental nature of the analysis stays the same: the examination of how a communicator uses language in a naturalistic setting. The final, and perhaps most important strength of language sample analysis over tests concerns the validity of the tool. While the validity of standardized language tests is questionable, analysis of carefully collected communication samples is the epitome of a valid measure of language production. The procedure, at its best, is describing a particular language skill in the exact context in which it naturally occurs.

Although language sample analysis appears to be an excellent method for the assessment tasks of setting therapy goals and measuring progress, it has been criticized by both researchers and clinicians within the scientific literature. One set of concerns has centred on the transcription process, in other words, the challenges of producing accurate and reliable transcriptions. The experience of the Language Analysis Laboratory at the University of Wisconsin is instructive at this point. Jon Miller (Leadholm & Miller, 1992) and his associates have demonstrated that by creating appropriate transcription systems and standards, many if not most of these difficulties can be circumvented.

Another criticism of language sample analysis within clinical circles involves its presumed heavy demand on time resources (Danwitz, 1981). The time factor was cited as the most significant reason for not completing language sample analyses among
Canadian clinicians, even though it was viewed as a valuable and useful assessment tool (Kerr et al., 2003). This same result came from a survey of American speech-language pathologists, which found that 86% of the S-LPs not using language sample analysis claimed it was due to a lack of time (Kemp & Klee, 1997). However, the labour-intensive processes of language sample analysis can be lessened by the use of computerized analysis programs such as Systematic Analysis of Language Transcripts (Miller & Chapman, 1993). Unfortunately, the use of such programs is decidedly less than widespread: only 8% of American clinicians use them (Kemp & Klee, 1997).

A concern regarding language sample assessment to measure change over time is the difficulty involved in creating environmental contexts that are truly the same from one sampling point to the next. The very lack of standardized elicitation contexts that make naturalistic procedures desirable for assessing language in use also creates a problem for direct comparison of data across points in time. Although it is admittedly difficult to keep these conditions the same, attempts to do so are made, such as having the same person interact with the child in the same place, and ideally having them play with the same toys across times.

To summarize the argument thus far, both formal and informal assessment approaches have limitations and have been justifiably criticized. It is nevertheless clear that formal assessment approaches are particularly well suited to the identification of language delays, and informal assessment approaches are particularly well suited to the detailed description of the child’s strengths and weaknesses in various language domains.
1.2.3 Empirical Measurement of Progress Using Formal and Informal Assessment

Despite concerns about the inappropriateness of using standardized testing to evaluate treatment efficacy, review of the evaluation research shows that this is the common method.

Few of these studies report criterion-referenced test scores or changes in language characteristics as shown via sample analysis. For example, a recent study used three formal tests to evaluate improvements in non-language impaired children's expressive language abilities as a result of a six-week long, home-based intervention program (Huebner, 2000). The three tests were one of comprehension of vocabulary (the Peabody Picture Vocabulary Test), one of expressive vocabulary (the Expressive One Word Picture Vocabulary Test – Revised), and a subtest of verbal expression from the Illinois Test of Psycholinguistic Abilities (Kirk, McCarthy & Kirk, 1968). The author utilized standard scores from all tests. For this study, toddlers’ caregivers were trained on a particular technique to employ while reading to their children. The children were assigned to either an experimental or a control group, and were tested three times in total; a baseline measure (taken before parent training and intervention), within 6 weeks after the intervention period, and a follow-up 3 months after the second testing date. Results showed an increasing trend in scores on both vocabulary measures, but this trend was not statistically significant.

Huebner noted that this is an unexpected result, as the program targeted vocabulary development as one of its major goals. As might be surmised from an earlier discussion in this paper, Huebner asserted that the tests might not have been sensitive enough to detect changes in the relatively short period of time (i.e., not more than 6
months in total) with children who were already functioning at a proficient language level.

Another recent study, however, indicates that even a longer time interval may not show gains in children’s scores on standardized tests (Van Tuijl et al., 2001). This Dutch study also used solely norm-referenced measures for home-based educational intervention program evaluation. This particular program spanned 2 years, and was implemented with young school-aged bilingual children who were at risk for educational difficulties. Compared to a control group, the authors found no significant improvements in language development after the 2-year program as shown by the standardized test results. They state that these test results may have represented an underestimation of the actual language abilities of the children insofar as the group of at-risk children was not acclimatized to “the typical social interaction format of psychological testing” (Van Tuijl et al., 2001, p. 156).

Fewer studies have used informal assessment tools to measure changes in language skills over time than have used formal measures. A study by Rescorla, Roberts, and Dahlsgaard (2000) incorporated both types of measures, and the authors found that greater positive change was detected by the informal tools than by the formal ones. These authors studied late talkers and comparison (language-typical) children longitudinally by measuring their expressive language through both formal and informal assessments. The children were assessed at age 2 and again at age 3 on the following measures: Expressive One Word Picture Vocabulary Test, Reynell Expressive Language Scale (Reynell, 1977), mean length of utterance and Index of Productive Syntax (Scarborough, 1990). The former two are formal tests, while the latter two are informal
measures, taken from analyzing 30-minute free-play language samples. It was found that the late talkers and comparison children both scored within the average range at follow-up on the two standardized measures, which looked mainly at vocabulary and verbal concepts. In contrast, the late talkers were found to score much lower than comparison children on the informal measures, which examined morphosyntactic abilities.

The authors interpreted this difference as the late talkers having made faster progress in lexical development than in syntactic development. While this is certainly a plausible explanation, the authors entertained other interpretations of these data in only a cursory manner. For example, it is also possible that the difference could be due to the methodology used to assess each domain of language performance, in other words, the difference between formal and informal measures. While they made brief mention of the outdated norms given by the EOWPVT, the authors did not directly address the possibility of the test scores having overestimated the children’s lexical capabilities.

An additional study has used solely language sample analysis measures in the evaluation of a therapy program. This study was designed to compare two language sampling situations for the assessment of semantic form development in young children with specific language impairment (Bain et al., 1992). Each sampling situation represented a different end of the continua of both structure and predictability. Low structure and predictability were embodied by a free-play situation with the clinician following the child’s lead, while the high end of both dimensions was represented through a situation known as a joint action routine. The joint action routine was highly structured, with the clinician playing a directive role in the interaction. The children were given therapy targeting the production of two semantic combinations (e.g. possessor +
The frequency and diversity of several semantic relations were computed for all children before and after treatment. The authors hypothesized that new forms would be more likely to emerge within familiar, predictable routines than in less structured situations. However, the opposite result occurred, with the children producing a greater number of multi-word utterances (as measured by MLU) as well as more diverse forms (as assessed via a semantic coding scheme) during free-play situations. They concluded that analysis of a sample arising from a situation that is of interest to the child represents an effective way to monitor language change over time.

A final comment on this issue is provided by Nye, Foster, and Seaman (1987). As noted, few studies use informal measures to document positive change as a result of therapeutic programs. The authors performed a meta-analysis of 61 studies that measured treatment efficacy, in order to determine whether language intervention has an overall beneficial effect across different children, therapy methods, and research designs. The studies encompassed a wide range of methods for assessing progress in the children, including both language tests and language sample analyses. While 43% of the studies used standardized tests to evaluate the effects of therapy, only 28% used language sample measures to address the same goal. Overall, it was found that the language sample measures showed a substantially larger effect of language improvement than formal tests. The authors suggest that this may be due to the fact that informal measures are tailored to the specific areas of language that are the focus of intervention.

In sum, few researchers have employed informal measures of language to evaluate treatment programs. Although the use of norm-referenced tests has formally been reported to be inappropriate for the measurement of progress, this methodology
continues to be the one of choice for this particular assessment task. This discussion will now move to the issue of formal versus informal measures as they function in the assessment of young children with autism.

1.3 Assessment Challenges Specific to Autism

1.3.1 Assessment Implications of Language Characteristics in Children with Autism

Communication impairment is a necessary component of the triad of impairment types used to diagnose a child with an autism spectrum disorder, as outlined by the Diagnostic and Statistical Manual of Mental Disorders. The triad also includes social interaction impairment and repetitive behaviour (APA, 1994). The term “communication impairment” in this case is rather all encompassing when we consider the type of communication difficulties as well as the range of severity of language problems across individuals with autism. A person with autism must have problems with language, but this person could either be nonverbal with little or no functional language, or he could be particularly verbose with only an occasional, slightly odd tendency to repeat verbatim what others have said to him. Both scenarios could reasonably be considered prototypical representations of the language profile of a child with autism.

Although language characteristics in people with autism are heterogeneous, several common tendencies do exist. Broadly, these may include delay or lack of development of speech, difficulty initiating and/or sustaining communication, problems in the semantic-pragmatic areas of language, and stereotyped/repetitive use of speech (Happe, 1994; Paul, 1995; and Schuler, 1995). Indeed, verbal children with autism tend
to have relatively spared language form (i.e., morphology, syntax), with occasional exceptions such as pronoun reversals (Paul, 1995).

Three aspects of expressive communication in children with autism are of particular interest in the context of the present study, as they tend to be difficult (if not impossible) to assess through traditional formal tests: nonverbal communication, echolalia, and pragmatics.

*Preverbal Communication*

The arguments proposed above for non-standardized language assessment in the measurement of change in language behaviour speak also to the assessment of communication and language in young children with autism. Many of these children are at a prelinguistic stage of communicative development well into their toddler years, and a significant proportion stay at this level of communication for life, never becoming verbal. If a clinician or researcher were to use only standardized forms of language assessment to measure the progress of therapy with these nonverbal communicators, they might show only very small improvements, or might never show improvements at all after months or years of therapy. Again, it would be erroneous to conclude that these children have never benefited from treatment, or that they have never improved in their communication skills. In the case of a nonverbal child with autism, norm-referenced test results are often not a valid reflection of her or his communication skills in everyday life. Assessment of these types of communication skills is best done through informal measures, which allow a clinician or researcher to examine modalities of communication other than speech. Other modes of communication include eye gaze, body language, object manipulation, or use of a picture exchange system.
Of verbal children who have autism, approximately 75% exhibit some form of echolalia, or repetitive/stereotyped speech (Fay & Schuler, 1980). Once thought to be meaningless in nature, echolalia has been shown to serve a variety of functions in the language of children with autism. Prizant and Duchan (1981) demonstrated that immediate echolalia (those repetitions directly following a preceding utterance) may be a way for the child to engage in turn-taking or self regulation, or even to request something of his communicative partner. Although immediate echolalia may have its function in the child with autism’s development of language, it is a characteristic that decreases as the child’s language skills become stronger. Thus, it would be valuable to be able to monitor the amount and this type of echolalia as a way of gauging a child’s language system maturity; a task that would best be completed via language sample analysis.

Pragmatics

The final aspect of expressive communication that is affected in children with autism is that of language in use, or pragmatics. This area encompasses a plethora of difficulties that include children of all levels of verbal functioning. In nonverbal children, pragmatic problems may still manifest as failure to initiate or maintain communication, or exhibition of a very restricted number of communicative functions. For example, many children with autism tend to communicate only to protest or request, very rarely to comment or acknowledge (Wetherby & Prutting, 1984).

In verbal children with autism, contextual language difficulties may also evidence themselves as an inability to maintain conversational topic, or perseveration on a favourite personal topic. The verbal child may cause conversational breakdowns frequently due to topic maintenance problems, and may be unable to repair these
breakdowns once they occur. Inflexibility of thought and overliteral interpretations of humour are additional examples of common autistic tendencies. Children with autism may also be impaired at a conversational level by their flat or stereotyped intonation patterns, which leave the listener without the usual pragmatic clues given by typical intonation. Thus, the phonological problem of incorrect intonation patterns may have pragmatic consequences for children with autism. Dysfunctional language patterns of this sort cannot be measured by formal testing, which is completely decontextualized. Undeniably, naturalistic assessment techniques are the only method to intensively study such characteristics of language, which need to be examined in a rich, varied, diverse interactive context. As this area of language is invariably affected in children with autism, it is essential that communication sample analysis be employed to adequately gauge a child's strengths and weaknesses as he engages in daily communicative activities. Further discussion of this point is included in Chapter 2, as the area of pragmatics is invariably affected in children with autism.

1.3.2 Behavioural Challenges in Formal Testing of Children with Autism

In addition to assessment problems arising from the language profile of children with autism, there are general problems stemming from their challenging behaviour during testing situations. Indeed, even children without developmental disabilities are disadvantaged by the meta-linguistic confound of test-taking behaviour when being assessed with tests of language. The phenomenon of test refusal encompasses a variety of behaviours, including an oppositional attitude, playing with the test materials inappropriately, or not responding to questions. Some researchers have found that a
tendency towards test refusal of any kind is correlated with lower concurrent
europsychological test scores and verbal comprehension ability (Mantynen et al., 2001).

For young children with autism, this problem is magnified greatly. Due to
potentially severe attentional, social, and behavioural challenges, many children with
autism have great difficulty during testing situations. A standardized testing situation
requires that a child sit still, often for long periods of time, and engage in artificial
exchanges with a person who is likely relatively unfamiliar to the child. This represents a
significant change in routine, a state of affairs that most children with autism will find
stressful. As a result, these children may have a tendency to refuse to respond, become
upset, or otherwise not conform to the scripted behaviour of test-taking.

An experiment by Koegel, Koegel & Smith (1997) examined the effects of
attentional and motivational factors on the standardized test scores of children with
autism. They administered a variety of language (e.g. PPVT) and intelligence tests
repeatedly to 6 children with autism, who varied widely in their cognitive and
communicative abilities. In some of the testing sessions, the tests were given just as the
standardized administration protocols dictated. In other conditions, the attentional and
motivational factors likely to be disruptive to the child’s test session were identified
ahead of time, and a strategy was put in place to minimize the effects of the disruptive
behaviours. For example, one child had a tendency to engage in the speaking register of
a cartoon voice incessantly, and thus the strategy for him was to allow him to engage in
this behaviour at pre-set intervals throughout the session (i.e., as a reward for good test-
taking behaviour). The authors found that with only one exception, test scores were
always higher in the sessions when the attentional and motivational behaviours were
taken into account. In fact, three of the children who did not receive a score in the standard administration condition were actually able to sometimes score within the normal range when the attentional and motivational techniques were applied. They conclude that for children with autism, “Functionally then, the test session becomes one of assessing motivation, attention, or compliance more than of assessing language or intelligence” (Koegel, Koegel & Smith, 1997, p. 241).

Children with autism are often summarized as being “untestable,” which provides clinicians and parents with no useful information as to the child’s level of communicative functioning, or what appropriate treatment strategies might be; and provides researchers with no idea of how the child has progressed in therapy. Indeed, Catherine Lord (2000) revised her first version of the Autism Diagnostic Observation Schedule (ADOS) after noting that “the context of sitting at a table for 30 minutes [was] not effective for eliciting a range of social-communicative behavior or play from very young children [with autism]” (p. 206).

To conclude, there are substantive reasons to question the adequacy of formal tests to assess the change over time in the expressive language of young children with autism. First there is the general concern that tests, because of their psychometric properties, are a poor choice for documenting progress (Lipsey, 1982). Second is the limited ability of tests to adequately describe the language behaviour of children with autism, either because their language characteristics are not addressed by such tests, or because the children’s behaviour during testing prohibits useful interpretation of results.

Interestingly, only one study in the scientific literature has focused on this methodological question as it applies to children with autism. Condouris, Meyer, and
Tager-Flusberg (2003) asked how standardized tests, such as the PPVT, the Expressive Vocabulary Test (Williams, 1997), and the Clinical Evaluation of Language Fundamentals correlated with measures derived from spontaneous language samples in their ability to diagnose impairment in lexical-semantic and grammatical domains. Informal measures in the study were number of different word roots, MLU, and IPSyn (Scarborough, 1990). All assessments were administered to 44 children with autism, whose mean age was 7;3 (range 4-14) and whose average IQ was in the normal range. They found that the standardized tests and informal assessments correlated significantly in the identification of language impairment in the areas of both content and structure. Surprisingly, given the emphasis of pragmatics with this population of children, language in use was not assessed in this study.

While the above study found that formal and informal measures provided similar information about the diagnosis of language impairment in children with autism of normal intelligence, at least in the areas of content and structure, it does not provide any answers about the measurement of progress with both types of assessment in these children, nor about the use of these assessments with preverbal children with autism. It is the aim of the current study to investigate these issues.

1.4 Current Research Question

In the expressive language measurement of young children with autism, do the data provided by informal measures (i.e. communication sample analyses) and formal measures converge on a single view of how a child’s language abilities change over time or do they diverge?
2. METHODS

2.1 Overview

2.1.1 General Description of Current Study

The purpose of the present research project was to explore the relationship between the scores obtained from formal language testing of children with information gathered from more informal language assessment procedures. A within-participants, longitudinal experimental design was employed to examine the research question.

Participants were seven preschool-aged children with autism spectrum disorders. The children varied greatly in their severity of autism, and consequently in their communicative abilities. All children were receiving early intensive behavioural intervention (EIBI) for up to 20 hours per week throughout the duration of this study. The present study included data collected as part of a larger, government-initiated and funded research project aimed at evaluating EIBI programs. Investigators involved in this evaluation project used a comprehensive battery of standardized tests in several developmental domains to assess the children’s progress at four successive data points over a 2-year span (Mirenda, 2003).

Videotaped behavioural samples were collected concurrently with a subset of the standardized testing dates. With few exceptions, these samples were collected during sessions with each child’s Autism Interventionist (AI), during which the child was engaged in intervention, free play, or a combination of the two. For the current project, two of these videotapes were obtained for each child, each approximately 30 minutes in length. Each sample was transcribed and analyzed for communication and language
patterns for each child. The analyses focused on the language and communication goals set by the child's speech-language pathologist for the time period being examined, but also included other areas of communication appropriate to the child's level. The data collected from these samples were then compared with the concurrent standardized language test scores.

2.1.2 Description of the Early Intensive Behavioural Intervention Evaluation Project

The evaluation of outcomes resulting from the Early Intensive Behavioural Intervention is being conducted by Dr. Pat Mirenda at the University of British Columbia. The study is a highly comprehensive one, including information measured in many domains of development: language, play and social skills, adaptive behaviour, sensory functioning, and temperament. The evaluation also included information on ratings of severity of the children's autism over time, as well as measures of parental stress and parent's perceptions of their child's progress as a result of the intervention program. The data were collected by several evaluation teams, each consisting of a speech-language pathologist, a psychologist, and a family interviewer (trained graduate students) (Mirenda, 2003).

The early intensive intervention program itself is site-based, although a large portion of the treatment occurred in the children's homes. The children in the present study received an average of 15 hours per week of intervention. Intervention consisted of largely one-on-one therapy with a trained Autism Interventionist, who provided highly structured therapy for the children based on principles of applied behavioural analysis. The children also had integrated, relatively infrequent direct therapy from speech-
language pathologists and occupational therapists. Some therapy was provided in a
group setting, both with other children with autism and with typically developing peers.
The measurement of progress in all developmental domains was conducted on the
following schedule: baseline testing was performed prior to each child’s entry into the
program, and follow-up testing was conducted at 6 months, 1 year, and 2 years post-
baseline.

2.2 Participants

2.2.1 Description of Participants

Participants in this study were seven preschool-aged children with a diagnosis of
autism spectrum disorder. The children were recruited from the Early Intensive
Behavioural Intervention program being run at Queen Alexandra Hospital in Victoria. At
the start of the present study, all children had been receiving intervention for a period of 4
to 12 months. They continued to receive the 15 hours per week of intervention for the
duration of this study.

All children were male. At the first data collection point, the children ranged in
age from 3;3 to 4;8. By the second testing / taping point, the age range was 4;2 to 5;11.
All children came from predominantly English-speaking families, and had normal
hearing.

Six of the seven boys had been given a diagnosis of autism prior to the start of the
current study. The remaining child had a diagnosis of Pervasive Developmental Disorder
Not Otherwise Specified (PDD-NOS) due only to the fact that he also had the possibility of a genetic syndrome.

2.2.2 Communication Characteristics of Participants

Although all seven children had a diagnosis on the autism spectrum, they varied widely in the severity of the autism, and their level of functioning with regard to expressive communication skills. Three of the children were at a preverbal stage of development at the first data collection point (Time 1), and even these children varied on the amount and type of communication in which they engaged. The remaining four children were at least somewhat verbal at Time 1, but differed in the degree to which they were verbal. "Verbal" ranged from being at the single word level of expressive language development, to being able to communicate in lengthy, complex, well-formed utterances. Overall, there were two distinct groups of children: those who were preverbal at Time 1 and those who were verbal at Time 1. Information about general communication skills, derived mainly from what the videotapes contained as well as some information provided by the children's Autism Interventionists, is summarized for each child below. All names of children in this study are pseudonyms.

EMMETT

This child was 3;3 when he began the early intervention program, and was 3;8 at Time 1. At the first data collection point, he was completely nonverbal. He had very minimal communicative abilities, and made few attempts to initiate communication with his Autism Interventionist (AI). Communicative acts were usually in the form of object
manipulation or vocalization, and typically occurred for the purposes of protesting or requesting. He was able to follow simple commands. He vocalized frequently, and often continuously for extended periods of time. Emmett displayed many of the stereotypic / repetitive behaviours (e.g. peripheral eye gaze) and body movements commensurate with a diagnosis of autism. He did not exhibit any destructive or problematic behaviours, and appeared to have a good relationship with his AI, as she easily managed to engage him in activities involving joint attention that he seemed to enjoy. It should be noted that Emmett also had the question of a possible genetic syndrome aside from his diagnosis of autism spectrum disorder within his medical history. This possibility was still being explored at the time of this writing. By the DSM-IV definition, this means that Emmett was given a formal diagnosis of PDD-NOS, rather than autism.

Several therapy targets were set for Emmett’s communication by his speech-language pathologist during the six-month interval between his two data points. These included the following:

1. Appropriate eye contact
2. Use of object manipulation to request
3. Use of sign for “more”
4. Use of social greetings (waving “hi” and “goodbye”)
5. Appropriate responses to general instructions
6. Increase in the number of meaningful (i.e. communicative) vocalizations
JACOB

Jacob was 3;3 at Time 1, and had been receiving intervention for 4 months at that time. He was completely nonverbal, with extremely limited occurrences of intentional communication. He could follow simple commands. He vocalized frequently, and exhibited several stereotypic behaviours (e.g. rocking, hand flapping). When he did communicate, it was through developmentally early means such as body language / movement, and these nonverbal communicative acts served the function of requesting. Jacob had fleeting moments of eye gaze toward his communication partner.

Jacob’s communication goals as set by his S-LP for the 12 month interval between the two assessment points included the following:

1. Increased receptive language skills (following directions)
2. Increased joint attention behaviours
3. Use of sign for “more”

WESLEY

This boy was 4;7 at Time 1, and had been in the program since the age of 4;1. At Time 1, he was almost completely nonverbal, with only an occasional one-word utterance that was typically an exact repetition of his AI’s previous utterance. He used the sign for “more,” but was usually prompted to do so. At the first assessment point in the current study, he was learning to use the Picture Exchange Communication System, and could use it reliably to request objects. Wesley tended to vocalize quite frequently, but usually not communicatively. He had fairly good eye contact with his communication partner. He displayed some repetitive behaviours, such as hand flapping, as well as some
maladaptive behaviours (hitting his AI to protest). He did not initiate communication or play often at this time. It is noted here that Jacob and Wesley are brothers.

Communication goals for the 18 months between assessment points for Wesley included the following:

1. Increased usage of PECS
2. Improved and more frequent joint attention
3. Increased signed vocabulary, with more spontaneous use of signs

GRAHAM

Graham was 4;2 at the time of initial data collection, and had been in the program for 6 months at that time. At Time 1, he had one-word utterances that were largely repetitions of the last word in his AI’s preceding utterance, giving him the appearance of being highly echolalic. Besides spoken words, he had many other modes of communication, including gestures and vocalizations. Graham was communicating for a variety of reasons, including making comments on the activities he was engaged in. He appeared to be a very active boy with a rather short attention span, but did not exhibit any repetitive or injurious behaviours. His AI often used tangible rewards to entice him to come to or stay at an activity. He followed commands given by the AI quickly and appropriately.

Communication and language goals for Graham during the 18 month interval covered by this study included the following:

1. Increased spoken labeling of objects/pictures
2. Use of “more” or the sign for “more”
3. Requests for help
4. Increased use of social language (e.g. "hi" and "bye")

CHRIS

This child was 4;8 at Time 1, and had been in the program since the age of 4;1. When seen at Time 1, he was verbal, speaking mainly in 2-3 word utterances. His sentences were often missing articles and auxiliary verbs. He tended to imitate others' utterances quite frequently. Chris’s speech was often unintelligible due his nonstandard phonology, as well as his tendency to speak quietly. He appeared to be a very sociable boy who was easily engaged in activities.

No communication goals were provided by EIBI staff for Chris during the 6 month of the current study.

BRANDON

Brandon was 4;3 at the time he was initially tested and taped, which was 4 months following his entry into the program. He was highly verbal at this time, and might even be more appropriately described as verbose. However, his speech was very frequently unintelligible due to a combination of factors: a nonstandard phonology, odd and misleading intonation patterns, occasional use of a very high-pitched voice, and a tendency to speak too rapidly. His utterances were well-formed, with some regular morphosyntactic errors, such as misassigment of pronoun case ("him" for "he"). He appeared to have a large vocabulary and was sometimes able to make complex semantic constructions ("I want to turn it backwards, please"). However, many of his longer
utterances lacked coherence and at times he seemed unable to maintain the topic of conversation. An example of an utterance lacking coherence was one which began with several false starts followed by, “I’ll do the puts this cool Waldo book back in the store.” Brandon was sociable, made good eye contact, and did not appear to have any behaviours which could be described as typical of autism. He did seem to have a rather short attention span, although he responded well to prompts to pay attention made by his Autism Interventionist.

Speech and language goals for Brandon for the 12 month interval covered by this study included:

1. Improved phonology
2. Increased correct use of pronouns and prepositions
3. Provision of specific information requested by questions
4. Increased number of relevant conversational turns

ALEX

This child was aged 4;7 at Time 1, a full year after he began receiving intervention in the EIBI program. He spoke in full sentences at this time, usually 3-4 words in length. He was frequently unintelligible due to his nonstandard phonology, as well as his propensity to switch to a “baby talk” register at times, where his intonation, morphosyntax, and phonology seemed to temporarily regress. Alex frequently perseverated on his own utterances, making him appear echolalic. Upon closer examination, he rarely repeated others’ utterances, but rather repeated his own utterances as many as 6 or 8 times each. It was also noted that it frequently took an exceptionally
Alex’s speech and language goals, as set by his S-LP for the 12 month period of this study included a range of phonological, syntactic, and pragmatic goals:

1. Increased proper use of pronouns
2. Increased correct word order in fluent speech
3. Use of language to talk about abstract phenomena such as feelings and ideas
4. Improved “s” and “f” sounds

2.3 Procedures: Communication Assessment

2.3.1 Overview of Procedures

Data for this project consisted of two types: scores from four different standardized tests and data obtained from language sample analyses. These data were collected twice for each child, over a time period of 6 to 18 months. The timeline for data collection from each child is given in Table 2.1 below. Time 1 and Time 2 are noted there as a function of cumulative months-in-program. The two data collection points for Emmett, for example, occurred 6 and 12 months after he began the program. Data collection points varied among children due to availability of tapes that corresponded in time with testing.
Table 2.1:

Dates for EIBI testing and language sample collection expressed as cumulative months-in-program

<table>
<thead>
<tr>
<th>CHILD</th>
<th>6 MONTHS</th>
<th>12 MONTHS</th>
<th>*18 MONTHS</th>
<th>24 MONTHS</th>
</tr>
</thead>
<tbody>
<tr>
<td>EMMETT</td>
<td>Time 1</td>
<td>Time 2</td>
<td>*Time 2</td>
<td></td>
</tr>
<tr>
<td>JACOB</td>
<td>Time 1</td>
<td></td>
<td>*Time 2</td>
<td></td>
</tr>
<tr>
<td>WESLEY</td>
<td>Time 1</td>
<td></td>
<td>Time 2</td>
<td></td>
</tr>
<tr>
<td>GRAHAM</td>
<td>Time 1</td>
<td></td>
<td>Time 2</td>
<td></td>
</tr>
<tr>
<td>CHRIS</td>
<td>Time 1</td>
<td>Time 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BRANDON</td>
<td>Time 1</td>
<td></td>
<td>*Time 2</td>
<td></td>
</tr>
<tr>
<td>ALEX</td>
<td>Time 1</td>
<td></td>
<td>Time 2</td>
<td></td>
</tr>
</tbody>
</table>

* Due to time constraints, these data were collected by a research S-LP assisting with this study, rather than by staff from the EIBI evaluation project.

2.3.2 Formal Measures Used in Current Study

The investigators in the EIBI evaluation project utilized several standardized tests in order to measure change in language comprehension and production skills over time. All of these selected tests were administered to each child at the specified data collection points. Each testing session took a half-day to complete (either a morning or an afternoon session). The tests were given to the children by the same examiner, a speech-language pathologist, on every occasion (with the exception of Jacob and Brandon, whose Time 2 tests were administered by a different S-LP). Tests were administered to all children in a private room, and testing sessions typically lasted for 2 to 3 hours.

Tests of comprehension abilities were the Peabody Picture Vocabulary Test (PPVT; Dunn, 1981), and subcomponents of both the Preschool Language Scale (PLS; Zimmerman, Steiner and Pond, 1992) and the MacArthur Communicative Development
Inventories (MCDI; Fenson, et al, 1993). Measures of productive language used in the evaluation project were: the Expressive One-Word Picture Vocabulary Test (EOWPVT; Gardner, 1990), the Expressive Communication subtest of the PLS, and subcomponents of the MCDI. Following are descriptions of the standardized tests used.

Peabody Picture Vocabulary Test (PPVT-R)

This test measures the comprehension of single words, primarily nouns and verbs that refer to classes of objects and events. The standard administration procedures for the test require the child to choose, by pointing at one of four pictures on a page, the correct depiction of a word the examiner has spoken. The child must be able to attend to, scan, and interpret the pictures in order to demonstrate knowledge of the lexical item (Dunn & Dunn, 1981).

Expressive One-Word Picture Vocabulary Test (EOWPVT)

The EOWPVT is a measure of a child’s expressive vocabulary abilities. In the test, the child is required to name pictures presented to him. The child must also be able to name the higher-order categories implicit in pictures that portray several members of the same category grouped together (e.g. child is asked “What word names all of these?” when shown a group of animals together including a bear, cow, and giraffe) (Gardner, 1990).
Preschool Language Scale – 3 (PLS-3)

The PLS is meant to be a comprehensive measure of language skills, and is comprised of two different subscales: Auditory Comprehension and Expressive Communication. It is comprised of observation-based items (either spontaneous or elicited behaviours), and can be used with children of a very young age (age range is 2 weeks to 6 years, 11 months). This early focus means that the PLS measures precursors to language such as social communication and vocal development. It purports to examine each of the areas of form (morphology and syntax), content (vocabulary and basic concepts), and use (integrative thinking skills, social communication) of language (Zimmerman, et al., 1992). However, the majority of the “use” area items are found only in a supplemental language sample checklist, not within the two main scales.

Since the present study is concerned mainly with the area of productive communication skills, only the Expressive Communication subtest will be covered in detail in this description. This subscale includes items measuring vocal development (“combines sounds to form syllables”), social communication (“imitates a word”), semantics (“names objects,” “has a vocabulary of at least 10 words”), structure (“uses pronouns,” combines 3 or 4 words in spontaneous speech”), and integrative thinking skills (“answers questions logically,” “names categories”). Children earn credit for these items, either by responding to standardized elicitation procedures, or by spontaneously demonstrating the specified language ability at any time during the testing session. The child is given scores for both subtests, as well as a composite “Total Language” score.
MacArthur Communicative Development Inventories (MCDI)

Although norm-referenced to the same degree as the three tests listed above, the administration procedures and nature of the data used for the MCDI are very different. Rather than a test, the MCDI is a parent report measure. There are two forms of the MacArthur scales: one for infants called the “Words and Gestures” form, and one for more linguistically advanced children (toddlers), known as the “Words and Sentences” form, each asking for data about both language comprehension and production. (Fenson, et al., 1993). The MCDI has been found to be a valid measure of young children’s expressive vocabulary, both in typically-developing and language-impaired populations (Dale, 1991; Thal, et al., 1999). Reliability also tends to be high (split-half and month-to-month in the 95% range) for the expressive portion of both the infant and toddler versions. However, validity for the comprehension part, as indexed by its correlation with other measures, tends to be poor (Bates, 2003).

The “Words and Gestures” form is a checklist format that gives parents a large number of words typically acquired early by children, and requires that they report when their child “understands” or “understands and says” each word. It also asks whether children engage in several social communicative games (e.g. peekaboo), produce early or late developing gestures, imitate, label and engage in certain play skills.

The more advanced version, meant for toddlers, gives parents a checklist of hundreds of early words, arranged in semantic categories such as “body parts.” The parent is asked to check off words that they have heard their child say. Although largely a measure of the child’s productive lexicon, this form also includes a section on morphology and syntax, asking parents to record whether their child produces the “-ed”
ending, for example. This form also asks parents to report the child’s three longest sentences as an approximation of mean length of utterance.

The infant version is normed on children aged 8 to 16 months and the toddler version is supposed to measure communication skills between the ages of 16 and 30 months. However, examiners may decide to use the form appropriate to the child’s communication level, rather than the one that is appropriate to chronological age. This was the case in the current study, where children who were well past the age limits for the infant form were administered that form due to the fact that they were still preverbal in their communication attempts. Similarly, children much older than 2½ years of age (30 months) were given the toddler form. These age differences made it inappropriate to use the norms provided with the test, and therefore raw scores were used for this project.

2.3.3 Informal Measures Used in Current Study

This study made use of pre-existing tapes created by staff in the EIBI program for their own purposes. All children were videotaped interacting in sessions with their Autism Interventionists. The sessions were taped by a second AI, or occasionally by one of the child’s parents. In all cases, the people in the room while the child was being videotaped were all highly familiar to the child, an important consideration when working with children with autism. The videotaped sessions lasted between 20 and 40 minutes. The Autism Interventionists always attempted to tape the children closer to the start of their sessions, noting that they often became tired and restless, and had decreased attention spans as they neared the end of their sessions.
Due to the fact that children were taped for purposes other than those of this particular project, there was considerable variation across children in the activities used and degree of structure in the interaction. Details of what the videotapes contained, including where they took place, who was present, how long they were, and which activities the children were engaged in, are contained in Table 2.2 and Table 2.3. The sessions took place either in the intervention classroom, or in the child’s own home. Within these two settings, children were either taped while they were on the floor or seated on a chair. Often when the children were sitting, they were at a table across from their interventionist. There were sometimes other people besides the child’s AI participating in the session, including mothers, peers, or speech-language pathologists.

As a result of the type of intervention being given to the children, the activities tended to be fairly structured and often changed frequently within a session. The clinical literature recommends that language sample collection be done while the child is engaged in less structured, lengthy and consistent activities (Miller, 1981). This was not possible given the source of the tapes used in this project. It is possible, however, that because the children in this study have autism, free play activities might not provide the best portraits of their communication capabilities. Many children with autism, when left to direct their own play, tend not to initiate communication with even familiar people.
Table 2.2: Contexts of each Child’s Videotaped Communication Sample – Time 1

<table>
<thead>
<tr>
<th>Child</th>
<th>Location</th>
<th>Duration</th>
<th>Partner</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emmett</td>
<td>Seated at table, classroom</td>
<td>18 min.</td>
<td>AI</td>
<td>Books, matching</td>
</tr>
<tr>
<td></td>
<td>Floor, Home</td>
<td>17</td>
<td>AI, mother</td>
<td>Play with microphone, colouring, puzzle</td>
</tr>
<tr>
<td>Jacob</td>
<td>Floor, classroom (in corner)</td>
<td>23</td>
<td>AI</td>
<td>Toys (shape sorter, ring stand), books, singing</td>
</tr>
<tr>
<td>Wesley</td>
<td>Standing at sink, classroom</td>
<td>12</td>
<td>AI</td>
<td>Water play</td>
</tr>
<tr>
<td></td>
<td>Seated at table, classroom</td>
<td>12</td>
<td>AI</td>
<td>Toy boat with balls and toy hammer</td>
</tr>
<tr>
<td></td>
<td>Seated at table, classroom</td>
<td>7</td>
<td>AI</td>
<td>Drawing</td>
</tr>
<tr>
<td>Graham</td>
<td>Seated at table, classroom</td>
<td>30</td>
<td>AI</td>
<td>Puzzles, picture matching, toy animals</td>
</tr>
<tr>
<td>Chris</td>
<td>Seated on couch, home</td>
<td>10</td>
<td>AI, sister</td>
<td>Books</td>
</tr>
<tr>
<td></td>
<td>Seated at table, home</td>
<td>12</td>
<td>AI</td>
<td>Bag of food toys</td>
</tr>
<tr>
<td></td>
<td>Floor, home</td>
<td>19</td>
<td>AI</td>
<td>Pretend play with school bus, playdough</td>
</tr>
<tr>
<td>Brandon</td>
<td>Seated at table, home</td>
<td>10</td>
<td>AI</td>
<td>Puzzle, magnets</td>
</tr>
<tr>
<td></td>
<td>Floor, Home</td>
<td>17</td>
<td>AI</td>
<td>Books</td>
</tr>
<tr>
<td></td>
<td>Seated at table, Home</td>
<td>5</td>
<td>AI</td>
<td>Playdough</td>
</tr>
<tr>
<td>Alex</td>
<td>Moving around classroom</td>
<td>6</td>
<td>AI, peer</td>
<td>Pretend play with cash register and food toys</td>
</tr>
<tr>
<td></td>
<td>Seated at table, classroom</td>
<td>2</td>
<td>AI</td>
<td>Brief conversation</td>
</tr>
<tr>
<td></td>
<td>Floor, classroom</td>
<td>9</td>
<td>AI, peer</td>
<td>Playing board game</td>
</tr>
<tr>
<td></td>
<td>Seated at table, classroom</td>
<td>12</td>
<td>S-LP</td>
<td>Speech flash cards, stickers</td>
</tr>
<tr>
<td>Child</td>
<td>Location</td>
<td>Duration</td>
<td>Partner</td>
<td>Activity</td>
</tr>
<tr>
<td>-----------</td>
<td>-------------------------</td>
<td>----------</td>
<td>---------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Emmett</td>
<td>Seated in chair, Home</td>
<td>5 min.</td>
<td>AI</td>
<td>Oral-motor exercises</td>
</tr>
<tr>
<td></td>
<td>Floor, Home</td>
<td>13</td>
<td>AI</td>
<td>Books, shape pegboard, singing</td>
</tr>
<tr>
<td></td>
<td>Floor, Home</td>
<td>7</td>
<td>AI</td>
<td>Gross motor play with a ball</td>
</tr>
<tr>
<td></td>
<td>Seated at table, Home</td>
<td>10</td>
<td>AI</td>
<td>Letter matching activity</td>
</tr>
<tr>
<td>Jacob</td>
<td>Seated at table, classroom</td>
<td>12</td>
<td>AI</td>
<td>Snack time</td>
</tr>
<tr>
<td></td>
<td>Floor, classroom</td>
<td>18</td>
<td>AI</td>
<td>Gross motor play with large toys</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>* child is moving around for most of this time, except for a brief session lying on his back on the floor with AI</td>
</tr>
<tr>
<td>Wesley</td>
<td>Seated at table, classroom</td>
<td>10</td>
<td>AI</td>
<td>Lunch time</td>
</tr>
<tr>
<td></td>
<td>Seated at table, classroom</td>
<td>20</td>
<td>AI</td>
<td>Puzzle, colour matching activity</td>
</tr>
<tr>
<td>Graham</td>
<td>Floor, classroom</td>
<td>5</td>
<td>2 AI’s, peer</td>
<td>Matching toy animals</td>
</tr>
<tr>
<td></td>
<td>Seated at table, class</td>
<td>14</td>
<td>AI</td>
<td>Matching toy animals, puzzle</td>
</tr>
<tr>
<td></td>
<td>Seated at table, classroom</td>
<td>5</td>
<td>AI</td>
<td>Snack time</td>
</tr>
<tr>
<td></td>
<td>Floor, classroom</td>
<td>5</td>
<td>AI</td>
<td>Free play (while moving around room)</td>
</tr>
<tr>
<td>Chris</td>
<td>Seated on couch, home</td>
<td>10</td>
<td>AI, Chris’s younger sister</td>
<td>Books</td>
</tr>
<tr>
<td></td>
<td>Floor, home</td>
<td>5</td>
<td>AI, Chris’s younger sister</td>
<td>Matching with toys</td>
</tr>
<tr>
<td></td>
<td>Floor, home</td>
<td>16</td>
<td>AI</td>
<td>Matching with cards, printing activity</td>
</tr>
<tr>
<td>Brandon and Alex</td>
<td>Seated in chairs, classroom</td>
<td>29</td>
<td>S-LP, both children</td>
<td>Free play with model train</td>
</tr>
<tr>
<td></td>
<td>Floor, classroom</td>
<td>15</td>
<td>S-LP, both children</td>
<td>Board game</td>
</tr>
<tr>
<td></td>
<td>Seated at table, home</td>
<td>12</td>
<td>AI, S-LP, both children</td>
<td>Books (playing “I spy”)</td>
</tr>
</tbody>
</table>
Samples were transcribed by trained graduate students of Speech-Language Pathology. The transcription was done using the conventions outlined for the Systematic Analysis of Language Transcripts (Miller & Chapman, 1984). Transcription included all verbal and vocal acts, and all nonverbal acts that were potentially communicative. A sample transcript with nonverbal acts transcribed and coded is included in Appendix A. Inter-rater reliability of transcription was determined for 15% of the samples. Utterances used for reliability checks were always drawn from the initial section of the transcript, were divided between the two data collection times and were taken from the data of four randomly selected children. Reliability was found to be 93% overall. The communication samples were then analyzed according to the descriptions below. This analysis was completed by the principal investigator for this study, a graduate student of Speech-Language Pathology.

2.3.3.1 Dependent Variables for All Children: Summary of Formal & Informal Measures

While all children received the same formal tests, informal measures varied across children. As can be seen from the entire range of analyses depicted in Table 2.4, however, there was a great deal of overlap in the measures used within preverbal children, and again within verbal children.
### Table 2.4:
Dependent Variables for All Children – Formal and Informal Measures

<table>
<thead>
<tr>
<th>Language Domain</th>
<th>Informal</th>
<th></th>
<th>Formal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Preverbal / 1-word stage</td>
<td>Verbal</td>
<td>All children</td>
</tr>
<tr>
<td>Communicative</td>
<td>Rate</td>
<td>N/A</td>
<td>- PLS-3</td>
</tr>
<tr>
<td></td>
<td>Mode Communicative vocalizations</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Function</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spontaneity</td>
<td>Imitativeness</td>
<td>Imitativeness</td>
<td></td>
</tr>
<tr>
<td>Phonology</td>
<td>Babbling level Intelligibility</td>
<td>- Intelligibility</td>
<td></td>
</tr>
<tr>
<td>Vocabulary</td>
<td>Different words</td>
<td>Different words Different verbs Total auxiliaries</td>
<td>PPVT EOWPVT MCDI: #words produced</td>
</tr>
<tr>
<td>Syntactic Length and Complexity</td>
<td>N/A</td>
<td>MLU IPSyn DSS</td>
<td>PLS-3</td>
</tr>
<tr>
<td>Semantics</td>
<td>N/A</td>
<td>Propositional complexity Predicate type</td>
<td>PLS-3</td>
</tr>
<tr>
<td>Relevance</td>
<td>Eye gaze Appropriateness of response</td>
<td>Topic maintenance</td>
<td></td>
</tr>
</tbody>
</table>

### Inter-rater Reliability for Coding of Dependent Variables

Inter-rater reliability was calculated for 15% of the data coded for each variable. This 15% was selected for dual coding, with the constraint that (1) the selected portions consisted of contiguous interactions, and (2) there was even representation from all of the pertinent children and both sample points. Coding was done in two separate, blind sessions by the principal investigator and another graduate student in speech-language pathology and then later compared. Reliability results will be reported for each relevant variable in the next two sections.
2.3.3.2 Dependent Variables for Preverbal Children: Communication Sample Analyses

Dependent variables for the preverbal children are listed according to the language skill area they measured in Table 2.5. Details for each variable, including criteria for coding schemes and the rationale for its inclusion, are outlined below. An example of a coded transcript for a preverbal child is given in Appendix A.

Table 2.5: Dependent Variables for each Preverbal Child

<table>
<thead>
<tr>
<th>Child's Name and CA at Time 1</th>
<th>Language Domain</th>
<th>Dependent Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emmett – 3;8</td>
<td>Communicative Acts</td>
<td>Rate of communication (# of communicative acts / minute)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mode of communicative acts (coding system)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Function of communicative acts (coding system)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Communicativeness of vocalizations (% of vocalizations that are communicative)</td>
</tr>
<tr>
<td></td>
<td>Phonological</td>
<td>Babbling level of vocalizations (percentage of each of three levels)</td>
</tr>
<tr>
<td>Development</td>
<td></td>
<td>Relevance Appropriateness of response (percentage of each of three levels of appropriateness)</td>
</tr>
<tr>
<td></td>
<td>Vocabulary</td>
<td>Macarthur CDI - Total # words (raw score)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total # of words</td>
</tr>
<tr>
<td></td>
<td>Expressive Communication</td>
<td>Expressive Communication subtest of PLS-3 (raw score)</td>
</tr>
<tr>
<td>Jacob – 3;3</td>
<td>Communicative Acts</td>
<td>Rate of communication (# of communicative acts / minute)</td>
</tr>
<tr>
<td>------------</td>
<td>--------------------</td>
<td>--------------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mode of communicative acts (coding system)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Function of communicative acts (coding system)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Expressive Communication subtest of PLS-3 (standard score)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Communicativeness of vocalizations (% of vocalizations that are communicative)</td>
</tr>
<tr>
<td>Phonological Development</td>
<td>Mean babbling level of vocalizations (% of each of 3 levels of babbling)</td>
<td></td>
</tr>
<tr>
<td>Relevance</td>
<td>Appropriateness of response (% of each of 3 levels of appropriateness)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Directed eye gaze (# of times child’s gaze is to communication partner)</td>
<td></td>
</tr>
<tr>
<td>Expressive Communication</td>
<td>Expressive Communication subtest of PLS-3 (standard score)</td>
<td></td>
</tr>
<tr>
<td>Wesley – 4;7</td>
<td>Communicative Acts</td>
<td>Rate of communication (# of communicative acts / minute)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mode of communicative acts (coding system)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Function of communicative acts (coding system)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Expressive Communication subtest of PLS-3 (raw score)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Communicativeness of vocalizations (% of vocalizations that are communicative)</td>
</tr>
<tr>
<td>Productive Vocabulary</td>
<td>Macarthur CDI - Total # words (raw score)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total # words produced in sample</td>
<td></td>
</tr>
<tr>
<td>Expressive Communication</td>
<td>Expressive Communication subtest of PLS-3 (standard score)</td>
<td></td>
</tr>
</tbody>
</table>

Children at Preverbal and One-Word Stage: Expressive Measures

The ultimate goal of any speech-language therapy program is to improve communication, particularly for children with autism, who have problems with functional communication at the core of their disability. Before describing the expressive communication measures used to gauge progress in the preverbal children, the meaning of the term “communicative act” will be defined in the context of this study. This is a
key concept since all of the other variables to be measured here have the communicative act as their basic unit of analysis.

A communicative act or intent is a familiar and straightforward concept that has proven to be a challenge to operationalize. This is due to many factors, including the population being studied, the different developmental levels of children being studied, and the degree to which social / conversational context is taken into account (Chapman, 1981).

In the context of this study, which deals with children who have significant impairments in social aspects of communication, communicative acts were identified based on the following definition. A communicative act is any act that meets at least one of the following criteria (drawn from early discussions of pragmatics, e.g. Searle, 1969, and Bates, 1976):

- meant to convey a message, either social or propositional in nature
- treated as communicative by the interaction partner

Accepting either of these criteria rather than requiring both meant that purely social actions, such as an action performed in order to take a turn, were counted as communicative in nature. While this decision is admittedly arguable, use of the same criteria at both assessment points, coupled with the fact the standardized tests provide no data in this area, suggest that it will have little impact on the outcomes of the study. The following inclusions and exclusions, based on the mode of communication used by the child, further describe the “communicative act:”

- Vocalizations that did not fit the above criteria were NOT considered communicative. Vocalizations that were qualitatively different than the child’s
usual vocalizations (e.g. louder, or including different phonemes) OR were combined with some other mode (e.g. directed eye gaze, change in facial expression or body orientation) were posited to meet the first criterion.

- Verbal attempts (as they are inherently social) were communicative, even if unintelligible or exact repetitions of the preceding utterance

Inter-rater reliability for judgements as to whether an act was communicative or not, was 88% overall.

**Communicative Acts: Frequency, Mode, & Function**

**Rate of Communication**

Clinicians and researchers have frequently noted that children with autism tend to communicate less frequently than their typically-developing peers (Wetherby & Prutting, 1984). Rate of communication was determined by counting the total number of communicative acts for each sample, and then dividing the total number by the length of the sample as measured in minutes. This resulted in a number with attached units of “number of communicative acts per minute.”

**Mode of Communication**

In typical development, children move from using nonverbal means of communication (e.g. vocalization with gesture) to using the verbal modality (Bates, 1976). In this study, children’s communicative acts were coded for the mode they used in order to see if they tended towards more advanced modes over time. The codes used for mode of communication were as follows:
[ME]   eye gaze toward communicative partner
[MF]   facial expression / body language / movement of body
[MG]   symbolic gesture (e.g. reach, point)
[MO]   object or other person manipulation
[MP]   picture symbols
[MS]   sign language (conventional or idiosyncratic)
[MVE]  verbal (including unintelligible attempts at words)
[MVO]  vocalization

The above codes are self-explanatory, with the possible exception of [MO], or object / person manipulation. This mode is exemplified by a child giving an object to a communication partner, by the child taking a partner’s arm and pulling her to an object of interest, or by the child pushing the partner’s outstretched hand away. Most often, as per the definition of “communicative act” given above, children received a combination of the above codes. For example, a particularly frequently occurring combination was [ME+MF+MVO], or, vocalization accompanied by changes in eye gaze and facial expression.

Data collected from the children in this area justified collapsing the codes across categories for actual analysis. For example, one child’s final coding scheme was Vocal, Picture, Sign, Verbal, and Other, because this scheme yielded the most accurate description of his actual mode of communication behaviour. Agreement between raters on this coding scheme was 92%.
Communicativeness of vocalizations

An additional analysis related to that of mode of communication was suggested by this particular group of preverbal children. All three of these children tended to vocalize frequently at Time 1, most often without being meaningful in a communicative sense. As an estimation of whether the children’s vocalizations became more meaningful over time, the number of vocalizations that were communicative in nature was divided by the total number of vocalizations, yielding a percentage of vocalizations that were communicative.

Function of Communication

Perhaps the most striking feature of communication in children with autism is the extremely restricted diversity in the reasons they communicate. Typically, they tend to communicate largely to protest or request, and much more rarely to acknowledge or comment. Many schemes have been proposed for the types of communicative functions. The scheme used in this study was chiefly derived from one used by Coggins and Carpenter (1981), as follows:

[FP] protest
[FR] request (action OR object)
[FG] greet
[FA] acknowledge
[FC] comment

While other functions of communication exist, these were the only ones that appeared in this particular set of data. Inter-rater agreement on communicative function coding was 85%.
Phonology

Babbling level

Vocal development occurs in stages, with babbling moving from being composed of vowels only to containing true, reduplicated consonants and finally to containing varied consonants within a given vocalization (as reviewed by Paul, 2001, and by McCathren, et al, 1996). This final stage of vocal development involves babbling being mixed with meaningful speech, and babbling level has indeed been found to predict the onset of speech (Stoel-Gammon & Cooper, 1984). The phonemic make-up and development of babbling, then, is roughly captured by the following three levels:

1) Level 1: utterances consist exclusively of vowels, with occasional glottal stops and/or glides

2) Level 2: vowels are interspersed with true consonants, and utterances contain at least one CV or VC sequence. Syllables may be reduplicated e.g. /ba ba ba ba/.

3) Level 3: utterances contain at least 2 true consonants differing in place or manner of articulation, e.g. /do ga do ga/.

In this study, every vocalization was phonetically transcribed for each preverbal child’s sample. Each vocalization was assigned a babbling level number, and relative frequencies of each level were computed. Broad phonetic transcription for babbling level had an inter-rater agreement of 97%.
Relevance

As discussed in the previous chapter, children with autism tend to have many difficulties in the pragmatic domain of language. In these non-speaking children, expressive measures of social relations, or relevance, are impossible to analyze. However, the roots of relevance can be seen in joint attention and in the ability to respond appropriately to others (Rollins and Snow, 1998). Therefore, communication samples of the preverbal children were analyzed in two areas of social communication: directed eye gaze and appropriateness of response.

Eye gaze

Reciprocal eye gaze between an infant and its mother is recognized as being part of normal development. Also known as eye-to-face-gaze, its main function is to regulate social interactions. This type of gaze later develops into triadic eye gaze, which occurs when a child looks at an object, looks at a communicative partner, and then looks back at the object. Triadic eye gaze is required for establishing and maintaining joint attention, and it is a skill that needs to be directly taught to many children with autism. Inadequate gaze behaviour is believed to reflect many of the impairments found in children with autism and other developmental disabilities, especially those involving social and communicative skills (Mirenda et al., 1983; Arnold et al., 2000).

In this study, rate of directed eye gaze was calculated for Jacob, who was at a preverbal level at Time I and communicated only infrequently. A moment of directed eye gaze was defined by the following criteria:
1) Child looks at face of communication partner, or at what communication partner is doing (e.g. AI throws hands in air, child looks at AI’s hands).

2) Child looks at an object his communication partner is holding. This gaze must involve either:
   - Looking at an object out of the child’s reach OR
   - Changing the direction of eye gaze towards the object

The total number of moments of directed eye gaze was tallied for each sample. These totals were divided by the length of each sample in minutes, giving a measure of the rate of directed eye gaze. Inter-rater reliability for identification of moments of eye gaze was 95%.

**Appropriateness of response**

In preverbal children, pragmatic communication problems may manifest as the child either not responding to, or responding inappropriately to, communicative attempts made by his partner. The child may appear as if he is unable to engage in turn-taking conversationally. The literature supports the notion that children with autism possess a general lack of social responsiveness (Wetherby & Prutting, 1989).

In this study, appropriateness of response was coded after the adult gave the child some command, such as “Give me red” (given when there is a red cube and a blue cube in front of the child) or “Put on” (given while AI points to the place she wants the object to be placed). To increase the likelihood that failures to respond were not failure to comprehend, multi-step commands, and non-specific questions such as “Can you do that again?” were excluded. The child’s reaction was coded according to the following scheme:
[RN] no response  
[RI] inappropriate response  
[RA] appropriate response

A code of “appropriate response” was given if the child carried out the command given by his communication partner. He had to do so before the AI gave the command again. A code of “no response” was given if the child either kept doing the behaviour he was engaged in prior to the command, or if he otherwise made no attempt to respond to the command. A code of “inappropriate response” was given if the child changed his behaviour from what he was doing prior to the command, but did not follow the command given.

An important distinction must be made here between responses that are “appropriate” and those that are “correct.” A response that is not correct may still be appropriate. Consider the example of the AI saying, “Give me red,” and the child gives her the blue cube in response. While an incorrect response, the child still followed the command in a pragmatically proper fashion. This can be contrasted with the child responding by the child doing nothing with the cubes, but rather beginning to rock back and forth and flap his hands. Such a response would be coded as inappropriate.

Despite attempts to avoid comprehension problems, it is likely that some responses that were not appropriate and/or timely still represented a language understanding-based difficulty. It is also possible that the children may have had delayed responses to the commands. However, it was noted that most of the time, these children responded within a few seconds of the direction if they were going to respond at all. As an attempt to minimize the effect of any delayed responses, the stipulation that
the child’s response (or lack thereof) had to be coded before the adult repeated the command. If there was a very short period (less than 2 seconds) between commands, that command was not included in the analysis. Percent agreement between coders for this measure was 87%.

**Vocabulary**

**Number of Different Words**

While it is true that the three preverbal children had very little, if any, productive vocabulary at Time 1, at least two of them showed that they were on the threshold of vocabulary growth. Change in this domain for these children was measured by noting the total number of words and / or number of different words produced over the course of each communication sample. It is important to note with preverbal children that “productive vocabulary” includes words produced in modes alternative to speech, namely, through the use of signs (both conventional and idiosyncratic) and picture symbols (such as are used with the Picture Exchange Communication System). Idiosyncratic signs were only coded as words if the sign was interpreted as such by the communication partner, and / or the child used the sign consistently for the same purpose, or in the same context.

2.3.3.3 Dependent Variables for Verbal Children – Language Sample Analyses

Dependent variables for the verbal children are listed according to the language skill area they measured in Table 2.5. Details for each variable, including criteria for
coding schemes and mention of use in previous research studies, are outlined below.

Coded examples of the transcripts for the verbal children are given in Appendix B.

Table 2.6: Dependent Variables for each Verbal Child

<table>
<thead>
<tr>
<th>Child Name and CA at Time 1</th>
<th>Language Domain</th>
<th>Dependent Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graham – 4;2</td>
<td>Communicative Acts</td>
<td>Rate of communication</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mode of communication</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Function of communication</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Communicativeness of vocalizations (% vocalizations that are communicative)</td>
</tr>
<tr>
<td></td>
<td>Spontaneity</td>
<td>% of verbal attempts that are imitative of preceding utterance</td>
</tr>
<tr>
<td></td>
<td>Phonology</td>
<td>Intelligibility (% of verbal attempts that are fully intelligible)</td>
</tr>
<tr>
<td></td>
<td>Vocabulary</td>
<td>MacArthur CDI (total # words produced)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Expressive One-Word Picture Vocabulary Test – standard score</td>
</tr>
<tr>
<td></td>
<td></td>
<td># of different words produced</td>
</tr>
<tr>
<td></td>
<td></td>
<td># of verbs produced</td>
</tr>
<tr>
<td></td>
<td>Expressive Communication</td>
<td>Expressive Communication subtest of PLS-3 (standard score)</td>
</tr>
<tr>
<td>Chris – 4;8</td>
<td>Spontaneity</td>
<td>% of verbal utterances that are imitative of preceding utterance</td>
</tr>
<tr>
<td></td>
<td>Phonology</td>
<td>Intelligibility (% of verbal utterances that are fully intelligible)</td>
</tr>
<tr>
<td></td>
<td>Vocabulary</td>
<td>MacArthur CDI (total # words produced)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Expressive One-Word Picture Vocabulary Test – standard score</td>
</tr>
<tr>
<td></td>
<td></td>
<td># different words</td>
</tr>
<tr>
<td></td>
<td></td>
<td># different verbs produced</td>
</tr>
<tr>
<td></td>
<td></td>
<td># auxiliary verbs produced</td>
</tr>
<tr>
<td></td>
<td>Length and Complexity</td>
<td>Mean length of utterance</td>
</tr>
<tr>
<td>-------</td>
<td>-----------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Index of Productive Syntax (IPSyn) – raw score</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Expressive Communication Subscale of PLS – standard score</td>
</tr>
<tr>
<td>Chris – 4;8</td>
<td>Relevance</td>
<td>Topic maintenance (% of utterances that maintain topic)</td>
</tr>
<tr>
<td>Brandon – 4;3</td>
<td>Phonology</td>
<td>Intelligibility (% of utterances that are fully intelligible)</td>
</tr>
<tr>
<td></td>
<td>Vocabulary</td>
<td>MacArthur CDI (total # words produced)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Expressive One-Word Picture Vocabulary Test – standard score</td>
</tr>
<tr>
<td></td>
<td></td>
<td># different words</td>
</tr>
<tr>
<td></td>
<td></td>
<td># different verbs produced</td>
</tr>
<tr>
<td></td>
<td></td>
<td># auxiliary verbs produced</td>
</tr>
<tr>
<td></td>
<td>Length &amp; Complexity</td>
<td>MLU</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DSS – raw score</td>
</tr>
<tr>
<td></td>
<td>Semantics</td>
<td>Propositional Complexity Index</td>
</tr>
<tr>
<td>Alex – 4;7</td>
<td>Spontaneity</td>
<td>% of speech consisting of exact self-repetitions</td>
</tr>
<tr>
<td></td>
<td>Phonology</td>
<td>Intelligibility (% of utterances that are fully intelligible)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Expressive One-Word Picture Vocabulary Test – standard score</td>
</tr>
<tr>
<td></td>
<td></td>
<td># different words</td>
</tr>
<tr>
<td></td>
<td></td>
<td># different verbs produced</td>
</tr>
<tr>
<td></td>
<td></td>
<td># auxiliary verbs produced</td>
</tr>
<tr>
<td></td>
<td>Length &amp; Complexity</td>
<td>MLU</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IPSyn – raw score</td>
</tr>
<tr>
<td></td>
<td>Semantics</td>
<td>Propositional Complexity Index</td>
</tr>
</tbody>
</table>
The transcripts for one verbal child were subjected to some of the analyses listed above for preverbal children's expressive communication. This was done because at Time 1, this child, Graham, still engaged in modes of communication other than verbal, and was only speaking in one-word utterances that tended to be echoic in nature. Therefore, many of the other measures (syntactic or semantic analyses) applied to the other verbal children had limited applicability to this particular child.

*Spontaneity*

**Percentage of imitative utterances**

This measure was completed with two of the children in an effort to measure whether echolalic utterances decreased over time. While it has been established that echolalia can serve many meaningful functions in the speech of children both with and without autism (Prizant & Duchan, 1981), it would nonetheless be a positive change if a child moved from a large percentage of his speech being repetitive to more of it becoming spontaneous in nature over time. The goal of intervention with children who are echolalic is to facilitate spontaneous language use. In the current study, a subset of the imitative utterances were identified for the children at both times. The selected “imitative” utterances were ones that were an exact repetition of the conversational partner’s immediately preceding utterance, either in its entirety or in its last portion. This strategy focused the subsequent analyses on those imitative utterances which were least likely to have been process at syntactic or semantic levels. By this definition, the following exchange would be coded as imitative:
AI: Okay, give it a shake.
Child: Shake.

In contrast, this exchange would be not selected as an imitative utterance.

AI: He is right here.
Child: He right here.

The total number of imitated utterances within a given sample was then divided by the total number of utterances to give a percentage of imitative speech. Previous research has shown that the amount of imitation present in even a typically-developing child's speech can be quite high at certain times in development, and that a wide range of degree of imitation exists across children. Bloom (1974) found that some young children had very low percentages of imitative utterances, while others imitated much more frequently: across the children in her study, this range was between approximately 4% and 40%. She also found that this imitation tended to be selective in nature: that each child only imitated certain lexical and syntactic forms. Reliability on identification of imitative utterances was 100%.

**Percentage of Self-Repetitions**

The data for Alex inspired this individualized analysis, as his perseveration on his own utterances was quite striking at Time 1. It was noted that it interfered in his ability be an effective communicator, and to maintain topic with his conversational partner. Spontaneity of speech in this case was determined based on the percentage of his speech that consisted of self-repetitions. Exact, immediate self-repetitions were tallied for the sample, and divided by the total number of utterances. This gave a percentage of
repeated utterances. Repetitive utterances that differed in prosody (e.g., for the sake of emphasis) were not included in this analysis.

**Phonology**

**Percentage of intelligible utterances**

All verbal children in this study had phonological problems, ranging from mild to severe, that negatively impacted their intelligibility, and therefore their ability to be adequate communication partners. As a general measure of intelligibility of speech, the children’s utterances containing unintelligible portions were tallied within each sample. Unintelligible syllables or words are coded within SALT as “X’s”. All utterances not containing any “X’s” were divided by the total number of utterances for that sample, yielding a percentage of completely intelligible utterances. Counts for both of the variables required for this calculation were performed by the SALT software, hence were considered reliable.

**Vocabulary**

**Number of different verbs**

Verbs have been found to play a crucial role in children’s syntactic and semantic development, as they provide a framework that allows the inclusion of other lexical and relational word classes. Other language-delayed populations, namely children with specific language impairment, have been found to have fewer different verbs in their productive lexicons than their typically-developing peers (Watkins, et al., 1993). Conti-Ramsden and Jones found further that children with SLI have particular difficulty with
auxiliary verbs (1997). This points towards a possibility that children with autism, who have language impairment as a core deficit, may also have a similar difficulty with verbs.

In this study, number of different words, number of different verbs, and total number of auxiliaries were used as productive vocabulary measures. These counts were obtained from SALT for all four verbal children.

Length & Complexity

MLU

Mean length of utterance has been a long-time preferred general indication of a child's morphosyntactic complexity. However, Klee and Fitzgerald (1985), among others, questioned the validity of this index of grammatical development, since their data indicated only a weak correlation between MLU level and a child's age. Also, they found that MLU did not differentiate young children's grammatical performance, leading the authors to caution against the use of MLU to separate groups of children in research programs.

A subsequent study contradicted these results, however (Rondal, et al., 1987). They demonstrated that MLU does in fact have a strong positive correlation with age, but also showed that this correlation may weaken progressively over an MLU of approximately 3.0. After this point, MLU may not be as valid a representation of grammatical complexity. This possibility was not a current concern given the anticipated language levels of the children in this study.

Part of the explanation for the declining validity of MLU can be attributed to situational and discourse effects on MLU. For example, it has been demonstrated that
MLU tends to be longest when a child is engaged in a free-play activity, as opposed to one with more inherent structure, such as picture description (Cole et al., 1989; Miller, 1981). It has also been demonstrated that children with language impairments are more likely to provide elliptical responses to examiner questions than are children with normally developing language. This means that MLU values in samples with high numbers of questions will not provide a representative picture of the child’s knowledge.

In the current study, MLU was calculated by SALT based on the entire set of each child’s utterances, for each sample.

Along with MLU, two measures of syntactic development were employed in this study: the Index of Productive Syntax (IPSyn) and Developmental Sentence Scoring (DSS). Three children’s samples were subject to one or the other of these analyses. Only one child met the DSS criterion of having 50 subject + verb combinations necessary to perform this analysis; therefore, the other two children’s samples were analyzed using the IPSyn. While both these measures are standardized, in this particular study, the coding systems of each were used without reference to the norms. Raw score values were used for the Time 1 and 2 comparisons, and, both the IPSyn and the DSS functioned in essence like criterion-referenced procedures.

Index of Productive Syntax

The IPSyn measures the grammatical complexity of preschool-aged children. It has been shown to be reliable and age-sensitive (Scarborough, 1990). It measures emergence of particular morphosyntactic forms via four subscales: noun phrases, verb phrases, questions / negations and sentence structures. This measure was developed using a language sample data set gathered from fifteen children at several points between
the age of 24 months to the age of 48 months. The IPSyn is used widely in child language research, even with its very small norming sample. As mentioned, this norming sample was not referred to in this study.

The index is relatively efficient, with the clinician or researcher only having to record two instances of each particular structure, of which there are 56. The instances are found within a 100-utterance corpus, which is specified as the “child’s first successive, intelligible utterances, excluding imitations, self-repetitions and routines” (Scarborough, 1990, p.3). For the purposes of the current study, it was necessary to broaden and extend these utterance inclusion criteria, due to the high frequency of unintelligible, imitated and/or self-repeated utterances that occurred with these two children. The modified criteria allowed: partially intelligible utterances to be included, but not those that were fully unintelligible; self-repetitions were also included unless the two utterances were completely identical, but repetitions of just the last word(s) of the conversational partner were treated as imitations and were not included. Reliability for IPSyn coding in this project was 88%.

Developmental Sentence Scoring

Although first developed more than 30 years ago, the DSS is still considered useful as an index of young children’s expressive syntax much more recently (Hughes et al., 1992). It was developed using language sample data gathered from 200 children in total: 20 children at each six-month interval between the ages of 2;0 and 6;11.

The developmental sentence score itself is derived in the following manner. Each utterance in a 50-utterance corpus (all of which must consist of a minimum of subject + verb) is assigned a weighted score based on developmental order of acquisition. The
DSS includes eight grammatical categories, which are: indefinite pronouns or noun modifiers, personal pronouns, main verbs, secondary verbs, negatives, conjunctions, interrogative reversals, and wh-questions (Lee, 1974). Early forms of each category might be scored as one or two, with the most advanced forms receiving a maximum score of eight. In addition, if a sentence is well-formed overall (i.e. no errors) both syntactically and semantically, the utterance receives a “sentence point.” Each of the 50 utterances is given a score, which are subsequently averaged, producing the developmental sentence score.

This measure has received some criticism, a large portion of which has to do with the level of skill required to complete the analysis. There have even been computer programs designed to help train individuals on developmental sentence scoring. Other criticisms tend to focus on the rules for sentence scoring employed. It has been noted that several of these rules are counter-intuitive, or weight certain morphosyntactic items too heavily, resulting in an overestimated representation of the child’s performance (Hughes, et al., 1992). Despite these criticisms, Hughes and her colleagues argue that the DSS remains a viable index of language competence. In their clinical study of ten children enrolled in a treatment program, they found the DSS to be a reliable, sensitive measure in the documentation of intervention efficacy. Reliability for DSS scoring in the current project was 89%.
Semantics

Propositional Complexity Index (PCI)

Analysis of the semantic complexity of a child’s utterances can be a useful complement to syntactic analyses in the assessment of language impaired children. This approach allows examination of the complexity of the meaning in the utterance regardless of whether the grammatical structure is correct. Johnston and Kamhi (1985), for example, found that the conversational samples of children with specific language impairment included fewer units of meaning, or propositions per sentence, than those of normally developing children at the same language level, as indexed by MLU.

The calculation of the propositional complexity index is straightforward: the number of propositions are counted for each utterance, totaled, and then divided by the number of utterances in the analysis set. In this study, PCI was calculated on a run of 75 consecutive, fully intelligible utterances. A further requirement was that the overall meaning of the utterance had to be determinable to be included in the analysis.

Relevance

Topic maintenance

As noted in Chapter 1, many children with autism have particular difficulty with the pragmatic domain of language (Tager-Flusberg & Anderson, 1991; Wilkinson, 1998). In the present study, there was one child who seemed to struggle greatly in the area of maintenance of the topic of a conversation. He was often noted to stray off topic, or produce utterances that seemed completely nonsensical. In order to investigate this, each of his utterances in both samples were coded as to whether or not they clearly failed to
maintain the current conversational topic. This approach to judging topic maintenance
avoided the difficulties associated with more finely graded systems that attempt to
identify the degree to which each utterance does maintain topic. It does not, of course,
fully describe the child's adherence to principles of relevance, but it does provide a viable
index of change in this domain. Reliability for coding this measure was 100%.

2.4 Procedures: Analysis Strategy

Each child in this set of data was given a unique analysis. While substantial
overlaps did occur in the analyses performed within each group (preverbal and verbal),
no two children received exactly the same analyses. This heterogeneity stemmed from
not only the fact that the children themselves were very different from one another in
their communication abilities (even within each group), but also from the attempt made to
tailor each child's analysis set to his communication therapy goals. This numerous and
complex set of analyses and tests, different for each child, posed a significant challenge
to synthesizing the data in order to answer the research question. The eventual strategy
used was to complete seven separate case studies, determine the degree to which each
one taken separately indicated congruence or divergence between the formal and
informal measures, and finally to summarize the seven answers to this question in a
quantified fashion.

The analysis of these data took place in four phases. These analyses were
essentially completed as seven separate case studies.

1) First, dependent variables were chosen for each individual child. As outlined
above, this was done individually in the following fashion in order to best
capture progress. Time 1 videotapes were observed for each child to obtain an overall picture of their communication styles, and the areas needing improvement. The information from these brief viewings was combined with lists of communication and language therapy targets set by each child's speech-language pathologist for the time interval in question. The final piece of information taken into consideration when choosing targets was that of the standardized tests chosen for the evaluation project. For example, since the battery of tests was heavily focused on vocabulary, measures of vocabulary were also assessed using the children's language samples in an attempt to optimize the comparison between formal and informal measures. All dependent variables were chosen before the investigator viewed the Time 2 videotapes or any of the standardized test results. In fact, the Time 2 tapes were not viewed until all analysis and coding of the Time 1 tapes was complete. This was done in an effort to avoid any bias on the choice of dependent variables as a result of how the children were communicating at Time 2.

2) The next phase of analysis consisted of describing in detail the changes that occurred for each of the seven children in both the standardized measures and the informal measures. This consisted of describing each child's performance at Time 1 on first formal, then informal measures, then repeating this description for Time 2.

3) Still on a case-by-case basis, the next phase of analysis was to describe the degree of change that occurred within each type of assessment, and to compare these degrees of progress qualitatively. With each child and each language
domain measured, the question “Do the two types of assessment provide convergent information?,” was asked.

4) Finally, the various answers to this question were collated and used to construct a more general answer to the research question.

2.5 Summary of Methods

In this chapter, the proposed design and procedures of this study were outlined. Each of the seven participants' communication characteristics was described, and the contents of their videotapes were revealed. The four formal measures chosen by the evaluation project investigators were listed. Informal measures were described in detail for both preverbal and verbal children. Finally, the procedure for analyzing the data resulting from the formal and informal measures was proposed. In the following chapter, results from all tests and informal assessments will be presented for each child in turn.
3. RESULTS

3.1 Overview

The purpose of the current study was to see whether results given by formal and informal assessment of expressive language provide convergent or divergent information regarding changes in communication behaviours over time. This study was carried out on a set of seven case studies. In this chapter, each child's results at both testing and sampling points will be presented in turn, each with a final qualitative comparison of the two data sets and a decision on the convergent-or-divergent question. Lastly, results of an analysis using non-parametric statistics to investigate the differences between the two sets of data (formal and informal) will be presented.

3.2 Preverbal Children

3.2.1 EMMETT

3.2.1.1 Time 1 Results

*Formal*

On all standardized tests at Time 1, Emmett received floor-level standard scores of 50 or less. On the MacArthur CDI measures, it was reported by Emmett's parent(s) that he understood 394 words, produced 11 words (it was not specified whether these were verbal or signed), and had acquired 48 gestures.
Informal

Communicative Acts: Frequency, Mode & Function

At Time 1, Emmett was communicating at a rate of just slightly less than once every 2 minutes on average, or .49 communicative acts per minute.

With regard to mode of communication, Emmett’s entire data set justified collapsing across codes to generate the following three ordered categories. Level 1 represented the least specific modes, or those lowest in developmental order: object manipulation and vocalizations that occurred alone. Eighty-two percent of Emmett’s communications fell into this category at Time 1. Level 2 included gestures, such as pointing or reaching. None of his communication came under this category at this time. Finally, Level 3 modes were the most specific and communicatively mature, including signs, picture use and verbal attempts. Eighteen percent of Emmett’s communicative acts were in this category, and these were all signs. Modes such as eye gaze, facial expression, and vocalizations were considered secondary, or augmentative, to these levels of analysis, and co-occurred frequently with modes contained in each of these three levels. The augmentative modes were considered as such because they seldom represented a communicative act in isolation. Although Emmett did not use the verbal modality at Time 1, more than 1 of his vocalizations bordered very closely on being verbal attempts (for example, he used the same reduplicated syllable on more than one occasion). It was decided, however, to take a conservative approach to this issue. Emmett tended to vocalize frequently, and it was found that only 13% of these vocalizations were actually communicative in nature based on the current definition.
After a communicative function diversity analysis was completed, it was found that Emmett, like many children with autism, communicated for the purposes of protesting and requesting actions over 75% of the time ("Protest" was 35%, "Request" was 41%). In contrast, Emmett never acknowledged his communication partner, and commented only 18% of the time. The remaining 6% of his communicative acts fell into the category of "Greet."

**Phonology**

When Emmett's vocalizations were transcribed phonetically, it was found that the vast majority (84%) of them were at Level 1, which consists of vowels with or without glides. Only 16% of his vocalizations were at Level 2 (with true consonants), and none were at Level 3.

**Vocabulary**

Emmett produced only 1 word in his Time 1 communication sample: he signed "more." However, as noted above, it appeared as if he was just on the threshold of making verbal attempts. The investigator took a conservative approach in coding such attempts, but on more than one occasion, it appeared as though Emmett tried to say, "I did it!" after he did completed an activity such as placing all the pieces of a puzzle.

**Relevance**

Approximately one-quarter of the time (23%), Emmett made no attempt to respond to commands given by his communicative partner. He responded inappropriately to 16% of commands (by looking away and/or beginning a stereotyped, repetitive behaviour), and appropriately to 61%.
3.2.1.2 Time 2 Results

Formal

Six months after his testing at Time 1, Emmett still received floor scores of 50 on all standardized measures. His standard scores, then, did not change from Time 1.

On the MCDI measures, surprisingly, Emmett’s number of words understood and produced both decreased. Possible reasons for this phenomenon will be discussed in Chapter 4. His number of words understood were approximately the same at Time 2, although did exhibit a slight decrease to 389, and number of words produced was 5. This was proportionately a good deal less than Time 1 (less than half). His number of gestures increased to 56 by Time 2.

Informal

Upon observation of Emmett’s Time 2 sample, he immediately appeared to have made many gains in communication and language. The most striking difference was the fact that he was now clearly making verbal attempts, with a few intelligible words. In the realm of social language, Emmett appeared to be deriving inherent pleasure from interacting with his AI, something that was not noted at Time 1. For example, he laughed quite often during a game of ball in the second sample.

Communicative Acts: Frequency, Mode & Function

Emmett’s rate of communication increased by about 20% to .61 communicative acts per minute.

Mode of communication was very different for Emmett at Time 2. Whereas he did not verbalize at all at Time 1, at Time 2, verbal utterances represented a great proportion of his communication. Level 1 communications decreased to only 32%, he
was using some gesture (Level 2 = 18%), and the category including sign and verbal utterances increased greatly to now account for half of all his modes of communication. Interestingly, with regard to signs, Emmett’s attempts at Time 2 tended to be much more spontaneous in nature than at Time 1. For example, without a model or any cueing from his AI first, Emmett produced the sign “go” when she held up a ball for him, which was a request for her to roll the ball. At Time 1, he required at least a model, and sometimes direct hand-over-hand physical cues to produce signs. Also, his signs at Time 2 were far more like approximations of American Sign Language forms than his idiosyncratic attempts at Time 1. The communicativeness of Emmett’s vocalizations did not change by Time 2, staying at 13% of the total number of vocalizations being communicative in nature.

The communicative function analysis revealed changes in a positive direction as well. Protests and requests made up less than half of his communicative acts (“Protest” was 14%, “Request” was 33%). The largest percentage of function was represented by “Comment,” which was 43%. Emmett also acknowledged his communicative partner at Time 2; this made up 10% of his communicative acts. He did not engage in greetings of any sort at Time 2.

**Phonology**

Emmett’s vocalizations had more phonological diversity at Time 2 than at Time 1. They less often contained only vocalic components (65%), and a full third of his vocalizations contained true consonants. He also had a smattering (2%) of Level 3 vocalizations, which consist of more than one consonant occurring in a babbling sequence. An example of this was the sequence /ti da di ga/. 
Vocabulary

In his Time 2 sample, Emmett produced 8 different words, some signed and some verbalized. This is a large increase from his one word at Time 1. Some of this vocabulary was in repetition of his AI, and some was completely spontaneous in nature.

Relevance

The domain of relevance showed no improvement at Time 2, as Emmett responded less often to commands by his AI (35%). He also responded appropriately slightly less often at this time, only 54% (although this still represents the highest proportion of response types). The one change in a positive direction within this domain was a slight decrease in inappropriate responses, which accounted for only 11% of Emmett’s responses at Time 2 (5% less than Time 1).

3.2.1.3 Comparison Summary: Formal and Informal Measures

Emmett’s formal measures indicated that his language ability did not change, at least positively, from Time 1 to Time 2 in the areas of general communication abilities and vocabulary. In fact, two of his scores on the MCDI actually showed a deficit at Time 2. Given the information from these tests alone, it might be tempting to conclude either that he had no functional communication skills at all, or that the therapy program was not effective in the areas of language and social communication.

The issue of what constitutes a sufficient difference, or enough difference to document that an actual developmental change has occurred, is difficult to solve with informal measures. With these types of assessment, there is no way to account for measurement
error, or random variation. Formal tools, on the other hand, do acknowledge the standard error of measurement.

In this study, absolute difference values were reported in the documentation of change from one measurement point to the next. However, a general rule of thumb was applied that if a difference was proportionally greater than 10%, then this was considered a substantial difference in the child’s score. While an improvement that was proportionally less than 10% could be due to random variation, in this study these were referred to as “slight” or “modest” changes. In Emmett’s case, for example, his rate of communication increased by an absolute value of .12 acts/minute, which translates to a percentage increase of 24% (a substantial change).

Overall, the language sample measures provided a very different portrayal of Emmett than the formal measures, at both Time 1 and at Time 2. By Time 2, he showed gains not only in the areas of general productive communication and vocabulary, but also in areas not addressed by the tests: phonology and pragmatics (as measured by communicative function analysis). Whereas both standardized measures of productive vocabulary indicated that the size of his lexicon was unchanging (at least in a positive direction), Emmett’s vocabulary as indicated by his communication sample evidenced a large percentage increase in this domain. In the area of phonology, the fact that Emmett’s babbling became more complex by Time 2, coupled with the fact that he was making verbal attempts, pointed towards him becoming verbal in his future communications. Finally, within the pragmatics domain, Emmett showed a more diverse and less “autistic” communicative function profile at Time 2. Emmett’s formal and informal results are summarized in Tables 3.1a and 3.1b.
In the case of Emmett, then, the formal and informal measures of communication provide us with *divergent* data on the effect of the therapy program.

Table 3.1a: Standard Scores on Formal Measures for Emmett (raw scores in parentheses)

<table>
<thead>
<tr>
<th>Language Domain</th>
<th>Standardized Test Name</th>
<th>TIME 1</th>
<th>TIME 2 (+ 6 mos.)</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Vocabulary</em></td>
<td>PPVT</td>
<td>40 (0)</td>
<td>40 (0)</td>
<td>No change</td>
</tr>
<tr>
<td></td>
<td>MCDI – # words understood</td>
<td>394</td>
<td>389</td>
<td>- 5 words</td>
</tr>
<tr>
<td></td>
<td>EOWPVT</td>
<td>&lt;55 (0)</td>
<td>&lt;55 (0)</td>
<td>No change</td>
</tr>
<tr>
<td></td>
<td>MCDI – # words produced</td>
<td>11</td>
<td>5</td>
<td>- 6 words</td>
</tr>
<tr>
<td><em>Receptive Language</em></td>
<td>AC subtest of PLS-3</td>
<td>50 (0)</td>
<td>50 (0)</td>
<td>No change</td>
</tr>
<tr>
<td><em>Expressive Communication</em></td>
<td>EC subtest of PLS-3</td>
<td>50 (0)</td>
<td>50 (0)</td>
<td>No change</td>
</tr>
<tr>
<td></td>
<td>MCDI – # gestures</td>
<td>48</td>
<td>56</td>
<td>+ 8 gestures</td>
</tr>
<tr>
<td><em>Global Language</em></td>
<td>TL subtest of PLS-3</td>
<td>50 (0)</td>
<td>50 (0)</td>
<td>No change</td>
</tr>
</tbody>
</table>
Table 3.1b: Communication Sample Analysis Results for Emmett

<table>
<thead>
<tr>
<th>Language Domain</th>
<th>Dependent Variable</th>
<th>TIME 1</th>
<th>TIME 2 (+ 6mos.)</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Communicative Acts</em></td>
<td>Rate of communication</td>
<td>.49 comm.acts / minute</td>
<td>.61 comm.acts / minute</td>
<td>+.12 acts/min</td>
</tr>
<tr>
<td>Mode of communication</td>
<td>Level 1</td>
<td>82%</td>
<td>32%</td>
<td>- 50%</td>
</tr>
<tr>
<td></td>
<td>Level 2</td>
<td>0%</td>
<td>18%</td>
<td>+ 18%</td>
</tr>
<tr>
<td></td>
<td>Level 3</td>
<td>18%</td>
<td>50%</td>
<td>+ 32%</td>
</tr>
<tr>
<td>Communicativeness of vocalizations</td>
<td>13%</td>
<td>13%</td>
<td>No change</td>
<td></td>
</tr>
<tr>
<td>Function of communication</td>
<td>Protest</td>
<td>35%</td>
<td>14%</td>
<td>- 21%</td>
</tr>
<tr>
<td></td>
<td>Request</td>
<td>41%</td>
<td>33%</td>
<td>- 8%</td>
</tr>
<tr>
<td></td>
<td>Greet</td>
<td>6%</td>
<td>0%</td>
<td>- 6%</td>
</tr>
<tr>
<td></td>
<td>Acknowledge</td>
<td>0%</td>
<td>10%</td>
<td>+10%</td>
</tr>
<tr>
<td></td>
<td>Comment</td>
<td>18%</td>
<td>43%</td>
<td>+25%</td>
</tr>
<tr>
<td><em>Phonology</em></td>
<td>Babbling level</td>
<td>Level 1</td>
<td>84%</td>
<td>65%</td>
</tr>
<tr>
<td></td>
<td>Level 2</td>
<td>16%</td>
<td>33%</td>
<td>+17%</td>
</tr>
<tr>
<td></td>
<td>Level 3</td>
<td>0%</td>
<td>2%</td>
<td>+ 2%</td>
</tr>
<tr>
<td><em>Vocabulary</em></td>
<td># different words</td>
<td>1</td>
<td>8</td>
<td>+ 7 words</td>
</tr>
<tr>
<td><em>Relevance</em></td>
<td>Response Appropriateness</td>
<td>No response</td>
<td>23%</td>
<td>35%</td>
</tr>
<tr>
<td></td>
<td>Inappropriate response</td>
<td>16%</td>
<td>11%</td>
<td>- 5%</td>
</tr>
<tr>
<td></td>
<td>Appropriate response</td>
<td>61%</td>
<td>54%</td>
<td>- 7%</td>
</tr>
</tbody>
</table>
3.2.2 JACOB

3.2.2.1 Time 1 Results

Formal

On formal tests of vocabulary, Jacob received the basal standard score of 50. On the PLS, he received a standard score of 50 on the Auditory Comprehension subscale and a 52 on the Expressive Communication subscale (this was slightly higher than floor level). Jacob was considered testable at Time 1.

The MCDI measures showed that Jacob could understand 59 words and produce 8 gestures, but he had not produced any words.

Informal

Communicative Acts: Frequency, Mode & Function

At this time, Jacob engaged in very little communication. His rate of communication was only approximately one communicative act every 8 minutes (0.13 communicative acts per minute).

Jacob communicated very infrequently at this time, but when he did, he did so through use of a combination of body language, directed eye gaze, and/or vocalizations. He did not use the more advanced modes of symbolic gesture, sign, or verbal utterances.

Communicative functions were highly restricted for Jacob at Time 1: he only communicated to request. Further, these were only requests for actions on the part of his AI, not for objects or information.
Phonology

At Time 1, 100% of Jacob’s vocalizations were purely vocalic in nature (i.e. all were at Level 1 of babbling). The vocalizations did vary slightly in the particular vowels used, and also in the acoustic dimensions of loudness and pitch.

Relevance

Because Jacob communicated so infrequently at Time 1, it was decided that it would be informative to also look at social behaviour considered to be a precursor to communication: eye gaze directed at his communication partner. At Time 1, Jacob’s rate of directed eye gaze was 1.7 times per minute.

It was, by far, most common for Jacob to give no response to commands given by his AI during his Time 1 communication sample (62%). He responded inappropriately only 15% of the time, and these responses tended to be that he began self-stimulating (usually vocalizing, moving his head back and forth) after the command was given. He responded appropriately to almost a quarter of the AI’s commands at this time (23%).

3.2.2.2 Time 2 Results

Formal

At the second data collection point, Jacob fared worse on his standardized measures than at Time 1. He was received a designation of “could not test” on his EOWPVT, as he did not attend to the pictures. On the PLS, he received floor scores of 50 on both subscales. A PPVT was not administered to Jacob at this time.
Similarly, his MCDI scores also generally decreased, with the number of words understood decreasing by half. His parent reported he still had no words, and he was producing only one more gesture.

Informal

In his Time 2 sample, a full year after Time 1, Jacob looked quite different in the area of social communication. He was communicating more often, was more easily engaged and directed by his AI, and showed far more affect than at Time 1 (both negative and positive). He still tended to display repetitive/stereotypic behaviour often. By this time, he was also using more advanced modes of communication (signs). On a disconcerting note, he was using injurious behaviour to communicate (biting his AI), a behaviour that was not evident in his Time 1 sample.

Communicative Acts: Frequency, Mode & Function

Jacob communicated at a quadrupled rate at Time 2 as compared to Time 1. His rate of communication was 0.57 communicative acts per minute, or just over one act every 2 minutes.

As stated above, Jacob used the more developmentally advanced mode of sign language to communicate at Time 2 (“Sign” accounted for 70% of his communicative acts). However, the signs were highly idiosyncratic and only consisted of one lexical item: “more.” The sign consisted of Jacob flickering the fingers of one of his hands against the palm of the other. Though idiosyncratic, the sign was unlike any other self-stimulatory behaviours that Jacob exhibited, and was therefore convincingly communicative. This sign for “more” did occur spontaneously, and in different communicative contexts. Twelve percent of Jacob’s communicative acts were in the
“Gesture” category, which included pointing. The mode of facial expression and body language (“Face / Body Language”), which was the only mode he used at Time 1, accounted for only 18% at Time 2.

At Time 2, Jacob communicated for three different purposes (in contrast to the one at Time 1). The majority of his communications had the function of requesting (82%). However, he also communicated for the purposes of protesting (12%) and acknowledgement (6%). Examples of these new functions of communication occurred when Jacob protested by pulling his body away and vocalizing loudly when his AI tried to help him get up from the table; or when he acknowledged his AI by pointing at something on the floor to which she had just pointed. Jacob’s communicative acts appeared to be moving in the direction of becoming more diversified by Time 2.

*Phonology*

Jacob still vocalized frequently at Time 2, and the vast majority (97%) of these continued to be vocalic in nature. He did make a few attempts at consonants (e.g. “ba”), however, something he did not do in his Time 1 sample. Level 2 of babbling, then, accounted for 3% of his total vocalizations. It was also noted at this time that Jacob’s vocalizations had more complex intonation patterns than they did at Time 1. His vocalizations tended to vary in pitch and loudness quite frequently at Time 2, which may have in part been due to an increase in the amount of protest he engaged in.

*Relevance*

Moments of directed eye gaze were more frequent at Time 2, occurring at a rate of 2.1 moments per minute. It was also noted that Jacob’s eye gaze was of a higher
“quality” at Time 2; that is, the moments were longer, and he was looking more directly at his communication partner’s eyes. He also used peripheral gaze slightly less often.

At Time 2, Jacob was responding to a higher percentage of his AI’s commands. He made some attempt to respond to 83% of the commands (inappropriate responses were 22%, appropriate responses were 61%). The remaining 17% were commands to which Jacob did not respond, but this is a lower percentage of this category than at Time 1.

3.2.2.3 Comparison Summary: Formal and Informal Measures

Similar to the case of Emmett above, Jacob’s standardized test scores suggested that he did not make progress in therapy over time. In the instance of his EOWPVT scores, we are unable to draw any conclusion at all about his skill level in this area due to the fact that he was not testable at Time 2. Curiously, like Emmett, one of his MCDI scores also decreased over time.

When we examine Jacob’s communication sample analyses, we see many areas of positive change: increased rate of communication, diversified modes and functions of communicative acts, slightly larger phonetic inventory, more frequent eye gaze directed at his communication partner, and a considerably higher percentage of responses to commands. Please refer to Tables 3.2a and 3.2b for summaries of Jacob’s formal and informal assessment results. In Jacob’s case, the information gleaned from his formal measures of language and his informal measures of language and communication diverged.
Table 3.2a: Scores on Standardized Tests for Jacob (raw scores in parentheses)

<table>
<thead>
<tr>
<th>Language Domain</th>
<th>Standardized Test Name</th>
<th>TIME 1</th>
<th>TIME 2 (+ 12 mos.)</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Vocabulary</strong></td>
<td>PPVT</td>
<td>40 (0)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>MCDI – # words understood</td>
<td>59</td>
<td>29</td>
<td>- 30 words</td>
</tr>
<tr>
<td></td>
<td>EOWPVT</td>
<td>&lt;55 (0)</td>
<td>CNT</td>
<td>Not possible to assess</td>
</tr>
<tr>
<td></td>
<td>MCDI – # words produced</td>
<td>0</td>
<td>0</td>
<td>No change</td>
</tr>
<tr>
<td><strong>Receptive Language</strong></td>
<td>AC subtest of PLS-3</td>
<td>50 (6)</td>
<td>50 (0)</td>
<td>No change</td>
</tr>
<tr>
<td><strong>Expressive Communication</strong></td>
<td>EC subtest of PLS-3</td>
<td>52 (8)</td>
<td>50 (1)</td>
<td>- 2 standard points</td>
</tr>
<tr>
<td></td>
<td>MCDI – # gestures</td>
<td>8</td>
<td>9</td>
<td>+ 1 gesture</td>
</tr>
<tr>
<td><strong>Global Language</strong></td>
<td>TL subtest of PLS-3</td>
<td>50 (14)</td>
<td>50 (1)</td>
<td>No change</td>
</tr>
</tbody>
</table>
Table 3.2b: Communication Sample Analysis Results for Jacob

<table>
<thead>
<tr>
<th>Language Domain</th>
<th>Dependent Variable</th>
<th>TIME 1</th>
<th>TIME 2 (+12 mos.)</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communicative</td>
<td>Rate of communication</td>
<td>.13 comm.acts / minute</td>
<td>.57 comm.acts / minute</td>
<td>+ .44 acts/min</td>
</tr>
<tr>
<td>Mode of communication</td>
<td>Face / Body Language</td>
<td>100%</td>
<td>18%</td>
<td>- 82%</td>
</tr>
<tr>
<td></td>
<td>Gestural</td>
<td>0%</td>
<td>12%</td>
<td>+ 12%</td>
</tr>
<tr>
<td></td>
<td>Sign</td>
<td>0%</td>
<td>70%</td>
<td>+ 70%</td>
</tr>
<tr>
<td>Function of communication</td>
<td>Protest</td>
<td>0%</td>
<td>12%</td>
<td>+ 12%</td>
</tr>
<tr>
<td></td>
<td>Request</td>
<td>100%</td>
<td>82%</td>
<td>- 18%</td>
</tr>
<tr>
<td></td>
<td>Acknowledgement</td>
<td>0%</td>
<td>6%</td>
<td>+ 6%</td>
</tr>
<tr>
<td>Phonological Development</td>
<td>Babbling level</td>
<td>Level 1</td>
<td>100%</td>
<td>97%</td>
</tr>
<tr>
<td></td>
<td>Level 2</td>
<td>0%</td>
<td>3%</td>
<td>+ 3%</td>
</tr>
<tr>
<td></td>
<td>Level 3</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Relevance</td>
<td>Rate of eye gaze</td>
<td>1.7 times / minute</td>
<td>2.1 times / minute</td>
<td>+ .40 times/min</td>
</tr>
<tr>
<td>Response Appropriateness</td>
<td>No response</td>
<td>62%</td>
<td>17%</td>
<td>- 45%</td>
</tr>
<tr>
<td></td>
<td>Inappropriate response</td>
<td>15%</td>
<td>22%</td>
<td>+ 7%</td>
</tr>
<tr>
<td></td>
<td>Appropriate response</td>
<td>23%</td>
<td>61%</td>
<td>+ 38%</td>
</tr>
</tbody>
</table>
3.2.3 WESLEY

3.2.3.1 Time 1 Results

Formal

At this data point, Wesley received a standard score of 50 on the PPVT and <55 on the EOWPVT, both of which are the lowest scores possible. On the PLS, he fared no better, receiving scores of 50 on both subscales. On the MCDI measures, Wesley was reported to comprehend 177 words, produce 4 words and use 16 gestures.

Informal

*Communicative Acts: Frequency, Mode & Function*

Wesley was found to communicate at a rate of 0.57 communicative acts per minute at this time, or just slightly more than one communicative act every 2 minutes. He was inclined to communicate in a diverse number of ways at Time 1, including a few verbal attempts that were imitative of his AI’s preceding utterance. Wesley also appeared to be just learning to use the Picture Exchange Communication System (Frost & Bondy, 1987), and was using pictures to request objects. On more than one occasion, Wesley used the mode of injurious behaviour to communicate (he struck his AI). Wesley’s mode of communication analysis was done in the same fashion as that for Emmett. At this time, Level 1 represented 37% of his communicative acts; Level 2, 21%; and Level 3, 42%. Additionally, one-quarter of Wesley’s vocalizations were communicative in nature, although as mentioned, they were frequently combined with other modes of communication.

With regard to function of communication, as might be expected, 79% of Wesley’s communications were either protests (42%) or requests (37%). An example of
a request was taking a picture of a hammer and giving it to his AI in exchange for a toy hammer. He was not observed to make comments or greetings of any kind. The remaining 21% of his acts were for the purpose of acknowledging his communication partner (e.g. his AI told him to look out the window and wave, demonstrated this behaviour, and Wesley imitated this).

**Phonology**

Some of Wesley’s vocalizations fell into each of the levels of babbling. The majority were Level 1 vocalizations, which accounted for 70%. One-quarter of Wesley’s vocalizations contained unvarying true consonants (e.g. /gu gu/). The remaining 5% were at Level 3 (i.e. contained consonants that varied within an utterance).

**Vocabulary**

At Time 1, Wesley produced only three different words; two were verbal (“hi” and “pen”), and one, “hammer,” was a Picture Communication Symbol. Each word was produced more than once, although the contexts in which each was used did not vary.

**Relevance**

Wesley most often responded in some way to his AI’s commands at Time 1. The “no response” category accounted for only 11%. Wesley did respond inappropriately quite frequently (one-third of the time). Inappropriate responses were usually engagement in repetitive or stereotypic behaviour, or looking away from the activity. He responded appropriately to the majority of commands (56%).
3.2.3.2 Time 2 Results

By Wesley's second communication sample, which was taken 18 months after his first one, he was communicating notably more frequently, and his injurious mode of communication had decreased substantially.

Formal

At the second testing point, one and a half years later, Wesley's standard scores were all unchanged compared to Time 1.

Wesley's MCDI results also showed decreases in comparison to the first time he was assessed. The reported number of words that he understood decreased by a large amount (58 words). His number of words produced changed from 4 to 0. The only formal measure that portrayed a positive change for Wesley was his reported use of gesture, which increased by a number of 8.

Informal

In his second sample, Wesley appeared to have matured both in his communication skills and his behavioural tendencies. For example, he used his Picture Exchange Communication System to request a variety of items spontaneously, and did this more often than in Time 1. He was noticeably calmer, and became upset less easily and less often. However, he did not produce any verbal language in this sample, which is perhaps surprising given the smattering of verbal attempts at Time 1, as well as the lengthy interval between the two samples for Wesley (18 months).
Communicative Acts: Frequency, Mode & Function

Wesley was communicating twice as often as at Time 1, at a rate of 1.2 communicative acts per minute. This increased frequency of communication was especially noticeable during activities that were highly reinforcing for Wesley, such as snack time.

This child also underwent changes in his mode of communication between the first and second data collection points; however, these were not all positive. Wesley’s Level 1 communicative acts actually increased to 58%, while both the other categories decreased: Level 2 to 6%, and Level 3 only slightly to 36%. Wesley used vocalizations in combination with other modes much more often. While this could be seen as a positive move towards becoming verbal, Wesley did not produce any verbal attempts at Time 2, even though they accounted for a significant proportion of his acts at Time 1. Delving a bit deeper into this analysis, there are some positive changes. For example, there was some change in the area of function in his use of Picture Communication Symbols, as he sometimes used the pictures to make choices between activities, something he did not do at Time 1. Commensurate with his increased use of vocalization, a larger proportion of Wesley’s vocalizations were produced with communicative intent: 39% of his vocalizations at Time 2 were meaningful in a communicative sense compared with 25% at Time 1. Finally, he was using signs, something Wesley did not do in his Time 1 sample.

Positive changes also occurred within the area of the reasons Wesley communicated. His profile was more diversified at Time 2, with the inclusion of the
category "Comment" (14%), a category that was not represented in his Time 1 sample. An example of this occurred when Wesley handed a food item back to his AI when it was given to him, as if to say, "I don’t want this." He also protested less: "Protest" accounted for only one-quarter of his communicative acts, in comparison to almost half of his Time 1 acts. An encouraging interaction occurred between mode and function for Wesley: not only did his protests decrease, the mode used for protesting became more socially appropriate. Instead of hitting his AI, he would whine or sometimes be more specific in his protest, like pushing an unwanted object away while looking at his AI. Wesley’s proportion of requests increased to 50%, but this may have been a function of the activity during which he was filmed at Time 2 (lunchtime). This will be discussed further in Chapter 4. Wesley acknowledged his communication partner less at Time 2 (only 11%), probably due to the fact he was no longer imitating verbal utterances.

**Phonology**

Wesley vocalized with less phonetic variation at Time 2. The overwhelming majority of his vocalizations fell into Level 1 babbling (96%). Each of the other two more advanced categories had 2% each. This was very different from Time 1, where 30% of his vocalizations contained at least one true consonant. This result is perhaps not so surprising, however, when examined in conjunction with the fact that Wesley also did not make verbal attempts at this time.

**Vocabulary**

Although Wesley did not verbalize at Time 2, he had acquired three different signs: “more,” “please,” and “help” (although this final one was idiosyncratic in form). His sign for “more,” especially, appeared to be completely spontaneous and occurred in
different communication contexts (eg. during lunch time and play). He also spontaneously used five different Picture Communication Symbols. This is an increase from his three different words at Time 1.

Relevance

Wesley also improved in this pragmatic area of communication. Over three-quarters (76%) of his responses to commands made by his AI were appropriate at Time 2. He did not respond at all only a very small percentage of the time (5%), and the remaining 19% were inappropriate responses.

3.2.3.3 Comparison and Summary: Formal and Informal Measures

Wesley’s formal test scores suggested that he made no progress over the course of 18 months of therapy. His MCDI measures were similarly discouraging, as the reported vocabulary measures actually showed a substantial decrease over this time.

While not all of Wesley’s informal measure results were entirely positive (i.e., no verbal utterances, less phonetic diversity in vocalizations), he did show some important gains in social communication over this time. He communicated more frequently, used sign language, had begun to comment as well as protested less, and usually responded appropriately to directions given by his communication partner. Perhaps the most heartening result was that Wesley was not hitting his AI in order to get his point across in his Time 2 sample. Please see Table 3.3a and 3.3b for a summary of Wesley’s formal and informal assessment results.

The formal tests and informal measures generally produced divergent results for Wesley. Interestingly, his MCDI raw score for reported number of words produced and
his mode of communication analysis produce surprising but convergent results: Wesley was no longer communicating verbally at Time 2. Possible explanations for this unexpected finding will be considered further in Chapter 4.

Table 3.3a: Scores on Standardized Tests for Wesley (raw scores in parentheses)

<table>
<thead>
<tr>
<th>Language Domain</th>
<th>Standardized Test Name</th>
<th>TIME 1</th>
<th>TIME 2 (+ 18 mos.)</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vocabulary</td>
<td>PPVT</td>
<td>40 (0)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>MCDI – # words understood</td>
<td>177</td>
<td>119</td>
<td>- 58 words</td>
</tr>
<tr>
<td></td>
<td>EOWPVT</td>
<td>&lt;55 (0)</td>
<td>&lt;55 (0)</td>
<td>No change</td>
</tr>
<tr>
<td></td>
<td>MCDI – # words produced</td>
<td>4</td>
<td>0</td>
<td>- 4 words</td>
</tr>
<tr>
<td>Receptive Language</td>
<td>AC subtest of PLS-3</td>
<td>50 (6)</td>
<td>50 (1)</td>
<td>No change</td>
</tr>
<tr>
<td>Expressive Communication</td>
<td>EC subtest of PLS-3</td>
<td>50 (8)</td>
<td>50 (6)</td>
<td>No change</td>
</tr>
<tr>
<td></td>
<td>MCDI – # gestures</td>
<td>16</td>
<td>24</td>
<td>+ 8 gestures</td>
</tr>
<tr>
<td>Global Language</td>
<td>TL subtest of PLS-3</td>
<td>50 (14)</td>
<td>50 (7)</td>
<td>No change</td>
</tr>
</tbody>
</table>
Table 3.3b: Communication Sample Analysis Results for Wesley

<table>
<thead>
<tr>
<th>Language Domain</th>
<th>Dependent Variable</th>
<th>TIME 1</th>
<th>TIME 2 (+18 mos.)</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communicative Acts</td>
<td>Rate of communication</td>
<td>.59 comm. acts /minute</td>
<td>1.2 comm. acts /minute</td>
<td>+.61 acts/min.</td>
</tr>
<tr>
<td></td>
<td>Mode of communication</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Level 1</td>
<td>37%</td>
<td>58%</td>
<td>+ 21%</td>
</tr>
<tr>
<td></td>
<td>Level 2</td>
<td>21%</td>
<td>6%</td>
<td>- 15%</td>
</tr>
<tr>
<td></td>
<td>Level 3</td>
<td>42%</td>
<td>36%</td>
<td>- 6%</td>
</tr>
<tr>
<td></td>
<td>Communicativeness of vocalizations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>25%</td>
<td>39%</td>
<td>+ 14%</td>
</tr>
<tr>
<td></td>
<td>Function of communication</td>
<td>Protest</td>
<td>42%</td>
<td>25%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Request</td>
<td>37%</td>
<td>50%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Acknowledgement</td>
<td>21%</td>
<td>11%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Comment</td>
<td>0%</td>
<td>14%</td>
</tr>
<tr>
<td>Phonology</td>
<td>Babbling level</td>
<td>Level 1</td>
<td>70%</td>
<td>96%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Level 2</td>
<td>25%</td>
<td>2%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Level 3</td>
<td>5%</td>
<td>2%</td>
</tr>
<tr>
<td>Vocabulary</td>
<td># different words</td>
<td></td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>Relevance</td>
<td>Response Appropriateness</td>
<td>No response</td>
<td>11%</td>
<td>5%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Inappropriate response</td>
<td>33%</td>
<td>19%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Appropriate response</td>
<td>56%</td>
<td>76%</td>
</tr>
</tbody>
</table>

3.2.4 Brief Summary: Preverbal Children

In the case of these three preverbal children, formal and informal measures of expressive communication diverged in the information each provided on their progress in
therapy. The informal measures, in fact, were the only ones to show any change at all in these children’s communication skills, most of which occurred in a positive direction.

3.3 Verbal Children

3.3.1 GRAHAM

3.3.1.1 Time 1 Results

Formal

At Time 1, it was not possible to administer either of the vocabulary tests to Graham due to his behavioural problems. He scored 50 on both the AC and EC portions of the PLS. His MCDI results at this time were as follows: 229 words understood, 81 words produced and 41 gestures used.

Informal

*Communicative Acts: Frequency, Mode & Function*

At Time 1, Graham was communicating at a rate of approximately twice per minute (2.3 communicative acts per minute).

Although the majority of Graham’s communicative acts were verbal (85%), he still used vocalizations fairly often to communicate (11%). He also vocalized at times other than the ones in which he was communicating; the percentage of vocalizations that were actually communicative was only 35%. His verbal utterances were usually only one word, and were frequently repetitions of the preceding utterance by his AI. The rest of Graham’s communicative acts (4%) were in the “Other” category, which consisted of object manipulation, gesture, facial expression and body language, and eye gaze.
Graham never used sign language to communicate (which is surprising given that his AI was adept at using sign with him).

Graham already possessed a fairly balanced communicative function profile at Time 1. “Protest” and “Request” only accounted for little more than one-quarter of his communicative acts. The bulk of his communications fell into the “Acknowledgement” category, as a direct result of Graham’s tendency to repeat his AI. As noted in Chapter 2, these repetitions were coded as communicative, as they were almost always coupled with eye gaze towards his AI, giving the imitation the distinct appearance of providing a turn-taking function. Graham also commented quite often; 18% of his acts of communication served this function. An example of a comment was Graham touching a red light that his AI was holding, and saying “light.”

Spontaneity

Graham was highly imitative of his AI’s utterances at Time 1. Over half (51%) of his utterances were exact repetitions of at least the last portion of his AI’s preceding sentence, and were quite often imitations of the whole utterance (his AI tended to speak in short, simple phrases).

Phonology

This child was fairly unintelligible in his first communication sample: only 57% of his utterances were fully intelligible. This is true despite the facts that he usually spoke only one or two words at a time, and that the transcribers were often aided by context since he was often repetitive.
Vocabulary

At Time 1, Graham produced 56 different words in his communication sample. Because he produced so few verbs, the measure of "number of different verbs" was altered to "total number of verbs" for this child. In this first sample, he produced nine verbs in total.

3.3.1.2 Time 2 Results

Formal

Although Graham received scores on the PPVT and EOWPVT at Time 2 (standard scores of 71 and 66, respectively), it is not possible to determine if they represent increases in vocabulary skills above what he was capable of at Time 1, due to the fact that he was untestable in these areas at that time. Had he not exhibited the behavioural problems, he might well have received the same scores as he achieved at Time 2. On the PLS, Graham scored the same as he did in Time 1 (50 on both subscales).

Because Graham had developed a productive vocabulary of more than 50 words, his parents were given the "words and sentences" form at Time 2. On this, he displayed an impressive 6-fold increase in his productive vocabulary. This is understandable given that Graham had a very long interval between Time 1 and Time 2 (18 months), and the fact that he was at the developmental stage when vocabulary growth is most rapid.

Informal

By Time 2, eighteen months after the first data collection point, Graham looked like a very different child in his communication sample. The most striking social
behavioural change was his greatly increased attention span. He was able to sit at an activity for an extended period of time without tangible reinforcement or a "movement" break. His speech was easier to understand and he seemed not to repeat as much. He really appeared to enjoy interactions with his AI, as he smiled frequently. He also exhibited a playful sense of humour a few times during this sample.

Communicative Acts: Frequency, Mode & Function

Graham communicated twice as frequently at Time 2 than he did at Time 1, at a rate of 4.5 communicative acts per minute.

Not a great deal of change occurred among Graham's mode of communication categories as compared to Time 1. We might expect this, given that his profile in this area already consisted mainly of advanced modes of communication. His verbal utterances actually decreased slightly, making up 80% of his communicative acts at Time 2. Nineteen percent of these acts were comprised of some combination including a vocalization. Many of these included gesture, as he was pointing quite frequently at Time 2. The remaining one percent of communicative acts was the less advanced modes, those included in the "Other" category.

Communicative functions remained quite diverse for Graham at Time 2, and included positive changes among the four categories. Protests and requests accounted for only 11% of communicative acts at this time. Acknowledgements, which for Graham represented his echolalia to a large extent, decreased substantially to 42% of this sample. Another encouraging change in this domain was that 47% of his communications fell into the "Comment" category at this time, a big jump from Time 1, and in a function category that is relatively rare in the communications of children with autism.
Spontaneity

Graham was decidedly less echolalic at Time 2. His imitative utterances decreased to only 28%, a value less atypical than at Time 1.

Phonology

Graham's intelligibility also substantially improved at Time 2. There was a 17% increase in intelligibility from Time 1 to 74%. It should be noted, however, that compared with age norms in SALT, this was still a significantly atypical intelligibility percentage.

Vocabulary

The number of different words produced in the Time 2 sample increased to 66. More impressive was that Graham used 35 verbs in this sample, in contrast to nine at Time 1.

3.3.1.3 Comparison and Summary: Formal and Informal Measures

Graham was not testable at Time 1. He did not receive scores on either vocabulary test, and as a result, his progress in these areas is not determinable by the use of these tests (even though he did receive scores at Time 2). The major issue here is that we cannot tell if Graham made progress or not in therapy by examining his test scores, at least in the domain of vocabulary. His MCDI measures gave much more information about his productive vocabulary: it increased greatly.

In contrast, his language sample analysis provides considerable information about his communication skills at both data points, and therefore about how he progressed with therapy. Graham made improvements in the areas of rate and function of
communication, spontaneity and intelligibility of speech, and vocabulary. Graham’s formal and informal results are summarized in Table 3.4a and 3.4b.

It is not determinable whether Graham’s vocabulary tests and commensurate language sample information converge or diverge, because he was not testable in these areas at Time 1. He did make substantial gains on his parent report measure of vocabulary, however. In other domains, Graham showed progress on his language sample measures, but did not on the PLS. For several reasons that will be discussed in Chapter 4, it appears in general that Graham’s two data sets provided divergent information.

Table 3.4a: Scores on Standardized Tests for Graham (raw scores given in parentheses)

<table>
<thead>
<tr>
<th>Language Domain</th>
<th>Standardized Test Name</th>
<th>TIME 1</th>
<th>TIME 2 (+ 18 mos.)</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Vocabulary</strong></td>
<td>PPVT</td>
<td>CNT</td>
<td>40 (71)</td>
<td>Not determinable</td>
</tr>
<tr>
<td></td>
<td>EOWPVT</td>
<td>CNT</td>
<td>30 (66)</td>
<td>Not determinable</td>
</tr>
<tr>
<td></td>
<td>MCDI – # words produced</td>
<td>81</td>
<td>529</td>
<td>+ 448 words</td>
</tr>
<tr>
<td><strong>Receptive Language</strong></td>
<td>AC subtest of PLS-3</td>
<td>50 (11)</td>
<td>50 (23)</td>
<td>No change</td>
</tr>
<tr>
<td><strong>Expressive Communication</strong></td>
<td>EC subtest of PLS-3</td>
<td>50 (10)</td>
<td>50 (19)</td>
<td>No change</td>
</tr>
<tr>
<td><strong>Global Language</strong></td>
<td>TL subtest of PLS-3</td>
<td>50 (21)</td>
<td>50 (42)</td>
<td>No change</td>
</tr>
</tbody>
</table>

CNT = Could not test
Table 3.4b: Communication Sample Analysis Results for Graham

<table>
<thead>
<tr>
<th>Language Domain</th>
<th>Dependent Variable</th>
<th>TIME 1</th>
<th>TIME 2 (+18 mos.)</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Communication Acts</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rate of communication</td>
<td>2.3 comm. acts / minute</td>
<td>4.5 comm. acts / minute</td>
<td>+ 2.2 acts / min.</td>
</tr>
<tr>
<td>Mode of</td>
<td>Vocal</td>
<td>11%</td>
<td>19%</td>
<td>+ 8%</td>
</tr>
<tr>
<td>communication</td>
<td>Verbal</td>
<td>85%</td>
<td>80%</td>
<td>- 5%</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>4%</td>
<td>1%</td>
<td>- 3%</td>
</tr>
<tr>
<td></td>
<td>Communicativeness of vocalizations</td>
<td>35%</td>
<td>46%</td>
<td>+ 11%</td>
</tr>
<tr>
<td>Function of</td>
<td>Protest</td>
<td>7%</td>
<td>7%</td>
<td>No change</td>
</tr>
<tr>
<td>communication</td>
<td>Request</td>
<td>21%</td>
<td>4%</td>
<td>- 17%</td>
</tr>
<tr>
<td></td>
<td>Acknowledgement</td>
<td>54%</td>
<td>42%</td>
<td>- 12%</td>
</tr>
<tr>
<td></td>
<td>Comment</td>
<td>18%</td>
<td>47%</td>
<td>+ 29%</td>
</tr>
<tr>
<td></td>
<td>Spontaneity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Imitativeness of utterances</td>
<td>51%</td>
<td>28%</td>
<td>- 23%</td>
</tr>
<tr>
<td>Phonology</td>
<td>Overall intelligibility of utterances</td>
<td>57%</td>
<td>74%</td>
<td>+ 17%</td>
</tr>
<tr>
<td>Vocabulary</td>
<td># different words</td>
<td>56</td>
<td>66</td>
<td>+ 10 words</td>
</tr>
<tr>
<td></td>
<td>Total # of verbs used</td>
<td>9</td>
<td>35</td>
<td>+ 26 verbs</td>
</tr>
</tbody>
</table>

3.3.2 CHRIS

3.3.2.1 Time 1 Results

Formal

Chris earned standard scores of 86 on his PPVT and 78 on his EOWPVT at the first assessment point. It is evident from these test results that vocabulary knowledge is an area of relative strength for Chris. More general measures of language skill were quite weak, however, as he received floor level scores of 50 on both subscales of the PLS.
Even within the vocabulary domain, comprehension scores were notably better than production for Chris (he was at the low end of normal on the PPVT, but below normal on the EOWPVT). On the MCDI, it was reported that Chris produced 418 words.

**Informal**

**Spontaneity**

At Time 1, Chris exhibited some tendency to repeat his AI's utterances. While he did not do this too often (11%), it was still a noticeable problem because of what he imitated and because of the quality of the imitations. That is, he repeated only long utterances, usually with a very similar intonation pattern to his AI or his sister, whomever he was imitating. He also repeated these utterances many times afterwards with the same intonation. Although this self-repetitive pattern was not considered imitative by the definition put forth in this study, it nevertheless had an aberrant quality.

**Phonology**

Chris was only moderately intelligible at Time 1, due mostly to minor articulatory problems, and sometimes to a tendency to mumble. Ninety-one percent of his utterances were completely intelligible.

**Vocabulary**

In the 30-minute sample, Chris produced 157 different words, 22 different verbs, and two auxiliaries in total. He produced the modal auxiliary “can” several times as well; however, this was excluded from the analysis due to the routinized nature of the utterances containing this form ("Can I have ___, please?").
**Length and Complexity**

Chris had an MLU at Time 1 of 2.86. This may have been inflated slightly due to a large number of routine utterances of the form “Can I have a ____ , please?” His IPSyn raw score was 41. Areas of weakness within the IPSyn centred around verbs: the “Verb Phrases” subscale had the lowest percentage of points given for items.

3.3.2.2 Time 2 Results

**Formal**

Chris made several improvements in his test scores over the 6-month interval. Most impressive were his gains in receptive language measures, both on the PPVT and the Auditory Comprehension subscale of the PLS. On the PPVT, he scored 10 more raw points than at Time 1, although this did not result in much of a change to his standard score. However, on the AC scale of the PLS, Chris increased his raw score by 13 points, which translated into an increase in his standard score of 20 points (i.e., it went from a floor value of 50 to a score of 70).

Formal expressive measures did not show as much of an increase. On both the EOWPVT and the Expressive Subscale of the PLS, he gained only three raw points, translating into either a decrease or no change in his standard scores. On the number of words he produced as reported on the MCDI, Chris increased a small amount (approximately 19%).

**Informal**

At Time 2, only 6 months later, Chris was imitating less, and was noted to be using less routinized speech. He also had more diversity in his sentence structure.
Spontaneity

Chris’s use of imitative speech decreased considerably in his Time 2 language sample. At this time, only 4% of his speech was imitative of preceding utterances.

Phonology

The intelligibility of Chris’s speech did not improve appreciably in his second sample (a 1% increase over Time 1 to 92%), but was essentially at ceiling at both Times.

Vocabulary

Chris improved on all three measures of vocabulary in his Time 2 language sample. The number of different words was 179 (a 22 word increase), number of different verbs was 25, and total number of auxiliaries used was 11. This increase in his use of auxiliaries was encouraging, as it increased from just two at Time 1.

Length and Complexity

At Time 2, Chris’s MLU decreased somewhat to 2.47. As noted above, it is possible that his MLU for Time 1 was slightly inflated. His IPSyn score, however, increased quite dramatically (to 54; an increase of 13 raw points). His score on the “Verb Phrases” subscale doubled, and the score on “Sentence Structures” increased considerably as well.

3.3.2.3 Comparison and Summary: Formal and Informal Measures

Overall, Chris’s standard scores on all tests suggested either no change over this 6-month period, or a slight change in either a positive or negative direction. The only exceptions to this were his Auditory Comprehension subscale score, which jumped up by 20 points, and his MCDI score, which to 496 words.
His language sample measures reveal that Chris progressed in a number of expressive language domains. Specifically, Chris made improvements in amount of spontaneous speech, vocabulary (especially auxiliary verb use), and certain areas of syntactic complexity (i.e., more diverse verb phrases and sentence structures). See Table 3.5a and Table 3.5b for a summary of Chris's formal and informal results. It seems that standardized measures and language sample measures are generally divergent in the case of Chris.

Table 3.5a: Scores on Standardized Tests for Chris (raw scores given in parentheses)

<table>
<thead>
<tr>
<th>Language Domain</th>
<th>Standardized Test Name</th>
<th>TIME 1</th>
<th>TIME 2 (+ 6 mos.)</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Vocabulary</strong></td>
<td>PPVT</td>
<td>86 (44)</td>
<td>89 (54)</td>
<td>+ 3 standard points</td>
</tr>
<tr>
<td></td>
<td>EOWPVT</td>
<td>78 (32)</td>
<td>76 (35)</td>
<td>- 2 standard points</td>
</tr>
<tr>
<td></td>
<td>MCDI – # words produced</td>
<td>418</td>
<td>496</td>
<td>+ 78 words</td>
</tr>
<tr>
<td><strong>Receptive Language</strong></td>
<td>AC subtest of PLS-3</td>
<td>50 (22)</td>
<td>70 (35)</td>
<td>+ 20 standard points (&gt;1 S.D.)</td>
</tr>
<tr>
<td><strong>Expressive Communication</strong></td>
<td>EC subtest of PLS-3</td>
<td>50 (19)</td>
<td>50 (22)</td>
<td>No change</td>
</tr>
<tr>
<td><strong>Global Language</strong></td>
<td>TL subtest of PLS-3</td>
<td>50 (41)</td>
<td>50 (57)</td>
<td>No change</td>
</tr>
</tbody>
</table>
Table 3.5b: Language Sample Analysis Results for Chris

<table>
<thead>
<tr>
<th>Language Domain</th>
<th>Dependent Variable</th>
<th>TIME 1</th>
<th>TIME 2 (+ 6 mos.)</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spontaneity</td>
<td>% of utterances that are imitative</td>
<td>11%</td>
<td>4%</td>
<td>- 7%</td>
</tr>
<tr>
<td>Phonology</td>
<td>% of utterances that are fully intelligible</td>
<td>91%</td>
<td>92%</td>
<td>+ 1%</td>
</tr>
<tr>
<td>Vocabulary</td>
<td># different words</td>
<td>157</td>
<td>179</td>
<td>+ 22 words</td>
</tr>
<tr>
<td></td>
<td># different verbs</td>
<td>22</td>
<td>25</td>
<td>+ 3 verbs</td>
</tr>
<tr>
<td></td>
<td>Total # auxiliaries</td>
<td>2</td>
<td>11</td>
<td>+ 9 auxiliaries</td>
</tr>
<tr>
<td>Length and Complexity</td>
<td>MLU</td>
<td>2.86</td>
<td>2.47</td>
<td>-.39</td>
</tr>
<tr>
<td></td>
<td>IPSyn (raw score)</td>
<td>41</td>
<td>54</td>
<td>+ 13 raw points</td>
</tr>
</tbody>
</table>

3.3.3 BRANDON

3.3.3.1 Time 1 Results

Formal

At his first data collection point, Brandon had standardized test scores that were in the average range on measures of comprehension (PPVT = 97; AC subscale of PLS = 98). Relative to this, his scores on measures of expressive vocabulary and language were lower (EOWPVT = 80; EC subscale of PLS = 63). His parent reported via the MCDI that he produced 462 words at this time.

Informal

Phonology

At Time 1, Brandon was fairly unintelligible. This was likely due to a combination of several factors: phonological differences, a propensity to mumble, a
tendency to speak quickly, aberrant prosodic patterns, and an occasional pitch change in his voice to a very high register. Only 80% of his utterances were fully intelligible. Many of his completely unintelligible utterances were very long strings, which only added to the peculiarity of his speech.

**Vocabulary**

Brandon produced many word roots, including verbs, at Time 1. He produced 122 different words in only 100 utterances. In this same set, he uttered 26 different verbs, and a total of 15 auxiliary verbs.

**Length and Complexity**

At this time, Brandon had an MLU of 3.61, which indicates an average utterance length of between three and four words. His sentences lacked variability and complexity as measured by the DSS, where he scored only 5.04. Brandon repeated himself (or at least repeated the same type of syntactic frame) quite frequently, and he had a good deal of difficulty with certain morphosyntactic structures such as pronouns (either using the wrong one or not using one at all) and the verb *to be*, both in copula and in auxiliary forms. Brandon only received half of the “sentence points” available, due to frequently incorrect structures or structures that were semantically odd.

**Semantics**

A propositional complexity index was assessed for Brandon. It was calculated to be 1.64, meaning that he produced well under two semantic propositions per utterance.

**Relevance**

In his first communication sample, Brandon was noted to produce several utterances that did not make sense either semantically or pragmatically. For example, he
often began speaking on a new topic without alerting his conversational partner that he was doing so. This resulted in a good deal of confusion on the part of his AI. These utterances often contained pronouns with no obvious prior reference. In one instance of this sort, his AI made a comment about the puzzle they were doing together, and Brandon said, "It's always on," referring to the light on the videocamera that they had not been talking about for quite some time. In an effort to capture this pragmatically disruptive trait, Brandon’s utterances were coded as maintaining topic or not maintaining topic. At Time 1, it was found that a fairly substantial percentage of his speech did not make sense in this manner: only 90% of his utterances at least partially maintained the ongoing topic. Stated another way, 1 out of every 10 utterances that Brandon spoke was potentially highly disruptive to the flow of the conversation. This was quite disconcerting, given that Brandon spoke at a rate of 11 utterances per minute.

3.3.3.2 Time 2 Results

Formal

Brandon showed considerable improvement in his standard scores, especially on the expressive measures. Although he did not make gains in his receptive language as shown by the PLS (AC subscale), he was already within the average range on both this measure and the PPVT. He gained more than 1 standard deviation in his score on the EOWPVT, and more than 2 standard deviations on his Expressive Communication score of the PLS.
Informal

In his Time 2 sample (1 year after Time 1), Brandon appeared to have matured in his behaviour, and he was far less verbose. His speech was also noticeably more intelligible.

Phonology

By Time 2, Brandon had improved quite substantially in this domain. Most notable, the long strings of unintelligible words were no longer evident. His longest string of “X’s” was only three words. He was 91% intelligible at this time.

Vocabulary

Brandon underwent very little change in this language area when his Time 1 and 2 samples were compared. However, this was already a relative area of strength for Brandon.

Length and Complexity

Brandon’s MLU had actually decreased quite considerably at Time 2 relative to Time 1, from 3.61 to 3.17. This may in part be due to the sampling context. In Time 1, Brandon was alone with his AI. In Time 2, he shared the session with a peer, an AI, and a speech-language pathologist. His AI had been especially good in Time 1 about not asking Brandon many questions; but in Time 2, the activities in which they participated made questions from both the AI and the S-LP more common. For example, the group played a board game through part of the sample. Although long utterances were less common in Brandon at this time, they were noted to be grammatically correct, an observation supported by his DSS score.
In contrast to MLU, Brandon's Developmental Sentence Score was much higher at Time 2: he achieved a score of 7.38. This was consistent with the comment above, in that his utterances were indeed better constructed syntactically. When the individual components of the DSS were examined, Brandon had made substantial gains in the "Secondary Verbs" portion, as well as the "Conjunctions" section. An example of this was, "But I want to break the whole track." While he had attempted many secondary verbs at Time 1, he rarely got points for them because, for example, they were often missing the infinitival "to" (e.g. "I want look at pictures"). His points in this area increased 6-fold at Time 2, because his attempts were always correct and because he was using more advanced forms (e.g. "How do you get to go back down?").

**Semantics**

Brandon had a substantially larger Propositional Complexity Index at Time 2, that of 2.51. This means that, on average, Brandon produced more than two propositions per utterance. For example, he said, "I worked at a farm in Thailand too." This is intriguing in light of his smaller MLU value at this time.

**Relevance**

Topic maintenance was also an area of improvement for Brandon in his second language sample. His utterances were found to maintain the conversational topic 96% of the time.
3.3.3.3 Comparison and Summary: Formal and Informal Measures

Brandon demonstrated sizable gains on all formal language measures. This was especially true for the Expressive subscale of the PLS. On both this measure and the EOWPVT, Brandon went from below average to average in his standard score.

He also demonstrated gains to a comparable degree in his language sample measures, especially on select measures of syntactic and semantic complexity. His informal measure results also showed impressive improvement in areas not addressed by the formal tests, but that were noted to be some of the most problematic for Brandon in the domain of communication: intelligibility and pragmatics. Refer to Table 3.6a and 3.7b for summaries of both Brandon’s formal and informal results.

In the case of Brandon, the information given by both formal and informal measures generally converge.

Table 3.6a: Standard Scores on Tests for Brandon (raw scores given in parentheses)

<table>
<thead>
<tr>
<th>Language Domain</th>
<th>Standardized Test Name</th>
<th>TIME 1</th>
<th>TIME 2 (+12 mos.)</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Vocabulary</strong></td>
<td>PPVT</td>
<td>97</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>EOWPVT</td>
<td>80 (31)</td>
<td>97 (54)</td>
<td>+ 17 standard points (&gt;1 S.D.)</td>
</tr>
<tr>
<td></td>
<td>MCDI – # words produced</td>
<td>462</td>
<td>478</td>
<td>+ 103%</td>
</tr>
<tr>
<td><strong>Receptive Language</strong></td>
<td>AC subtest of PLS-3</td>
<td>98</td>
<td>95</td>
<td>- 3 standard points</td>
</tr>
<tr>
<td><strong>Expressive Communication</strong></td>
<td>EC subtest of PLS-3</td>
<td>63</td>
<td>99</td>
<td>+ 36 standard points (&gt;2 S.D.)</td>
</tr>
<tr>
<td><strong>Global Language</strong></td>
<td>TL subtest of PLS-3</td>
<td>78</td>
<td>97</td>
<td>+ 19 standard points (&gt;1 S.D.)</td>
</tr>
</tbody>
</table>
Table 3.6b: Language Sample Analysis Results for Brandon

<table>
<thead>
<tr>
<th>Language Domain</th>
<th>Dependent Variable</th>
<th>TIME 1</th>
<th>TIME 2 (+12 mos.)</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Phonology</strong></td>
<td>% of intelligible utterances</td>
<td>80%</td>
<td>91%</td>
<td>+11%</td>
</tr>
<tr>
<td><strong>Vocabulary</strong></td>
<td>* # different words</td>
<td>122</td>
<td>128</td>
<td>+ 6 words</td>
</tr>
<tr>
<td></td>
<td>* # different verbs</td>
<td>26</td>
<td>28</td>
<td>+ 2 verbs</td>
</tr>
<tr>
<td></td>
<td>* Total # auxiliaries</td>
<td>15</td>
<td>14</td>
<td>-1 auxiliary</td>
</tr>
<tr>
<td><strong>Length and Complexity</strong></td>
<td>MLU</td>
<td>3.61</td>
<td>3.17</td>
<td>- .44</td>
</tr>
<tr>
<td></td>
<td>DSS (raw score)</td>
<td>5.04</td>
<td>7.38</td>
<td>+ 2.34</td>
</tr>
<tr>
<td><strong>Semantics</strong></td>
<td>* Propositional Complexity Index</td>
<td>1.64</td>
<td>2.51</td>
<td>+ .87</td>
</tr>
<tr>
<td><strong>Relevance</strong></td>
<td>% of utterances completely maintaining topic</td>
<td>90%</td>
<td>96%</td>
<td>+ 6%</td>
</tr>
</tbody>
</table>

- These measures based on 100-utterance set

### 3.3.4 ALEX

#### 3.3.4.1 Time 1 Results

**Formal**

At his first testing point, most scores that Alex achieved were within the average range (standard scores: PPVT = 94, EOWPVT = 91, PLS-AC = 88). The one exception was his PLS Expressive Communication subscale score (66), which was surprisingly low compared to the rest of his results (i.e., more than two standard deviations below average). His parent reported that he used 443 words at this time.
Informal

Spontaneity

Alex was inclined to repeat his own utterances often at Time 1. Eighteen percent of his utterances were self-repetitions.

Phonology

During this first sample, Alex’s speech was very difficult to understand. This was due almost entirely to severe phonological difficulties. Interestingly, at times Alex’s speech seemed even less intelligible, as he had a propensity to resort to high-pitched “baby talk.” An example of this occurred when he was counting items several times. During part of the counting segment, Alex was able to pronounce the numbers correctly, but on another portion, he recited, “One, two, wee, pour, pie.” Alex’s intelligibility was only 78% at Time 1.

Vocabulary

Alex produced 113 different words, 18 different verbs, and five auxiliary verbs in a subset of utterances at Time 1. It seems, then, that verbs were an area of relative weakness for Alex.

Length and Complexity

Alex had an MLU of 2.68 at Time 1. His IPSyn score was 64. This total score was relatively evenly distributed across the four subscales of the measure, with the exception of receiving points for a higher proportion of the “Noun Phrases” subscale.

Semantics

At this time, Alex had a Propositional Complexity Index of 1.75 (under two propositions per utterance). A typical utterance for him at Time 1 was “I like ketchup.”
3.3.4.2 Time 2 Results

**Formal**

In contrast to the other children in this study, Alex made large gains in all his standard test scores. This was especially remarkable on the Expressive Communication portion of the PLS, where Alex went from two standard deviations below the mean to well within the average range. His number of words produced as measured by the MCDI increased moderately to 555.

**Informal**

Commensurate with his test scores, Alex had made clear gains in his language sample. He was appeared to have much improved intelligibility, spoke in more complex sentences, and had the appearance of having matured both behaviourally and in his speech patterns.

**Spontaneity**

In his Time 2 sample, Alex exhibited much improvement with regard to the spontaneity of his speech. Not only had his percentage of self-repetitive speech decreased considerably overall (by 10%); in addition, when he repeated himself, it was only once or twice. This is in comparison to Time 1, when it was not uncommon for him to repeat the same utterance four or five times.

**Phonology**

In Alex’s second language sample, he was far more intelligible than in his first sample, with 91% of his utterances being complete and fully intelligible. Further, he did not exhibit any of the high-pitched “baby-talk” that was present at Time 1.
Vocabulary

Alex showed increases in two of his language sample vocabulary measures at the second data point: both verb measures (diversity and auxiliary count). His verb repertoire doubled, and his use of auxiliary verbs quadrupled. His number of different words did not increase appreciably, but we might expect this given that Alex’s vocabulary was not a particular area of weakness at Time 1 (based on the language sample measure, his MCDI score and his EOWPVT score).

Length and Complexity

At Time 2, Alex’s MLU was 3.17, a sizeable increase over his Time 1 value of .49. He also evidenced a large increase in his IPSyn raw score from 64 to 84. In this IPSyn, his “Verb Phrases” scale score came out particularly high, which might be expected considering his verb diversity and usage measures improvements outlined above.

Semantics

Alex’s PCI value showed a slight increase at Time 2, to 1.83 propositions per utterance.

3.3.4.3 Comparison and Summary: Formal and Informal Measures

Alex’s formal measures demonstrated the substantial improvements he made in several different areas of language production. He exhibited gains in standard scores on all of the formal measures.

His language sample measures in general showed a similar degree of improvement in the variables addressed by the tests, those of vocabulary and syntax. He
also showed progress in measures not covered by tests: spontaneity and intelligibility of speech. Please refer to Table 3.7a and 3.7b for a summary of these results.

In the case of Alex, then, the information gleaned from formal and informal data on his progress over time were generally convergent.

Table 3.7a: Standard Scores on Tests for Alex

<table>
<thead>
<tr>
<th>Language Domain</th>
<th>Standardized Test Name</th>
<th>TIME 1</th>
<th>TIME 2 (+ 12 mos.)</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Vocabulary</strong></td>
<td>PPVT</td>
<td>94 (51)</td>
<td>99 (72)</td>
<td>+ 5 standard points</td>
</tr>
<tr>
<td></td>
<td>EOWPVT</td>
<td>91 (41)</td>
<td>106 (64)</td>
<td>+ 15 standard points (1 S.D.)</td>
</tr>
<tr>
<td></td>
<td>MCDI – # words produced</td>
<td>443</td>
<td>555</td>
<td>+ 112 words</td>
</tr>
<tr>
<td><strong>Receptive Language</strong></td>
<td>AC subtest of PLS-3</td>
<td>88 (40)</td>
<td>103 (45)</td>
<td>+ 15 standard points (1 S.D.)</td>
</tr>
<tr>
<td><strong>Expressive Communication</strong></td>
<td>EC subtest of PLS-3</td>
<td>66 (29)</td>
<td>93 (43)</td>
<td>+ 27 standard points (&lt;2 S.D.)</td>
</tr>
<tr>
<td><strong>Global Language</strong></td>
<td>TL subtest of PLS-3</td>
<td>74 (69)</td>
<td>96 (88)</td>
<td>+ 22 standard points (1.5 S.D.)</td>
</tr>
</tbody>
</table>
Table 3.7b: Language Sample Analysis Results for Alex

<table>
<thead>
<tr>
<th>Language Domain</th>
<th>Dependent Variable</th>
<th>TIME 1</th>
<th>TIME 2 (± 12 mos.)</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spontaneity</td>
<td>% of perseverative speech (self-repetitions)</td>
<td>18%</td>
<td>8%</td>
<td>-10%</td>
</tr>
<tr>
<td>Phonology</td>
<td>% of intelligible utterances</td>
<td>78%</td>
<td>91%</td>
<td>+13%</td>
</tr>
<tr>
<td>Vocabulary</td>
<td>* # different words</td>
<td>113</td>
<td>114</td>
<td>+1 word</td>
</tr>
<tr>
<td></td>
<td>* # different verbs</td>
<td>18</td>
<td>32</td>
<td>+14 verbs</td>
</tr>
<tr>
<td></td>
<td>* Total # auxiliaries</td>
<td>5</td>
<td>22</td>
<td>+17 auxiliaries</td>
</tr>
<tr>
<td>Length and</td>
<td>* MLU</td>
<td>2.68</td>
<td>3.17</td>
<td>+.49</td>
</tr>
<tr>
<td>Complexity</td>
<td>IPSyn (raw score)</td>
<td>64</td>
<td>84</td>
<td>+20 raw points</td>
</tr>
<tr>
<td>Semantics</td>
<td>Propositional Complexity Index</td>
<td>1.75</td>
<td>1.83</td>
<td>+.08</td>
</tr>
</tbody>
</table>

* Samples yielded very different numbers of utterances; these measures were all based on a 130-utterance run

3.3.5 Brief Summary: Verbal Children

Within the group of children who were already using verbal language at Time 1, there were mixed results with regard to the convergence or divergence of the two types of measures of progress in the therapy program. Two of the children’s (Graham and Chris) data generally diverged from one assessment mode to the other, while for the other two children (Brandon and Alex), data from the two assessment approaches converged. Mixed results also occurred within children, but across language skill domains.
3.4 Therapy Goal Results

As previously outlined (see Chapter 2), each child's communication goals for therapy as set by their speech-language pathologist helped to determine which communication variables would be measured through informal assessment. An interesting question was whether or not the specific therapy targets improved over the interval in question, as measured by communication sample analysis. This question is answered below.

The number of therapy targets, across all children, that were directly addressed in this study was 15. Of these, 12 showed at least some degree of improvement, and many showed large amounts of improvement over the course of this study. Therefore, 80% of therapy goals set by speech-language pathologists that were directly measured by informal assessment showed improvement. For example, one therapy goal for Jacob was to improve eye gaze behaviours necessary for joint attention, and his informal communication sample results indicated that he did improve in this targeted area.
3.5 Summary of Results

Table 3.8 provides a general summary of the findings in this study. The table addresses the following question: *Do formal and informal measures of expressive language converge in the information each provides about progress in therapy?*

As can be seen from Table 3.8a, even if the phonological domain is excluded from the analysis, the two assessment methods lead to divergent conclusions about progress in 59% of the comparison points. It appears as though this lack of convergence occurs especially with preverbal children, and within language domains not addressed by formal tests. These apparent differences will be explored further in the following chapter.
Table 3.8a: Summary of Results – Do formal and informal expressive language measures *converge* on the presence / direction of progress?

<table>
<thead>
<tr>
<th>CHILD</th>
<th>COM</th>
<th>SPON</th>
<th>REL</th>
<th>PHON</th>
<th>VOCAB</th>
<th>L / C</th>
<th>SEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emmett</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jacob</td>
<td>N</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wesley</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Graham</td>
<td>N</td>
<td>N</td>
<td></td>
<td></td>
<td>N</td>
<td>Y</td>
<td></td>
</tr>
<tr>
<td>Chris</td>
<td>N</td>
<td></td>
<td>N</td>
<td>N</td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brandon</td>
<td></td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Alex</td>
<td></td>
<td>N</td>
<td></td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
</tbody>
</table>

**Light grey** = Formal tests not available  
**Dark grey** = Formal tests available but not given  
**Y** = Convergence in presence / direction of progress indicated by both measures  
**N** = No convergence in presence / direction of progress indicated by both measures  
**Blank square** = area not assessed for that child
4. DISCUSSION

4.1 Summary of Results

4.1.1 Brief Summary of Findings

This study was designed to determine whether formal and informal language production assessment tools would provide convergent or divergent information about speech and language therapy outcomes for a group of young children with autism. For children who were preverbal, the data provided a relatively straightforward answer: formal and informal measures always diverged. In every case, informal measures showed improving communication abilities, whereas formal measures revealed either no change or a decrease in skills over time. However, for verbal children, the results were not so clear-cut. The two data sets were sometimes convergent and sometimes divergent. This was true across the four verbal children, as well as across domains for a given child.

It is noted here that this study was not designed to determine program efficacy per se, but only to evaluate different assessment procedures that could contribute to such studies in the future. In this context, it is important that a high percentage of specific therapy targets, as set by each child’s speech-language pathologist, showed progress as measured by informal assessment tools.

A second caveat in the interpretation of these results is the focus of this study on the extremes of assessment measures. For the sake of argument, only the most formal of formal assessments, standardized tests, and the most informal of informal measures, communication sample analysis, were examined in this study. Assessment methods
falling within the gray area between these two types of methodologies, such as observational checklists, were not addressed.

4.1.2 Comparison with Other Studies

As reviewed in Chapter 1, portions of the present study have been addressed in the research literature. For example, measurement of progress in therapy using both types of assessment has been documented within the literature (Fey et al., 1993), as has measurement of autistic children using both types of measures (Condouris et al., 2003). In this latter study, the authors focused directly on the methodological aspect of assessment. They asked how standardized tests, such as the PPVT, the Expressive Vocabulary Test, and the Clinical Evaluation of Language Fundamentals correlated with measures derived from spontaneous language samples in their ability to detect language impairment in lexical-semantic and grammatical domains. Their chosen informal measures were number of different word roots, MLU, and IPSyn. All assessments were administered to 44 children with autism, whose mean age was 7;3 (range 4-14) and whose average IQ was in the normal range. They found that the standardized tests and informal assessments correlated significantly in the identification of language impairment in the areas of content and structure. In the present study, findings for the verbal children are generally in line with those of Condouris et al (2003) in that there was some degree of agreement between formal and informal measures in the areas of vocabulary and sentence structure. However, overall, the paper by Condouris and her colleagues differs from the present study in four key ways: (1) they did not measure progress over time as a result of treatment, (2) the children were older, (3) the children were all verbal (with an average
IQ that was on the low end of normal), and (4) the researchers did not examine pragmatic skills in the children.

There has not yet been a published study comparing the two types of assessment in the measurement of progress in children with autism.

**4.1.3 Discussion Overview**

In the first section of this chapter, methodological characteristics particular to this study will be considered and their impact on the validity of the data will be presented. Features to be discussed will include the reliability of measures, communication sample characteristics and timing of samples, and standardization. This section will also consider the significance of several peculiar findings. The second major section of the paper will address the possible factors contributing to whether the two data sets converged. These factors involve the nature of the assessment measures themselves, and the language level of the children. Last, the clinical implications of this research as well as limitations of this study and potential courses for future research will be addressed.

**4.2 Methodological Issues**

The discrepancies found between formal and informal measures of progress could be attributed to any one or more of five factors: the nature of the chosen variables, unreliability of the coding schemes, sampling context characteristics, idiosyncrasies of particular children or data, or inherent differences in standard and raw scores. The following section considers each of these possibilities in turn. The factors considered in
this section are inherent to the design of this particular study, and do not speak to the nature of the assessment methodologies themselves.

4.2.1 Range and Selection of Variables

The variables chosen for the current study were fairly comprehensive, in that they addressed both treatment goals plus the language domains covered by standardized tests. They also allowed a thorough description of each child’s language functioning in several different domains. However, the measures chosen did not approximate an exhaustive set of measures. Rather, the variables were limited to those conforming to the three criteria outlined previously: (1) they were in accordance with how the child was communicating in his first sample; (2) they directly assessed the children’s treatment goals; and, (3) when possible, they examined the language behaviours also measured by the formal assessments. For example, the “Communicative Acts” domain was not addressed in any of the children who were at the sentence level of expressive language development. This does not mean that it could not have been assessed; the children’s rate of verbal attempts, modes of communication, and speech act functions could have been analyzed. However, within this particular group of children, neither the investigator nor the children’s speech-language pathologists had any particular concerns about the amount they were communicating, nor their purposes for doing so. Note, however, that if “Communicative Acts” variables had been assessed for the older children, this would have increased the number of discrepant findings since no formal measures were available in this domain. Further, variables were chosen with clinical applicability in mind: the measures were simple to complete, and many took a relatively short period of time (as in the topic
maintenance variable where utterances were coded either as maintaining or clearly not maintaining topic). A possible disadvantage to this approach was that, by keeping things simple, some amount of detail in the information was inevitably lost.

4.2.2 Coding Reliability

No problems of reliability arose during the course of this study, either in transcription or in the coding analysis. Within the coding of the different variables, the inter-rater reliability agreements varied quite considerably, although all were well within an acceptable range (from 83% to 100%). This was due to the nature of the variables themselves. Some measures were clearly objective, such as identification of utterances that were repetitions of the previous ones, and thus yielded perfect reliability values. Others were not so straightforward: identification of communicative acts, for example, included taking into consideration a few different criteria before making a judgment. The lowest percent agreement value (83%) occurred with the “Appropriateness of Response” variable, where raters sometimes disagreed upon whether an incorrect response was inappropriate or not.

4.2.3 Data Idiosyncrasies

Some surprising results occurred that might call the credibility of some of the measures into question. These issues will be discussed here as they occur within the language domains of vocabulary and length and structural and semantic complexity.

**Vocabulary**

In all three of the preverbal children, number of words either understood or
produced (or both) decreased as measured by parent report on the MCDI. This is especially surprising given that all three children improved in several domains of language, including vocabulary in one case, on the informal measures. Decrease in the number of words understood could be accounted for by the difficulties inherent in the evaluation of language comprehension in very young children, especially by their parents. Bates (2003), for example, noted that the comprehension portion of the measure does not stack up as well against other measures of language comprehension (i.e. poor construct validity) as the production portion. Decline in expressive language scores is more difficult to dismiss. Why might there be this result, where the number of words produced on the MCDI opposes both the natural course of development and the information gleaned from the other measures employed? Research to date (Bates, 2003) indicates that while the MCDI does show the same variability in toddler language that has been seen in other measures over the longer term (Fenson, 2000), month-to-month and split-half reliability for expressive vocabulary are in the 95% range. These findings suggest we must look beyond the psychometric properties of the measure to answer this question.

Two plausible explanations exist for this unexpected result. First, the two children whose productive vocabulary decreased may actually have regressed with regard to expressive language skills over the time interval. This occurrence has been documented in a small percentage of children within the scientific literature (Werner, 2003; Wilson et al., 2003). The data do not, however, support this explanation for the vocabulary decreases observed here. At one level, the MCDI results for Wesley (which went from four words produced to zero) are congruent with his informal measures and
seem to suggest regression. In his first communication sample, he was making definite verbal attempts and this was not the case in his Time 2 sample. However, at Time 2 he actually increased in his use of alternative modes of communication (sign, picture communication symbols). This pattern of change seems to indicate some new preference in communication mode rather than a true language regression since symbolic communication continued to occur. For Emmett, there is little uncertainty. At Time 1 he was using signs and at best, one unanalyzed verbal phrase; in his Time 2 sample, he continued to use signs and produced several intelligible words. Although phonological analysis indicated increased syllabic complexity, his speech remained immature and many utterances, thought potentially verbal, were not recognizable. These data, while indicating significant delays, are clearly incompatible with developmental regression.

The second possible explanation for the decline in MCDI vocabulary scores is that parents may have simply overestimated their child’s productive lexicon at the first data collection point and/or underestimated it at the later point. Given the communication patterns for Emmett and Wesley described above, the latter possibility seems very real. Emmett’s phonological limitations could have prevented his parents from recognizing his words. Even more likely, since the MCDI instructions focus particularly on “words”, parents of both of these children did not report use of unconventional symbols, such as picture symbols or signs (they were not instructed to do so).

Length & Complexity / Semantics

A second unexpected result occurred in the interaction between Brandon’s sentence length and his propositional complexity. While his MLU decreased quite substantially at Time 2, his propositional complexity index value increased considerably.
If we take sentence length as a rough index of grammatical complexity, this finding is unexpected since usually the two measures would be positively correlated. To attempt to explain this occurrence, we must revisit Brandon’s Time 1 sample, which was characterized by long, winding, nonsensical utterances (e.g. “And the guy’s (is) scared of in and a bugged no it (and a what) and a what is these?”). Although the utterances made little sense semantically, they were nonetheless included in Brandon’s MLU calculation as long as they were “complete and intelligible” by SALT’s definition. In contrast, utterances that did not make sense were not included in the PCI analysis, which means that many of his long utterances were excluded, making the ones left in the analysis short and low in their number of propositions. While this analysis indicates that MLU can be misleading, other measures of grammatical growth proved to be more robust. At Time 2, Brandon had very few long, nonsensical utterances but showed evidence of improved syntactic (as seen by his DSS) and semantic complexity overall. A growing literature argues that caution is needed in the use and interpretation of MLU as an index of language development, especially with children who have language impairments (Klee et al., 1989; Johnston et al., 1993; Johnston, 2001).

4.2.4 Communication Sample Characteristics

Upon initial analysis, the videotaped communication samples appeared to be less than ideal in a variety of ways. First, they were designed for a purpose other than this research, namely, for the private use of the speech-language pathologists and autism interventionists at the children’s EIBI program. Due to this, the usual clinical guidelines about types of materials and activities to be employed in the samples were not controlled.
Materials tended to vary from the first sample to the second and at times did not seem ideal for encouragement of language and communication. An example of this was the use of toys that were not novel or interesting to the child. Activities also tended to be quite structured, especially for the preverbal children. Most tapes consisted of mainly intervention sessions, rather than free play situations.

Ideal samples for the documentation of progress should remain the same from one collection time to the next, and should be highly representative of the child's communication in everyday contexts. At first glance, the above description of these samples seems to point away from both these recommendations, indicating that the samples were less likely to capture progress. However, although the samples did vary with regard to the activities from one data collection point to the next, the degree of structure imposed did not change overall. That is, the children seemed to have a sampling of both structured and unstructured activities at both data collection points. In fact, the preverbal children’s activities were almost always structured in nature. For instance, if there was a case in which the child had a highly structured situation at Time 1, and a free-play situation at Time 2, this would indeed necessitate caution in the interpretation of communication improvements as being due to progress in therapy. In this study, this was never the case.

Ensuring representativeness is the other criterion to consider in the collection of samples that will best showcase gains in therapy over time. The characteristics outlined above (non-novel materials and structured activities) seem to contradict this principle of representativeness. However, there were also positive aspects in this regard to the contents of the videotapes. For instance, children were always interacting with a highly
familiar adult (their AI’s or their parents), something that is of great importance to obtaining a truly natural communication sample in children with autism. In most cases, this familiar other stayed the same from Time 1 to Time 2. Also, as previously noted, children with autism probably interact more when some degree of structure is imposed. This is especially true for the preverbal children, who all had low rates of spontaneous communication. Structure might have benefited some of the verbal children as well, especially the ones who tended to have prominent behaviour problems. The sampling of a few different activities and contexts across the two data collection points can even be seen as an advantage in the maximization of representativeness. That is, the children were seen as they interacted within a variety of activities, that ranged from structured matching tasks to free play with water toys or a favourite model train; as well as in a few different contexts, such as lunch time, sessions with peers, and intervention sessions carried out both at home and at the centre where the program was based. These characteristics of the samples demonstrate that they were, in three important ways, sufficiently representative of the children’s communicative functioning in many different real-life contexts.

On consideration of the previous points, it is difficult to determine whether the sampling context was suitable or not. The fact that all children showed a direction of movement from Time 1 to Time 2 that followed a developmental path does seem to indicate, however, that the nature of the samples was not a serious problem. Therefore, the samples were adequate in that the children were all, upon both observation and analysis, unequivocally more mature in many areas of their language and communication at the second data point.
As mentioned, these sample characteristics were due to the tapes having been made for a purpose other than the present study. Another consequence of this was that the availability of samples that were usable for this study was governed by when the testing dates were set for the EIBI project. Tapes were not obtained at every testing period, and as a result, some children had short intervals between their two data collection times and others had long intervals. Since the children varied quite widely in the amount of time between the first and second data collection points (between 6 and 18 months), it is possible that this difference had an impact on their ability to show progress on the measures, perhaps with the formal measures being more compromised in their ability to detect change by a short time interval than the informal ones. To examine this, we asked whether the children who did not show progress on standardized measures but did so on informal measures, were also the ones with a shorter interval between Time 1 and Time 2. In fact, this was not the case. All five children who clearly showed more progress on informal than formal measures were representative of the whole range of time intervals. Two had intervals of 6 months between samples, another had a one year interval, and the remaining two had intervals of 18 months.

4.2.3 The Effects of Standardization

In this study, standardized tests were used in the exact way that they were meant to be used. For instance, all standardized administration procedures were strictly adhered to for all tests. Also, in presenting the data for the children derived from the tests (MCDI excluded), only standard scores were reported. There is a possible inherent problem with this, however, since standard scores tend to hide change over time. Because the child
gets chronologically older, their later scores are compared against a set of norms for children who are at an even higher stage of language development. Therefore, the child must actually accelerate in rate of learning over time in order to produce an increase in standard scores. If there is no change in a child’s rate of learning, this will be reflected as a decrease in his standard score. Although two of the children in this study (Brandon and Alex) showed impressive gains in many of their standard scores, the majority of the children did not. Some even showed slight decreases in their scores. To what degree can the findings of this study be attributed to the inherent nature of standard scores? One way to answer this question is to see whether or not the children showed change in their raw scores; in other words, treat standardized scores in the same way we treat informal measures.

Five children showed little or no change in the majority of their standard scores. Examination of the raw scores for these children indicates that some of the preverbal or one-word stage verbal children’s raw scores on standardized tests increased from Time 1 to Time 2, while others decreased. For example, Graham’s raw score on the Expressive Communication subtest of the PLS went from 10 to 19, while Jacob’s score on this same subscale decreased from 8 to 1. Additionally, Chris (one of the children who was verbal at Time 1) showed small improvements in raw scores across most of his tests, but not enough to translate into changes in standard scores. So, even though these children showed no change in formal scores as they are meant to be reported, their raw scores reflect the fact that change did occur and could be measured with the formal assessment tools.
For comparison's sake, what happens if we treat informal measures as formal ones? This would only be possible when looking at the informal measures that include norms, such as the IPSyn and the DSS. The IPSyn was used to analyze two children's samples in this study, Chris and Alex. They both made proportionally large raw score gains from their first samples to their second, with Alex making a larger amount of gain than Chris. When Chris's scores are compared to the age equivalence data, we find that he made about 2 months of progress over the 6-month interval. Alex, on the other hand, made 13 months of gain in the year between his two scores. If we treat these children's raw IPSyn scores in a standardized fashion (by comparing their scores to the age-equivalence norms), we see the same general pattern of improvement as was seen in the raw test scores. In addition, Brandon's samples were analyzed using the DSS, on which he showed a substantial gain in his overall sentence score from Time 1 to Time 2. If these scores are compared to the developmental norms (i.e. treated in a standardized fashion), we also see an impressive gain for this child. At Time 1, his score falls well below the 10th percentile; while at Time 2, his score lies in the middle of the 10th percentile and the 25th percentile. Essentially, these percentiles indicate that Brandon's developmental sentence score went from below average to approximately the low end of average over the course of a year.

In sum, it appears as though roughly the same estimates of gain are obtained if we treat informal (i.e. targeted) measures as formal ones, while this is less true if tests are looked at in an informal manner. This suggests that, at least for these children, tests do not capture progress as dependably as the informal tools. It also implies that the results
of this study cannot be viewed as simply a by-product of the use, or lack of use, of standard scores.

4.2.4 Summary of Methodological Issues

The possible roles of three methodological factors in contributing to the findings of this study were investigated. These factors, related to the chosen variables, the characteristics of the samples, and the effects of using standard scores, were specific to the design of this study and did not involve the underlying nature of the assessment methods themselves. The variables chosen here obviously influenced the findings to some degree, but this may have more to do with the inherent nature of the tools, as discussed in the next section. We found that all data idiosyncrasies had viable explanations, and so the credibility of the measures themselves need not be questioned. Communication sample characteristics as an influencing factor on the results were more difficult to reject, but since a positive developmental trajectory from one sample to the next was observed for all children, whatever influence occurred does not appear to have jeopardized the purposes of this study. Variant time interval factors were found not to be a data-influencing factor. Finally, upon comparison, it was found that use of standard test scores (as opposed to raw) did not substantially sway the results one way or another. From this discussion, then, we can reasonably conclude that differences in the data probably did not result from this particular research design, but more from the intrinsic characteristics of the assessment methods and children themselves. We will examine these possibilities below.
4.3 Issues of Convergence vs. Divergence

In the majority of cases in this study, data produced by formal and informal measures did not converge with regard to the information each provided about the children's progress in therapy. The previous section dealt with possible contributing factors resulting from the design of this study itself, which did not seem to fully explain the pattern occurring in the data. We will now focus on the essential characteristics of the assessment tools, as well as factors related to the language level of the children as they both contribute to this question of convergence.

4.3.1 Factors Contributing to Convergence Question

4.3.1.1 Assessment tool factors

Availability and use of tests across domains

In many cases where the data between the two types of measurement did not converge, it was due to one or the other of the types not even addressing a particular expressive language domain. For example, the formal measures did not address the domain of "Phonology." This area is measurable by formal tests, such as the Goldman-Fristoe Test of Articulation (Goldman & Fristoe, 2000), but the investigators of the EIBI evaluation project chose not use these measures. This might have been due to the high level of skill and time required to transcribe language samples phonetically, or perhaps because the researchers chose to focus exclusively on content and form aspects of language.

The second domain omitted from analysis by formal measures was that of "Pragmatics". The emphasis that the researchers put on the content and structure of
language was not an oversight; as mentioned previously, formal tests are not available to measure true pragmatic function. In this study, this included all of the variables under the “Communicative” and the “Relevance” domains. For the preverbal children, these were the two main components of their communication sample analyses, as these types of variables were the most relevant level of analysis for them. In Chapter 1, it was briefly mentioned that there is a certain incompatibility inherent in the notion of testing language in use, or pragmatics. An adequate measure of pragmatics requires that there be communicative purpose, as exists in everyday conversation – a requirement that is difficult to meet in a testing situation. Language-in-use behaviours, by their very nature, cannot be elicited in the standardized manner of a test. Tests such as the CSBS are arguably successful in eliciting requests, but fall short in creating compelling elicitation contexts for comments and other speech acts.

If a standardized test of pragmatics could even be designed, its creators would run into more difficulties upon selecting a norming sample. Keeping in mind that pragmatics is the use of language in a particular time and place for a specific purpose, norms would conceivably have to encompass all the possible characteristics of different examiners (e.g. separate norms for both genders), or different testing locations (home, clinic, etc.). This would be required because acceptable social communicative behaviour varies widely depending on the situation. For example, far less communication would be expected from a child in an unfamiliar location, or with an unfamiliar person in the room (i.e., it is completely appropriate to be quiet around a stranger), than if the situation was comfortable and known to the child. Given all this, it seems highly unlikely that a satisfactory option for testing within the domain pragmatics will ever be possible.
For children with autism, then, whose core communicative deficits centre around this domain, informal assessments will continue to be the measures of choice in this area. Again, children with autism have difficulties not with just one or two areas within the domain of pragmatics, but in many of these areas. Problematic areas may include a general deficiency in amount of spontaneous communication, a restricted repertoire of communicative purposes, lack of precursor behaviours to social communication (such as eye gaze), and difficulty maintaining conversational topics or perseveration on a favourite topic. It is in the area of pragmatics that assessment is most crucial.

Finally, although this study focused on language production, it should be noted that the formal measures did address one domain that was not covered by the informal measures, that of comprehension of language. This is an essential part of language functioning, and one that is not easily assessed using language and communication sample analyses, at least not with completely naturalistic data. Comprehension probes, however, may be an ideal solution to this problem. Utilized in context, such probes can tell us a great deal about a child’s everyday comprehension skills (Chiat, 2000).

Comprehension probes could not be constructed and evaluated for this study, as the sample tapes were collected for other purposes.

In sum, if we examine lack of convergence between the measures in all areas of expressive language domains, a major contributing factor lies in the content of the formal tests. Most especially, the tests chosen did not include an analysis of the important area of pragmatics, whereas informal measures were well suited to assess many aspects of this domain.
Sensitivity

Another possible reason for lack of convergence between the two methods, also related to the methods themselves, could be that informal measures are more sensitive to change. The data certainly support this suggestion, as introduced above in the section where we treated formal scores as informal ones, and vice versa, and found that the informal methodology did capture differences that the standardized tests failed to see. Two reasons may explain this greater sensitivity. First, the communication sample analyses focused directly on areas of children’s language that were most likely to improve, based on information about their child’s level of communicative functioning at the start of the study and goals set by their S-LPs.

Second, within the samples, the children all had numerous opportunities to produce each of the targeted communication behaviours; this is in striking contrast to the situation in most standardized tests. A demonstration of this point can be provided from the current study. Brandon had only two opportunities to produce the progressive morpheme /-ing/ on the Preschool Language Scale. He produced this same form 15 times in a 25-minute language sample. This highlights the fact that one of the disadvantages of formal assessment is that children do not always produce the language behaviours in question when we try to elicit them, as we do in a testing format. Instead, children are quite complex and inconsistent in their communication behaviours, due to their varying control over a particular language skill. For example, new or infrequent forms are less likely to appear spontaneously, and are therefore highly variable in their appearance in a child’s speech. Variability may also be due to a child’s changing moods or varying states of anxiety or distractibility (these may be especially likely factors contributing to the
unpredictability of children with autism). This variability necessitates a larger sampling of behaviour in order to maximize the likelihood of capturing the language form of interest. Children with autism are almost certainly more variable in their communicative behaviours than their typically-developing peers, and so require this large sample of behaviour to an even greater extent.

4.3.1.2 Child factors: language level

Two groups of children emerged in this study based on language and communication level: those who were at a preverbal stage of development, and those who used speech as their main mode of communication. The difference found between the groups in amount of convergence was at least partly due to the children's language level. All three nonverbal children had divergent results. The child at the one-word verbal level, as well at the child at the 2- to 3-word utterance level, had a small proportion of their results converge. The two highest-level children, both of whom produced complex sentences, had relatively more of their results converge. These differences are likely due to a few factors, including the nature of the two assessment types, the language level of the children, and behavioural issues.

Although at least two of the differences between the two assessment measures were outlined previously (content and sensitivity of the measures), the differences as they relate directly to the children in this study will be addressed now. There are two issues here: the prerequisite language level needed to be successful in a testing session, and the prerequisite behaviours that are necessary. The ability to complete formal language tests is based on a prerequisite level of both language and behaviour. In other words, the children must at least be able to understand the directions given to them in order to be
able to succeed on any items of a test. At least three of the children in this study may not have been at this level. In most of the formal measures in this study (e.g. the EOWPVT), language production is also required to attain a score that is comparable to a norming sample of age-matched peers. In the one test where no spoken language is required from the children (the PPVT), they must be able to use a communicative gesture, that of pointing their finger at a picture. This ability presumes a developmental level that some of the preverbal children had not yet attained. By definition, if a child is verbal, they almost certainly have the prerequisite knowledge needed to earn a score above basement level on language tests. We cannot confidently make the same claim about preverbal children, who likely do not possess all such prerequisite skills.

Behaviourally, children must be able to initiate and maintain attention to the task, and must be able to respond correctly to directions given to them. Test-taking behaviour was documented as a problem with the preverbal children in this study on several occasions: the S-LP could not conduct testing because the child “would not attend” to the test materials, or because the child had a general lack of responsivity. Behavioural problems such as tantrumming or engaging in stereotyped, repetitive behaviours would also make a reliable testing session an impossibility. For Wesley and Graham in the present study, these behaviours resulted in either the children not being tested at all, or in too many of their test scores ending up at floor level. These children both exemplified cases where formal and informal measures did not converge.

4.3.1.3 Summary of factors

The above discussion does not provide any simple answers about the factors contributing to whether or not formal and informal measures converge on treatment
progress information. However, we can conclude the following: factors related to the methodological design of the study probably did not contribute to the outcomes to a great extent, whereas factors related to the assessment measures and language domains explored, as well as to the language and behavioural levels of the children each made substantial contributions to whether or not the two data sets converged.

4.3.2 Does Convergence Equal Redundancy?

The cases where the two data sets converged were highly verbal children, in language domains that were addressed by both formal and informal measures. It might be tempting to conclude from these results that the two methods provide equivalent views of the content and structure of language in children with autism who are functioning at a higher level of verbal competency, and by extension that the two methods would provide equivalent data on progress. However, such conclusions would be premature.

The issue is one of *convergence versus redundancy*. In other words, does the fact that the data converge necessarily mean that the two measures produce the same, or redundant, information about a particular child's progress? Can one measure simply be substituted for another in the case of children with autism who are verbally competent? If so, a clinician or researcher may always choose to use a formal test rather than a language sample, since such tests are readily available and relatively easy to administer. To examine this issue further, an example from the current study may be helpful.

Alex arguably made the greatest gains (at least as measured by formal assessments) of any child in this study. His standard scores increased in every measured domain, sometimes by more than one standard deviation. From these positive changes in
standard scores, we might conclude that both Alex’s receptive and expressive vocabulary (as measured by the PPVT, the EOWPVT, and the MCDI) improved, and that his general auditory comprehension and expressive communication skills also improved (as indicated by the PLS). Indeed, it is true that Alex moved from being well below average in comparison to his peers on the Expressive Communication subscale of the PLS to being within the average range over the course of the year of intervention.

Alex’s informal measures also showed impressive amounts of gain over the time interval that the study took place. Again the data indicate improvements in the area of vocabulary. But with the measures taken from his language sample, we now see that the improvement was not in the total size of his productive lexicon, but occurred more specifically in a sizable increase in his number and use of both lexical and auxiliary verbs. Alex also showed great improvements in the structural aspects of language on his informal measures. In particular, we can see from his IPSyn analysis that while many areas of his syntactic complexity improved, the most notable was again his use of verb phrases. The informal measures go beyond the findings of the formal measures to indicate that Alex also made progress in areas that were deemed very high priorities within his speech and language therapy: spontaneity and phonology. Although he tended to be quite self-repetitive at Time 1, he used more spontaneous speech at Time 2. While phonological issues impeded Alex’s ability to be an effective communicator a great deal at Time 1, his intelligibility improved considerably by Time 2.

As noted previously, Alex’s formal and informal measures generally provided convergent information. However, as is clear from the summary of his improvements above, these measures, while convergent, did not provide redundant information. The
informal measures provided detailed, rich and specific information about the areas in which he improved. Formal tests, on the other hand, provided a general summary of how Alex performed relative to his peers in a sampling of broad areas of language skill. These two very different sorts of information may have each indicated progress, but the equivalency ends there. Clinical use of one or the other assessment methods will largely depend upon the purpose of the assessment.

4.3.3 Summary of Convergence Issues

It seems that the most satisfying explanation for the divergent set of data comes from two sources: the inability of tests to measure pragmatic domain of language, and the language level and behavioural tendencies of the children.

4.4 Usage of Standardized Tests

4.4.1 Usage in the Measurement of Progress

The findings in this study strongly suggest that standardized tests have limited value as measures of progress in therapy programs, especially for the population of preschool-aged, preverbal children with autism, for two reasons. First, although this project made use of several of the most widely used tests, they provided very limited information about the communicative functioning of children with autism at a preverbal stage of development. Furthermore, their lack of sensitivity and descriptive power at the early stages made them ineffective in assessing these children's progress in therapy over time. Finally, since the tests provided no useful information about the preverbal
children's communicative functioning, they could not provide a profile of children's strengths and deficits to guide a treatment plan.

4.4.2 Usage with Children who have Autism

A second very real limitation of standardized tests lies in the necessity of the child's compliance. Many children with autism, preverbal or verbal, have behavioural problems that make test-taking difficult. A case in point is that two of the seven children in this study (Wesley and Graham) were not able to be tested at all for at least their baseline sessions for the EIBI evaluation project. This made it impossible to assess the children's communicative functioning over time. With communication sample analysis, on the other hand, a great deal of information was obtained about both of these children's language and communication skills. This may in part reflect the fact that with informal assessment procedures, the child does not know he is being assessed. In fact, it was often noted that within the samples in this study, the children appeared to be enjoying themselves. This point is especially valuable for those young children with autism who are inclined to confrontational social interactions.

Despite these factually based arguments, many clinicians continue to rely on tests for measuring progress and guiding therapy programs. Kerr and her colleagues (2003), found that among clinicians working with school-aged students, formal testing was ranked as the most important and frequently used assessment tool for establishing therapy goals as well as for measuring progress in therapy. At the least, these practice patterns would seem to compromise the clinician's ability to address accountability. If a clinician wants to definitively show that a particular type of therapy has been effective for a young,
preverbal child with autism, she or he will be more successful with informal assessment tools.

For both preverbal and verbal children in this study, informal measures were superior for providing rich, detailed, clinically helpful portraits of children’s language capabilities, as well as how they changed over time. This was due to the fact that the profile of measures used was crafted for each individual child by drawing upon observational information and his own communication goals. In this way, clinicians can provide meaningful, ongoing assessment of communication behaviours and language skills that are of direct relevance to an individual child without requiring compliance. Standardized tests cannot give this type of information, as they are designed to measure a broad, shallow sampling of a larger number of language skills in ways that require the child’s cooperation.

If the clinician’s purposes require standardized assessment with a child with autism, methods other than tests should also be used. For example, observational checklists such as the *Functional Communication Profile* are available. Further, if the clinician uses a test, some may be more appropriate for this population of children than others. An example is the Communication and Symbolic Behavioural Scales (Wetherby & Prizant, 1993), which includes an analysis of precommunicative behaviours and uses a videotaped sample component as part of the tool. This tool is standardized (both in its administration protocol and in the inclusion of norms), but was specifically designed to be used with young children who show atypical patterns of communicative development.
4.4.3 Best Practice Guidelines

A few clinically relevant points from the previous discussion are summarized below:

➢ As the sole measure of language progress as a result of therapy, standardized tests are not the ideal choice. Informal assessments such as criterion-referenced measures and communication sample analysis are the best choices for measuring language change in therapy.

➢ Formal testing provides little useful information about the communicative functioning of children who are preverbal, particularly those who have autism. Communication sample analysis should be employed with these children.

➢ Informal assessment is a good choice with children who exhibit behaviours that make test-taking difficult.

➢ If a clinician or researcher must use a test, some may be more appropriate than others for the population of preschoolers who have autism; these include, for example, observational checklists or the standardized Communication and Symbolic Behavioral Scales.

4.5 Limitations of Current Study & Directions for Future Research

4.5.1 Sample Size

An obvious limitation of this study is the very small sample size used. In order to determine whether these results represent generalizable differences between data
obtained by formal versus informal assessments, a larger sample of children with autism is required. Also, measurement of progress at several data points would allow for growth curve analysis. In other words, this study needs to be replicated with a larger sample, and in a way that formal statistical analyses can be performed in order to determine if these results represent real differences between the two types of measures, or if they occurred by chance.

4.5.2 Methodological Concerns

Another criticism of the design of this study has to do with the nature of the videotaped communication samples. As discussed earlier, the fact that these were created for purposes other than this particular project, while not a fatal flaw, does raise questions of representativeness. This research, if replicated, should make use of samples that are more ideal in the areas of activities and materials used. The samples should also employ these same (or similar, taking into account children’s changing interests as they develop) activities and materials at each data collection point. It is probable that this study, which included more structured activities than is ideal for a language sample, negatively impacted the verbal children more than the preverbal ones. On observation, it was noted that some of the verbal children talked more and used longer utterances in general when engaged in free play situations than when in more structured activities. In contrast, the preverbal children, as mentioned previously, probably did engage in more communicative acts while guided by their Autism Interventionists than they would have if left to their own devices.
Consistency in activities and materials (from one data collection point to another) was another factor that was not controlled within this study. Therefore, we must be cautious in interpretation of the results, as a child who appeared to have greatly improved from the first sampling session to the second may have, in part, had a sampling situation on the latter occasion that resulted in slightly more optimal communicative opportunities.

In sum, more research should be conducted where suitable materials and unstructured activities are kept consistent for the duration of the study.

4.6 Conclusion

The findings of the present study were that informal and formal measures of assessment most often produced divergent information about the progress made in therapy by young children with autism. Two reasons likely accounted for this result. First, there are inherent differences in the nature of the two measures, including the inability of tests to adequately measure the domain of pragmatics and the ability of informal procedures to assess in a focused, indepth manner, ultimately allowing them to be more sensitive to change over time. The other major contributing factor to the results was the level of the children's development in both language and behaviour. Some children were simply not able to perform on standardized tests for one or both of these reasons.

These results suggest that the measurement of progress as a result of therapy in young children with autism is optimally captured by informal procedures, rather than formal tests. This is especially true for children at very early, preverbal stages of
language development, as well as for children whose behaviour makes testing sessions
difficult and unproductive. For all children, the depth of information gleaned from
informal measures makes them a preferred tool for assessment of progress in therapy.
REFERENCES


Appendix A
Sample Coded Transcript for a Preverbal Child:
Wesley, Time 2

The following sample transcript contains segments in bold type, which are the communicative acts on the part of the child. Codes are defined in Chapter 2 of this manuscript.

=W is seated at a table facing the camera, B is sitting at right angles to W, sideways to the camera. It is lunchtime. B has a picture symbol book in front of him and is setting out picture symbol options on the cover page for W to choose from representing the different kinds of food in his lunch. W watches him. B has W's lunch kit.

W {picks up the picture representing cheese and hands it to Brian} [C]
[FR] [MP].
B Cheese.
W {signs "more" then touches the container with both hands in "more" position and signs "more" again with brief moment of eye contact} [C]
[FR] [MS].
B Say help {takes W's hands and makes the sign for "help"} please {guides W's hand to sign please i.e. pat on chest}.
= taping paused (there is a break)
B That's what he asked for.
B He asked for cheese {unclear who he is talking to, but it's not W}.
=W is eating the cheese
= tape appears to have been paused momentarily twice, W is still eating cheese
O Hey, Brian?
O You want to just flash me the book?
O Just show me the front of it.
B Okay, sure.
B I'll make it pretty.
= B holds book up and proceeds to turn page by page, W is still eating the cheese
O Perfect.
B The extras are in there {points to the book}.
B These are the rest of William's pictures for his daily schedule.
B There you go.
O Perfect.
= B sets the book back down facing William.
W {immediately pulls off a picture of ham, hesitates, looks toward B (no eye contact) then puts it back; then pulls off another picture, B holds out his hand, W hesitates then hands it to B} [C] [FR] [MS].
B Ham.
B All right.
B Here's your ham {hands W a baggie with ham in it}.
=W opens the baggie and eats the ham
- 00:07:03
= B opens book and pulls out another picture and puts on the front of the book
= B appears to adjust W's chair it is out of camera view, however
B then rubs W's shoulder which seems to elicit a smile from W although it is difficult to be certain because he is chewing a mouthful of food.

W {finishes the food in his mouth, looks at the book and pulls off a picture of a granola bar, looks at it for a moment then glances at B and hands it to him} [C] [FR] [ME+MP].

B William's granola bar.

W watches B getting the granola bar, then looks at the bar when B holds it up alongside the picture.

B Here's the granola bar.

W {takes the granola bar and sets it on the table then pulls a picture of a juice box off the book and hands it to B; brief eye gaze to B when he hands the picture to B} [C] [FR] [ME+MP].

W flaps his arms briefly.

B sets the juice box on the table in front of W.

W {hands the granola bar back to B} [C] [FC] [MO].

B You don't want that {B puts the granola bar picture back on the book}.

W {pulls off the straw from the juice box, looks to B and makes eye contact; hands B the straw then signs "please" by patting chest once then signs it again followed by the sign for "more"} [C] [FR] [ME+MO+MS].

W {B puts down the straw and as he does so, W pats his hands together once} [C] [FR] [ME+MS].

B Sign help.

W {B signs "help" and W looks at B's hands. At the same time W pats his hands together twice} [RA].

B Help {B moves W's hands upward to complete sign for "help"} please.

W {has his hands in the "more" position but pats his chest with them three times, then pats his hands together twice then signs "more"} [C] [FR] [ME+MS].

B Please {B guides W's hand to pat his chest}.

B tears open the plastic on the straw while W watches.

W takes it from B and pulls the plastic off the straw, puts it into the juice box and begins to drink.