CONVERSATIONAL SPEECH IN MEASURING LANGUAGE PERFORMANCE

THE USES OF CONVERSATIONAL SPEECH IN MEASURING LANGUAGE PERFORMANCE AND PREDICTING BEHAVIOURAL AND EMOTIONAL PROBLEMS

By

JOHANNA K. LAKE, B.A., M.Sc.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements

for the Degree

Ph.D.

McMaster University

© Copyright by Johanna K. Lake, date 2010

DOCTORATE OF PHILOSOPHY(2010) McMaster University
(Psychology) Hamilton, Ontario

TITLE: The uses of conversational speech in measuring language performance and predicting behavioural and emotional problems

AUTHOR: Johanna K. Lake, B.A., M.Sc. (McMaster University)

SUPERVISOR: Karin R. Humphreys

NUMBER OF PAGES: ii, 172

Abstract

Challenges to the diagnostic accuracy of standardized tests of language can make the utility of these measures on their own, problematic. Consequently, this research program uses tools of conversational analysis to study the speech of preschoolers and young adults.

In the first of three studies we examine, from a purely data-driven approach, how conversational measures relate to one another and compare with WPPSI-III expressive and receptive vocabulary scores in assessing preschoolers' language. Mean length of utterance (MLU) was found to be the only conversation measure strongly related to WPPSI-III language scores. However, other conversation measures constituted reasonably stable factors that may have utility for children's language assessment.

The second study uses the same sample of children to investigate what features of language best predict behavioural and emotional problems and whether conversation measures provide better prediction of these symptoms than standardized scores. Results indicated that conversation measures of language significantly improved prediction of Child Behavior Checklist (CBCL C-TRF) DSM-oriented and syndrome scales beyond that accounted for by WPPSI GLC scores.

Finally, the third study uses conversational analysis to study the role of disfluencies in the speech of young adults with and without autism spectrum disorders (ASDs) to determine whether these features of speech serve listener or speaker-oriented functions. Individuals with ASD were observed to produce fewer filled pause words (ums and uhs) and revisions than controls, but more silent pauses. Filled-pause words, therefore, appear to be listener-oriented features of speech.

Taken together, findings of this program of research highlight the importance of using conversational analysis as an alternative or in addition to standardized tests of language as well as inform what specific measures of language are best suited for this purpose.

Preface

The three manuscripts composing this thesis were all authored by myself and Dr. Karin Humphreys. Two of these manuscripts were also co-authored by Dr. Tracy Vaillancourt and by Shannon Cardy. For the two manuscripts in Chapter 2 and 3, all data were collected previously as part of a larger study examining executive function in preschoolers. In collaboration with my supervisor, Dr. Humphreys, we decided to analyze the data first to examine the relationship between WPPSI-III vocabulary scores and conversational measures, and then to use these measures of language to look at the prediction of behavioural and emotional problems. I completed all analyses, background research, writing, and served as a second reliability coder for both manuscripts under the guidance of Dr. Humphreys.

The manuscript for Chapter 4 has been accepted for publication in *Psychonomic Bulletin and Review,* pending final changes. I collected part of the data for this study (testing adults with and without autism spectrum disorders) along with Shannon Cardy and members of Dr. Humphreys' lab for my Master's thesis. I transcribed and coded the large majority of speech samples using SALT software and used these data to conduct analyses with the guidance of Dr. Humphreys. Half of my Master's thesis was based on these data and half from another experiment with these same participants. Preliminary findings from the analysis of the conversational data were presented in the Master's thesis, but the current paper presents the same data analyzed somewhat differently, as well as a series of further analyses and interpretations of the conversational findings than were in the Master's. I wrote the first draft of this manuscript, and Dr. Humphreys kindly provided the edits necessary for its re-submission.

All of my committee members, including Dr. Karin Humphreys, Dr. Tracy Vaillancourt, and Dr. Scott Watter, provided invaluable edits and suggestions to the three manuscripts.

Acknowledgements

It is a pleasure to thank the many people who made this thesis possible.

It is difficult to put into words the gratitude I feel toward my Ph.D. supervisor, Dr. Karin Humphreys. It has been an honour to be her student. I have learned so much from her fierce intelligence, humor, and unwavering support of all that I do. Her enthusiasm and passion for language is contagious and she never failed to leave me inspired by the wonders of the spoken word. Karin always encouraged me to pursue my goals, even when they veered from the traditional; and her belief in my abilities gave me the confidence to reach for them. She has helped me become a better writer, statistician, and most importantly, an independent researcher. Karin; I am enlightened by your knowledge and spirit, and so grateful for your friendship.

I am very thankful for the excellent advice, ideas, and expertise of my committee members Dr. Scott Watter, Dr. Tracy Vaillancourt, and Dr. Judith Shedden. I would have been lost without you. Your tireless efforts on my behalf and willingness to work on a tight deadline are truly commendable. I have gained and grown so much from your knowledge and insightful comments. I feel forever grateful of your counsel and for pushing me to do my very best.

The members of the Cogsci Lab have contributed immensely to my personal and professional time at McMaster. I am so appreciative of your support, feedback, and advice. I wish to thank Mel Rutherford and the members of his lab for help with recruiting participants. I also thank Megan Keilty, Shannon Cardy, and members of the Vaillancourt lab for assistance with data collection. I am grateful to Elizabeth Hall and Elizabeth McAllister for helpful comments and suggestions and will forever be indebted to Esther Manoian, Maya Gupta, Shauna Stokely, and Stephanie Pile for their timely and thorough assistance coding and transcribing data. A special thank you to Aaron for always cheering me on and for your endless assistance with graphs, tables, and printing. I share this success with you.

Lastly, I would like to thank my family and friends for all their love and encouragement. For my parents who raised me to work hard, love what I do, and supported me in all my pursuits.

TABLE OF CONTENTS

Chapter 1

Introduction

Challenges to diagnostic accuracy of standardized tests of language have raised important concern over their ability to effectively assess language. As an alternative or in addition to standardized tests of language, the current program of research studies tools of conversational analysis in assessing the speech of preschoolers and young adults.

Language learning during the preschool years is a highly variable social and developmental process. Language acquisition at this age tends to occur experientially as children interact with peers and adults, vocabularies grow and children learn to modify and elaborate sentences (Bukatko & Daehler, 1995; Peterson & McCabe, 1992). We also begin to see the emergence of conversation skills and awareness of how to negotiate, take turns, and make intelligible contributions (Schickedanz, Schickedanz, Forsyth, & Forsyth, 1998). Rates of language development, however, can vary enormously between children, which makes assessing this population particularly challenging (Fenson, et al., 2003).

To study children's spontaneous speech we must first have a basic understanding of how children develop and produce language. Compared to studies of adult language production, research on the development of spoken language in children has been slower to progress. Most evidence comes from studies of phonological and lexical speech errors in spontaneous utterances where children have been observed to preplan, self-monitor, and revise their speech as it is being produced (Jaeger, 1992; MacWhinney & Osser, 1977; Wijnen, 1992).

Language Development

Acquiring and developing language is an incredibly complex and intricately coordinated task, yet these processes are largely effortless for most children. Language acquisition typically occurs within the first four years of life as children become aware of the basic phonological properties of language, vocabulary grows, and articulation of sounds becomes more refined. At this time children also develop syntax and learn the pragmatic and social attributes of language such as turn taking and understanding the listener's perspective. The vocabularies of young children largely reflect their early experiences. Names of people, food, animals, and clothing are typical first words reflective of children's daily routines (Hoff, 2005). Daily routines also tend to be the source of children's first verbs, including eat, drink, and kiss. For English-speaking children with vocabularies between 20 and 50 words, nouns form nearly 45% of a child's vocabulary and verbs only 3% (Caselli, et al., 1995). After a child's first words, children add an average of 10 words to their vocabularies each month. At the 50-word mark, the rate of new words tends to rapidly increase to 22 to 37 words per month (Goldfield & Reznick, 1990).

One of the more traditional accounts of language acquisition is that of Roger Brown's (1973) five stages of language development. Brown divides children's language acquisition into stages based on mean length of utterance. Mean length of utterance, or MLU, is a measure of linguistic productivity and refers to the mean number of morphemes produced per utterance. Stage I occurs when children are between 15 and 30 months, having achieved a vocabulary of 50 to 60 words, and an MLU of approximately 1.75 morphemes (smallest units of meaning in language). At this stage, children are observed to produce simple noun and verb, or action and object utterances. Between 28 and 36 months, the child reaches stage II, where children begin producing present tense verbs (e.g. going) as well as regular plural nouns. Average MLU at stage II is 2.25 morphemes and utterances tend to be longer than two words. Brown also documents some use of locative prepositions, such as *on* or *in*. Stage III generally occurs when children are between 36 and 42 months and MLU is at an average of 2.75 morphemes. At stage III children begin to use possessives with nouns (e.g. doggie's bone), as well as some irregular past tense use. By stage IV, children are between 40 and 46 months, have an MLU of 3.50, and are beginning to use articles such as *the* and *a*. Regular past tense use also appears at this stage (e.g. played) as does the use of third person (e.g. he jumps). The final stage, stage V, occurs between 45 and 52 months with the appearance of contractions (e.g. she's here, they're going). MLU at stage V is on average 4 morphemes. The utility of these stages has been debated over time as new ways to classify language development emerge (Gold, 1967; Osherson, Stob, & Weinstein, 1985; Pinker, 1979; Seidenberg, 1997; Wexler & Culicover, 1980). However, at a basic level, the stages provide us with some guidelines as to the progression of normal language development.

Individual Differences in Vocabulary Acquisition

For the general population, language development follows a fairly normative track, however, rates and course of acquisition can vary greatly from child to child. Individual differences in the lexical development of children are surprisingly high, particularly in the early years of life. Language involves many different domains including syntax, semantics, morphology, phonology, and pragmatics, with children acquiring competencies in these domains at different times (Pence & Justice, 2008). As a result, preschoolers' language tends to grow more rapidly in some areas and more slowly in others. Some researchers refer to these areas of strength and weakness as language profiles (Fey, 1986). For example, some children's vocabularies consist largely of referential words, whereas others use mostly context-bound words. Some of these differences can be explained by the context in which children first hear words (Hoff, 2005). Similarly, social, gender, personality, genetic, and cultural factors can greatly influence a child's motivation to communicate as well as the types of words likely to be learned (Hoff, 2005; Stromswold, 2001). Much of the research in this area has focused on differences between referential and expressive language users. In a study conducted by Nelson (1973) large discrepancies were found in children's use of nominals (e.g., Daddy, cats). Furthermore, Nelson discovered that children who used fewer nominals tended to use more personal/social words. From this, it was suggested that children form two kinds of vocabularies; one consisting largely of object labels (referential) and the other personal/social words (expressive).

A large body of research also supports the notion that rates of lexical development, or the size of children's vocabularies, vary enormously. Data from the MacArthur-Bates Communicative Development Inventory, assessing children's comprehension and production vocabularies, demonstrates how vocabulary size of 16 month-olds can vary from 0 to 160 words, and for 24 month olds from 50 to 550 words (Fenson, et al., 2003). Not surprisingly, a variety of influences can account for these differences. Some researchers have found that children exposed to more speech add words to their vocabularies more quickly (Hoff & Naigles, 2002; Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1991). Other studies have documented the role of socioeconomic status (Fenson, et al., 1994; Hart & Risley, 1995; Hoff, Laursen & Tardif, 2002; Pine 1994) as well as maternal verbal responsiveness and mother-child interactions (Tomasello & Todd, 1983).

Discourse Development

Developing conversation skills is an important step in language acquisition. During the preschool years, children begin to use language for more complex purposes, including logical and participatory functions (Halliday, 1977, 1978). Logical functions express logical relations between ideas, and participatory functions express wishes, feelings, attitudes and judgments (Pence & Justice, 2008). At this stage, preschoolers tend not to participate in true dialogue, as they have yet to fully understand the role and perspective of others. Piaget (1926) first labeled this type of interaction as *collective monologues*, where children take turns in conversation but each turn has little to do with the previous turn. Children at this age are also observed to engage more frequently in private speech, talking to themselves when alone or while playing (Winsler, Carlton, & Barry, 2000). Private speech or solitary monologues are thought to provide opportunities for children to explore and practice language (Gallagher & Craig, 1978). Most preschoolers, however, can maintain a conversation for two or more turns, although they likely still have some difficulty recognizing when communication breaks down or in giving listeners enough information to facilitate understanding (Pence & Justice, 2008).

Disfluencies in Children's Speech

There is some evidence that children preplan, self-monitor, and revise their speech as it is being produced with much of this research coming from studies of phonological and lexical speech errors in spontaneous utterances (Jaeger, 1992; MacWhinney & Osser, 1977; Wijnen, 1992). As speakers we are often found to repeat and revise phrases, words, and even parts of words, to effectively communicate our thoughts. Similarly, we frequently produce 'ums,' 'uhs,' and other seemingly undesirable pauses or filled-pause words. Most researchers collaboratively refer to these acts of speech as disfluencies; since they interrupt fluent speech. Disfluencies in the speech of children can be a frequent and normal part of language development. In other instances, disfluencies can be part of a larger issue as in the case of children who stutter. Disfluent speech is thought to originate from planning problems, occurring when a speaker halts their speech or inserts fillers before formulating the remainder of the utterance (Levelt, 1983; 1989). Many researchers have studied this process and found that for children between the ages of 3 and 5, it is the complexity of the sentence, not whether the child stutters or does not stutter, which influences the likelihood that a disruption in speech or disfluency will occur (Bernstein, Ratner, & Costa Sih, 1987; Gordon, Luper, & Peterson, 1986; McLaughlin & Cullinan, 1989). For example, Yaruss, Newman and Flora (1999) examined the relationships between utterance length, syntactic complexity, and disfluency, in the spontaneous speech of 12 nonstuttering children between 44 and 64 months. Results indicated that disfluent utterances were longer and more syntactically complex than fluent utterances and that utterance length was the most important factor in predicting a disfluent utterance. Bernstein, Ratner, and Costa Sih (1987) document similar findings in a task varying utterance length and complexity with fluency and accuracy of sentence reproduction in the speech of 8 nonstuttering and stuttering children between 47 and 76 months. Both groups were found to produce greater rates of disfluent speech as syntactic complexity increased. Length was also observed to increase disfluency rates; however, it was not as strong a predictor as syntactic complexity (Bernstein Ratner & Costa Sih, 1987). Significantly higher rates of disfluency as utterances become more complex has also been documented by several other researchers (Colburn & Mysak, 1982; McLaughlin & Cullinan, 1989; Pearl & Bernthal, 1980). It is important to note, however, that these studies do not distinguish between specific types of disfluencies, rather they use the term disfluency to encompass all forms.

Stalls and revisions

In an attempt to better understand the processes underlying disfluencies, some researchers have further categorized disfluencies into stalls and revisions. Stalls are interruptions that do not change the linguistic structure produced (e.g., ums, uhs, silent pauses, etc), whereas revisions change the structures produced by adding or deleting words, phrases, or parts of words (Rispoli, Hadley, & Holt, 2008). Some researchers propose that stalls and revisions represent discrete types of disfluencies, originating from different problems in language production. Rispoli, Hadley and Holt (2008) set out to test this theory by studying the sentence disruptions of 20 typically developing children. Results demonstrated that revisions occurred in approximately 1% of children's utterances at 27 months, and that this increased with age. Stall rates, by contrast, did not change with age. The number of stalls did, however, increase with sentence length, whereas the number of revisions remained constant. These findings lend support for the stall-revision dichotomy, and the authors argue that increased rates of revisions with age coincide with children's ability to monitor language production. Stalls, they assert, arise from glitches in sentence production, the rate of which increases as sentences become longer (Rispoli, Hadley, & Holt, 2008). In another study, Rispoli and Hadley (2001) examined the relationship between stalls and revisions, this time investigating grammatical development and stall and revision rates in 52 typical children. Revision rate was found to increase with level of grammatical development; however, stall rate did not. Rispoli suggested therefore, that the capacity for self-monitoring in children increases during grammatical development, while stall rates remain unaffected by this process.

Filled pauses

Filled-pause words, including um and uh, represent a unique sub-category in the broader classification of disfluencies. Recent research suggests that filled-pause words can actually be quite helpful in conversations as they indicate that the speaker is not finished speaking yet and is trying to put together their next thought or find the correct word (Fox Tree, 2001). Fox Tree (2001) studied this effect by examining the role of ums and uhs during on-line processing of speech. Results indicated that um and uh may be utilized by a listener to facilitate conversations. Uh appeared to signal an upcoming short delay, while um an upcoming longer delay. The use of uh was found to increase the speed at which listeners were able to recognize words in upcoming speech, however, um had no effect on listeners' speech recognition (Clark & Fox Tree, 2002; Fox Tree, 2001). Fox Tree suggests that ums and uhs help listeners by alerting them that the speaker is still speaking (that it is not the listener's turn yet) and indicating the length of the upcoming delay in speech.

In a similar study, this time with 3 and 4 year old children, Hudson Kam and Edwards (2008) document the role of delay markers such as um and uh in the spoken language of preschoolers. Children in this study were observed to produce longer pauses following delay markers, but no differences were found in pauses following um or uh. The authors suggest that children at this age understand the basic use of delay markers, but not how to differentiate them. According to the findings of Rispoli, Hadley, and Holt (2008) ums and uhs fall into the category of stalls, and unlike revisions do not increase with grammatical development or age. It may be that as children get older and acquire more sophisticated grammar so does the ability to differentiate between the use of um and uh without actually increasing the number of these disfluencies. It is the process of monitoring one's speech and determining appropriate use of ums and uhs use that increases with time rather than the amount produced.

Measuring Children's Language

There are many reasons why researchers and clinicians test children's language. Speech language pathologists frequently administer language tests as a way to identify or rule out language impairments and disorders as well as to determine appropriate treatment. Similarly, clinicians use language assessments to evaluate school readiness, aptitude, and literacy. From a research perspective, language tests provide a measure of verbal or cognitive ability that can be used to define or compare groups. The preschool years represent a particularly critical time for language assessment since significant language development occurs during this time and the impact of intervention for children with language delay or impairments is heavily influenced by early identification and treatment.

There is much debate concerning techniques to measure children's language. Over time, researchers and clinicians have developed a variety of methods to assess different aspects of language. Some common measures of vocabulary development include total number of words, number of different words, and type token ratio (ratio of number of different words to total words within a language sample). Measures of syntax generally include mean length of utterance (MLU), while measures of pragmatics code for communicative functions used by the child, such as requesting, commenting, or responding. Similarly, the types of assessment tools can range from standardized tests of receptive and expressive vocabulary, to spontaneous language sampling and discourse analysis (Pence & Justice, 2008).

It is well established that the production of specific classes of words and the use of syntax and pragmatics convey essential information about language progression (Klee, 1992; Redmond 2004; Rice & Wexler, 1996). As children mature, their language becomes increasingly complex as sentences get longer and are arranged in new, more complex combinations (Scott & Stokes, 1995). Collaboratively, researchers refer to this linguistic progression as advances in syntactic complexity. Syntactic complexity can be challenging to measure as the changes that occur over time are subtle. To tackle this issue, researchers have sought to determine what particular aspects of language are 'markers' of linguistic complexity or specific language deficits. For example, Scott and Stokes (1995) identified two language indices as important in the assessment of syntactic complexity in school-age children and adolescents; sentence length and clause density. Their findings showed that as sentence length increased and began to contain subordinate clauses, syntactic structure also became more complex. Similarly, Rice and Wexler (1996) found tense-marking morphemes to be a clinical marker of specific language impairment in English-speaking children.

Other researchers have taken a word usage approach to studying language. The Weintraub approach was of these methods, whereby researchers coded words and phrases into 15 different categories (Weintraub, 1981). These categories were then compared across people with different psychiatric diagnoses and atypical behaviours (e.g., depression, binge eating, alcoholism) to determine if any patterns of linguistic profiles emerged. A similar approach is to use word-based counting systems. This method assumes that one can create linguistic profiles based on the specific words people use (Stone, Dunphy, Smith, & Oglivie 1966). Most of these strategies have been adopted to test a particular theoretical hypothesis, with few studies systematically examining how measures of spontaneous language sampling group together or align with standardized tests of language.

Standardized Measures of Language

Many researchers and clinicians seek to describe a child's language by comparing his or her language to the language of other children of the same age. On a large scale, this is what standardized measures of language enable individuals to do. Researchers interested in children's language frequently use standardized measures to describe the children they are studying, whereas practitioners in the field of communicative disorders use such measures to assess children for diagnosis and treatment purposes (Hoff, 2005). There currently exists a wide range of standardized tests to measure the language skills of preschoolers. The efficacy and accuracy of some norm-referenced tests, however, remains debatable. In particular, standardized psychometric measures have been criticized for their restrictive assessment of language and decreased sensitivity in identifying language impairments (Dunn, Flax, Sliwinski, & Aram, 1996; Gray, Plante, Vance & Henrichsen, 1999; McCauley & Swisher, 1984; Merrell & Plante, 1997; Plante & Vance, 1994). Sensitivity of a test refers to how many cases of a particular disorder or disease a particular test can find. Specificity, on the other hand, refers to how often a test diagnoses a particular disorder or disease when the individual does not have one. Ideally, a test is both highly sensitive and specific. In the case of standardized tests of language, concerns typically focus on how sensitive they are at identifying or picking up on cases of language impairment or delay. These issues are discussed in greater detail below. Nevertheless, despite their shortcomings, standardized tests continue to be widely used. One of the major reasons for their popularity stems from the fact that they tend to be easy to administer quickly and relatively inexpensively on a broad scale, even when dealing with a large population, and they have the benefits of well-documented distributions and reliability. As a result, standardized tests appear more objective, particularly within the medical model of disability, where disability is viewed as an inherent part of the individual rather than being contextually mediated. Finally, standardized tests are easily utilized in a variety of ways, including providing diagnostic criterion, determining program entry, and evaluating treatment efficacy.

Types of standardized measures of language

Standardized language tests are frequently used by teachers and early childhood educators as tools to evaluate children's skills on entry into preschool. These assessments also serve to monitor children's progress during school and to tailor instruction to meet early literacy and language needs (Pence & Justice, 2008). Some of the most commonly used standardized language and literacy measures for children include: the Peabody Picture Vocabulary Test- Third Edition (PPVT-III Dunn & Dunn, 1997); the Preschool

Language Scale – Fourth Edition (PLS-4 Zimmerman, Steiner, & Pond, 2002); the Test of Early Language Development, Third Edition (TELD-3 Hresko, Reid, & Hammill, 1999); the Clinical Evaluation of Language Fundamentals – Preschool, Second Edition (CELF- Preschool-2 Wiig Secord, & Semel, 2004); the Test of Early Ready Ability- Third Edition (TERA-3 Reid, Hresko, & Hammill, 2002); and the Wechsler Preschool and Primary Scale of Intelligence- third edition (WPPSI-III Wechsler, 2002). More naturalistic standardized measures of language, such as the Renfrew Bus Story (Cowley & Glasgow, 1994), a measure of narrative ability, also exist.

PLS-4. The PLS-4 is a norm-referenced measure of vocabulary, grammar, morphology, and language reasoning. The PLS-4 generates scores on two scales; the Auditory Comprehension scale and the Expressive Communication Scale. The Auditory Comprehension scale measures language comprehension abilities, including receptive vocabulary, comprehension of concepts, grammatical markers, and the ability to make comparisons and inferences. The Expressive Communication scale measures language production abilities, such as expressive vocabulary, segmenting words, using grammatical markers, completing analogies, and telling a story in order (Pence & Justice, 2008).

Reliability of the PLS-4 has been demonstrated using test-retest reliability data showing that PLS-4 scores are dependable and stable across repeated administrations. Test-retest coefficients for the PLS-4 vary from 0.82 to 0.95 for subscale scores and from 0.90 to 0.97 for the Total Language Score. Reports of internal consistency vary from 0.66 to 0.96, and inter-rater reliability shows 99% agreement between scorers. Content validity of the PLS-4 was obtained through a broad literature review and user survey about language skills the test should address. Internal consistency of subscales is high, with a correlation of 0.80 between Auditory Comprehension and Expressive Communication scales across ages. In a representative sampling of 150 children (75 with a language disorder and 75 typically developing), sensitivity and specificity was high for both subscales and total language scores (Zimmerman, Steiner, & Pond, 2002).

TELD-3. The TELD-3 is an early language test assessing receptive, expressive, and overall spoken language in young children. The test consists of two subtests; receptive and expressive language, which form an overall composite score.

Reliability of the TELD-3 is well documented across content, time, and scorer, with all coefficients rounding to or exceeding .90. Content validity was established through careful selection of items, controlled vocabulary, construct review by a panel of language experts, conventional item analysis, differential item functioning analysis, and form equivalence. Similarly, criterion validity was established by correlating TELD-*3* scores with a variety of widely recognized measures of language ability (e.g., CELF Preschool-2, Woodcock-Johnson Psycho-Educational Battery, Revised). The relationship of the TELD-3 standardized scores with age, IQ, and academic achievement has also been extensively studied (Hresko, Reid, & Hammill, 1999).

PPVT-III. The PPVT-III is a measure of receptive vocabulary. Children are presented with a page of four pictures and asked to point to one of the pictures. Since the PPVT-III only measures receptive vocabulary, it is often used in conjunction with other more comprehensive measures of language (Pence & Justice, 2008). The PPVT-III is co-normed with the Expressive Vocabulary Test, allowing for direct comparisons (Dunn & Dunn, 1997).

The PPVT-III reports high reliability with median internal consistency of 0.95 and test-retest reliability of .92. Validity of the PPVT-III is well documented, with an average correlation of .69 with the Oral and Written Language Scales Listening Comprehension scale and .74 with the Oral and Written Language Scales Oral Expression scale. Similarly, it correlates well with other measures of verbal ability including the Wechsler Intelligence Scale for children-III VIQ (r=.91), the Kaufman Adolescent and Adult Intelligence Test (KAIT Kaufman & Kaufman, 1993) Crystallized IQ (r=.89), and the Kaufman Brief Intelligence Test (K-BIT Kaufman & Kaufman, 1990) Vocabulary score (r=.81).

CELF-Preschool-2. The CELF-Preschool-2 contains eight subtests including sentence structure, word structure, expressive vocabulary, concepts /following directions, recalling sentences, basic concepts, and word classes. Sentence structure, word structure, basic concepts, and expressive vocabulary, together, form a core language score used by clinicians as an overview of a child's key language abilities (Pence & Justice, 2008).

Reliability of the CELF-2 is well documented in the literature. Several studies report high levels of test-retest reliability and internal consistency (Allen & Yen, 1979; Crocker & Algina, 1986; Magnusson, 1967). Similarly, clinical validation studies show that the CELF-2 is very sensitive to language difficulties in a variety of clinical groups. Diagnostic validity statistics demonstrate excellent sensitivity at 1 *SD* below the mean (Wiig, Secord, & Semel, 2004).

TERA-3. The TERA-3 is a standardized measure of children's early reading skills. The test consists of three subtests including, alphabet knowledge, conventions, and meaning. Alphabet knowledge measures children's knowledge and use of the alphabet, and conventions measures knowledge of print conventions. Meaning provides a measure of children's ability to construct meaning from print. Taken together, the three subtests form an overall reading quotient (Pence & Justice, 2008).

Reliability coefficients have been computed for subgroups of the normative sample (e.g. African Americans, Hispanic Americans, females) as well as for the entire normative sample. Reported internal consistency coefficients are high, ranging from 0.83 to 0.95 across subtests. Similarly, coefficients for the reading quotient varied from 0.91 at 3 years, to 0.97 at 4 years. Using a two week interval, test-retest reliability varied from 0.86 to 0.98 for two age groupings (4-6 years and 7-8 years) across all subtests and including reading quotient (Reid, Hresko, & Hammill, 2002). Content validity of the TERA-3 was established by thorough literature and test review as well as consultation with an expert panel. Criterion prediction validity is demonstrated through moderate to high correlation with Stanford Achievement Test Series – Ninth Edition (Psychological Corporation, 1996), and Woodcock Reading Mastery-Revised (Woodcock, 1998) (Reid, Hresko, & Hammill, 2002).

The Renfrew Bus Story. The Renfrew Bus Story is a screening measure of a child's ability to retell relevant information about a story. The story is read aloud to the child, who follows along with a series of 12 pictures. The child uses these pictures to retell the story, which is recorded and then later transcribed and scored. The child's performance depends on the integration of a variety of skills, including coordinating auditory and visual input, attention, listening, comprehension, memory, and sentence formulation (Cowley & Glasgow, 1994).

Reliability of the Renfrew Bus Story is documented through a pre and post test sample of 27 children. Tests were administered at a four week interval with coefficient scores of 0.79 for information, 0.73 for sentence length, and 0.58 for complexity. Narrative recall requires the child to integrate many skills and is considered a valid integrative test (Paul & Smith, 1993). Information, sentence, length, and complexity scores are observed to increase as a function of age. The authors report mean scores for information beginning at 15 and rising to 31 with age. Similarly sentence length increased from 6 to 10, and for complexity form 0 to 3 (Cowley & Glasgow, 1994). Several studies report the Bus Story as an effective tool in the identification of individuals with language delay (Bishop & Edmundson, 1987; Howland & Kendall, 1991)

WPPSI-III. The WPPSI-III is a test of general intelligence for children between the ages of 2 years 6 months and 7 years 3 months. The battery consists of fourteen subtests including: information, vocabulary, word reasoning, comprehension, similarities, block design, picture concepts, picture completion, object assembly, symbol search, matrix reasoning, coding, receptive vocabulary and picture naming (Rock & Stenner, 2005). Out of these fourteen subtests, the WPPSI yields three distinct IQ scores; Full Scale IQ (FIQ), Verbal IQ (VIQ), and Performance IQ (PIQ). Depending on the age of the individual, the test may also provide a General Language Composite (GLC) or a Processing Speed Quotient (PSQ). For the purpose of this chapter, I focus on WPPSI language subtests; receptive and expressive (picture naming) vocabulary, as well as block design.

Receptive vocabulary measures a child's vocabulary knowledge. This subtest also tests the ability to comprehend verbal directions, auditory and visual discrimination (the ability to differentiate objects based on their individual characteristics or sounds), auditory memory and auditory processing (Sattler, 1992). During the test, the examiner says a word and the child must choose the matching picture from a set of four possible pictures. This information provides a measure of word knowledge, language development and concept formation. Expressive vocabulary requires the child to name pictures displayed in a stimulus book. This subtest measures word knowledge and language development. It also provides the examiner with information on the child's knowledge and everyday experiences (Sattler, 1992). Block Design is a nonverbal task measuring visual-spatial, visual organization, and visual-motor coordination abilities. The child is required to build blocks according to a model constructed by the examiner within a specified time. The test begins with single coloured blocks and goes on to include blocks with both red and white sides. The child must re-create the design first based on an actual model and later based on pictures from a stimulus book. In order to do this the child must be able to analyze abstract visual stimuli (model or picture) and then break that design down into its individual parts- a process called analysis and synthesis. Success requires the child to apply logic and reasoning to spatial relationship problems (Sattler, 1992).

The WPPSI is one of the most commonly used research measures of preschoolers' cognitive and language skills (Rice, Burh, & Nemeth, 1990; Rice, Buhr, & Oetting, 1992). Although the WPPSI is used less frequently in speech pathology settings, research findings based on the WPPSI often guide and inform clinical practice of speech language pathologists.

Traditionally, the WPPSI has been used to predict academic achievement in a variety of domains including reading, math, and general knowledge. The test is used as a preschool estimate of IQ that has been shown to reliably and validly predict achievement scores years later (Kaplan, 1996; Tew & Laurence, 1983; Yule, Gold, & Busch, 1982). Many studies document the predictive validity of the WPPSI in effectively predicting achievement on standardized test scores (Bishop & Butterworth, 1979; Kaplan, 1996; Lieblich & Shinar, 1975). In one of several longitudinal studies examining predictive validity of the WPPSI, 85 children were administered the WPPSI at 5 ½ years of age. These same children were then tested again, 11 years later, using the Wechsler Intelligence Scale for Children (WISC-R). Correlations between the two intelligence tests were high, ranging from $r = 0.73$ to 0.86 for verbal, performance, and full scales (Yule, Gold, & Busch, 1982). Correlations between the WPPSI and three achievement tests; the Sentence Reading Test NS6, Vernon's Graded Spelling (0.61), and Vernon's Graded Arithmetic –Mathematics Test (0.72), were also high. Another 11 year longitudinal study investigated a sample of 51 Welsh children with spina bifida. Children were administered the WPPSI at 5 years and the WISC-R at 16 years. Again, correlations were high, with a correlation of 0.88 for verbal, 0.92 for performance, and 0.92 for full scale. Additionally, mean IQ differences between the tests were less than 6 points (Tew & Laurence, 1983). The effectiveness of the WPPSI in predicting mathematics and reading scores on the Metropolitan Achievement Test, reading sections of the Stanford Binet and Gilmore Oral Reading Paragraphs Test, and spelling, reading, and arithmetic scores on the Wide Range Achievement Test (WRAT) are well documented (Kaufman, 1973; Krebs 1969; Pasewark, Scherr, & Sawyer, 1974; Reynolds, Wright & Dapper, 1981). The majority of these studies test children with the WPPSI at baseline, and then again several years later with the WISC or a similar standardized measure. The issue with much of this research is that while it demonstrates the reliability of the WPPSI in predicting scores on the same or similar tests years later, it does not determine whether the constructs themselves are effective in identifying children with specific deficits or impairments. See Table 1 for a summary of standardized and conversational language assessment tools.

Issues with standardized measures of language

Many norm-referenced tests of language lack key components of language development and consequently, are very poor at identifying language impairments. Gray et al. (1999) conducted a study examining the diagnostic accuracy of four vocabulary tests; Peabody Picture Vocabulary Test-III (Dunn & Dunn, 1997), Receptive One-Word Vocabulary Test (Gardner, 1985), Expressive Vocabulary Test (Williams, 1997), and the Expressive One-Word Vocabulary Test-Revised (Gardner, 1990). The four vocabulary tests were administered to preschool age children to determine their ability to screen or identify specific language impairment (SLI). Results demonstrated that none of the aforementioned tests were strong identifiers of SLI. In fact, children with SLI typically fell within the normal range. Similarly, Plante and Vance (1994) examined the ability of 21 commonly used tests of language skills in meeting ten psychometric criteria based on standards set forth by McCauley and Swisher (1984). The ten criteria included description of normative sample, sample size, item analysis, means and standard deviations, concurrent validity, predictive validity, test-retest reliability, interexaminer reliability, description of test procedures, and description of tester qualifications. Language tests were chosen on the basis of having norms for children between the ages of 4 and 5 and were reviewed on the psychometric criteria. Results indicated that only 38% of these tests met half or more of the psychometric criteria and all tests had low correlations with measures of nonverbal skills. Furthermore, only one of the aforementioned tests provided acceptable accuracy in discriminating between children with SLI and age-matched controls. Marini, Tavano, and Fabbro (2008) took this research one step further by comparing the linguistic skills of a large group of SLI participants against a group of typically developing peers using a narrative task, a standardized battery of tests assessing linguistic function, and the WPPSI or WISC-R. The battery, "Batteria della valutazione del linguaggio in bambini dai 4 ai 12 anni" ("Battery for linguistic assessment of children from 4 to 12 years", Fabbro, 1999) provides an overview of children's language function by examining phonological, lexical, and syntactic skills in all modalities (comprehension, production, and repetition). Their results demonstrated that while SLI participants produced a similar number of words as controls, their narratives were filled with significantly more omissions and/or substitutions of bound and free morphemes. These data are consistent with what is currently known about SLI, where morphosyntax and syntax are particularly impaired. The standardized tests could account for the number of words spoken by both groups, however, it failed to recognize the more minute grammatical details. In this experiment, the standardized tests would have failed to differentiate individuals with SLI from controls.

Aram, Morris, and Hall (1993) observed similar results in their study examining the ability of psychometrically derived criteria and spontaneous measures in identifying 252 preschoolers with SLI. Children were tested using the Test of Early Language

Development (TELD Hresko, Reid, & Hammil, 1981), a series of more specific language tests, and measures of spontaneous language sampling including mean length of utterance. Results demonstrated that no single criterion of the psychometric tests applied alone could identify more than 71% of the clinical group. Mean length of utterance, derived from spontaneous language sampling, however, was the most sensitive measure, capturing 80% of the SLI group (Aram, Morris & Hall, 1993). Again, in this study, spontaneous measures of language were better than standardized measures at identifying children with language impairment.

Overall, standardized tests present considerable shortcomings in their ability to identify language impairments and to more generally assess language competence. Many features critical to language development are not even tested for in standardized tests of language. Additionally, differences in terms of context, rather than language structure, can also lead to differences in associated skills (DeThorne & Watkins, 2006; Ukrainetz & Blomquist, 2002). Furthermore, recent studies demonstrate how a broad range of linguistic skills including phonological awareness, narrative ability, and knowledge of syntax and grammar, are critical predictors of language acquisition, reading, and future academic success (Barnhart, 1991; Boudreau & Hedberg, 1999; Roth, Speece, & Cooper, 1997; Scanlon & Vellutino, 1996).

Spontaneous Measures of Language

Spontaneous speech samples and narrative discourse are widely used tools in speech and language therapy to document the language of young children (Botting, 2002). Conversation and narratives form the basis of many childhood speech acts and therefore represent a natural and practical way to examine speech. Spontaneous language sampling occurs in a naturalistic context (e.g., child's preschool classroom or home) with trained research examiners or individuals with whom the child is familiar. An unstructured conversation between child and examiner, ideally of at least 200 utterances, is then recorded (Bowen, 1999). Topics of interest such as books or videos the child has recently viewed can be used to help guide the conversation, but the procedure is meant to be as natural and unstructured as possible. Recorded language samples are then thoroughly coded and transcribed to obtain a detailed and comprehensive picture of a child's language. This process affords clinicians and researchers the unique opportunity to gain valuable linguistic insight including measures of word frequency, length, mean length of utterance (MLU), word and utterance level errors, disfluencies, as well as indices of syntax and pragmatics.

Some researchers have taken these tools of conversational analysis to study the language of clinical and non clinical groups of children. Findings of this work have shown some aspects of spoken language to be indicative of development or disorder including sentence length, clause density, revisions, and filled-paused words (Redmond, 2004; Scott & Stokes, 1995). More generally, what these preliminary results indicate is the need for more studies using broader and more comprehensive tools of linguistic analysis in addition to standardized test of language, to better understand the language of young children. Clinically, effective treatment of individuals with language impairment or developmental disability is largely dependent on the use of accurate and sensitive screening tools.

The psychometric properties of spontaneous language and word use have been well established in adult populations. In 1999, Pennebaker and King analyzed a broad range of text samples obtained from diaries, journal abstracts, and college assignments. Results confirmed strong internal consistency for 36 language dimensions. Similarly, Mehl and Pennebaker (2003) analyzed the conversations of students twice for two days over the course of four weeks. The study utilized an electronically activated recording device (EAR), demonstrating, again, that students spontaneous word usage was stable over time. Average test-retest correlation for standard linguistic variables was, $r = 0.41$ and for psychological processes, $r = 0.24$. Consistency was also observed across social context (e.g., word use at home, public places, work, etc). Several other studies also document the finding that people's word choices are sufficiently stable over time and consistent across topic or context (e.g., Gleser, Gottschalk, & Watkins, 1959; Pennebaker & King, 1999; Pennebaker, Mehl, & Niederhoffer, 2003).

Taken together, most researchers agree that language can reliably and accurately serve as an individual difference measure in adults. There is less research on the consistency of word use and spontaneous language in young children; however, some researchers document findings similar to what we observe in adults. For example, Bishop et al. (2000) found reasonable stability in conversations with children with pragmatic language impairment (PLI) between differing conditions of stimuli and interlocutor. Adams and Lloyd (2005) report similar findings of small variation in conversation indices compared with baseline measures.

Narratives as a clinical tool

Oral narratives are highly related to later literacy ability, particularly in children with SLI (Gillam & Johnston, 1992; Stothard, Snowling, Bishop, Chipchase, & Kaplan, 1998). For example, Kaderavek and Sulzby (2000) conducted a study examining oral narratives of typically developing and language impaired preschool children. Narratives of typically developing children were more complex and showed more devices of written language than narratives of children with language impairment, suggesting an association between early reading skills and oral language. Cain and Oakhill (1996) document similar findings linking comprehension skills and storytelling ability in 7 and 8 year old children. Children with poor comprehension skills were found to produce stories with less sophisticated narratives and story structure. In another study, oral retelling narratives were found to be a powerful predictor of long-term language skill (Bishop & Edmundson, 1987).

In the case of children with language impairments, oral narratives play a critical role in the identification and diagnosis of impairment. Children with language impairment are frequently observed to produce poor narratives when telling and retelling stories (Merritt & Liles, 1987; Tager-Flusberg, 1995, Van der Lely, 1997). To study this, Liles and Duffy (1995) conducted an experiment to determine what particular aspects of narratives distinguished children with language disorders from nondisordered peers. Examination of narratives revealed two distinct factors; global organization of content (how events are logically related) and linguistic structure (text specific organization). From this, they observed that poorer narrative production of language disordered children was primarily the result of deficits in text structure rather than knowledge about how events were related. Additionally, language impairments tend to persist over time, with 88% of children identified as having SLI at 7 years of age, still presenting with communication difficulties at age 11 (Conti-Ramsden, Botting, Simkin & Knox, 2001).

Type token ratio and D

Type Token Ratio (TTR) has long served as a critical measure of lexical diversity. Recently, however, considerable debate has surfaced regarding its use. TTR is defined as the ratio between the number of different words used in a sample of language and the total number of words in that sample (Watkins & Kelly, 1995). Consequently, type token ratio is dependent on the number of tokens in a given language sample. As a result, samples with more tokens will naturally produce smaller TTR values, and samples with fewer tokens will produce larger TTR values (Richards & Malvern, 1997a). The resulting TTR numbers, therefore, distort the true value of vocabulary diversity. Watkins and Kelly (1995) conducted a study evaluating the extent to which measures of lexical diversity such as TTR and the number of different words, differentiated children with specific language impairment and children with typically developing language. TTR, in analyses of 50 and 100 utterance samples, did not differentiate between groups, however, children with SLI were observed to produce significantly fewer different words. As a result, type token ratio did not differ between children with SLI and children with typical language. Several other studies document similar issues with TTR as an index of language development or impairment (Haas, Haug, & Landry, 1989; Rice & Bode, 1993). In an effort to avoid the problems of TTR, many researchers have turned to a new measure of lexical diversity: D. D is calculated by a program called the Computerized Language Analysis (CLAN), developed by the Child Language Data Exchange System (CHILDES) (MacWhinney, 2000). CLAN calculates the value of D by random sampling of tokens to plot the curve of TTR against increasing token size for the transcript under investigation by using the probability of new vocabulary as it is introduced to larger samples of speech. This function is then used to create a mathematical model of the way TTR varies with token size. Finally, the function is compared with empirical data to produce the value of D. The value of D, therefore, is not based on the number of words in a given sample; rather, it uses mathematical models in conjunction with empirical data to evaluate vocabulary diversity making it a more valid measure of lexical diversity.

Conclusion

Challenges to diagnostic accuracy of standardized tests of language raise important concerns over their ability to assess language and identify impairment without additional linguistic information. This is particularly apparent in a child's early years when language development is highly variable. Spontaneous language sampling and discourse analysis has enabled researchers and clinicians to obtain a more comprehensive and arguably valuable, linguistic picture. Acquiring a more in-depth assessment of a child's language will help us better understand how language develops in young children and whether variability simply reflects individual differences versus disorder. Despite issues related to standardized tests of language, very few studies examine how spontaneous language sampling aligns with these measures. The present thesis, to our knowledge the first of its kind, investigates this relationship from an exploratory approach. The goal of this program of research is three-fold. Firstly, using the conversation samples of a group of 46 preschool children, we examine how vocabulary measures obtained from spontaneous language sampling align with Wechsler Preschool and Primary Scale of Intelligence (WPPSI-III Wechsler, 2002) receptive and expressive vocabulary scores in the assessment of preschoolers' language abilities. For the purpose of our research we examine issues of processing, primarily fluency and speech errors, rather than measures of syntactic complexity. Additionally, we implement factor analysis to examine the factor structure of measures of conversational analysis to determine which indices are most useful in measuring children's language and how these features of speech co vary. Secondly, using this same sample of children, we examine the ability of language measures obtained from spontaneous speech to predict maladaptive behaviour and emotional problems as measured by The Child Behavior Checklist Caregiver-Teacher Report Form (C-TRF 1½-5: Achenbach & Rescorla, 2000). Specifically, we study how measures of spontaneous speech predict DSM-oriented and total problems scales, beyond that afforded by WPPSI general language composite scores. We examine C-TRF maladaptive behaviour and emotional problems as a continuum of problems rather than a specific clinical cut-off for disorder, allowing greater understanding of how these issues vary developmentally.

Thirdly, this time studying a sample of young adults with and without autism, we investigate the language role of disfluencies such as "um" or "uh," in conversation. The goal of this phase of research is to discern whether these disfluencies serve listener or speaker- oriented functions by looking at their occurrence (or lack of occurrence) in the speech of young adult participants with autism.

Taken together, this research program uses tools of conversational analysis to study the spoken language of preschoolers and young adults. Specifically, we investigate what features of language best predict behavioural or emotional problems and whether conversation measures provide better prediction of these issues than standardized scores. Additionally, using a group of individuals with autism, we examine what features of conversation serve listener or speaker-oriented functions. Finally, from a purely data-driven approach we explore the relationship between standardized measures of language (WPPSI-III) and indices obtained from spontaneous language sampling, with the aim to identify differences in their measurement, as well as determine specific measures of language best suited for discourse analysis.

Table 1

Standardized and Conversational Language Assessment Tools

Language Assessment Tool	Age	What it measures
Standardized Tests of Language		
PLS-4	birth- 6 years	Receptive & expressive language
TELD-3	2-7 years	Receptive, expressive & spoken language
PPVT-III	2-90 years	Receptive vocabulary
CELF-Preschool-2	3-6 years	Sentence structure, word structure & expressive vocabulary
TERA-3	3-8 years	Reading skills
WPPSI-III	2-7 years	Receptive & expressive vocabulary
Discourse Analysis		
The Renfrew Bus Story	3-6 years	Sentence length, complexity, & information
Spontaneous Language Sampling		
Conversation Measures	2 – 90 years	Quantitative & qualitative

Chapter 2

Running Head: SPONTANEOUS VERSUS STANDARDIZED

Spontaneous versus standardized: Examining the relationship between the WPPSI-III and measures from spontaneous language sampling

Johanna K. Lake[1*]

Karin R. Humphreys[1]

Tracy Vaillancourt[1,2]

[1]Department of Psychology, Neuroscience & Behaviour, McMaster University

[2]Faculty of Education and School of Psychology, University of Ottawa

*Corresponding author: Johanna K. Lake, Department of Psychology, Neuroscience & Behaviour, McMaster University, 1280 Main St West, Hamilton, ON L8S 4K1, Canada; phone: (905) 525-9140 x22853, fax: (905) 529-6225; krh@mcmaster.ca

Abstract

Standardized measures are challenged by the need for efficiency without sacrificing sensitivity or specificity. From the perspective of language evaluation, diagnostic tests must accurately identify language impairments and assess often subtle grammatical features of language. Out of the need for better, sensitive tools, researchers and clinicians have turned to more naturalistic forms of testing including conversational and discourse analysis. The present study examined how various language indices obtained from spontaneous language samples relate to one another and how they compare with the Wechsler Preschool and Primary Scale of Intelligence (WPPSI-III) receptive and expressive vocabulary scores in the assessment of preschoolers' language abilities. Results indicated that of the conversational measures only mean length of utterance (MLU) was strongly related to WPPSI language measures. However, there were other measures from conversational speech that appeared to constitute reasonably stable factors that may have utility for children's language assessment. The implications of these findings for both the WPPSI, and for the use of conversational measures in language assessment are discussed.

Key words: language assessment; standardized measure; receptive vocabulary; expressive vocabulary; spontaneous language sampling, WPPSI-III

Spontaneous versus standardized: Examining the relationship between the WPPSI-III and measures from spontaneous language sampling

Many researchers and clinicians seek to describe a child's language by comparing his or her language to the language of other children the same age. On a large scale, this is what standardized measures enable individuals to do. Researchers interested in children's language, or general intellectual abilities, frequently use standardized measures to describe the children they are studying. Practitioners in the field of communicative disorders use these measures to assess children for diagnosis and treatment purposes (Hoff, 2005).

There currently exists a wide range of standardized tests designed to measure the language skills of preschoolers. Some of the most frequently used measures include the Peabody Picture Vocabulary Test- Revised (PPVT-R; Dunn & Dunn, 1981), the Preschool Language Scale – Fourth Edition (PLS-4 Zimmerman, Steiner, & Pond, 2002), and the Clinical Evaluation of Language Fundamentals – Preschool, Second Edition (CELF- Preschool-2 Wiig Secord, & Semel, 2004). For psychological research purposes, the Wechsler Preschool and Primary Scale of Intelligence- third edition (WPPSI-III; Wechsler, 2002), is one of the most commonly used measures of preschoolers' language (and intelligence). Since the WPPSI-III is well standardized and administered regularly by researchers, it represents a useful tool to measure the language and cognitive skills of groups of interest (Rice, Burh, & Nemeth, 1990; Rice, Buhr, & Oetting, 1992). Although the WPPSI-III is used extensively by clinical and school psychologists, and can be the basis of referral to a speech-language pathologist, is not used as frequently by speech-language pathologists themselves.

Despite their widespread use, the efficacy and accuracy of standardized tests in measuring language skills continues to be widely debated. Specifically, standardized psychometric measures have been criticized for their restrictive assessment of language and decreased sensitivity in identifying language impairments (Dunn, Flax, Sliwinski, & Aram, 1996; Gray, Plante, Vance & Henrichsen, 1999; McCauley & Swisher, 1984; Merrell & Plante, 1997; Plante & Vance, 1994). Benefits such as ease of administration and cost effectiveness can sometimes come at the expense of accurately measuring children's language and effectively identifying language impairments. It is out of this need for better, more sensitive tools to assess language, that the notion of using spontaneous language sampling has been gaining recognition (Dunn, Flax, Sliwinski, & Aram, 1996; Klee, 1992; Liles & Duffy, 1995). While spontaneous language sampling and conversational analysis of spoken language has gained popularity over the years, what specific features of spoken language are best suited to analyzing children's speech is not well understood. Of course, some measures from spontaneous speech, notably MLU (mean length of utterance) have been long-recognized markers of language development (Brown, 1973; Miller & Chapman, 1981). However, there are a large number of other potential measures to be taken from spontaneous speech samples. The question to be addressed here is which measures might prove the most useful, and how do the various measures relate to each other.

Studying language in preschool age children is particularly important as speech and language delays are known to affect between 5% and 8% of preschool children, often persisting into school years, and associating with lowered school performance and psychosocial problems (US Preventive Services Task Force, 2006). Prior to age 2 or 3, children's language development is highly variable, making it difficult to assess accurately. By age 3, most children have acquired basic language skills and variability among children's language is dramatically reduced (Hoff, 2005). Consequently, the preschool years represent an ideal time to screen for language impairment in children. Measuring and assessing the language of young children tells us important information about development. Specifically, the production of particular classes of words and disfluencies conveys critical information about language progression as well as deficits (Klee, 1992; Redmond 2004; Rice & Wexler, 1996). One of the goals of this study, therefore, was to examine how elements of conversational speech related to underlying linguistic skills in a non-clinical population. Eventually, if we can determine what features of spoken language are most highly related to language impairment, we can more easily identify and treat these deficits early on, leading to improved outcomes.

WPPSI-III

The WPPSI-III is a test of general intelligence for children between the ages of 2 years 6 months and 7 years 3 months. The complete battery consists of fourteen subtests, including information, vocabulary, word reasoning, comprehension, similarities, block design, picture concepts, picture completion, object assembly, symbol search, matrix reasoning, coding, receptive vocabulary and expressive vocabulary (picture naming) (Rock & Stenner, 2005). For the purpose of this study, we chose to examine receptive vocabulary, expressive vocabulary, and block design subtests. Receptive vocabulary requires the child to look at a group of four pictures and point to the one the examiner names aloud. Expressive vocabulary involves naming pictures displayed in a stimulus book. Block design requires the child to re-create designs using coloured blocks, first based on an actual model and later based on pictures from a stimulus book. Block design also serves as an index of executive function and a proxy of non-verbal IQ without expressive language demands. In this study, raw scores on WPPSI-III receptive vocabulary, expressive vocabulary, and block design were utilized to compare with measures obtained from spontaneous language samples.

Traditionally, the WPPSI-III has been used to predict academic achievement in a variety of domains including reading, math, and general knowledge. Many studies document the predictive validity of the WPPSI-III in effectively predicting achievement on standardized test scores (Bishop & Butterworth, 1979; Kaplan, 1996; Lieblich & Shinar, 1975). In one of several longitudinal studies examining predictive validity of the WPPSI (Wechsler, 1967), 85 children were administered the WPPSI at 5 ½ years of age, and again, 11 years later, using the Wechsler Intelligence Scale for Children (WISC-R). Correlations between the two intelligence tests were extremely high, ranging from 0.73 to 0.86 for verbal, performance, and full scales (Yule, Gold, & Busch, 1982). Similarly, correlations between the WPPSI and a variety of achievement tests have been well documented in the literature (Kaufman, 1973; Krebs 1969; Pasewark, Scherr, & Sawyer, 1974; Reynolds, Wright & Dapper, 1981; Tew & Laurence, 1983). The issue with much of this research is that while it demonstrates the reliability of the WPPSI in predicting scores on the same or similar tests years later, it does not determine whether the constructs themselves are effective in identifying children with specific disorders.

The popularity of standardized tests stems from the fact that they are easy to administer quickly and inexpensively on a broad scale, even when dealing with large populations. Additionally, standardized tests have the benefit of well-documented distributions and reliability and are easily utilized in a variety of ways, including providing diagnostic criteria, determining program entry, and evaluating treatment efficacy.

Spontaneous Language Sampling

Spontaneous language sampling occurs in a naturalistic context (e.g., child's preschool classroom or home) with trained research examiners or individuals with whom the child is familiar. An unstructured conversation between child and examiner, ideally of at least 200 utterances, is then recorded (Bowen, 1999). Topics of interest such as books or videos the children has recently viewed can be used to help guide the conversation, but the procedure is meant to be as naturalistic as possible. Recorded language samples are then transcribed and coded for measures such as word frequency, disfluencies, utterance length, and errors. This process affords clinicians and researchers the opportunity to gain valuable linguistic insight and to study language in far greater detail (Dunn, Flax, Sliwinski, & Aram, 1996; Paul & Smith, 1993)

Evaluating the language of young children can yield important information about development. Specifically, the production of specific classes of words and the use of syntax and pragmatics conveys critical information about language progression as well as deficits (Klee, 1992; Redmond 2004; Rice & Wexler, 1996). As children mature, their language becomes increasingly complex. Sentences get longer and are arranged in new, more complex, combinations (Scott & Stokes, 1995). Researchers refer to this linguistic progression as advances in syntactic complexity. Syntactic complexity can be challenging to measure as the changes that occur over time are subtle. To tackle this issue, researchers have sought to determine what particular aspects of language are 'markers' of linguistic complexity or specific language deficits. For example, Scott and Stokes (1995) identified two language indices as important in the assessment of syntactic complexity in school-age children and adolescents: sentence length and clause density. Their findings showed that as sentence length increased and began to contain subordinate clauses, syntactic structure also became more complex. Similarly, Rice and Wexler (1996) found tense-marking morphemes to be a clinical marker of specific language impairment in English-speaking children.

Spontaneous language sampling is capable of detecting these subtle markers; however, most standardized tests of language are not by nature sensitive enough to pick up these indices. For example, the WPPSI-III is designed to measure single word production and comprehension, not the more complex features of conversational speech such as word frequency, errors, and utterance length.

Many norm-referenced tests of language ability are also very poor at identifying language impairments. Gray et al. (1999) conducted a study examining the diagnostic accuracy of four vocabulary tests; Peabody Picture Vocabulary Test-III [Dunn & Dunn, 1997], Receptive One-Word Vocabulary Test [Gardner, 1985], Expressive Vocabulary Test [Williams, 1997], and the Expressive One-Word Vocabulary Test-Revised [Gardner, 1990]. The four vocabulary tests were administered to preschool age children to determine their ability to screen or identify specific language impairment. Results demonstrated that none of the aforementioned tests were strong identifiers of SLI. In fact, children with SLI typically fell within the normal range. Similarly, Plante and Vance (1994) examined the ability of 21 commonly used tests of language skills in meeting 10 psychometric criteria based on standards set forth by McCauley and Swisher (1984).

Language tests were chosen on the basis of having norms for children between the ages of 4 and 5 and were reviewed on 10 of the psychometric criteria. Results indicated that only 38% of these tests met half or more of the psychometric criteria and all tests had low correlations with measures of nonverbal skills. Furthermore, only one of these tests provided acceptable accuracy in discriminating between children with SLI and age-matched controls. Marini, Tavano and Fabbro (2008) took this research one step further by comparing the linguistic skills of a large group of SLI participants against a group of typically developing peers using a narrative task and two forms of the Wechsler Intelligence Scales; the WPPSI (Wechsler, 1996) for children under 6 years and the Wechsler Intelligence Scale for Children- Revised (WISC-R; Wechsler, 1993) for children above 6. Results demonstrated that both forms of the Wechsler failed to differentiate between individuals with SLI and controls. SLI participants were observed to produce the same quantity of words as controls; however, their narratives were simpler in form and contained significantly more omissions and/or substitutions of bound and free morphemes. Impaired morphosyntactic and syntactic processing are well established linguistic features of SLI, features which the Wechsler tests lacked the sensitivity to capture. The WPPSI and WISC-R were able to account for the number of words spoken by both groups; however, they failed to recognize the more minute grammatical details sometimes critical to SLI assessment and diagnosis. Additionally, in a study examining differences in the conversational performance of 6-year-old children with ADHD but no language impairment, specific language impairment, and typical age-matched controls, Redmond (2004) found utterance formulation measures to be the only conversational indices that showed statistically significant group differences between children with

ADHD and controls. Groups were not observed to differ on the standardized language test administered (Test of Language Development Primary-Third edition (TOLDP-3); Newcomer & Hammill, 1997).

Overall, standardized tests present considerable shortcomings in their ability to identify language impairment and to more generally assess language competence. We propose that spontaneous language sampling may represent a beneficial alternative or addition to standardized tests that allows researchers and clinicians to capture a large amount of linguistic data. Presently, there are few studies that examine what measures of conversational analysis are best suited to assessing children's spoken language and how patterns of these features of speech might emerge developmentally. Traditionally, studies have investigated the structure of conversation samples from a linguistic or theoretical perspective.

The present study, however, took an alternative, data-driven approach to conversational analyses. To do this we transcribed and coded conversation samples from a group of typically developing preschoolers aged 3 to 5 years to see what measures of spontaneous language sampling statistically grouped together and which did not. These findings provide an empirical basis for examining how various elements of conversational speech may relate to some set of underlying linguistic skills. Furthermore, we examined how indices obtained from spontaneous language samples related to one another and compared these with children's WPPSI receptive and expressive vocabulary scores in the assessment of preschoolers' language abilities. While it will take very large scale studies to create diagnostic-grade scales from conversational measures, the data and analyses presented here represent a first step towards this goal.

Method

Participants

Participants included 46 children (24 girls and 22 boys) ranging from 3 years, 4 months to 5 years, 9 months with a mean age of 4 years 6 months (SD= 7.3 months). Recruitment took place at several licensed childcare centers in Hamilton, Ontario, Canada as part of a larger longitudinal study examining language development, executive functioning, and behaviour in preschool children. Parents of preschoolers in the centers were approached and asked to participate in a study examining how executive function and language skills in preschool children relate to other behaviours such as aggression. Informed consent was obtained from all parents and each child's teacher was asked to participate in the study. Children took part in the conversational portion of the study as part of their participation in the larger longitudinal study of intellectual and behavioural development. Inclusion criteria for the study required children to be between 3 and 5 years to participate.

Spontaneous language samples

Sampling Procedure. Spontaneous language samples were obtained by means of a brief recorded conversation. Sampling took place in participants' classrooms at the childcare centers. Samples were recorded using a Marantz CD recorder CDR300, and a Shure omnidirectional boundary microphone. Spontaneous conversation was generated by the examiner discussing either a Dora the Explorer or Bob the Builder book with the child. Examiners were instructed to use open-ended questions to help facilitate and encourage language. For example, examiners might have asked questions such as "what's happening in this picture?" or "tell me more about this character." Sample length typically lasted for a minimum of three minutes and was constrained by the number of other measures gathered at the same session as part of the larger longitudinal study.

Transcription. Trained researchers listened to the recordings and transcribed the conversations using Systematic Analysis of Language Transcripts (SALT) computer software (SALT Version 8.0.4, Miller, 2004). Transcription was completed by several researchers independently and then compared, with discrepancies resolved by one of the original transcribers. As per SALT conventions, utterances produced by each participant were analyzed using SALT guidelines with regard to syntactic, phonological, semantic, and pragmatic properties. Utterances were further categorized according to the number of revisions, repetitions, pauses, etc. Additionally, individual dictionaries were created for each child based on the words produced during language sampling. These words were then categorized into parts of speech (number of different nouns, verbs, adjectives, etc), and coded for spoken and written frequency (Brown, 1984; Kucera & Francis, 1967), word length, number of phonemes, number of syllables, and concreteness ratings (Coltheart, 1981a). To obtain a measure of the number of different words produced as a proportion of the total words in a given sample, we chose to use the computation "D." Traditionally, type token ratio (TTR) has been used to obtain this value, however, several studies document issues with using TTR as an index of language development or impairment (Haas, Haug, & Landry, 1989; Rice & Bode, 1993; Watkins & Kelly, 1995). D is calculated by a program called the Computerized Language Analysis (CLAN), developed by the Child Language Data Exchange System (CHILDES; MacWhinney, 2000). CLAN calculates the value of D by using the probability of new vocabulary as it is introduced to larger samples of speech. This probability is then used to create a mathematical model of the way TTR varies with token size. Finally, the probability is compared with empirical data to produce the value of D. The value of D, therefore, is not based on the number of words in a given sample; rather, it uses mathematical models in conjunction with empirical data to evaluate vocabulary diversity. For a complete list of all 29 spontaneous language measures refer to Table 1.

Coding . In forming our spontaneous language sample items, utterances were determined based on a combination of prosodic boundaries and determinations of thought completion. Contra SALT conventions, utterances were not split at conjunctions ("and", "or", "because," "unless"), in order to be able to detect children's longer, multi-clause sentences. Mean Length of Utterance (MLU) was calculated using the number of words, not morphemes. As per SALT conventions, mazes were coded into the following classifications; whole word repetitions, revisions, pauses, and filled pauses. Revisions were further subdivided into short revisions (1-3 words) and long revisions (4-7 words). Previous research provides some evidence that short and long revisions reflect different processes, (Brennan & Schober, 2001) and that these disfluencies cause less disruption to listeners than filled pauses or repetitions (Fox Tree, 1995). Similarly, Fox Tree, (2001, 2002) reports findings differentiating filled pauses from silent pauses, whereby filled pauses alert listeners that the speaker is still speaking. Refer to Table 2 for examples of disfluency categories. The number of complete words produced varied greatly, ranging from 30 to 225, with a mean of 113.8 (SD = 52.8). To deal with this issue, the present study analyzed the majority of spontaneous indices per every 100 complete words in the sample. Additionally, individual dictionaries were created for each child based on the words produced during language sampling. Given the exploratory nature of the study, we endeavored to capture as much information as possible in selecting items for spontaneous language sampling. All words produced during conversation were categorized into parts of speech (e.g. number of different nouns, verbs, adjectives, etc), as well as syntactic category (e.g. spoken and written frequency, word length, number of phonemes, number of syllables, and concreteness ratings). It should be noted that individual word types were counted, not tokens, and that examination of data distributions deemed whether analyses based on median or mean values were most appropriate. In cases where the data were fairly normally distributed mean values were used. For instances where the data were more skewed, as in word frequency counts, the median was considered a more appropriate measure. All coding was performed by the same research assistant.

To calculate the reliability of our coding, an independent transcriber also transcribed and coded a randomly selected subset of the conversations (8 of 42 children). Since correspondence between the two coders was high, the original coding was used. Correlations between the various measures were as follows: MLU, $r = 0.97$; number of utterances, $r = 0.90$; number of mazes, $r = 0.93$; number of maze words, $r = 0.99$.

Standardized Measures

As part of the larger study, children were tested using the WPPSI-III. In this study raw scores of the WPPSI-III receptive vocabulary, expressive vocabulary, and block design subtests were utilized.

Analytic Plan

Initial bivariate correlation analyses were conducted on all spontaneous language sampling items and WPPSI-III vocabulary and block design scores to examine how our measures relate to standardized test scores. Then, regression analysis on statistically significantly correlated items was then conducted to see what variables from our conversational analysis predicted WPPSI scores over and above MLU, a well-established measure of language performance. To examine the underlying factor structure of this large group of spontaneous language indices, factor analysis was used to determine what aspects of this measure grouped together, and to see whether a meaningful (from a psycholinguistic perspective) set of factors could be seen. These factors were then compared to WPPSI-III receptive vocabulary, expressive vocabulary, and block design scores to determine where they did and did not overlap. Exploring the factor structure of spontaneous measures and how these align with WPPSI-III vocabulary scores may give us a better understanding of the capabilities, and limitations of some standardized tests.

Results

Of all participants in the study, only one child received a WPPSI-III score categorized as "borderline," with the remainder meeting criteria for "low average" or above. Scaled scores, unlike raw scores, have the advantage of accounting for age in their interpretation. Since scaled scores factor in the performance of other children of the same age, direct comparisons can be made more easily. In the case of our study, however, we view our conversational measures as measures of development and predicted that better conversational performance would also be related to higher WPPSI-III raw vocabulary scores. Not accounting for age in this particular instance enabled us to determine whether better WPPSI-III raw vocabulary scores were indicative of better language performance regardless of age. Additionally, since we do not yet know the scaling for our measures (e.g. how quickly and at what age children develop these aspects of conversational spoken language) it becomes more challenging to use scaled scores. Despite these issues, we recognize the importance and value of accounting for age in the interpretation of our findings. To address this, we examined how our conversational measures changed with age to better understand what conversational indices vary with time and which ones are more stable characteristics. Descriptive statistics for intelligence (WPPSI-III Receptive, Expressive, and Block Design) and conversational language measures are shown in Tables 3 and 4.

Bivariate Correlations

To first understand how our indices of spontaneous language related to one another, correlation analysis was conducted on the 29 measures (see Table 5 for full correlation matrix.) As expected, many variables were highly related. For example, MLU in words correlated highly with the number of different nouns $r(44) = 0.43, p < .01$, verbs, $r(44) = 0.59, p < .01$, adjectives $r(44) = 0.33, p < .01$, adverbs, $r(4) = 0.39, p < .01$, proper nouns, $r(44) = 0.38, p < .01$, conjunctions, $r(44) = 0.66, p < .01$, pronouns, $r(44) = 0.47, p < .01$, total words (tokens) $r(44) = 0.71, p < .01$, and total different words (types), $r(44) = 0.63, p < .01$. Similarly, pauses and filled pauses were highly related $r(44) = 0.30, p < .05$. For the category of revisions, we observed high inter-correlations between total number of revisions, average words per revision, number of short revisions, and number of long revisions. Short and long revisions, however, were not correlated. Diversity of vocabulary within various parts of spoken language, including the number of different verbs, nouns, adjectives, adverbs, proper nouns, conjunctions, prepositions, and pronouns, were highly correlated with one another. Finally, morphological errors and word errors were highly correlated $r(44) = 0.91, p<.01$.

Partial Correlations

To then understand how our conversational language measures related to standardized measures of language, partial correlations, controlling for age, were conducted between WPPSI-III raw scores on block design, expressive vocabulary, and receptive vocabulary as well as all conversational measures. Results are depicted in Table 6. Mean length of utterance and items related to types of word counts (e.g. number of different nouns, prepositions, etc) formed the majority of significant correlations with WPPSI-III vocabulary and block design scores.

Regression Analyses

To further understand the relationship between WPPSI-III scores and spontaneous language sampling, we examined the prediction of our conversation measures. Given the strong correlation between MLU and WPPSI-III scores, and the known utility of MLU in measuring language development, we thought it useful to examine whether a subset of our measures improved prediction of WPPSI-III receptive, expressive, or block design scores. To do this, linear regression analysis was conducted on the twelve conversational measures observed to correlate with the WPPSI-III in the previous analysis, to see whether after accounting for MLU, these conversational items improved prediction of WPPSI-III scores. In all analyses, none of the conversational items improved prediction of WPPSI scores beyond that afforded by MLU (all F's < 1).

Factor Analyses

To determine what aspects of our conversational measures grouped together, and whether a meaningful set of psycholinguistic factors could be observed, preliminary principal factors extraction with varimax rotation was conducted on the 29 spontaneous language sampling items. Principal components extraction was used prior to principal factors extraction to estimate the number of factors, presence of outliers, and factorability of the correlation matrices. Factor analysis was then conducted to further determine how measures obtained from spontaneous language sampling group together and to use these factors to determine where spontaneous language measures do and do not overlap with standardized (WPPSI-III) scores. In the interpretation of these analyses, it is important to note that the sample size is smaller than what is normatively desirable for a factor analysis. Nonetheless, this represents a useful preliminary step toward understanding the factors underlying conversational speech in this age group. Given the highly correlated nature of our spontaneous language measures, oblique rotation was also conducted. Results were extremely similar to that of varimax rotation.

Nine factors were extracted with an eigenvalue cutoff of 0.45 for inclusion of a variable in interpretation of a factor. The first factor included 11 of the 29 measures including total number of words, MLU, and number of different words (types) within each part of speech (e.g. nouns, verbs, adjectives, etc). This factor accounted for 25.7% of the total variance. The second factor was related to word structure, specifically, number of syllables, phonemes, and concreteness ratings. This factor accounted for 12.8% of the total variance. The third factor was associated with word frequency and explained 8.6% of the total variance. The fourth factor was word errors, accounting for 7.5% of the total variance. The fifth factor yielded measures associated with fewer pauses and more long revisions. This factor accounted for 7.1% of the total variance. The sixth factor related to revisions, both per 100 words and in terms of short revisions. This factor accounted for 6.1% of the total variance. The seventh factor consisted of repetitions and explained 4.8% of the total variance. The eighth factor, greater vocabulary diversity and fewer phonological errors, accounted for 4.4% of the total variance. The ninth factor pertained to filled pauses and explained 3.7% of the total variance. All items positively correlated with the aforementioned factors except for phonological errors and pauses. Refer to Table 7 for the percentage of variance accounted for by each factor and Table 8 for the order in which variables contributed to the factors.

Correlation analysis was performed to address the relationship between WPPSI-III vocabularies and block design raw scores and measures obtained from spontaneous language sampling. Nine factors extracted from the previous analysis were used to predict standardized test scores (see Table 9). Only one factor was found to correlate with any of the standardized scores. WPPSI-III receptive and expressive vocabulary raw scores correlated positively with factor 1, the amount and diversity of words (MLU/total number of words/total number of different word types of each part of speech) $r(45)=.40$ $p< .01$ and $r(44)= .61$, $p<.01$, respectively.

To further investigate these relationships, partial correlations to control for non-verbal intelligence were completed. After controlling for block design, receptive vocabulary no longer correlated with factor 1. This was also observed for expressive vocabulary.

Discussion

Results obtained from the above analyses provide important information on the progression of language development and the best ways to measure this process. Results obtained from bivariate correlation analyses of 29 spontaneous language measures suggest a reduced list of items as several measures were highly correlated. Many of the correlated variables related to word counts (e.g. MLU and total words) indicating some redundancies in our measures. Individual parts of speech (e.g. total different nouns, verbs, etc) demonstrated high inter-correlations with each other and also with MLU and total words/different words. What this means is that children who have longer MLU's also produced more different nouns, and people who produce more different nouns also tended to produce more unique prepositions, pronouns, etc. This suggests, not surprisingly, that MLU increases with vocabulary, but also interestingly, that vocabulary widens in terms of parts of speech. Additionally, MLU was also highly related to the total number of different words. We would expect MLU to be related to total words, since longer utterances likely mean using more words. However, longer utterances do not necessarily imply greater numbers of different words. This finding adds further support to the notion that as MLU increases, so does vocabulary diversity. Previous research has demonstrated mixed findings on the relationship between MLU and vocabulary diversity as measured by TTR or D. Some authors report highly significant correlations between MLU and D (Duran, Malvern. Richards, & Chipere, 2004), whereas others, document negative findings when examining MLU and TTR (Hansson, 1996). It has been proposed by some that issues over the validity of TTR may explain these negative results (Haas, Haug, & Landry, 1989; Rice & Bode, 1993; Watkins & Kelly, 1995). Variations in the category of revisions (e.g. short, long, and average words per revision) were not highly related, suggesting they may have different functions in spoken production. This adds support to findings obtained from factor analysis, as noted below, suggesting that short and long revisions represent unique and separate factors. Brennan and Schober (2001) also document some evidence that short and long revisions have different effects on listeners. Despite evidence of fundamental differences in short and long revisions, they seem to group together as predictors. It should be noted, however, that individual indices of revisions may contribute different amounts of variance in terms of predicting language performance. This would add support to the concept of using one, most highly predictive measure, but still needs further study.

From a developmental perspective, results of correlation analyses also provide a snapshot of these indices at this point in development and which develop separately. MLU and the number of different nouns, verbs, etc, are different vocabulary measures, however, they do appear to develop together. Similarly, short, long, and average words per revision, are different measures of language, yet developmentally they tend to emerge together.

Correlation analysis between WPPSI-III raw scores and all spontaneous language items revealed a strong relationship between MLU and related word counts with WPPSI-III receptive vocabulary, expressive vocabulary, and block design. Since MLU appeared to be the strongest predictor of WPPSI-III scores, linear regression analysis was conducted on the remaining twelve measures, after accounting for MLU, to determine whether any of these measures improved prediction of WPPSI receptive, expressive, or block design scores beyond that accounted for by MLU. None of these twelve variables were found to improve prediction. These findings provide some confirmation of redundancies observed in our original correlational analysis of 29 spontaneous language items, as many of the variables that were found to correlate with WPPSI-III scores were also highly related to MLU.

Preliminary factor analysis of 29 spontaneous language items revealed nine distinct factors that cumulatively accounted for 80.8% of the variance in our sample. These factors included; parts of speech/number of words, word structure, word frequency, word errors, pauses/long revisions, short revisions, repetitions, and filled pauses. All items positively contributed to the factors, except for phonological errors and pauses. Phonological errors were negatively related to factor eight, indicating that children with greater vocabulary diversity, as measured by D, had fewer phonological errors. One explanation of these findings could be that children with larger and more diverse vocabularies likely have more sophisticated spoken language and produce fewer phonological errors. Similarly, pauses were negatively related to factor five, suggesting that children with longer revisions and more words per revision produced fewer pauses. Since pauses in spoken language are sometimes indicative of planning utterances, children who are observed to produce longer revisions may take less opportunity to pause and plan resulting in longer revisions. Critically, our findings demonstrate how individual categories of disfluencies including pauses, short revisions, long revisions, and repetitions, are different features of spoken language. Specifically, short and long revisions appear to represent distinctive factors or predictors, and as previous research indicates, may have differing impacts on speakers and listeners (Brennan & Schober, 2001). Clinically, these findings highlight the importance of analyzing specific categories of disfluencies on their own, rather than as a whole category.

Data from the factor analysis also revealed important information about how specific indices of language group together. For example, factor 1, parts of speech and number of words, consists of 11 individual language measures. Some of these measures, however, are more highly correlated with this factor than others. For factor 1, MLU and total words were better predictors than number of pronouns or prepositions. This would suggest that some measures, while contributing to the factor, may not represent the best indicators. However, it is important to note that pronouns and prepositions are less likely to be as highly correlated since people produce fewer of them as compared to nouns or verbs. It is interesting, then, that they still appear as predictors despite limitations in their range or production. In the case of our analysis, it may be more efficient and valuable to collect information on MLU and total words instead of the number of pronouns, prepositions and adjectives. Results of correlational analysis also suggest the creation of a pared down list that would be more economical and efficient in evaluating preschool children's language, but that appear to load on separable underlying factors. This list would likely include a measure of short revisions, long revisions, pauses, MLU, total words, word frequency, and morphological error, as related items were highly correlated with one another.

To determine where conversational and standardized measures overlap, correlation of the nine factors with WPPSI-III expressive, receptive, and block design raw scores was conducted. Results of correlation analysis identified several important distinctions. Of the 29 spontaneous items, WPPSI-III receptive and expressive vocabulary scores correlated with only one, MLU/number of specific parts of speech/number of words. Block design scores, however, did not correlate with any factors. Not surprisingly, our results suggest that children who produce more words, longer utterances, and more types of words (nouns, verbs, adverbs, etc) have higher receptive and expressive scores. The key finding, however, is that only one of the nine critical factors were captured by the WPPSI-III since many linguistic measures, including word errors, word structure, revisions, pauses, filled pauses, and repetitions, were not captured by this standardized test. Previous research has clearly demonstrated the crucial role these subtle features of spoken language play in effectively identifying language impairments (Marini, Tavano, & Fabbro, 2008; Redmond, 2004; Rice & Wexler, 1996). Given that many of these features did not overlap with WPPSI-III vocabulary measures, our findings provide evidence of the need for additional, more comprehensive, testing to accurately assess the language of preschoolers. Partial correlations, controlling for block design, revealed that receptive and expressive vocabulary no longer correlated with factor 1. These findings indicate that the variance that the WPPSI-III receptive and expressive vocabulary raw scores and factor 1 share is likely attributable to general intelligence rather than language-specific factors.

Overall, our findings suggest two possibilities. The first is that our measures of spontaneous language and subsequent factor structures are not reliable, explaining the lack of significant correlations with the WPPSI-III. Alternatively, our results may demonstrate real individual differences in language skills that are not well captured by the WPPSI-III.

Several limitations to the study should be noted. First, the sample only consisted of 46 children which may have generalizability issues when attempting to scale these results to a broader population, including other age groups, or children with frankly diagnosed language disorders. Secondly, we do not have any actual proof that our factors predict specific behaviours such as success in school or problems with language development. However, they do give us an empirical way to look at how these measures align.

Factor analysis of spontaneous language measures in our database revealed a nine-factor solution. This is relatively large, and even though all measures met cut-off criteria and do seem interpretable according to our a priori theories about language processing, it would be necessary to see if it these factors are still evident in a larger population. It is also important to note that the study utilized a general population and although some children had low WPPSI-III scores that may warrant SLP evaluation, it is not a clinical population. We acknowledge that it would take very large scale studies to create diagnostic-grade scales from conversational measures, however, the data and analyses presented here represent a first step towards that goal.

Despite limitations of sample size, findings shed important light on the process of assessing the spoken language of young children. Disparities between measures of spontaneous language sampling and standardized tests of language support the use of conversational analysis in addition to standardized testing. Similarly, findings inform what indices of conversational analysis would be most appropriate and useful to measure. Finally, knowledge obtained from analyzing samples of spontaneous language reveal important information about the developmental process and emergence of language.

References

Bishop, D., & Butterworth, G.E. (1979). A longitudinal study using the WPPSI and WISC-R with an English sample. *British Journal of Educational Psychology, 49,* 156-168.

Bowen, C. (1999). *Structural analysis of a language sample.* Retrieved from: http://members.tripod.com/caroline_bowen/minisample.htm

Brennan, S. E., & Schober, M. F. (2001). How listeners compensate for disfluencies in spontaneous speech. *Journal of Memory and Language, 44*, 274-296.

Brown, G.D.A. (1984). A frequency count of 190,000 words in ther London-Lund Corpus of English Conversation. *Behavioural Research Methods Instrumentation and Compu- taters, 16* 502-532.

Brown, R. (1973). *A first language: The early stages.* London: George Allen & Unwin Ltd.

Coltheart, M. (1981a). MRC Psycholinguistic Database User Manual: Version 1. [Available from Professor Coltheart at Birkbeck College, London, U.K.]

Dunn, L.M., & Dunn, L. (1981). *Peabody Picture Vocabulary Test Revised.* Circle Pines, MN: American Guidance Service.

Dunn, L.M., & Dunn, L. (1997). *Peabody Picture Vocabulary Test* (3rd ed.). Circle Pines, MN: American Guidance Service.

Dunn, M., Flax, J., Sliwinski, M., & Aram, D. (1996). The use of spontaneous language measures as criteria for identifying children with specific language impairment: An attempt to reconcile clinical and research incongruence. *Journal of Speech, Language, and Hearing Research, 39,* 643-654.

Duran, P., Malvern, D., Richards, B., & Chipere, N. (2004). Developmental trends in lexical diversity. *Applied Linguistics, 25,* 220-242.

Gardner, M. (1985). *Receptive One-Word Picture Vocabulary Test.* Novato, CA: Academic Therapy Publications.

Gardner, M. (1990). *Expressive One-Word Picture Vocabulary Test-Revised.* Novato, CA: Academic Therapy Publications.

Gray, S., Plante, E., Vance, R., & Henrichsen, M. (1999). The diagnostic accuracy of four vocabulary tests administered to preschool-age children. *Language, Speech, and Hearing Services in Schools, 30,* 196-206.

Haas, C.W., Haug, H.T., & Landry, R.G. (1989). Sample size and type-token ratios for oral language of preschool children. *Journal of Speech and Hearing Research, 32,* 536-540.

Hansson, K. (1996). Verb type token ratio in Swedish children with SLI: A reply to Richards and Malvern. *Logopedics Phoniatrics Vocology, 21,* 112-113.

Hoff, E. (2005). *Language Development.* 3rd Ed. Belmont, CA: Thomson Wadsworth.

Kaplan, C. (1996). Predictive validity of the WPPSI-R: A four year follow-up study. *Psychology in the Schools, 33,* 211-220.

Kaufman, A.S. (1973). Comparison of the WPPSI, Stanford-Binet, and McCarthy Scales as predictors of first- grade achievement. *Perceptual and Motor Skills, 36,* 67-73.

Klee, T. (1992). Developmental and diagnostic characteristics of quantitative measures of children's language production. *Topics in Language Disorders, 12,* 28-41.

Krebs, E.G. (1969). The Wechsler Preschool and Primary Scale of Intelligence and prediction of reading achievement in first grade. *Dissertation Abstracts International, 30,* 4279-A (University Microfilms, No. 70-3361).

Kucera and Francis, W.N. (1967). *Computational Analysis of Present-Day American English.* Providence: Brown University Press. Lieblich, A., & Shinar, M. (1975). The predictive validity of the WPPSI with Israeli children. *Educational and Psychological Measurement, 35,* 473-475.

Liles, B.Z., & Duffy, R.J. (1995). Measurement of narrative discourse ability in children with language disorders. *Journal of Speech, Language, and Hearing Research, 38,* 415-425.

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk. 3rd Edition.* Mahwah, NJ: Lawrence Erlbaum Associates.

Marini, A., Tavano, A., & Fabbro, F. (2008). Assessment of linguistic abilities in Italian children with Specific Language Impairment. *Neuropsychologia, 46,* 2816-2823.

McCauley, R.J., & Swisher, L. (1984a). Psychometric review of language and articulation tests for preschool children. *Journal of Speech and Hearing Disorders, 49,* 34-42.

Merrell, A.W., & Plante, E. (1997). Norm-referenced test interpretation in the diagnostic process. *Language, Speech, and Hearing Services in Schools, 28,* 50-58.

Miller, J.F. (2004). *Systematic Analysis of Language Transcripts* v. 8.0.4. Madison, WI: University of Wisconsin-Madison

Miller, J.F., & Chapman, R.S. (1981). The relation between age and mean length of utterance in morphemes. *Journal of Speech and Hearing Research, 24,* 154-161.

Newcomer, P. L., & Hammill, D. D. (1997). *Test of Language Development-Primary, Third Edition.* Austin, TX: Pro-Ed.Paul, R., & Smith, R.L. (1993). Narrative skills in 4-year-olds with normal, impaired, and late-developing language. *Journal of Speech, Language, and Hearing Research, 36,* 592-598.

Pasewark, R., Scherr, S., & Sawyer, R. (1974). Correlations of scores on the Vane Kindergarten, Wechsler Preschool and Primary Scale of Intelligence, and Metropolitan Reading Readiness tests. *Perceptual Motor Skills, 38*, 518.

Paul, R., & Smith, R.L. (1993). Narrative skills in 4-year-ols with normal, impaired, and late-developing language. *Journal of Speech, Language, and Hearing Research, 36,* 592-598.

Plante, E., & Vance, R. (1994). Selection of preschool language tests: A data-based approach. *Language, Speech, and Hearing Services in Schools, 25,* 15-24.

Redmond, S.M. (2004). Conversational profiles of children with ADHD, SLI and typical development. *Clinical Linguistics and Phonetics, 18,* 107-125.

Reynolds, C.R., Wright, D., & Dappen, L., (1981). A comparison of the criterion-related validity (academic achievement) of the WPPSI and the WISC-R. *Psychology in the Schools, 18,* 20–23.

Rice, M.L., & Bode, J.V. (1993). GAPS in the verb lexicons of children with specific language impairment. *First Language, 13,* 113-131.

Rice, M.L., Buhr, J.C., & Nemeth, M. (1990). Fast mapping word-learning abilities of language-delayed preschoolers. *Journal of Speech and Hearing Disorders, 55,* 33-42.

Rice, M.L., Buhr, J.C., & Oetting, J.B. (1992). Specific-language-impaired children's quick incidental learning of words: The effect of a pause. *Journal of Speech and Hearing Research, 35,* 1040-1048.

Rice, M.L., & Wexler, K. (1996). Toward tense as a clinical marker of a specific language impairment in English-speaking children. *Journal of Speech and Hearing Research, 39,* 1239-1257.

Rock, D.A., & Stenner, A.J. (2005). Assessment issues in the testing of children at school entry. *The Future of Children, 15,* 15-34.

Scott, C.M., & Stokes, S.L. (1995). Measures of syntax in school-age children and adolescents. *Language, Speech, and Hearing Services in Schools, 26,* 309-319.

Tew, B.J., & Laurence, K.M. (1983). The relationship between spina bifida's children intelligence scores on school entry and at school leaving: A preliminary report. *Child: Care, Health, and Development, 9,* 13-17.

US Preventative Services Task Force. (2006). Screening for speech and language delay in preschool children: Recommendation statement. *Pediatrics, 117,* 497-501.

Watkins, R.V., & Kelly, D.J. (1995). Measuring children's lexical diversity: Differentiating typical and impaired language learners. *Journal of Speech & Hearing Research, 38,* 1349-1356.

Wechsler, D. (1967). *Manual of the Wechsler Preschool and Primary Scale of Intelligence.* New York: Psychological Corporation.

Wechsler, D. (1993). *Wechsler Intelligence Scale for Children- Revised. (WISC-R).* New York: Psychological Corporation.

Wechsler, D. (1996). *WPPSI: Scala d'intelligenza Wechsler per bambini. Firenze.* Firenze, Italia: Organizzazioni Speciali.

Wechsler, D. (2002). *Wechsler Preschool and Primary Scale of Intelligence, 3rd. ed. (WPPSI-III).* New York: Psychological Corporation.

Williams, K. (1997). *Expressive Vocabulary Test.* Circle Pines, MN: American Guidance Service.

Yule, W., Gold, D.R., & Busch, C. (1982). Long-term predictive validity of the WPPSI: An 11-year follow-up study. *Personality and Individual Differences, 3,* 65-71.

Table 1

Spontaneous Language Sampling Items

Items Coded
Pause total per 100 words
Filled pause total per 100 words
Word level error total
Phonological error total
Morphological error total
Total number of words
D Average
Complete words total
MLU in words
Number of repetitions per 100 words
Number of revision per 100 words
Number of short revisions per 100 words
Number of long revisions per 100 words
Average words per revision
Average words per repetition
Total number of different words
Total different nouns
Total different verbs
Total different adjectives
Total different adverbs
Total different proper nouns
Total different conjunctions
Total different prepositions
Total different pronouns
Average concreteness ratings
Average phonemes
Average syllables
Median log written frequency
Median log spoken frequency

Table 2

Disfluency Category Examples

Disfluency category	Example
Stalls	
Pause	"I went (0:05)* to the store."
Filled pause	"I went to the (um) store"
Repetition	"This is (the) the bed."
	"I saw a deer (a deer) in the forest."
Revisions	
One word revision	"I went to go (upstairs) downstairs."
Long revision (2-7 words)	"Sarah was walking and then she (walk home to go) had to run to the store."

*Numbers mark the length of the silent pause. Following SALT coding, all disfluencies are contained within brackets.
** Curly brackets contain the actual production of the word preceding the brackets.

Table 3

Descriptive Statistics for raw scores of WPPSI Block Design, Receptive and Expressive Vocabulary

	Minimum	Maximum	Mean	Std. Deviation	N
Block Design subtest – WPPSI[a]	13	28	20.64	3.41	42
Receptive Vocabulary subtest – WPPSI[a]	14	32	24.11	4.48	45
Expressive Vocabulary subtest – WPPSI[a]	14	27	20.30	3.53	44

Table 4

Descriptive Statistics for Conversational Language Measures

Conversational measures (N = 46)	Min	Max	Mean	S.D.
Length of speech				
Mean Length of Utterance (MLU) in words	1	11	4.13	1.86
Total words	30	225	113.87	52.79
Average word length	1.27	5.42	4.31	0.57
Maze measures				
Pause per 100 words	0	7.69	2.97	1.88
Filled pause per 100 words	0	2.80	0.23	0.60
Revision per 100 words	0	2.30	0.33	0.59
Repetition per 100 words	0	3.92	1.25	1.27
Short revisions per total words	0	4	.002	.005
Long revisions per total words	0	.02	.001	.003
Average words per revision	0	7	0.82	1.57
Average words per repetition	0	4	1.2	1.12
Error measures				
Phonological error total	0	41	10.76	9.22
Morphological error total	0	5	0.91	1.28
Word level error total	0	5	1.15	1.37
Lexical diversity measures				
Total different words	16	116	61.73	23.40
Total different nouns	3	31	14.22	6.75
Total different verbs	0	35	12.60	6.89
Total different adjectives	0	12	5.51	3.12
Total different adverbs	0	11	4.38	2.64
Total different proper nouns	0	9	3.49	2.28
Total different conjunctions	0	4	1.58	1.01
Total different prepositions	0	6	3.13	1.41
Total different pronouns	1	15	6.64	3.39
Average concreteness ratings	274.45	456.33	341.34	29.51
Average phonemes	2.58	3.60	2.94	0.21
Average syllables	1.05	1.33	1.18	0.07
Median log written frequency	1.64	2.34	2.0	0.21
Median log spoken frequency	1.34	2.60	2.02	0.34
D Average	17.77	86.58	46.78	18.39

Table 5

Correlation Matrix of Conversational Language Measures

	Pause	Filled Pause	Repetition	Revision	D Average	Total Words	Total Different Words	Average Concrete	Average Phonemes
MLU in Words	.22	.12	-.11	.13	.03	.71**	.63**	-.20	-.01
Pause		.30*	.03	-.17	.00	.09	.15	-.12	.00
Filled Pause			-.19	-.04	.11	.04	.07	-.05	.03
Repetition				-.34*	-.03	-.05	-.02	-.08	.10
Revision					-.10	.10	.07	-.03	.03
D Average						.26	.52**	.01	-.07
Total Words							.94**	-.17	-.06
Total Different Words								0	-.05
Average Concrete									.49**

*$p < .05$. **$p < .01$.

	Median Written Frequency	Median Spoken Frequency	Average Words per Revision	Average Words per Repetition	Short Revisions	Long Revisions	Noun	Verb	Adjective
MLU in Words	-.17	-.08	.14	.02	.14	.07	.43**	.59**	.33*
Pause	-.18	-.12	-.23	-.02	-.06	-.21	.30*	.15	.09
Filled Pause	.05	.04	-.04	-.16	.03	-.12	.02	.10	-.09
Repetition	.01	.01	-.29	.68**	-.23	-.26	.01	-.01	.03
Revision	-.26	-.35*	.67**	-.19	.84**	.51**	.01	.06	.06
D Average	.27	.12	.06	-.02	-.05	-.13	.32*	.44**	.49**
Total Words	-.15	-.13	.32*	.23	.13	.00	.61**	.83**	.64**
Total Different Words	-.12	-.14	.28	.18	.10	-.04	.67**	.87**	.73**
Average Concrete	-.06	-.02	-.03	-.22	.01	-.08	.18	-.28	.02
Average Phonemes	-.29*	-.23	0	.03	-.02	.07	.09	-.18	.25
Average Syllables	.02	.06	-.09	.01	-.18	.16	-.10	-.32*	.06
Median Written Frequency		.90**	-.20	.07	-.13	-.28	-.29	-.12	.04
Median Spoken Frequency			-.26	.04	-.22	-.29	-.25	-.19	.01
Average Words per Revision				-.09	.34*	.70**	.11	.31*	.12
Average Words per Repetition					-.06	-.24	-.08	.24	.06
Short Revisions						-.04	.07	.05	.09
Long Revisions							-.08	.03	-.03
Noun								.37*	.53**
Verb									.48**

$*p < .05$. $**p < .01$.

	Adverb	Proper Noun	Conjunction	Preposition	Pronoun	Phonological Error	Morphological Error	Word Error
MLU in Words	.39**	.38**	.66**	.67**	.47**	.11	.03	.13
Pause	-.01	.08	.16	.18	-.07	-.01	-.01	-.06
Filled Pause	-.09	.17	.15	.10	0	.12	.14	.13
Repetition	-.06	-.19	.10	-.13	-.02	-.01	.01	-.01
Revision	.30*	-.16	.01	.18	.10	.01	-.29	-.19
D Average	.34*	.50**	.14	.05	.41**	-.15	.25	.20
Total Words	.60**	.62**	.67**	.65**	.70**	.15	.26	.36*
Total Different Words	.66**	.67**	.61**	.60**	.74**	.09	.31*	.36*
Average Concrete	-.27	-.10	-.13	-.27	-.30*	-.02	-.06	-.09
Average Phonemes	-.01	-.12	-.09	-.18	-.19	.03	-.12	-.17
Average Syllables	-.24	-.14	-.17	-.19	-.25	.05	.01	-.02
Median Written Frequency	-.04	.15	-.15	-.24	.02	-.04	.10	-.02
Median Spoken Frequency	-.11	.13	-.10	-.16	-.03	-.02	.08	.03
Average Words per Revision	.42**	.15	.29	.13	.32*	.01	-.24	-.13
Average Words per Repetition	.12	.05	.23	-.01	.23	-.11	.04	.07
Short Revisions	.23	-.09	.02	.13	.05	.09	-.21	-.13
Long Revisions	.16	-.16	-.01	.14	.10	-.09	-.20	-.13
Noun	.21	.42**	.46**	.44**	.30*	.11	.22	.20
Verb	.56**	.57**	.55**	.50**	.72**	.10	.29	.39**
Adjective	.62**	.42**	.32*	.41**	.41**	.10	.36*	.33*
Adverb		.39**	.40**	.40**	.47**	-.01	.04	.07
Proper Noun			.49**	.32*	.39**	.03	.10	.11
Conjunction				.42**	.41**	.04	.05	.15
Preposition					.36*	.22	.07	.19
Pronoun						-.10	.27	.29*
Phono-logical Error							.16	.31*
Morpho-logical Error								.91**

$*p < .05$. $**p < .01$.

Table 6

Correlation between Conversational Language Measures and WPPSI raw scores (controlling for age)

Conversational Language measure	WPPSI Receptive Vocabulary (raw scores)	WPPSI Expressive Vocabulary (raw scores)	WPPSI Block Design (raw scores)
	N = 45	N = 44	N = 42
MLU (words)	.35*	.33*	.33*
Total words	.36*	.48**	
Total different words	.36*	.56***	
Pronoun Total		.39*	
Preposition Total	.35*	.41**	.34*
Conjunction Total		.34*	
Other Total	.43**		
Adverb Total	.32*	.48**	
Adjective Total		.48**	
Proper Noun Total		.46**	
Noun Total		.36*	
Verb Total		.43**	
Phonological Error Total	-.43**		
Pause per 100 words			.45**

*$p < .05$. **$p < .01$. ***$p < .001$

Table 7

Percent of Variance Explained by each of the Rotated Orthogonal Factors

Factors	Percent of variance
1	25.72
2	12.84
3	8.55
4	7.53
5	7.11
6	6.13
7	4.81
8	4.44
9	3.71

Table 8

Order (by Size of Loadings) in which Variables Contribute to Factors

Factor 1	**Factor 2**	**Factor 3**	**Factor 4**	**Factor 5**	**Factor 6**	**Factor 7**	**Factor 8**	**Factor 9**
Parts of Speech/# Words	***Word Structure***	***Word Frequency***	***Word Error***	***Pause/Long Revision***	***Short Revision***	***Repetition***	***Vocabulary Diversity/Phonological Error***	***Filled Pause***
Total Different Words	Avg Phonemes	Written Frequency	Word Error	Long Revisions	Short Revisions	Avg Word per Repetition	D Avg	Filled Pause
Total Words	Avg Syllables	Spoken Frequency	Morphological Error	Avg Word per Revision	Revisions	Repetitions	Phonological Error	
Verbs	Avg Concreteness			Pause				
MLU								
Conjunctions								
Adjectives								
Proper Nouns								
Nouns								
Adverbs								
Prepositions								
Pronouns								

Table 9

Correlations between Spontaneous Language Factors and WPPSI raw scores

Factors	WPPSI Receptive Vocabulary (raw scores)	WPPSI Expressive Vocabulary (raw scores)	WPPSI Block Design (raw scores)
	N = 45	N = 44	N = 42
Parts of Speech/ # Words	.40**	.61**	.28
Word Structure	-.12	.06	-.20
Word Frequency	-.16	.04	-.30
Word Error	-.11	-.19	-.11
Pause/ Long Revision	-.11	.02	-.25
Short Revision	-.08	.03	-.07
Repetition	-.10	-.09	-.01
Vocabulary Diversity/Phonological Error	.02	.25	-.05
Filled Pause	.07	-.07	.09

*$p < .05$. **$p < .01$.

Chapter 3

Running head: CONVERSATION AND BEHAVIOURAL AND EMOTIONAL PROBLEMS

Using conversational measures of language to predict behavioural and emotional problems in preschool children

Johanna K. Lake[1]

Karin R. Humphreys[1*]

Tracy Vaillancourt[1,2]

[1]Department of Psychology, Neuroscience & Behaviour, McMaster University

[2]Faculty of Education and School of Psychology, University of Ottawa

*Corresponding author: Karin Humphreys, Department of Psychology, Neuroscience & Behaviour, McMaster University, 1280 Main St West, Hamilton, ON L8S 4K1, Canada; phone: (905) 525-9140 x23011, fax: (905) 529-6225; krh@mcmaster.ca: Natural Sciences and Engineering Research Council of Canada grant #293145 to KRH.

Abstract

Previous research suggests a link between language impairment and psychiatric symptoms in children. There are few studies, however, that examine the relationship between conversational measures of language and the continuum of behavioural and emotional problems, ranging from normal behaviour to clinically significant psychiatric symptoms. The present study utilized spontaneous language sampling and analysis of conversational transcripts, to study the relationship between language and maladaptive behavioural and emotional problems in a non-clinical, and therefore highly variable, sample of preschool children. The utility of using language measures obtained from spontaneous speech in predicting mental health problems as indexed by the Child Behavior Checklist (CBCL C-TRF) problems scales was also examined and compared to the predictability of the WPPSI-III general language composite scores (GLC). Speech samples were collected from 46 preschoolers from a general daycare population, along with standardized measures of intelligence (WPPSI-III), as well as behavioural and emotional problems (CBCL C-TRF). Results indicated that conversational measures of language significantly improved prediction of CBCL C-TRF DSM-oriented and syndrome scales beyond that accounted for by WPPSI GLC scores. Revisions, specifically, appeared to play a critical role in the prediction of behavioural and emotional problems in preschoolers.

Key words: language assessment; revisions; spontaneous language sampling; WPPSI-III; child psychopathology; CBCL C-TRF

Introduction

Children with language impairments are over-represented in the clinical population, with half of children in mental health clinics presenting with language impairments, and half of children in speech and language clinics diagnosed with comorbid behavioural or emotional disorders (Cohen, 2001; Cohen, Barwick, Horodezky, Vallance, & Im, 1998a; Cohen et al., 1998b). Preschoolers with behavioural or emotional problems also frequently present with and show increased risk for comorbid psychiatric conditions. For example, anxiety disorders in children are frequently comorbid with depression, ADHD, oppositional defiant disorder (ODD) or conduct disorder (Angold, Costello, & Erkanli, 1999; Beesdo, et al., 2007; Crick & Grotpeter, 1995; Last, Perrin, Hersen, & Kazin, 1996; Rockhill, Kodish, DiBattisto, Macias, Varley, & Ryan, 2010). Critically, of these children, only 10% are referred for further evaluation and treatment, mainly due to diagnostic difficulties and the fact that there is currently no gold standard for screening mental health disorders (Rockhill, Kodish, DiBattisto, Macias, Varley, & Ryan, 2010; Tervo, 2007).

Language and Clinical Disorders in Children

Variability in early language development largely contributes to the challenge of clarifying the complex relationship between language proficiency, behavioural difficulties, and emotional problems. Debate over how best to define and assess clinical populations has led clinicians and researchers to use qualitative assessment tools such as structured interviews or teacher and parents reports, since they are more sensitive to differentiation of disorders and more helpful in determining appropriate treatment (Aylward, 1985; Jensen, Hoagwood, & Zitner, 2006; Mash & Hunsely, 2005a).

Over the years, many researchers have investigated the association between language and childhood disorders with mixed findings. Rescorla and Alley (2001) found no significant association between language delay as measured by the Language Development Survey (LDS) (Rescorla, 1989) and behavior problems as measured by the Child Behavior Checklist (CBCL; Achenbach & Rescorla, 2000) total problems scale, in children under 3 years of age. In another study, this time using a national probability sample of 278 children between the ages of 18 and 35 months, again, no correlations between LDS and CBCL scores were observed (Rescorla & Achenbach, 2002). Several other researchers document similar findings (Carson, Klee, Perry, Donaghy, & Muskina, 1997; Rescorla, Ross, & McClure, 2007).

Despite these results, there remains a large body of research linking language delay, as measured by standardized tests of language, to a multitude of problems, including ADHD, anxiety, social problems, and poor academic achievement (Beitchman et al., 1996; Cohen, Barwick, Horodezky, Vallance, & Im, 1998; Redmond & Rice, 1998). For example, Caulfield, Fischel, DeBaryshe and Whitehurst (1989) compared the language of 34 late talkers between the ages of 24 and 32 months, to a group of typically developing children matched for age, sex, and receptive language. Late talkers were observed to be more fearful or shy in new situations, exhibited higher levels of crying, screaming, hitting, and throwing toys, and also had more behaviour problems at bedtime than their typically developing peers. In another study comparing 14 late talking toddlers with 14 typically developing toddlers, late talkers were 17 times more likely to exhibit symptoms of depression/withdrawal and deficits in social relatedness than children who were not language delayed (Irwin, Carter, & Briggs-Gowan, 2002). Social-emotional functioning of toddlers in this study was assessed by the Infant-Toddler Social and Emotional Assessment (Briggs-Gowan, Carter, Irwin, Wachtel, & Cicchetti, 2004), the CBCL, Vineland Adaptive Behavior Scales-Expanded Form (Sparrow, Cicchetti, & Balla, 2005) and the Parent-Child Early Relational Assessment (Clark, 1999). Using a sample of 1,189 children between 22 and 29 months of age, Horwitz, et al., (2003) reported poor attention, noncompliance, and low social competence in children with language delay between 18 and 29 months. Additionally, children with language delay were four times more likely to have externalizing behaviour problems by age 30 months than typically developing peers. Several other studies document similar findings of a relationship between language delay and behaviour problems in preschool and school-age children (Beitchman et al., 1996; Cohen, Barwick, Horodezky, Vallance, & Im, 1998; Redmond & Rice, 1998; Stevenson, Richman, & Graham, 1985; Qi & Kaiser, 2004).

The large majority of studies investigating language and behavioural or emotional problems rely heavily on standardized tests of language to assess communicative competence. Part of the reason for these disparate findings, therefore, could be that standardized tests of language are not sensitive enough to pick up on the more subtle features of language distinguishing these groups. To tackle this issue, some researchers have turned to analyzing conversation samples of children's spoken language to obtain a more comprehensive linguistic picture.

In a study examining the relationship between conversational measures of spoken language and attention deficit hyperactivity disorder (ADHD), Barkley, Cunningham, and Karlsson (1983), compared the conversation samples of 9-year-old children with and without ADHD. No differences were observed between groups in terms of the number of utterances and average number of syllables. Similarly, Zentall (1988) examined the total number of words, sentences, grammatical errors, repetitions and revisions in the conversation samples of 9-year-old children with and without ADHD. Again, results indicated no statistically significant differences between groups regarding the number of disfluencies (repetitions and fillers) produced.

Still, some researchers document the use of linguistic analysis of speech samples in reliably and accurately classifying adults into diagnostic categories including schizophrenia, depression, paranoia, and somatization disorders (e.g., Oxman, Rosenberg, & Tucker, 1982; Rude, Gortner, & Pennebaker, 2002; Tucker & Rosenberg, 1975). Further support for the use of conversational analyses in differentiating clinical groups is reported in a study examining the language characteristics of 11 children with ADHD and 11 typically developing children between the ages of 6 and 8 years (Kim & Kaiser 2000). Children were tested using standardized measures (Peabody Picture Vocabulary Test, Dunn & Dunn, 1997;Test of Pragmatic Language, Phelps-Terasaki, & Phelps-Gunn 1992; and Test of Language Development, Hresko, Reid, & Hammill, 1999) combined with a spontaneous language sample obtained during free play with an adult. Results indicated no differences between groups on receptive vocabulary as measured by the PPVT or pragmatic knowledge as measured by the TOPL, however, children with ADHD were found to produce more inappropriate pragmatic behaviors in conversational interactions.

In another study, Redmond (2004) reports similar findings when examining differences in the conversational performance of 6-year-old children with ADHD but no language impairment (N = 10), with specific language impairment (SLI; N = 10), and typical age-matched controls (N = 13). Conversational measures used to analyze group differences included the category maze, defined as the part of an utterance containing a false start, filler, repetition or revision (Redmond, 2004). Results indicated that utterance formulation indices were the only conversational measures that showed statistically significant group differences between children with ADHD and controls. Children with ADHD produced on average more words per maze and had a greater percentage of maze words (out of the total number of words) compared to typically developing, age-matched controls. Interestingly, the groups did not differ on the standardized language test administered (Test of Language Development Primary-Third edition (TOLDP-3); Newcomer & Hammill, 1997).

Lake, Humphreys, and Cardy (2010) document related findings in a population of young adults with autism spectrum disorders (ASDs). Analysis of spontaneous language samples of young adults with and without ASD showed that individuals with ASD produced fewer filled-pause words (ums and uhs) and revisions than those without ASD, but more silent pauses. In a similar study, Wetherby and Prutting (1984) examined the use of pragmatics in the language of children with autism. Results of their research determined that compared to the language of typically-developing children, children with autism showed higher frequencies of requesting objects and actions, protesting, and non-focused utterances. Furthermore, unlike typically-developing children, those with autism were not observed to request information, acknowledge others, show off, comment, or label (Wetherby & Prutting, 1984). To capture conversational impairments in ASD, de Villiers, Fine, Ginsberg, Vaccarella, and Szatmari (2007) developed a rating scale measuring pragmatic impairment in the conversation of individuals with autism. Based on analysis of semi-structured conversations of children and adolscents with ASD, five constructs were developed including atypical intonation, pedantic speech, terseness, semantic drift, and perseveration.

Overall, many of these studies document the role of fairly subtle aspects of spoken language such as revisions, repetitions, and pauses, in discriminating clinical and non clinical groups (Belser & Sudhalter, 2001; Redmond, 2004; Zentall, 1988). Little is known about why these disfluencies may represent linguistic markers of disorder, but some researchers have studied the processes underlying the production of disfluencies in typically developing children. Rispoli and colleagues (2008) have conducted extensive research on revisions in children's spoken language, caftegorizing disfluencies into stalls and revisions. Stalls are interruptions that do not change the linguistic structure produced (e.g., ums, uhs, silent pauses, etc), whereas revisions change structures produced by adding or deleting words, phrases, or parts of words (Rispoli, Hadley, & Holt, 2008). It is proposed that stalls and revisions represent discrete types of disfluencies originating from different problems in language production. In 2008, Rispoli, Hadley and Holt set out to test this theory by studying the sentence disruptions of 20 typically developing children.

Results demonstrated that revisions occurred in approximately 1% of children's utterances at 27 months and increased with age. Stall rates, by contrast, did not change with age. The number of stalls did, however, increase with sentence length, whereas the number of revisions remained constant. These findings lend support for the stall-revision dichotomy, and the authors argue that increased rates of revisions with age coincide with children's ability to monitor language production. Stalls, they assert, arise from glitches in sentence production, the rate of which increases as sentences become longer. In 2003, Rispoli conducted a further study, this time observing revision rates to increase with level of grammatical development; while stall rates did not. Rispoli suggests therefore, that the capacity for self-monitoring in children increases during grammatical development, with stall rates remaining unaffected by this process.

Despite these findings of connections between aspects of conversational spoken language and childhood disorders, little conclusive research exists examining conversation samples of children with a wide range of potential behavioural and emotional problems. Similarly most studies examine the association between language impairment and mental health in school-age children, even though research indicates that school-age is sometimes too late for meaningful intervention. As a result, we know little about, for example, what the linguistic profiles of oppositional defiant or hyperactive preschoolers' looks like. Part of the goal of this research, therefore, was to examine whether there are measures of language that can be used to predict, early on, behavioural and emotional problems in children.

One of the challenges in assessing children with psychiatric disorders stems from the fact that many children with behavioural or emotional problems also have comorbid language impairment (Cohen, 2001; Cohen, Barwick, Horodezky, Vallance, & Im, 1998a; Cohen et al., 1998b). Consequently, it becomes difficult to separate what issues are related to language problems and what issues are related to behavioural or emotional problems. This is compounded by the fact that clinicians are often reluctant to diagnose young children with psychiatric disorders and will only do so in extreme cases (Vaillancourt & Boylan, 2010). Additionally, many mental health professionals are not knowledgeable in speech pathology and mistakenly attribute language issues to other psychiatric disorders or early variability in young children's spoken language. But, if we can begin to understand these patterns of comorbidity, we will be better equipped to accurately identify the primary problem and ultimately treat these conditions. One of the goals of this study, therefore, was to use spontaneous language sampling and subsequent analysis of conversational transcripts, to study the relationship between language and maladaptive behavioural and emotional problems in preschool children. To suit this purpose, we utilized a non-clinical sample of preschool children and viewed emotional and behavioural problems on a continuum ranging from typical to clinically significant disorder. By examining problem behaviour and emotions on a continuum we maximize variability in our sample while also furthering our understanding of the relationship between language and a range of symptomatology.

To achieve this, regression analysis was utilized to examine the ability of measures obtained from spontaneous language sampling to predict DSM-oriented and total problems scales of the CBCL C-TRF, beyond that afforded by WPPSI language indices. We chose to use a standardized language measure such as the WPPSI, to replicate present research findings as well as determine how standardized measures of language relate to conversational measures. Results of these findings could help shed light on language production problems in clinical groups, while also having the potential to eventually inform the creation of early screening tools.

Method

Participants

Participants in the study included 46 children (24 girls) ranging from 3 years, 4 months to 5 years, 9 months with a mean age of 4 years 6 months (SD= 7.3 months). Children in the study were native English speakers, predominantly Caucasian (85%), and from a wide range of socioeconomic statuses. Recruitment took place at several high quality licensed childcare centers in Hamilton, Ontario, Canada as part of a larger longitudinal study examining language development, intelligence, executive functioning, and behaviour in preschool children. Informed consent was obtained from all parents and each child's teacher was asked to participate in the study.

Standardized Measures

WPPSI-III

The Wechsler Preschool and Primary Scale of Intelligence- third edition (WPPSI-III; Wechsler, 2002), is one of the most commonly used research measures of preschoolers' language. Traditionally, the WPPSI has been used to predict academic achievement in a variety of domains including reading, math, and general knowledge.

Many studies have documented the predictive validity of the WPPSI in effectively predicting achievement on standardized test scores (Bishop & Butterworth, 1979; Kaplan, 1996; Lieblich & Shinar, 1975). In one of several longitudinal studies examining predictive validity of the WPPSI, 85 children were administered the WPPSI at 5 ½ years of age, and again, 11 years later, using the Wechsler Intelligence Scale for Children (WISC-R). Correlations between the two intelligence tests were extremely high, ranging from 0.73 to 0.86 for verbal, performance, and full scales (Yule, Gold, & Busch, 1982). Similarly, correlations between the WPPSI and a variety of achievement tests have been well documented in the literature (Kaufman, 1973; Krebs 1969; Pasewark, Scherr, & Sawyer, 1974; Reynolds, Wright & Dapper, 1981; Tew & Laurence, 1983). The issue with much of this research is that while it demonstrates the reliability of the WPPSI in predicting scores on the same or similar tests years later, it does not determine whether the constructs themselves are effective in identifying children with specific disorders.

Children were tested using the WPPSI-III and for the purpose of our study, assessing language, scaled scores of the WPPSI receptive and expressive vocabulary were utilized. WPPSI receptive vocabulary is designed to measure word knowledge, learning ability, memory, language development, concept formation and the ability to use language to express ideas while expressive vocabulary measures word knowledge and language development (Sattler, 2001). Taken together, WPPSI expressive and receptive scaled scores form the General Language Composite (GLC), a measure of basic receptive and expressive language development which can be used as a proxy for general verbal IQ.

Administration of the WPPSI-III took place in a quiet section of each classroom by trained research assistants.

Child Behavior Checklist C-TRF

Although there is no singular tool for assessing behaviour and emotional problems in children, the Child Behavior Checklist Caregiver-Teacher Report Form (C-TRF 1½-5; Achenbach & Rescorla, 2000) and Parent Form (CBCL 1½-5; Achenbach & Rescorla, 2000), are the most widely used standardized measures for evaluating maladaptive behavioural and emotional problems in preschool children. The CBCL C-TRF has excellent psychometric properties and corresponds well with clinical diagnoses of mental health disorders (e.g., Achenbach, Dumenci & Rescorla, 2003; Achenbach & Rescorla, 2000). Reliability, validity, and temporal stability of the CBCL C-TRF scales are well documented and researchers continue to use the CBCL C-TRF in clinical populations as a tool for predicting Diagnostic and Statistical Manual of Mental Disorders (DSM, (Fourth Edition, 1994; "DSM-IV") diagnoses (Arend, Lavigne, Rosenbaum, Binns & Christofeel, 1996; Biederman, Faraone, Mick, Moore, & Lelon, 1996; Edelbrock & Costello, 1988; Jensen Salzberg, Richters, & Watanabe, 1993; Weinstein, Noam, Grimes, Stone, & Schwab-Stone, 1990). Additionally, there is substantial evidence supporting the use of teacher ratings in capturing aspects of a child's behaviour that are useful in clinical assessments (Hinshaw, Han, Erdhardt, & Huber, 1992; Stanger & Lewis, 1993). The CBCL C-TRF is a rating form profiling children between the ages of 18 months and 5 years. It obtains teacher ratings on 99 problem items rated as 0 (not true), 1 (somewhat or sometimes true), and 2 (very true or often true), of the child's behaviour and abilities over the last 2 months. Based on these items, the CBCL C-TRF produces scores on two scales: syndrome and DSM-oriented scales.

Syndrome scales are generated from factor analyses clustering items into six distinct scales: (1) Aggressive Behavior (items such as stubborn, uncooperative, destroy own things, easily frustrated, defiant, and not liked); (2) Anxious/Depressed (items such as clings, nervous, self-conscious, fearful, sad, and upset by separation); (3) Attention Problems (items such as can't concentrate, difficulty with directions, inattentive, fidgets, and fails to carry out tasks); (4) Emotionally Reactive (items such as disturbed by change, moody, worries, twitches, whining, and sulks); (5) Somatic Complaints (items such as aches, too concerned with neatness or cleanliness, nausea, and can't stand things out of place); and (6) Withdrawn (items such as apathetic, unresponsive to affection, daydreams, little interest, avoids eye contact, and doesn't answer). Using large representative samples of children, norms were constructed from distributions of syndrome scores to form *T* scores. Borderline clinical ranges are defined as *T* scores ranging from 65 (93rd percentile) to 69 (97th percentile). *T* scores $\geq$ 70 (98th percentile) are considered clinical.

Scores obtained from CBCL C-TRF syndrome scales group together to form more global internalizing, externalizing and total problems scales. Internalizing reflects problems within the self including emotional reactivity, anxiety, depression, somatic complaints without known medical cause, and withdrawal from social contacts (Achenbach & Rescorla, 2000). Externalizing reflect conflicts with others and their expectations of children's behaviour (Achenbach & Rescorla, 2000). In terms of profiles, emotionally reactive, anxious/depressed, somatic complaints and withdrawn syndrome scores form internalizing problems scales, whereas, attention problems, and aggressive behaviour form externalizing problem scales (Achenbach & Rescorla, 2000).

There are five DSM-oriented scales: (1) Affective Problems (items consistent with Dysthymia and Major Depressive Disorder); (2) Anxiety Problems (items consistent with Generalized Anxiety Disorder, Separation Anxiety Disorders, and Specific Phobia); (3) Pervasive Developmental Problems (items consistent with Asperger's Disorder and Autistic Disorder); (4) Attention Deficit/Hyperactivity Problems (items consistent with Hyperactive-Impulsive and Inattentive types of ADHD); and (5) Oppositional Defiant Problems (Achenbach & Rescorla, 2000). Specific scores on DSM-oriented scales of the CBCL C-TRF are not directly equivalent to DSM diagnoses, however, high scores on DSM-oriented scales indicate consideration of diagnosis and further consultation with the DSM to determine whether criteria are met (Achenbach & Rescorla, 2000).

Content validity of the CBCL C-TRF is well established as nearly all items discriminate between children referred for mental health or special education services, and demographically similar children who were not referred (Achenbach & Rescorla, 2000). Criterion validity of all scales is also supported by significant discrimination between referred and non-referred children (Achenbach & Rescorla, 2000). Finally, construct validity is well established through concurrent and predictive association with a variety of other measures, including DSM criteria (Arend, Lavigne, Rosenbaum, Binns, & Christoffel, 1996; Keenan & Wakschlag, 2000), the Richman Behavior Checklist (BCL) (Richman, 1977; Richman, Stevenson, & Graham, 1982; Spiker, Kraemer, Constantine, & Bryant, 1992), the Toddler Behavior Screening Inventory (TBSI) (Mouton-Siemen, McCain, & Kelley, 1997); and later problem scores on the CBCL/4-18 (Achenbach, Howell, Aoki, & Rauh, 1993). Overall, reliability of the CBCL C-TRF is high for most scales, with test-retest correlations ranging from 0.80 to 0.90. Test-retest correlation for total problems average 0.88 and 0.81 across all scales (Achenbach & Rescorla, 2000).

In this study, teacher reports of the CBCL C-TRF were used to measure total problems and DSM-oriented scales pertaining to pervasive developmental problems, oppositional defiant problems, and attention/hyperactivity problems.

Spontaneous language samples

Sampling Procedure

Spontaneous language samples were obtained by means of a three to five minute recorded conversation. Sampling took place in participants' classrooms at the childcare centers. Samples were tape recorded using a Marantz CD recorder CDR300, and a Shure omnidirectional boundary microphone. Spontaneous conversation was generated by the examiner discussing either a Dora the Explorer or Bob the Builder book with the child. Examiners were instructed to use open-ended questions to help facilitate and encourage language. For example, examiners might have asked questions such as "what's happening in this picture?" or "tell me more about this character." Sample length lasted for a minimum of three minutes and was constrained by the number of other measures gathered at the same session as part of the larger longitudinal study.

Transcription

Trained researchers listened to the recordings and transcribed the conversations using Systematic Analysis of Language Transcripts (SALT) computer software SALT

Version 8.0.4, Miller, 2004). Transcription was completed by several researchers independently and then compared, with discrepancies resolved by one of the original transcribers. As per SALT conventions, utterances produced by each participant were analyzed using SALT guidelines with regard to syntactic, phonological, semantic, and pragmatic properties.

Coding

Utterances were determined based on a combination of prosodic boundaries and determinations of thought completion. Contra SALT conventions, utterances were not split at conjunctions ("and", "or", "because," "unless") and mean length of utterance (MLU) was calculated using the number of words, not morphemes. As per SALT conventions, mazes were coded into the following classifications: whole word repetitions, revisions, and silent pauses. Since there is some evidence that short and long revisions have different effects on listeners, measures of average words per revision and average words per repetition were also calculated (Brennan & Schober, 2001). As per Rispoli's 2003 study, repetitions and pauses were combined and classified into the category of stalls. Refer to Table 1 for examples of disfluency categories.

Individual dictionaries were also created for each child based on the set of unique words (i.e. word types) produced during language sampling. These dictionaries were then used to generate measures including the total number of different words produced, measures of the word frequency of the words produced, (Brown, 1973), and average word length. To obtain a measure of the number of different words produced as a proportion of the total words in a given sample, we chose to use the computation "D." Traditionally, type token ratio (TTR) has been used to obtain this value, however, several studies document issues with using TTR as an index of language development or impairment (Haas, Haug, & Landry, 1989; Rice & Bode, 1993; Watkins & Kelly, 1995). D is calculated by a program called the Computerized Language Analysis (CLAN), developed by the Child Language Data Exchange System (CHILDES) (MacWhinney, 2000). CLAN calculates the value of D by using the probability of new vocabulary as it is introduced to larger samples of spoken language. This probability is then used to create a mathematical model of the way TTR varies with token size. Finally, the probability is compared with empirical data to produce the value of D. The value of D, therefore, is not based on the number of words in a given sample; rather, it uses mathematical models in conjunction with empirical data to evaluate vocabulary diversity.

Examination of data distributions for child dictionary measures deemed whether analyses based on median or mean values were most appropriate. In cases where the data were fairly normally distributed mean values were used. For instances where the data was more skewed, as in word frequency counts, the medians were considered a more appropriate measure.

To calculate the reliability of our coding, an independent transcriber also transcribed and coded a randomly selected subset of the conversations (8 of 42 children). Since correspondence between the two coders was high, the original coding was used. Correlations between the various measures were as follows: MLU, $r = 0.97$; number of utterances, $r = 0.90$; number of mazes, $r = 0.93$; number of maze words, $r = 0.99$.

Results

WPPSI Receptive and Expressive Vocabulary scores were used to calculate a GLC score (proxy for full scale verbal IQ). The GLC has a mean of 100, and a standard deviation of 15. The mean GLC score for children in our study was 102 and varied from 74-129. Only one child received a score categorized as "borderline," with the remainder meeting criteria for "low average" (i.e. 80) or above. Descriptive statistics for intelligence (WPPSI-III Receptive, Expressive, and GLC), conversational language measures are shown in Tables 2 and 3. The number of complete words produced varied greatly, ranging from 30 to 225, with a mean of 113.8 (SD = 52.8). To deal with this issue, the present study analyzes the majority of spontaneous indices per every 100 complete words in the sample.

To understand how our conversational measures change with age, correlation analyses were conducted on all spontaneous language sampling items and age in months (see Table 4). The majority of speech categories (e.g. total different nouns, verbs, etc) increased with age, as did vocabulary diversity (D average), word errors, morphological errors, filled pauses, MLU, and revisions. The only item that significantly correlated with age was total different nouns $r(43)=0.32, p < .05$. Spoken and written word frequency, phonological errors, repetitions, stalls, syllables, phonemes, and concreteness ratings were all negatively correlated with age.

Sequential regression analysis was employed to determine if adding conversational language measures improved prediction of DSM-oriented and total problems scales of the CBCL C-TRF beyond that afforded by WPPSI GLC scores. To evaluate assumptions and determine appropriateness of spontaneous language measures, preliminary screening of normality, linearity, and homoscedasticity of residuals was conducted. None of the variables were deemed inappropriate and measures were only removed that appeared redundant or highly correlated (e.g. spoken and written word frequency).

Examination of correlations between CBCL C-TRF DSM-oriented and syndrome subscales indicated substantial overlap between scales. This was not surprising since both DSM and syndrome items utilize the same items in their composites. Since many of the CBCL C-TRF syndrome subscales were highly correlated with one another, the decision was made to use the total problems scale rather than individual syndrome scale measures. Similarly, DSM-oriented subscales were also highly correlated with one another and with syndrome subscales; therefore, we chose to eliminate anxiety and affective problems since items were already captured in syndrome scales (anxious/depressed, withdrawn). Table 5 displays correlations between CBCL C-TRF DSM-oriented and total problems *T* scores.

To determine how much unique variance, beyond that accounted for by WPPSI GLC scores, our conversational measures contributed to CBCL C-TRF subscales, simple, part, and partial correlations were conducted (see Table 6, 7, 8, 9). Partial correlations enabled us to control for the effects of GLC, whereas part correlations allowed us to see whether there were any effects left over after removal of variance from GLC. The majority of conversation measures did not correlate with C-TRF total problems, except for average words per revision $r(32)= .42$ $p < .05$ and revisions per 100 words $r(32)= .43$

$p < .05$. After controlling for effects of GLC, average words per revision and revisions per 100 words accounted for 3% and 5% of the variance respectively. None of our conversational items were significantly correlated with CBCL C-TRF pervasive developmental problems, nor did they account for much of the variance in its prediction (3% of revisions and 7% of average words per revision). For attention deficit/hyperactivity problems, again, average words per revision $r(32)= .38$ $p < .05$, and revisions per 100 words $r(32)= .46$ $p < .01$ correlated with these CBCL C-TRF scores. This time, average words per revision accounted for 3% of the unique variance and revisions per 100 words 2%. Average words per revision and revisions per 100 words also correlated with oppositional defiant problems $r(32)= .56$ $p < .01$ and $r(32)= .63$ $p < .01$ respectively. Average words per revision accounted for 15% of unique variance and revisions per 100 words 20%.

Regression Analyses

Stepwise regression analysis was conducted for ten language measures including MLU, total words (tokens, not types), average word length, stalls, revisions, average words per revision, average words per repetition, morphological error, median spoken word frequency, and D average, to determine which measures, after accounting for WPPSI GLC scores, improved prediction of DSM-oriented and total problems scales. Stepwise regression analysis was used to examine which of our conversational measures added statistically significant variance beyond WPPSI GLC scores. Since we did not have any initial predictions about out variables, and the number of variables was relatively large, stepwise regression was chosen as an initial screening process to determine what variables were most useful in predicting DSM-oriented and total problems scales. On its own, the GLC was observed to correlate with pervasive developmental problems $r(32) = -.49, p < .01$ and attention deficit/hyperactivity problems $r(32) = -.41, p < .05$, but not with total problems or oppositional defiant problems.

Total Problems. Revisions per 100 words, stalls per 100 words, D average, morphological error and average word length, were observed to significantly improve prediction of CBCL C-TRF total problems *T* scores beyond that afforded by WPPSI GLC scores. Children producing more revisions $\beta = 0.24$, $t(23) = 1.76$, *ns* , fewer stalls, $\beta = -.31$, $t(23) = -2.46, p < 0.05$, shorter words $\beta = -53$, $t(23) = -3.93, p < .01$, fewer morphological errors $\beta = -.51$, $t(23) = -3.73, p < 0.01$, and with more diverse vocabularies as measured by D average $\beta = .30$, $t(23) = 2.19, p < 0.05$, scored higher on measures of total problems. Revisions per 100 words also explained a significant proportion of variance in total problems scores, $R^2 = .28$, $F(1, 29) = 5.4, p < .05$, as did stalls per 100 words, $R^2 = 0.58$, $F(1, 29) = 6.84, p < .001$, average word length, $R^2 = 0.38$, $F(1, 29) = 5.43, p < .01$, morphological errors $R^2 = .49$, $F(1, 29) = 6.33, p < .01$, and D average $R^2 = .65$, $F(1, 29) = 7.36, p < .001$. Refer to table 10 for summary of hierarchical regression analysis.

DSM-Oriented Scales. Revisions per 100 words significantly improved prediction of CBCL C-TRF pervasive developmental problems *T* scores beyond that afforded by WPPSI GLC scores, $\beta = 0.32$, $t(23) = 2.09, p < .05$. Findings indicate, therefore, that children scoring more highly on pervasive developmental problems were found to revise their speech more often. Revisions per 100 words also explained a significant proportion of variance in pervasive developmental problems scores, $R^2 = 0.35$, $F(1, 29) = 7.39$, $p < .01$. Refer to table 11 for summary of hierarchical regression analysis.

For attention deficit/hyperactivity problems, revisions per 100 words and average words per repetition significantly improved prediction beyond that afforded by WPPSI GLC score. Children producing more revisions, $\beta = 0.45$, $t(23) = 3.26$, $p < .01$, and having fewer words per repetition, $\beta = -0.29$, $t(23) = -2.06$, $p < .05$ tended to receive higher scores on attention deficit/hyperactivity problems. Revisions per 100 words and average words per repetition also explained a significant proportion of variance in attention deficit/hyperactivity problems scores, $R^2 = 0.42$, $F(1, 29) = 9.98$, $p < .01$ and $R^2 = 0.50$, $F(1, 29) = 8.85$, $p < .01$, respectively. Refer to table 12 for summary of hierarchical regression analysis.

Revisions per 100 words significantly improved prediction of CBCL C-TRF oppositional defiant problems *T* scores beyond that afforded by WPPSI GLC scores, $\beta = 0.68$, $t(23) = 5.06$, $p < .001$, meaning that children rating highly on oppositional defiant problems, also revised their speech more frequently. Revisions per 100 words also explained a significant proportion of variance in oppositional defiant problems scores, $R^2 = 0.50$, $F(1, 29) = 13.87$, $p < .001$. Refer to table 13 for summary of hierarchical regression analysis.

To examine the relationship between conversational measures, and behavioural or emotional problems, correlational analysis was conducted on CBCL C-TRF subscales and conversational measures, while controlling for age. Results revealed that total problems, attention deficit hyperactivity problems, and oppositional defiant problems all correlated with revisions and average word per revision (see Table 14). Pervasive developmental problems did not correlate significantly with any conversation items.

Since CBCL C-TRF subscales were highly correlated, regression analysis using the three DSM-oriented and total problems subscales was conducted to determine if the number of revisions was accounting for the same or unique variance. Results indicated that oppositional defiant problems were mostly highly correlated with the number of revisions, explaining 39% of unique variance in the outcome. The next mostly highly related scale, attention deficit/hyperactivity problems, contributed no unique variance outside of that accounted for by oppositional defiant problems. Total problems explained 3.8% unique variance and pervasive developmental problems 2.6%, beyond that afforded by the previous subscales.

Discussion

The present study used conversational analysis of language to study the relationship between language and maladaptive behavioural and emotional problems in preschool children. We also examined the utility of spontaneous language sampling in predicting mental health problems and compared this prediction to that of WPPSI-III GLC scores. Results pertaining to our initial analysis of how conversational measures changed with age revealed several important findings. It was not surprising that MLU, parts of speech, and vocabulary diversity were observed to improve with age. Greater numbers of filled pauses, however, was less expected. It could be that as children age they become more aware of their listener and begin to utilize filled pauses in a pragmatic sense to inform the audience that they are not finished speaking or preparing their next utterance. The finding of revisions increasing with age parallels results of Rispoli, Hadley and Holt's (2008) study where the authors propose that these increases coincide with children's ability to monitor language production. In terms of negative relationships, it makes sense that speech would become less repetitious and less errorful, and that word frequency would decline as children get older and language becomes more sophisticated. Taken together, these findings add validity to our set of conversational measures since these features of language appear to progress according to developmental theory.

In all analyses, conversational measures of language significantly improved our ability to predict scores on CBCL C-TRF total problems and DSM-oriented scales beyond that accounted for by WPPSI GLC scores. Additionally, for the large majority, GLC scores alone did not significantly predict any of the subscales. The exception to this was two DSM-oriented subscales: pervasive developmental problems $\beta = -0.49$ $t(23) = -3.06$, $p < .01$, and attention deficit/hyperactivity problems $\beta = -0.41$ $t(23) = -0.25$, $p < .05$.

Generally, these findings suggest that spontaneous language sampling, and specific indices derived from this, might serve as critical predictors of maladaptive emotional and behavioural problems in children. Results also add support to the growing body of literature where conversational studies of children with behavioural and emotional problems are differentiated from typically developing peers on measures of language such as maze words, revisions, repetitions, and silent pauses (Belser & Sudhalter, 2001; Dibner, 1958; Kasl & Mahl, 1965; Lake, Humphreys, & Cardy, 2010; Mahl, 1956; Pope, Blass, Siegman, & Raher, 1970; Redmond, 2004).

The number of revisions, specifically, appeared as a predictor in all analyses. Revisions were positively associated with all DSM-oriented and total problems scales, indicating that children rating highly on symptoms of psychopathology also produced more revisions. Revising speech involves self-repair, whereby an individual detects a problem and formulates a revision or replacement to correct it (Levelt, 1983). This process requires input from two sources: overtly produced language and internally represented spoken language. The internal route is responsible for phonological repairs (Postma & Kolk, 1993) and must keep pace with the rate of spoken language. The external route, by contrast, is slower and uses overtly produced language to compare with the communicative intention after the speaker has spoken them. Revisions, therefore, appear to be a product of this slower external route, identified after the speaker has produced language (Rispoli, Hadley, & Holt, 2008). Consequently, the language of children with behavioural or emotional problems appears to face unique language planning and production challenges, resulting in greater numbers of revisions.

In our study, children who scored high on total problems were found to revise their speech more often, produce shorter words, fewer stalls, less morphological errors, and had greater vocabulary diversity as measured by D average. Total problems, as measured by the CBCL C-TRF, is an overall indicator of a child's behavioural and emotional adjustment. Given, the significant association between language impairment and a range of childhood disorders, it is not surprising that we found children with greater total problems to have less fluent spoken language (Cohen, Barwick, Horodezky, Vallance, & Im, 1998; Cohen, et al., 1998; Kim & Kaiser, 2000; Redmond, 2004;

Rockhill, Kodish, DiBattisto, Macia, Varley & Ryan, 2010). Since structural language skills play a critical role in the ability to represent and understand others' perspectives in a social problem solving situation, it makes sense that children with language impairment often have co-morbid mental health problems (Astington & Jenkins, 1999). Findings of fewer morphological errors and stalls and greater vocabulary diversity, however, were more unexpected. Fewer stalls and morphological errors could result from children in the more clinical range producing less complex utterances and thereby fewer errors and stalls. These findings also lend support to the research of Rispoli, Hadley, and Holt (2008), where the number of stalls was found to increase with sentence length.

Our findings indicated that children with higher scores on pervasive developmental problems revised their speech more frequently. Pervasive developmental disorders affect children's social skills, language, attention, and perception, resulting in qualitative abnormalities atypical for any developmental stage (Sattler, 1992). In children with autism spectrum disorders, one of the most profound characteristics is language impairment or pragmatic conversational impairment, often manifesting itself as abnormalities in the production (volume, pitch, and rate), form (repetitive, stereotyped), and or ability to initiate or sustain conversations with others (Villiers, Fine, Ginsberg, Vaccarella, & Szatmari, 2007; Sattler, 1992). Although we did not code for some of these pragmatic aspects of speech, it could be that challenges in using pragmatics in speech also translates to problems with speech planning and formulation. Interestingly, our findings are at odds with those of Lake, Humphreys, and Cardy (2010), where high-functioning individuals with ASD were found to produce fewer revisions than controls. An important distinction between the two studies, however, is the age and diagnosis of individuals. Our study examined a non-clinical sample of preschoolers, whereas Lake, Humphreys, and Cardy (2010), examined a population of young adults with a diagnosis of ASD. The present findings refer to children rating high on the CBCL C-TRF pervasive developmental problems scale and there may be some features of language that are unique to individuals with a clinical diagnosis of ASD, particularly within a high-functioning adult subset of this population.

Children rating high on oppositional defiant problems produced greater numbers of revisions. Oppositional defiant problems refer to problems whereby children frequently lose their temper, deliberately defy rules and requests, blame others for mistakes, and are spiteful and vindictive (Achenbach & Rescorla, 2000). Several studies report links of language impairment in children with conduct problems (e.g., Moffitt 1993; Moffitt & Lynam 1994). Language processing skills are critical components of self-regulation and affective modulation, and children with oppositional defiant disorder have difficulty labeling emotions and using language to create adaptive behavior strategies for responding to such emotions (Bronowski, 1967, 1976; Greene & Doyle, 1999; Vygotsky, 1987). As such, children with oppositional defiant problems may have difficulties expressing themselves and communicating with others and in this frustration are found to revise their speech more often. Additionally, aggressive children are by nature, more impulsive, angry, uncooperative, and easily frustrated than peers (Achenbach & Rescorla, 2000). Impulsivity and frustration could make aggressive children less likely to plan an utterance before speaking, resulting in greater numbers of revisions.

Children who rated higher on attention deficit/hyperactivity problems produced greater numbers of revisions. These findings replicate Redmond's (2004) observations of children with ADHD producing a greater percentage of maze words. Redmond's category of mazes included false starts, fillers, repetitions, and revisions, while we measured revisions separately from stalls (repetitions and pauses). Still, these results do corroborate Redmond's findings in terms of revisions. Children with ADHD are also known to perform worse than typically developing children on tasks involving sentence imitation, word articulation and language quotient (Kim & Kaiser, 2000). Since their language may not be as highly developed as peers without ADHD, children rating higher on attention deficit/hyperactivity problems could be more prone to producing revisions. The CBCL C-TRF defines attention problems as children who have difficulty concentrating, carrying out tasks, and who quickly shift their attention (Achenbach & Rescorla, 2000). From a language perspective, difficulty focusing and changing tasks can create linguistic processing and fluency challenges. Speaking is a complicated task requiring coordination of lexical (lemma) and production (phoneme) stages of processing. A child with attention problems could encounter difficulties planning and coordinating these stages, resulting in longer revisions.

Results of the present study provided validation of Rispoli's (2003) findings of a dichotomy between stalls and revisions. In our study, children rating high on the CBCL C-TRF total problems scale produced more revisions and fewer stalls. Revisions and stalls also added different amounts of prediction to total problems scores. Since these processes appear to operate independently and in the case of total problems, in opposing ways, our findings support the notion that stalls and revisions may represent different features of spoken language. As noted above, our findings of children with greater total problems producing fewer stalls could also be a reflection of children with more problems producing shorter and less complex utterances, making stalls less likely to occur. To determine if the relationship between stalls and total problems was confounded by utterance length, partial correlations controlling for MLU were conducted. As predicted, stalls did not correlate with total problems. Additionally, revisions were found to be the only predictor in regression analysis forcing MLU first. Together these findings suggest that utterance length may be driving the relationship between stalls and total problems.

Findings of the present study did, however, contradict some previous research. Zentall (1988) observed no difference in the number of repetitions and fillers produced by children with and without ADHD. In coding for repetitions in our study, we chose to measure the average number of words per repetition and the category of stalls (repetitions and pauses). Results demonstrated that children rating high on ADHD produced shorter repetitions and children rating high on total problems produced fewer stalls. Since our measures were not identical to those in Zentall's study it is difficult to equate findings, however, part of the reason Zentall observed no differences in groups may be related to experimental tasks. Some of the verbal data recorded during the study were taken during storytelling tasks where children without ADHD were observed to talk more than children with ADHD. Zentall notes that because children with ADHD, in this task, produced less language than controls, they had fewer opportunities to produce disfluencies. The authors argue that the level of language produced in children with ADHD appears to be highly dependent on task, particularly since in a previous study involving a listening task, hyperactive children were observed to initiate more verbalizations than controls (Zentall et al., 1983).

No significant differences were observed in terms of MLU and total number of words between children who did and did not meet clinical cut-offs (*T* scores >70) for total problems, pervasive developmental problems, oppositional defiant problems, and attention deficit/hyperactivity problems. What this means for our study is that differences in the amount of language produced by children meeting clinical cut-offs and those who did not, does not explain why some groups produced more revisions, or fewer stalls. These findings are in line with results of Barkley, Cunningham, and Karlsson's study (1983), where no differences were observed between children with and without ADHD in terms of the number of utterances produced.

MLU, total words, and spoken word frequency did not predict or correlate with any of the CBCL C-TRF subscales in our study. This is particularly interesting in light of the fact that WPPSI GLC scores and MLU are highly correlated $r=.32, p<.05$. Additionally, in the previous chapter, none of the conversational items improved prediction of WPPSI vocabulary and block design raw scores beyond that afforded by MLU (all *F*'s < 1). Since MLU did not correlate with or predict any CBCL C-TRF subscales, it logically follows that WPPSI language scores may also be poor indicators of maladaptive behaviour and emotional problems in children. Similarly, conversational measures such as total words and word frequency do not appear to be related to childhood disorders as measured by CBCL C-TRF DSM-oriented and total problems scales.

Generally, these findings provide us with a basic understanding of how our conversational measures of language change with age and highlight the importance of utilizing comprehensive tools of language assessment when investigating the spoken language of children with suspected behavioural or emotional problems. Specifically, examining the production of revisions, stalls, morphological errors, and average words per repetition and revision, appear to be predictive of some childhood dysfunction. Critically, these measures of language capture important information beyond that afforded by WPPSI language scores, since, in most cases, WPPSI GLC scores alone could not account for a significant amount of the prediction of CBCL C-TRF DSM-oriented and total problems scales.

Determining linguistic predictors of specific disorders, such as increased numbers of revisions in children at risk for pervasive developmental disorders, could assist early identification and treatment resulting in improved outcomes. From an interventionist perspective, understanding the relationship between language and childhood disorders could also facilitate the use of language skills targeted to promote positive interactions with other children. Since variability in rates of development makes assessing young children particularly challenging, our results emphasize the importance of obtaining a thorough evaluation prior to diagnosis. Spontaneous language sampling, particularly in children of preschool age, represents a logical assessment tool as children may not have achieved the vocabulary or insight to complete more commonly used self-report measures. Taken together, these findings lend important evidence of the need to utilize more comprehensive language assessment tools, as an alternative or in addition to standardized tests of language.

References

Achenbach, T.M., Dumenci, L., & Rescorla, L.A. (2003). Are American children's problems still getting worse? A 23-year comparison. *Journal of Abnormal Child Psychology, 31*, 1-11.

Achenbach, T.M., Howell, C.T., Aoki, M.F., & Rauh, V.A. (1993). Nine-year outcome of the Vermont intervention program for low birth weight infants. *Pediatrics, 91,* 45-55.

Achenbach, T. M., & Rescorla, L. A. (2000). Caregiver-Teacher Report Form 1 ½-5. Burlington, VT: ASEBA.

Angold, A., Costello, E.J., & Erkanli, A. (1999). Comorbidity. *Journal of Child Psychology and Psychiatry, 40,* 57-87.

Arend, R., Lavigne, J. V., Rosenbaum, D., Binns, H. J., & Christoffel, K. K. (1996). Relation between taxonomic and quantitative diagnostic systems in preschool children: Emphasis on disruptive disorders. *Journal of Clinical Child Psychology, 25*, 388-387.

Astington, J.W., & Jenkins, J.M. (1999). A longitudinal study of the relation between language and theory-of-mind development. *Developmental Psychology, 35,* 1311-1320.

Aylward, G.P. (1985). Understanding and treatment of childhood depression. *Pediatrics, 107,* 1-9.

Barkley, R. A., Cunningham, C. E., & Karlsson, J. (1983). The speech of hyperactive children and their mothers: comparison with normal children and stimulant drug effects. *Journal of Learning Disabilities, 16,* 105-110.

Beesdo, K., Bittner, A., Pine, D., Stein, M.B., Hofler, M., Lieb, R., et al. (2007). Incidence of social anxiety disorder and the consistent risk for secondary depression in the first three decades of life. *Archives of General Psychiatry, 64,* 903-912.

Beitchman, J.H., Brownlie, E.B., Inglis, A., Wild, B., Ferguson, B., Schacter, D., et al. (1996). Seven-year follow-up of speech/language impaired and control children: Psychiatric outcome. *Journal of Child Psychology and Psychiatry, 37,* 961-970.

Belser, R.C., & Sudhalter, V. (2001). Conversational characteristics of children with fragile X syndrome: Repetitive speech. *American Journal of Mental Retardation, 106,* 28-38.

Biederman, J., Faraone, S., Mick, E., Moore, P., & Lelon, E. (1996). Child Behavior Checklist findings further support comorbidity between ADHD and major depression in a referred sample. *Journal of the American Academy of Child & Adolescent Psychiatry, 35,* 734-742.

Bishop, D.V.M., & Butterworth, G.E. (1979). A longitudinal study using the WPPSI and WISC-R with an English sample. *British Journal of Educational Psychology, 49,* 156-168.

Brennan, S. E., & Schober, M. F. (2001). How listeners compensate for disfluencies in spontaneous speech. *Journal of Memory and Language, 44,* 274-296.

Briggs-Gowan, M.J., Carter, A.S., Irwin, J.R., Wachtel, K., & Cicchetti, D.V. (2004). The Brief Infant-Toddler Social and Emotional Assessment: Screening for social-emotional problems and delays in competence. *Journal of Pediatric Psychology, 29,* 143-155.

Bronowski, J. (1967). Human and animal languages. In *To honor Roman Jakobson* (Vol. 1). The Hague, Netherlands: Mouton.

Bronowski, J. (1976). *The ascent of man.* Cambridge, MA: MIT Press

Brown, R. (1973). *A first language: The early stages.* London: George Allen & Unwin Ltd.

Carson, D.K., Klee, T., Perry, C.K., Donaghy, T., & Muskina, G. (1997). Measures of language proficiency as predictors of behavioral difficulties, social and cognitive development in 2-year-old children. *Perceptual and Motor Skills, 84,* 923-930.

Caulfield, M.B., Fischel, J.E., DeBaryshe, B.D., & Whitehurst, G.J. (1989). Behavioral correlates of developmental expressive language disorder. *Journal of Abnormal Child Psychology, 17,* 187-201.

Clark, R. (1999). The Parent-Child Early Relational Assessment: A factorial validity study. *Educational and Psychological Measurement, 59,* 821–846

Cohen, N.J. (2001). *Language impairment and psychopathology in infants, children, and adolescents.* Thousand Oaks, California: Sage Publications.

Cohen, N.J., Barwick, M.A., Horodezky, N.B., Vallance, D.D., & Im, N. (1998). Language achievement, and cognitive processing in psychiatrically disturbed children with previously identified and unsuspected language impairments. *Journal of Child Psychological Psychiatry, 39,* 865-877.

Cohen, N.J., Menna, R., Vallance, D.D., Barwick, M.A., Im, N., & Horodezky, N.B. (1998). Language, social cognitive processing and behavioural characteristics of psychiatrically disturbed children with previously identified and unsuspected language impairments. *Journal of Child Psychology and Psychiatry, 39,* 853-864.

Crick, N.R., & Grotpeter, J.K. (1995). Relational aggression, gender, and social-psychological adjustment. *Child Development, 66,* 710-722.

De Villiers, J., Fine, J., Ginsberg, G., Vaccarella, L., & Szatmari, P. (2007). Brief report: A scale for rating conversational impairment in autism spectrum disorder. *Journal of Autism and Developmental Disorders, 37,* 1375-1380.

Dibner, A.S. (1958). Ambiguity and anxiety. *Journal of Abnormal Psychology, 56,* 165-174.

Dunn, L.M., & Dunn, L. (1997). *Peabody Picture Vocabulary Test* (3rd ed.). Circle Pines, MN: American Guidance Service.

Edelbrock, C., & Costello, A.J. (1988). Convergence between statistically derived behavior problem syndromes and child psychiatric diagnoses. *Journal of Abnormal Child Psychology, 16,* 219-231.

Greene, R.W., & Doyle, A.E. (1999). Toward a transactional conceptualization of oppositional defiant disorder: Implications for assessment and treatment. *Clinical Child and Family Psychology Review, 2,* 129-148.

Haas, C.W., Haug, H.T., & Landry, R.G. (1989). Sample size and type-token ratios for oral language of preschool children. *Journal of Speech and Hearing Research, 32,* 536-540.

Hinshaw, S. P., Han, S. S., Erhardt, D., & Huber, A. (1992). Internalizing and externalizing behavior problems in preschool children: Correspondence among parent and teacher ratings and behavior observations. *Journal of Clinical Child & Adolescent Psychology, 21,* 143-150.

Horwitz, S.M., Irwin, J.R., Briggs-Gowan, M.J., Heenan, J.M.B., Mendoza, J., & Carter, A.S. (2003). Language delay in a community cohort of young children. *Journal of the American Academy of Child & Adolescent Psychiatry, 42,* 932-940.

Hresko, W.P., Kim Reid, D., & Hammil, D.D. (1999). *The Test of Early Language Development.* (3rd ed.). Austin, TX: PRO-ED.

Irwin, J., Carter, A., & Briggs-Gowan, M. (2002). The social-emotional development of late-talking toddlers. *Journal of the American Academy of Child & Adolescent Psychiatry, 41,* 1324-1332.

Jensen, P.S., Hoagwood, K., & Zitner, L. (2006). What's in a name? Problems versus prospects. In D. Cicchetti and D.J. Cohen (Eds.), *Developmental Psychopathology* (pp. 24-40). Hoboken, NJ: John Wiley & Sons, Inc.

Jensen, P.S., Salzberg, A.D., Richters, J.E., & Watanabe, H.K. (1993). Scales, diagnoses, and child psychopathology: I. CBCL and DISC relationships. *Child and Adolescent Psychiatry, 32,* 397-406.

Kaplan, C. (1996). Predictive validity of the WPPSI-R: A four year follow-up study. *Psychology in the Schools, 33,* 211-220.

Kasl, S.V., & Mahl, G.F. (1965). The relationship of disturbances and hesitations in spontaneous speech to anxiety. *Journal of Personality and Social Psychology, 5,* 425-433.

Kaufman, A.S. (1973). Comparison of the WPPSI, Stanford-Binet, and McCarthy Scales as predictors of first- grade achievement. *Perceptual and Motor Skills, 36,* 67-73.

Keenan, K., & Wakschlag, L.S. (2000). More than the terrible twos: The nature and severity of behavior problems in clinic referred preschool children. *Journal of Abnormal Child Psychology, 28,* 33-46.

Kim, O.H., & Kaiser, A.P. (2000). Language characteristics of children with ADHD. *Communication Disorders Quarterly, 21,* 154-165.

Krebs, E.G. (1969). The Wechsler Preschool and Primary Scale of Intelligence and prediction of reading achievement in first grade. *Dissertation Abstracts International,* 30, 4279-A (University Microfilms, No. 70-3361).

Lake, J.K., Humphreys, K.R., & Cardy, S. (in press). Listener vs. speaker-oriented aspects of speech: Studying the disfluencies of individuals with autism spectrum disorders. *Psychonomic Bulletin and Review.*

Last, C., Perrin, S., Hersen, M., & Kazdin, A. (1996). A prospective study of childhood anxiety disorders. *Journal of the American Academy of Child and Adolescent Psychiatry, 35,* 1502-1510.

Levelt, W.J.M. (1983). Monitoring and self-repair in speech. *Cognition, 14,* 41-104.

Lieblich, A., & Shinar, M. (1975). The predictive validity of the WPPSI with Israeli children. *Educational and Psychological Measurement, 35,* 473-475.

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk. 3rd Edition.* Mahwah, NJ: Lawrence Erlbaum Associates.

Mahl, G.F. (1956). Disturbances and silences in patient's speech in psychotherapy. *Journal of Abnormal and Social Psychology, 53,* 1-13.

Mash, E.J., & Hunsley, J. (Eds.). (2005a). Developing guidelines for the evidence-based assessment of child and adolescent disorders. *Journal of Clinical Child and Adolescent Psychology, 34,* 362-558.

Miller, J. F. (2004). *Systematic Analysis of Language Transcripts* v. 8.0.4. Madison, WI: University of Wisconsin-Madison.

Moffitt, T. E. (1993). The neuropsychology of conduct disorder. *Development and Psychopathology, 5,* 135-151.

Moffitt, T. E., & Lynam, D. (1994). The neuropsychology of conduct disorder and delinquency: Implications for understanding antisocial behavior. In D. C. Fowles, P. Sutker, & S. H Goodman (Eds.), *Experimental personality and psychopathology research 1994* (pp. 233-262). New York: Springer.

Mouton-Siemen, P., McCain, A.P., & Kelley, M.L. (1997). The development of the Toddler Behavior Screening Inventory. *Journal of Abnormal Child Psychology, 25,* 59-64.

Newcomer, P. L., & Hammill, D. D. (1997). *Test of Language Development-Primary, Third Edition*. Austin, TX: Pro-Ed.

Oxman, T.E., Rosenberg, S.D., & Tucker, G.J. (1982). The language of paranoia. *American Journal of Psychiatry, 139,* 275-282.

Pasewark, R., Scherr, S., & Sawyer, R. (1974). Correlations of scores on the Vane Kindergarten, Wechsler Preschool and Primary Scale of Intelligence, and Metropolitan Reading Readiness tests. *Perceptual Motor Skills, 38*, 518.

Phelps-Terasaki, D., & Phelps-Gunn, T. (1992). *The Test of Pragmatic Language (TOPL).* Austin, TX: PRO-ED.

Pope, B., Blass, T., Siegman, A.W., & Raher, J. (1970). Anxiety and depression in speech. *Journal of Consulting and Clinical Psychology, 35,* 128-133.

Postma, A., & Kolk, H. (1993). The covert repair hypothesis: Prearticulatory repair processes in normal and stuttered disfluencies. *Journal of Speech, Language, and Hearing Research, 36,* 472-487.

Qi, C.H.., & Kaiser, A.P. (2004). Problem behaviors in low-income children with language delays: An observation study. *Journal of Speech, Language, and Hearing Research, 47,* 595-609.

Redmond, S. M. (2004). Conversational profiles of children with ADHD, SLI and typical development. *Clinical Linguistics and Phonetics, 18*, 107-125.

Redmond, S.M., & Rice, M.L. (1998). The socioemotional behaviors of children with SLI: Social adaptation or social deviance? *Journal of Speech, Language, and Hearing Research, 41,* 688-700.

Rescorla, L.A. (1989). The Language Development Survey: A screening tool for delayed language in toddlers. *Journal of Speech, Language, and Hearing Research, 54,* 587-599.

Rescorla, L.A., & Achenbach, T.M. (2002). Use of the Language Development Survey in a national probability sample of children aged 18-35 months. *Journal of Speech, Language, and Hearing Research, 45,* 733-743.

Rescorla, L.A., & Alley, A. (2001). Validation of the Language Development Survey (LDS): A parent report tool for identifying language delay in toddlers. *Journal of Speech, Language, and Hearing Research, 44,* 434-445.

Rescorla, L.A., Ross, G.S., & McClure, S. (2007). Language delay and behavioral/emotional problems in toddlers: Findings from two developmental clinics. *Journal of Speech, Language, and Hearing Research, 50,* 1063-1078.

Reynolds, C.R., Wright, D., & Dappen, L., (1981). A comparison of the criterion-related validity (academic achievement) of the WPPSI and the WISC-R. *Psychology in the Schools, 18,* 20–23.

Rice, M.L., & Bode, J.V. (1993). GAPS in the verb lexicons of children with specific language impairment. *First Language, 13,* 113-131.

Richman, N. (1977). Behaviour problems in preschool children: Family and social factors. *British Journal of Psychiatry, 131,* 523-527.

Richman, N., Stevenson, J., & Graham, P.J. (1982). *Pre-school to school: A behavioural study.* New York: Academic Press.

Rispoli, M. (2003). Changes in the nature of sentence production during the period of grammatical development. *Journal of Speech, Language, and Hearing Research, 46,* 818-830.

Rispoli, M., Hadley, P., & Holt, J. (2008). Stalls and revisions: A developmental perspective on sentence production. *Journal of Speech, Language, and Hearing Research, 51,* 953-966.

Rockhill, C., Kodish, I., DiBattisto, C., Macia, M., Varley, C., & Ryan, S. (2010). Anxiety disorders in children and adolescents. *Current Problems in Pediatric and Adolescent Health Care, 40,* 66-99.

Rude, S.S., Gortner, E.M., & Pennebaker, J.W. (2002). Language use of depressed and depression-vulnerable college students. *Psychological Bulletin, 39,* 811-845.

Sattler, J. (1992). *Assessment of children.* (3rd ed.). San Diego, California: Jerome M. Sattler, Publisher, Inc.

Sattler, J. M. (2001). *Assessment of children: Cognitive applications*. San Diego, CA: Jerome M. Sattler.

Sparrow, S.S., Cicchetti, D.V., & Balla, D.A. (2005). *Vineland Adaptive Behavior Scales* (2nd ed.). Minneapolis, MN: Pearson Assessment.

Spiker, D., Kraemer, H. C., Constantine, N. A., & Bryant, D. (1992). Reliability and validity of behavior problem checklists as measures of stable traits in low birth weight preschool children. *Child Development, 63*, 1481-1496.

Stanger, C., & Lewis, M. (1993). Agreement among parents, teachers, and children on internalizing and externalizing behavior problems. *Journal of Clinical Child and Adolescent Psychology, 22,* 107-116.

Stevenson, J., Richman, N., & Graham, P. (1985). Behaviour problems and language abilities at three years and behavioral deviance at eight years. *Journal of Child Psychology and Psychiatry, 26,* 215-230.

Tervo, R.C. (2007). Language proficiency, development, and behavioral difficulties in toddlers. *Clinical Pediatrics, 46,* 530-539.

Tew, B.J., & Laurence, K.M. (1983). The relationship between spina bifida's children intelligence scores on school entry and at school leaving: A preliminary report. *Child: Care, Health, and Development, 9,* 13-17.

Tucker, G.J., & Rosenberg, S.D. (1975). Computer content analysis of schizophrenic speech: A preliminary report. *American Journal of Psychiatry, 132,* 611-616.

Vygotsky, L. S. (1987). Thinking and speech. In L. S. Vygotsky & R. W. Rieber (Series Eds.), & A. S. Carton (Vol. Ed.), *The collected works of L. S. Vygotsky, Vol. 1: Problems in general psychology.* New York: Plenum.

Watkins, R.V., & Kelly, D.J. (1995). Measuring children's lexical diversity: Differentiating typical and impaired language learners. *Journal of Speech & Hearing Research, 38,* 1349-1356.

Weinstein, S.R., Noam, G.G., Grimes, K., Stone, K., & Schwab-Stone, M. (1990). Convergence of DSM-III diagnoses and self-reported symptoms in child and adolescent inpatients. *Journal of the American Academy of Child and Adolescent Psychiatry, 29,* 627-634

Wetherby, A., & Prutting, C. (1984). Profiles of communicative and cognitive social abilities in autistic children. *Journal of Speech and Hearing Research, 27,* 364-377.

Yule, W., Gold, D.R., & Busch, C. (1982). Long-term predictive validity of the WPPSI: An 11-year follow-up study. *Personality and Individual Differences, 3,* 65-71.

Zentall, S.S. (1988). Production deficiencies in elicited language but not in the spontaneous verbalizations of hyperactive children. *Journal of Abnormal Child Psychology, 16,* 657-673.

Zentall, S. S., Gohs, D. E., & Culatta, B. (1983). Language and activity of hyperactive and comparison children during listening tasks. *Exceptional Children, 50,* 255-266.

Table 1

Disfluency Category Examples

Disfluency category	Example
Stalls	
Pause	"I went (0:05)* to the store."
Repetition	"This is (the) the bed."
	"I saw a deer (a deer) in the forest."
Revisions	
	"I went to go (upstairs) downstairs."
	"Sarah was walking and then she (walk home to go) had to run to the store."

*Numbers mark the length of the silent pause. Following SALT coding, all disfluencies are contained within brackets.
** Curly brackets contain the actual production of the word preceding the brackets.

Table 2

Descriptive Statistics for scaled scores of WPPSI GLC, Receptive and Expressive Vocabulary

	Minimum	Maximum	Mean	Std. Deviation	N
GLC – WPPSI[a]	74	129	102.47	12.81	45
Receptive Vocabulary subtest – WPPSI[a]	5	16	10.69	2.55	45
Expressive Vocabulary subtest – WPPSI[a]	6	17	10.43	2.42	44

Table 3

Descriptive Statistics for Conversational Language Measures

Conversational measures (N = 46)	Min	Max	Mean	S.D.
Length of speech				
Mean Length of Utterance (MLU) in words	1	11	4.13	1.86
Total words	30	225	113.87	52.79
Average word length	1.27	5.42	4.31	0.57
Maze measures				
Stalls per 100 words	0	8.89	4.22	2.30
Revision per 100 words	0	2.30	0.33	0.59
Average words per revision	0	7	0.82	1.57
Average words per repetition	0	4	1.2	1.12
Error measure				
Morphological error total	0	5	0.91	1.28
Lexical diversity measures				
Median log spoken frequency	1.34	2.60	2.02	0.34
D Average	17.77	86.58	46.78	18.39

Table 4

Correlations between Conversation Language Measures and Age in months

Conversation Measure	Age in Months
D Average	.14
Total words	.16
Total different words	.23
Total nouns	.32*
Total verbs	.24
Total adjectives	.05
Total adverbs	-.07
Total proper nouns	.13
Total conjunctions	.17
Total prepositions	.18
Total pronouns	.22
Median written word frequency	-.06
Median spoken word frequency	-.06
Total phonological error	-.14
Total morphological error	.20
Total word error	.08
MLU in words	.23
Pauses per 100 words	.08
Filled pauses per 100 words	.12
Average words per revision	.24
Average words per repetition	-.17
Repetitions per 100 words	-.20
Revisions per 100 words	.19
Stalls per 100 words	-.05
Average concreteness	-.11
Average phonemes	-.056
Average syllables	-.01

*$p < .05$.

Table 5

Correlations between C-TRF DSM-oriented and Syndrome scale T scores

	Anxious / Depressed	Somatic Complaints	Withdrawn	Aggressive Behaviour	Internalizing Problems	Externalizing Problems	Total Problems	Affective Problems	Anxiety Problems	Pervasive Developmental Problems	Attention Deficit/ Hyperactivity Problems	Oppositional Defiant Problems
Emotionally Reactive	.77**	.68**	.67**	.62**	.56**	.88**	.62**	.78**	.72**	.82**	.87**	.67**
Anxious / Depressed		.55**	.67**	.63**	.47**	.86**	.54**	.72**	.72**	.92**	.79**	.61**
Somatic Complaints			.50**	.49**	.33	.68**	.40**	.55**	.53**	.60**	.59**	.51**
Withdrawn				.79**	.58**	.83**	.67**	.80**	.79**	.54**	.85**	.77**
Aggressive Behaviour					.59**	.74**	.78**	.83**	.78**	.57**	.68**	.92**
Internalizing Problems						.62**	.91**	.83**	.52**	.33	.57**	.75**
Externalizing Problems							.72**	.91**	.81**	.78**	.90**	.75**
Total Problems								.93**	.68**	.44**	.64**	.85**
Affective Problems									.81**	.63**	.81**	.87**
Anxiety Problems										.59**	.77**	.75**
Pervasive Developmental Problems											.79**	.54**
Attention Deficit / Hyperactivity Problems												.70**

*$p < .05$, **$p < .01$

Table 6

Part and Partial Correlations between Total Problems and Conversation Language Measures

	Simple Correlation	Total Problems Partial Correlation	Part Correlation
D Average	-.05	.01	.01
Total words	.12	.27	.26
Average word length	-.34	-.36	-.35
Median spoken word frequency	.00	.032	.03
Morphological error	-.33	-.34	-.33
MLU in words	.02	.15	.15
Average words per revision	.42*	.47	.46
Average words per repetition	-.11	-.16	-.16
Revisions per 100 words	.43*	.49	.48
Stalls per 100 words	-.26	-.28	-.27

*$p < .05$. **$p < .01$.

Table 7

Part and Partial Correlations between Pervasive Developmental Problems and Conversation Language Measures

	Pervasive Developmental Problems		
	Simple Correlation	Partial Correlation	Part Correlation
D Average	-.11	-.01	-.01
Total words	-.12	.16	.14
Average word length	-.05	-.10	-.08
Median spoken word frequency	-.28	-.31	-.27
Morphological error	-.22	-.30	-.26
MLU in words	-.19	.08	.07
Average words per revision	.16	.24	.21
Average words per repetition	-.06	-.14	-.12
Revisions per 100 words	.28	.38	.33
Stalls per 100 words	-.09	-.08	-.07

*p <. 05. **p < .01.

Table 8

Part and Partial Correlations between Attention Deficit/Hyperactivity Problems and Conversation Language Measures

	Attention Deficit/Hyperactivity Problems		
	Simple Correlation	Partial Correlation	Part Correlation
D Average	-.01	.08	.08
Total words	-.13	.09	.08
Average word length	-.13	-.17	-.15
Median spoken word frequency	-.23	-.25	-.23
Morphological error	-.26	-.33	-.30
MLU in words	-.29	-.11	-.10
Average words per revision	.38*	.47	.43
Average words per repetition	-.29	-.38	-.35
Revisions per 100 words	.46**	.55	.50
Stalls per 100 words	-.26	-.27	-.25

$^*p < .05$. $^{**}p < .01$.

Table 9

Part and Partial Correlations between Oppositional Defiant Problems and Conversation Language Measures

	Oppositional Defiant Problems		
	Simple Correlation	Partial Correlation	Part Correlation
D Average	-.11	-.06	-.06
Total words	.17	.34	.33
Average word length	-.29	-.31	-.30
Median spoken word frequency	-.09	-.07	-.07
Morphological error	-.28	-.29	-.28
MLU in words	.10	.24	.23
Average words per revision	.56**	.62	.61
Average words per repetition	-.20	-.24	-.24
Revisions per 100 words	.63**	.69	.67
Stalls per 100 words	-.30	-.32	-.31

*$p < .05$. **$p < .01$.

Table 10

Summary of Hierarchical Regression Analysis for Variables Predicting Total Problems T Scores on the C-TRF (N =34)

Total Problems	*B*	*SE B*	Β
Step 1			
(Constant)	74.95	18.01	
GLC	-.19	.17	-.20
Step 2			
(Constant)	77.31	15.90	
GLC	-.24	.15	-.25
Revisions per 100 words	12.69	4.17	.49**
Step 3			
(Constant)	139.86	33.90	
GLC	-.25	.14	-.26
Revisions per 100 words	11.40	3.99	.44**
Avg word length	-14.12	6.86	-.32*
Step 4			
(Constant)	158.99	32.10	
GLC	-.25	.13	-.26
Revisions per 100 words	8.54	3.85	.33*
Avg word length	-17.61	6.46	-.40*
Morphological error	-3.34	1.36	-.36*

Step 5			
(Constant)	171.50	30.41	
GLC	-.24	.12	-.26
Revisions per 100 words	6.43	3.71	.25
Avg word length	-19.07	6.05	-.43**
Morphological error	-3.90	1.29	-.43**
Stalls per 100 words	-1.56	.70	-.30*
Step 6			
(Constant)	190.31	29.61	
GLC	-.31	.12	-.33*
Revisions per 100 words	6.09	3.46	.24
Avg word length	-23.54	5.99	-.53**
Morphological error	-4.69	1.26	-.51**
Stalls per 100 words	-1.60	.65	-.31*
D Avg	.19	.09	.30*

Note. $R^2 = 0.04$ for Step 1; $\Delta R^2 = 0.24$ for Step 2; $\Delta R^2 = 0.1$ for Step 3; $\Delta R^2 = 0.1$ for Step 4; $\Delta R^2 = 0.09$ for Step 5; $\Delta R^2 = 0.07$ for Step 6. (*ps* < .05); **p* < .05, ***p* < .01, ****p* < .001

Table 11

Summary of Hierarchical Regression Analysis for Variables Pervasive Development Problems T Scores on the C-TRF (N =34)

Pervasive Development Problems	*B*	*SE B*	Β
Step 1			
(Constant)	92.53	11.85	
GLC	-.34	.11	-.49**
Step 2			
(Constant)	93.67	11.23	
GLC	-.37	.11	-.53**
Revisions per 100 words	6.13	2.94	.32*

Note. $R^2 = .24$ for Step 1; $\Delta R^2 = 0.11$ for Step 2 (*ps* < .05). *$p < .05$, **$p < .01$, ***$p < .001$

Table 12

Summary of Hierarchical Regression Analysis for Variables Predicting Attention Deficit/Hyperactivity Problems T Scores on the C-TRF (N =34)

Attention Deficit/Hyperactivity Problems	*B*	*SE B*	B
Step 1			
(Constant)	86.24	11.86	
GLC	-.28	.11	-.41*
Step 2			
(Constant)	87.94	10.14	
GLC	-.31	.10	-.47**
Revisions per 100 words	9.11	2.66	.50**
Step 3			
(Constant)	93.24	9.94	
GLC	-.33	.09	-.50**
Revisions per 100 words	8.30	2.55	.45**
Avg words per repetition	-2.13	1.03	-.29*

Note. $R^2 = .17$ for Step 1; $\Delta R^2 = 0.25$ for Step 2 (*ps* < .05); $\Delta R^2 = 0.08$ for Step 3 (*ps* < .05) **p* < .05, ***p* < .01, ****p* < .001

Table 13

Summary of Hierarchical Regression Analysis for Variables Predicting Oppositional Defiant Problems T Scores on the C-TRF (N =34)

Oppositional Defiant Problems	*B*	*SE B*	Β
Step 1			
(Constant)	75.14	15.72	
GLC	-.16	.15	-.20
Step 2			
(Constant)	78	11.58	
GLC	-.23	.11	-.28
Revisions per 100 words	15.35	3.04	.69***

Note. R^2 = .04 for Step 1; Δ R^2 = 0.46 for Step 2 (*ps* < .001). **p* < .05, ***p* < .01, ****p* < .001

Table 14

Correlation Matrix of Conversation Language Measures and C-TRF DSM-oriented and Total Problems scales (controlling for age)

	Revisions per 100 words	Average words per revision
Total Problems	.47**	.51**
Attention Deficit/ Hyperactivity Problems	.47**	.44**
Oppositional Defiant Problems	.65***	.61***

*$p < .05$. **$p < .01$. ***$p < .001$.

Chapter 4

Listener vs. speaker-oriented aspects of speech: Studying the disfluencies of individuals with autism spectrum disorders

Johanna K. Lake, Karin R. Humphreys & Shannon Cardy

Department of Psychology, Neuroscience & Behaviour, McMaster University

Word Count: 3697

Running Head: Listener vs. speaker-oriented speech

Correspondence should be addressed to:

Karin Humphreys
Department of Psychology, Neuroscience & Behaviour
McMaster University
1280 Main St West
Hamilton, Ontario
L8S 4K1
Canada
phone: 905-525-9140 x23011
fax: 905-529-6225
krh@mcmaster.ca

Abstract

This study investigates the role of disfluencies such as "um" or "uh," in conversation to discern whether these features of speech serve listener or speaker- oriented functions by looking at their occurrence, (or lack of occurrence) in the spoken language of participants with autism. Since the characteristic egocentricity of individuals with autism means they should engage in minimal listener-oriented behaviour, they are a useful group to differentiate these functions. Transcription, analysis and categorization of 26 spontaneous language samples were derived from age-matched native English speaking controls and high-functioning individuals with Autism Spectrum Disorders (ASDs). Results showed that individuals with ASD produced fewer filled-pause words (ums and uhs) and revisions than controls, but more silent pauses. Filled-pause words therefore appear to be listener-oriented features of speech.

KEYWORDS: autism, pragmatics, disfluencies, language production, conversation

Listener vs. speaker-oriented speech: Studying the disfluencies of individuals with autism spectrum disorders

The language mechanisms underlying conversation are highly complex and intricately coordinated, yet they are largely effortless processes. One question that arises is to what extent this ease of communication is due to the work being done by the speaker or by the listener, the latter often in spite of a speaker's failure to be helpful. The goal here is to identify what aspects of a speaker's behaviour can be identified as cooperative, or otherwise. In the past, some approaches to this problem have involved referential communication tasks to better understand the roles of speakers and listeners in communicative exchanges (see Clark & Fox Tree, 2002; Fox Tree 2001; Fox Tree & Schrock 1999; Girbau, 2001; Horton & Keysar, 1996; Mangold & Pobel, 1988). This research has had success in determining which aspects of spoken language are helpful for a listener, but it remains unclear to what extent this is being done by the speaker for the benefit of the listener (i.e. is listener-oriented), or whether it is merely a regularity in the speaker's behavior that a listener may be able to exploit, and is not performed by the speaker with the listener's needs in mind (speaker-oriented) (see Bock, 1996; Brennan & Clark, 1996; Fox Tree & Schrock, 1999). For example, recent research has suggested that disfluencies such as filled-pause words serve a useful discourse function, indicating that the speaker is not finished speaking, and is trying to compose their next thought or find the correct word, and that listeners are able to utilize this information (Fox Tree, 2001). However, it is still unclear whether this is being done for the benefit of the listener.

In this work, we make the assumption that individuals with autism spectrum disorders will not tend to display elements of spoken language that are listener-oriented, thereby providing an alternate form of evidence as to whether certain discourse behaviours within typical populations are listener- or speaker-oriented. Individuals with autism display a variety of impairments in many areas including social skills, and language development. One of the most profound characteristics of individuals with autism is egocentricity, which manifests as a lack of interest in interacting with other people, a failure to develop social relationships and difficulty with social interactions. It has been suggested that underlying these problems may be a problem of "theory of mind" which describes how individuals with autism are unable to form representation of another's mental state (American Psychiatric Association, 2000; Baron-Cohen, 1995). A critical pragmatic aspect of conversational language use is the ability and willingness to recognize the listener's perspective and knowledge, a task often referred to as establishing 'common ground' (Clark, 1996). Common ground involves understanding the speaker's intention and beliefs such that a shared understanding of mental state is developed (Clark, 1996). Typical speakers are generally very good at using this knowledge and carry on conversations with little effort (e.g. Brennan & Clark, 1996). In contrast, individuals with autism tend to take egocentric approaches to conversation, and usually have poor pragmatic skills; this is true even of very high functioning individuals with autism, who are not judged to be language impaired by usual measures (Baltaxe, 1977; Bishop, 1998; de Villiers, Fine, Ginsberg, Vaccarella, & Szatmari, 2007; Fine, Bartolucci, Ginsberg, &

Szatmari, 1991; Wetherby & Prutting, 1984; Young, Diehl, Morris, Hyman & Bennetto; Ziatas, Durkin, & Pratt, 2003).

This group, therefore, presents an opportunity to explore which functions of language are produced by a speaker for the benefit of the listener, and which are independent of the perceived needs of the listener. Specifically, we predict that if high-functioning individuals with autism are seen to produce specific pragmatic aspects of spoken language at a normal rate, that feature is likely *not* being done for the benefit of the listener. Conversely, for those pragmatic aspects of spoken language that are relatively absent in an individual with autism, this constitutes some evidence that this feature may be listener-oriented in normal speech.

In this study we specifically examine the role of disfluencies in spoken language. It has been suggested that the use of filled-pause words or disfluencies in normal spoken language, such as um and uh, may represent an important role in conversation. Fox Tree (2001) examined the effect of ums and uhs during on-line processing of speech, and showed that um and uh may be utilized by a listener to facilitate conversations. Uh appeared to signal an upcoming short delay, while um a longer delay. The use of uh increased the speed at which listeners were able to recognize words, however, um had no effect on listeners' speech recognition (Fox Tree, 2001, 2002). Fox Tree suggests that ums and uhs help listeners by alerting them that the speaker is still speaking (that it is not the listener's turn yet) and indicating the length of the upcoming delay in speech. However, we do not know if speakers are intentionally using this function of spoken language to aid the listener or if this is merely a regularity in the speaker's behaviour that listeners are able to take advantage of. By studying the spoken language of individuals with autism, who by definition are unlikely to engage in listener-oriented functions of speech, this can provide evidence as to whether these types of disfluencies appear to be a speaker- or listener-oriented function.

Method

Participants

Participants with autism were recruited from a facility in Hamilton, Ontario, Canada, providing services to high-functioning individuals with autism spectrum disorders. Fourteen native English speaking individuals with ASD (thirteen male) took part in the experiment, all of whom had been diagnosed by an outside agency. One male participant was subsequently excluded for a verbal IQ score that fell below the normal range. According to the Autism Diagnostic Observation Schedule (ADOS: WPS Version Lord, Rutter, DiLavore, & Risi, 1999), as administered by the facility, six of the remaining participants had a prior diagnosis of autism spectrum disorder, four Asperger's syndrome, and three autism. The ADOS was repeated by a psychiatrist at McMaster University for four of the participants (others were not retested due to time constraints), and in all cases, the original diagnosis was confirmed. Six of the remaining participants with autism had a diagnosis of autism spectrum disorder, four Asperger's syndrome, and three autism. The mean age of participants with autism was 27 years, with a range of 19-35. Wechsler Adult Intelligence (WAIS Wechsler, 1939) scores were obtained for participants with ASD with an average verbal IQ of 99, and a range of 83 to 117, all within normal ranges. Thirteen age- and gender-matched control participants also took part in the experiment. Control participants were native English speaking students of McMaster University and members of the community who volunteered to participate. IQ information was not obtained for controls, nor were the participants matched in terms of education level.

Materials

A spontaneous language sample was obtained from a 5-10 minute recorded conversation. Participants were asked a variety of general questions related to their interests and hobbies. Following each question, participants were given roughly five seconds (as estimated by the trained experimenter) to respond before the experimenter used further prompting to achieve a reply. The same set of questions was used for both groups and all conversations were digitally recorded. Two experimenters listened to these recordings and transcribed the conversations using SALT software (Systematic Analysis of Language Transcripts, Miller & Chapman, 1983). Transcriptions were completed independently and then compared, with discrepancies resolved by one of the original transcribers. As per SALT conventions, the first 49 utterances produced by each participant were analyzed using SALT guidelines with regard to syntactic, phonological, semantic, and pragmatic properties. Transcripts were further categorized according to the rate per hundred words of revisions, filled-pauses (ums and uhs) and silent pauses (greater than two seconds).

Results

Conversation Sample

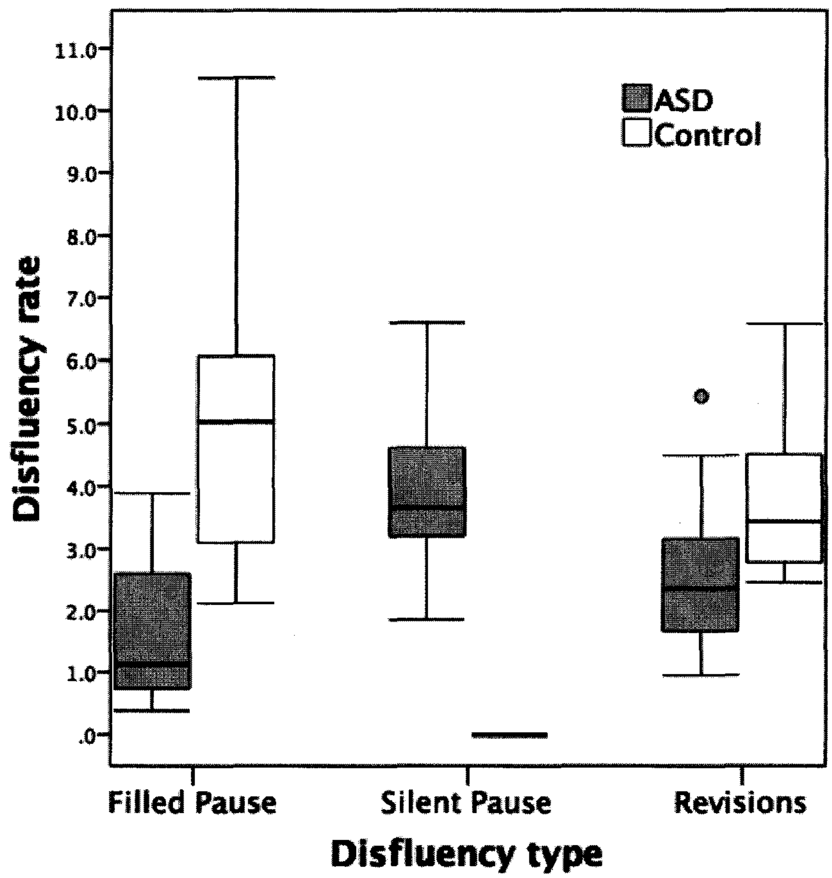
The mean length of utterance (MLU) for control participants of 9.1 words, and for ASD participants, 5.7 words; the latter tended to answer questions with shorter responses, particularly one-word replies. However, even when one-word utterances were excluded from the analysis, the MLU for control participants was still larger, ranging from 7.7 to 11.5 compared to 4.6 to 8.8 for ASD participants. Informal analysis of the conversation samples of ASD participants revealed no obvious deficits, and semantic and syntactic aspects of spoken language were comparable between the groups.

Throughout the conversation samples, ASD participants responded to 84.5% of questions, compared to 99% for control participants. The experimenter frequently had to pose and re-phrase the questions several times before obtaining a response from participants with autism. Disfluencies were coded into three categories: revisions, silent pauses, and filled pauses. Table 1 gives examples of each of these types of disfluencies. Figure 1 shows boxplots of the disfluency rates per hundred words by group, demonstrating a striking lack of overlap between the group's distributions, particularly for filled and silent pauses. A series of independent samples t-tests revealed significant differences between the control group and individuals with ASD with respect to disfluencies per hundred words. Participants with ASD were found to produce fewer filled-pause words (ums and uhs) than control participants, $t(24) = 4.3$, $p < .001$ (means of 1.7 vs. 5.0). Conversely, participants with ASD produced more silent-pauses than control participants $t(24) = 10.8$, $p < .001$ (with means of 4.0 and zero respectively). Of the silent pauses produced by individuals with ASD, 68% occurred at the beginning of an utterance, and 32% within utterances. Participants with ASD produced significantly fewer revisions than controls $t(24) = 2.3$, $p = .03$, (2.7 vs. 3.8).

Table 1. Examples of Disfluencies

Disfluency category	Example utterance
Revision	"My fat…favourite…best animal is a dog…my favourite animal is a cat"
Filled pause	"My favourite...um...animal...uh…is an um…cat"
Silent pause >2s	"My favourite…(>2 s silence) animal is a …(>2 s silence) cat"

Figure 1. Boxplots comparing number of disfluencies per hundred words for ASD vs. control groups.



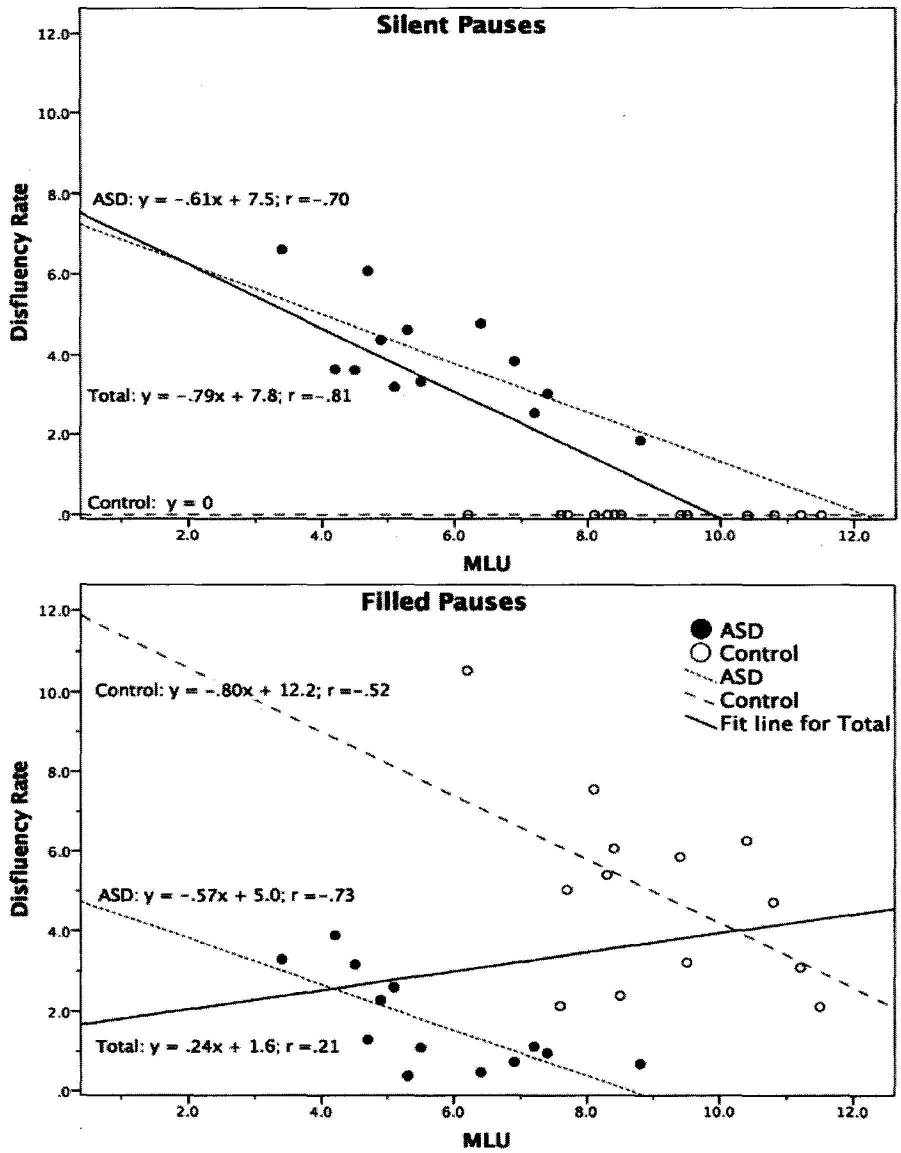
Note: Figure 1 demonstrates the general lack of overlap between the disfluency measures for ASD and control groups. For filled pauses, the median score for the ASD group was less than the minimum for the control group; this was true even for the median score of only the six ASD participants with VIQs above 100 (1.7). For silent pauses, there was no overlap between the distributions; median for ASD participants with VIQs above 100 was

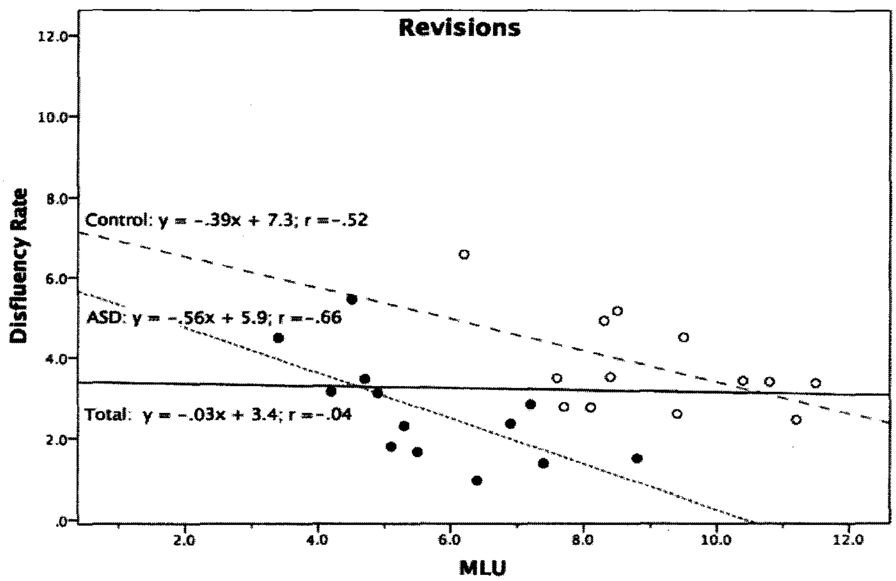
3.7. Revisions showed somewhat more overlap, although the median number of revisions for the ASD participants was still below the minimum score for the control participants (2.4 vs. 2.5), and the median for the speakers with VIQs over 100 was slightly above the first quartile for the control speakers (2.9 vs. 2.8).

Although the groups were matched for age and gender, they were not matched for education, or IQ, which was unavailable for the control group. While the ASD group showed a normal range of VIQ scores (83-117, mean = 99), it is probable that the control group had a higher mean VIQ, although we would expect at least some overlap in the distributions. To examine whether IQ underlies these effects on disfluencies we can look at the relationship between VIQ and performance within the ASD group. Results reveal that VIQ is positively correlated with filled pauses and revisions (in both cases, $r = .41$), but not with silent pauses ($r = -.03$). None of these correlations were significant, but given the small sample size, this is unsurprising. However, even though there was some relationship with VIQ in the ASD group, the distribution of disfluencies between the two groups is markedly different, as can be seen in Figure 1.

Another important way in which the groups differed was MLU, which could potentially account for the effects seen here (i.e. a greater MLU may lead to more opportunities for disfluencies). To investigate this possibility we plot MLU against the various disfluency measures, organized by group in Figure 2. In all three disfluency measures, while we see a clear relationship with MLU, it should be noted that the two groups show separate regression lines, with different intercepts, even when MLUs overlap.

Figure 2. Scatterplots of disfluency rates by MLU.





Note: For each disfluency type, regression lines are plotted for ASD and control groups individually, as well as for the ungrouped total. In all cases, while there is a strong linear relationship with MLU and disfluency rate, the two groups show notably different regression lines, with different intercepts.

Discussion

There are several aspects of conversational speech which could be listener or speaker-oriented. The use of filled pauses during disfluencies, including ums and uhs appear to help listeners (Fox Tree 2001), but it is unclear whether these are produced for their benefit.

Their characteristic egocentricity, perhaps due to challenges in understanding the perspective of another – theory of mind deficits –makes individuals with autism by definition unlikely to engage in listener-oriented behaviour. Therefore, if individuals with autism employ certain aspects of spoken language in the same way as typical speakers, we argue that this feature must not be listener-oriented and if they do not employ it, this is some evidence that it may be listener-oriented.

Results of the present experiment demonstrated that participants with ASD produced far fewer filled-pause words than controls. Interestingly, ASD participants appeared to use silent pauses in the place of filled-pauses. ASD participants used far more silent pauses than controls and engaged in these silent pauses at virtually the same rate as control participants used ums and uhs. However, unlike filled-pauses, silent pauses made it difficult for the speaker to know when the listener was finished speaking. In this sense, silent-pauses may reflect the same speaker-originating disfluencies in production, but do not attempt to remediate the potential confusion they cause to an interlocutor.

Participants with ASD also revised their speech significantly less often than controls. Belser and Sudhalter (2001) also found low levels of revisions in lower-functioning young adults with ASD, as well as in the spoken language of individuals with mental retardation, although this was not in comparison to typical controls. Revising speech involves self-repair, whereby a speaker detects a problem and formulates a revision or replacement to correct it (Levelt, 1983). Given this information, one could conclude that participants with autism either make fewer mistakes, or don't detect problems in their own speech the same way as controls do. One other alternative we suggest is that they may be able to detect their own formulation problems adequately, but may be less aware of the problems this may have caused a listener, and are therefore less likely to attempt to clarify and revise their utterance to aid a listener. The data here cannot distinguish between these possibilities, but it remains an intriguing possibility for follow-up work, and coincides with the findings of Clark and Wasow (1998).

While we argue that these results demonstrate the use of filled pauses as a listener-oriented behavior, the question remains, to what extent this is a volitional choice. The experience of at least many speakers seems to be a severely limited ability to inhibit ums and uhs, even if speaking only to themselves, making a purely volitional account problematic. We suggest instead that ums and uhs may become a habitual part of speech in typically developing children resulting from a responsiveness to interlocutors' states of mind. When one is interrupted when pausing before finished speaking, it seems likely that theory of mind reasoning would be required to understand that the interlocutor mistook the silence for the end of the utterance, and that filling the pause with verbal material would be required to hold one's turn. Similarly, lengthy silent pauses typically make one's interlocutors quite uncomfortable. Anecdotal experience from this study was that experimenters found it very awkward to simply wait for participants with ASD to resume speaking, and that at times it was difficult to follow experimental protocol and not fill the silence themselves. On the other hand, participants with ASD appeared to either not perceive, or at least not be concerned by any potential discomfort on the part of their conversational partners.

One limitation of this study is the fact that while the participants with ASD were high functioning, and had good verbal skills, the control participants were not matched to them on IQ or education, only on age and gender. While we do not have IQ information on our control participants, it seems likely that their scores would have been higher than those of the ASD participants, although there should be at least some overlap. Furthermore, there was a sizable difference in MLU, which could allow different levels of disfluency opportunities. However, inspection of the distributions suggests that while both VIQ and MLU do show some associations with measures of disfluency, they alone do not appear to account for the large differences we see between the groups.

We do note, however, that these findings are not entirely consistent with some other reports in the literature. Shriberg et al. (2001) found that in a sample of high functioning males with ASD aged 10-50, the ASD participants showed an increased rate of disfluencies, which they describe as an increased rate of one word repetitions and revisions, as compared to controls. However, their data actually show that while the ASD participants did indeed show significantly higher rates of one word repetitions, it was in fact the controls who showed significantly higher rates of revisions than the ASD participants (Figure 1, Panel C, p1105). Although the discussion in Shriberg et al. (2001) glosses over the distinction between the revisions and repetitions, and states that the ASD participants showed greater rates of disfluency, their data are very much in line with the findings that we present here, in that our ASD participants also showed decreased rates of revisions as compared to controls. Thurber and Tager-Flusberg (1993) showed that a sample of 12-year-old children with autism showed *lower* rates of silent pauses within phrases than did typical children matched for verbal mental age (approximately 8-year-olds), which they attribute to lower levels of communicative and cognitive demand from the stories told by the children with autism. However, they did not report filled pauses. This discrepancy between findings has several possible sources – these were from children with autism, who were also relatively lower-functioning. Furthermore, our silent pauses were defined to be significantly longer than the brief hesitations described in this study. However, a more complete understanding will require an investigation of the developmental trajectory of disfluencies in children with ASD.

These results add further support to the findings of Fox Tree and colleagues (Clark & Fox Tree, 2002; Fox Tree 2001; Fox Tree & Schrock 1999) in showing the useful nature fof ums and uhs for both speakers and listeners in conversation. We also provide convergent evidence for the idea that ums and uhs are not simply meaningless fillers that listeners have opportunistically discovered how to make use of. Instead, we find that speakers with ASD who have normal verbal IQs, who are by definition egocentric (and therefore not likely to be listener-oriented) do not use ums and uhs, but instead appear to use silent pauses. We therefore argue that ums and uhs have likely become part of normal speaking as a response to listeners' needs, even if we eventually lose some volitional control over their usage.

Acknowledgements

We thank Mel Rutherford and the Human Development Lab for help with recruiting participants for this study. We also thank Megan Keilty for assistance with data collection, and Elizabeth Hall, Amy Beth Warriner, Scott Watter and Judith Shedden, as well as members of the Cognitive Science Lab, for helpful comments and suggestions. This work was supported by Natural Sciences and Engineering Research Council of Canada grant #293145 to KRH. Correspondence regarding this article should be addressed to either Johanna Lake <lakejk@mcmaster.ca> or Karin Humphreys <krh@mcmaster.ca>, Department of Psychology, Neuroscience & Behaviour, McMaster University, 1280 Main St West, Hamilton, Ontario, Canada, L8S 4K1.

References

American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders Fourth Edition.* Washington, DC: American Psychiatric Association.

Baltaxe, C. (1977). Pragmatic deficits in the language of autistic adolescents. *Journal of Pediatric Psychology, 2,* 176-180.

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind.* Cambridge, MA: MIT Press.

Belser, R. C., & Sudhalter, V. (2001). Conversational characteristics of children with fragile X syndrome: Repetitive speech. *American Journal on Mental Retardation, 106,* 28-38.

Bishop, D. V. M. (1998). Development of the children's communicative checklist (ccc): A method for assessing qualitative aspects of communicative impairment in children. *Journal of Child Psychology and Psychiatry, 39,* 879-891.

Bock, K. (1996). Language production: Methods and methodologies. *Psychonomic Bulletin & Review, 3,* 395-421.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory and Cognition, 22,* 1482-1493.

Clark, H. H. (1996). *Using Language.* New York, NY: Cambridge University Press.

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition, 84,* 73-111.

Clark, H. H., & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology, 37,* 201-242.

De Villiers, J., Fine, J., Ginsberg, G., Vaccarella, L., & Szatmari, P. (2007). A scale for rating conversational impairment in autism spectrum disorder. *Journal of Autism and Developmental Disorders, 37,* 1375-1380.

Fine, J., Bartolucci, G., Ginsberg, G., & Szatmari, P. (1991). The use of intonation to communicate in pervasive developmental disorders. *Journal of Child Psychology and Psychiatry, 32*, 771-782.

Fox Tree, J. E. (2001). Listeners' uses of um and uh in speech comprehension. *Memory & Cognition, 29,* 320-326.

Fox Tree, J. E. (2002). Interpreting pauses and ums at turn exchanges. *Discourse Processes, 34,* 37-55.

Fox Tree, J. E., & Schrock, J. C. (1999). Discourse markers in spontaneous speech: Oh what a difference an oh makes. *Journal of Memory and Language, 40,* 280-295.

Girbau, D. (2001). Children's referential communication failure: The ambiguity and abbreviation message. *Journal of Language and Social Psychology, 20*, 81-89.

Horton, W. S., & Keysar, B. (1996). When do speakers take into account common ground? *Cognition, 59,* 91-117.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition, 14,* 41-104.

Lord, C., Rutter, M., Goode, S., Heemsbergen, J., Jordan, H., Mawhood, L., & Schopler, E. (1989). Autism diagnostic observation schedule: A standardized observation of communication and social behavior. *Journal of Autism and Developmental Disorders, 19,* 185-212.

Mangold, R., & Pobel, R. (1988). Informativeness and instrumentality in referential communication. *Journal of Language Social Psychology, 7,* 181-191.

Miller, J. F., & Chapman, R. S. (1983). *Systematic analysis of language transcripts (SALT).* San Diego, College Hill Press.

Shriberg, L. D., Paul, R., McSweeny, J. L., Klin, A., Cohen, D. J., % Volkmar, F. R. (2001). Speech and prosody characteristics of adolescents and adults with high-functioning autism and Asperger syndrome. *Journal of Speech, Language and Hearing Research, 44*, 1097-1115.

Thurber, C., & Tager-Flusberg, H. (1993). Pauses in the narratives produced by autistic, mentally retarded, and normal children as an index of cognitive demand. *Journal of Autism and Developmental Disorders, 23*, 309-322.

Wechsler, D. (1939). *The measurement of adult intelligence.* Baltimore: Williams & Wilkins.

Wetherby, A., & Prutting, C. (1984). Profiles of communicative and cognitive social abilities in autistic children. *Journal of Speech and Hearing Research, 27,* 364-377.

Young, E. C., Diehl, J. J., Morris, D., Hyman, S. L., & Bennetto, L. (2005). The use of two language tests to identify pragmatic language problems in children with autism spectrum disorders. *Language, speech, and hearing services in schools, 36*, 62-72.

Ziatas, K., Durkin, K., & Pratt, C. (2003). Differences in assertive speech acts produced by children with autism, Asperger syndrome, specific language impairment, and normal development. *Development and Psychopathology, 15,* 73-94.

Chapter 5

Conclusion

Language development during the preschool years is highly variable and to date there is little consensus over the best way to measure children's language. Researchers and clinicians have developed a variety of methods to assess different aspects of language from standardized tests of language to more comprehensive discourse analysis of speech. Concerns regarding the diagnostic accuracy of standardized tests of language have raised important questions about their ability to assess language and identify impairment without additional linguistic information. Additionally, previous research indicates that the production of specific classes of words and the use of syntax and pragmatics conveys essential information about language progression in childhood and later in life (Fox Tree, 2001; Klee, 1992; Redmond 2004; Rice & Wexler, 1996).

The present research program, therefore, used tools of conversational analysis to study the spoken language of preschoolers and young adults. Specifically, we used language indices obtained from spontaneous language samples to study how these measures related to one another and how they compared with Wechsler Preschool and Primary Scale of Intelligence (WPPSI-III) receptive and expressive vocabulary scores in the assessment of preschoolers' language abilities. Of our conversational measures, only mean length of utterance (MLU) was strongly related to WPPSI-III language measures. However, other measures from conversational speech appeared to constitute reasonably stable factors that may have utility for children's language assessment.

Developmentally, MLU was found to increase with vocabulary size and also with vocabulary diversity. Variations in the category of revisions (e.g. short, long, and average words per revision) were not highly related, suggesting they may have different functions in spoken production. These findings were also supported by factor analysis, suggesting that short and long revisions represent unique and separate factors.

Correlational and factor analyses suggested the use of a reduced list of core conversational measures since some measures were highly correlated with one another and also with MLU and total words/different words. Factor analysis determined a set of nine language factors, of which only one correlated with WPPSI receptive and expressive vocabulary scores. The remaining eight factors constituted features of spoken production that have been well documented in the literature as measures of spoken language crucial to the identification of language impairments (Marini, Tavano, & Fabbro, 2008; Redmond, 2004; Rice & Wexler, 1996).

Some researchers have used tools of conversational analysis to study the language of clinical and non clinical groups of children. Findings of this work have demonstrated some aspects of spoken language to be indicative of development or disorder including sentence length, clause density, revisions, and filled-paused words (Beitchman et al., 1996; Qi & Kaiser, 2004; Redmond, 2004; Scott & Stokes, 1995), while other researchers document no relationship between language and childhood disorder (Rescorla & Alley, 2001; Rescorla & Achenbach, 2002). Few studies, however, examine the relationship between conversational measures of language and the continuum of behavioural and emotional problems, ranging from normal behaviour to clinically significant psychiatric symptoms. Taken together, what these results indicated was the need for more studies using broader and more comprehensive tools of linguistic analysis in addition to standardized tests of language, to better understand the language of young children with and without psychiatric symptoms. To fill this knowledge gap, the current program of research used spontaneous language sampling and analysis of conversational transcripts, to study the relationship between language and maladaptive behavioural and emotional problems in a non-clinical, and therefore highly variable, sample of preschool children. The utility of using language measures obtained from spontaneous speech in predicting mental health problems as indexed by the Child Behavior Checklist (CBCL C-TRF) problems scales was also examined and compared to the predictability of WPPSI-III general language composite scores (GLC). Results indicated that conversational measures of language significantly improved prediction of CBCL C-TRF DSM-oriented and syndrome scales beyond that accounted for by WPPSI-III GLC scores. Revisions, specifically, appeared to play a critical role in the prediction of behavioural and emotional problems in preschoolers.

Given the significance of disfluencies in the spoken language of children with and without mental health symptoms, this program of research examined the role of disfluencies such as "um" or "uh," in the conversation of young adults with and without ASD. The characteristic egocentricity of individuals with autism meant they should engage in minimal listener-oriented behaviour, and made this population a useful group to differentiate listener and speaker-oriented functions. Results showed that individuals with ASD produced fewer filled-pause words (ums and uhs) and revisions than controls, but more silent pauses. Filled-pause words therefore appear to be listener-oriented features of speech.

In its entirety, this program of study adds valuable and original research to our knowledge of language assessment in young children and adults. Our conversational measures contributed significant prediction to standardized tests of language and also identified limitations of standardized tests breadth of measurement. These results are particularly profound in the context of treatment for children with language impairment or psychopathology since it is largely dependent on accurate and sensitive screening tools. Overall, our findings confirm the importance of using conversational measures of language as an alternative or in addition to standardized tests of language.

References

Achenbach, T.M., & Rescorla, L.A. (2000). *Manual for the ASEBA Preschool Forms and Profiles,* Burlington: VT: University of Vermont.

Adams, C., & Lloyd, J. (2005). Elicited and spontaneous communicative functions and stability of conversational measures with children who have pragmatic language impairments. *International Journal of Communication Disorders, 40,* 333-347.

Allen, M.J. & Yen, W.M. (1979). *Introduction to measurement theory.* Monterey, CA: Brooks/Cole.

Aram, D.M., Morris, R., & Hall, N.E. (1993). Clinical and research congruence in identifying children with specific language impairment. *Journal of Speech and Hearing Research, 36,* 580-591.

Barnhart, J.E. (1991). Criterion-related validity of interpretations of children's performance on emergent literacy tasks. *Journal of Reading Behavior, 23,* 425-444.

Bernstein Ratner, N., & Costa Sih, C. (1987). Effects of gradual increases in sentence length and complexity on children's disfluency. *Journal of Speech, Language, and Hearing Research, 52,* 278-287.

Bishop, D.V.M., & Butterworth, G.E. (1979). A longitudinal study using the WPPSI and WISC-R with an English sample. *British Journal of Educational Psychology, 49,* 156-168.

Bishop, D.V.M., Chan, J., Adams, C., Hartley, J., & Weir, F. (2000). Conversational responsiveness in specific language impairment: Evidence of disproportionate pragmatic difficulties in a subset of children. *Development and Psychopathology, 12,* 177-199.

Bishop, D.V.M., & Edmundson, A. (1987). Specific language impairment as a maturational lag: Evidence from longitudinal data on language and motor development. *Developmental Medicine and Child Neurology, 29,* 442-459.

Botting, N. (2002). Narrative as a tool for the assessment of linguistic and pragmatic impairments. *Child Language Teaching and Therapy, 18,* 1-21.

Boudreau, D.M., & Hedberg, N.L. (1999). A comparison of early literacy skills in children with specific language impairment and their typically developing peers. *American Journal of Speech-Language Pathology, 8,* 249-260.

Bowen, C. (1999). *Structural analysis of a language sample.* Retrieved from: http://members.tripod.com/caroline_bowen/minisample.htm

Brown, R. (1973). *A first language: The early stages.* London: George Allen & Unwin Ltd.

Bukatko, D. & Daehler, M. W. (1995). *Child Development: A thematic approach.* Boston: Houghton Mifflin Company.

Cain, K., & Oakhill, J. (1996). The nature of the relationship between comprehension skill and the ability to tell a story. *British Journal of Developmental Psychology, 14,* 187-201.

Caselli, M.C., Bates, E., Casadio, P., Fenson, J., Fenson, L., Sanders, L., & Weir, J. (1995). A cross-linguistic study of early lexical development. *Cognitive Development, 10,* 159-199.

Clark, H.H., & Fox Tree, J.E. (2002). Using uh and um in spontaneous speaking. *Cognition, 84,* 73-111.

Colburn, N. & Mysak, E. (1982). Developmental disfluency and emerging grammar: II. Co-occurrence of disfluency with specified semantic-syntactic structures. *Journal of Speech and Hearing Research, 25,* 421- 427.

Conti-Ramsden, G., Botting, N., Simkin, Z., & Knox, E. (2001). Follow-up of children attending infant language units: Outcomes at 11 years of age. *International Journal of Language and Communication Disorders, 36,* 207-219.

Cowley, J., & Glasgow, C. (1994). *The Renfrew Bus Story – U.S. Edition.* Centreville, DE: The Centreville School.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modem test theory.* Fort Worth, TX: Holt, Rinehart, and Winston.

DeThorne, L.S., & Watkins, R.V. (2006). Language abilities and nonverbal IQ in children with language impairment: Inconsistency across measures. *Clinical Linguistics and Phonetics, 20,* 641-658.

Dunn, L.M., & Dunn, L. (1997). *Peabody Picture Vocabulary Test* (3rd ed.). Circle Pines, MN: American Guidance Service.

Dunn, M., Flax, J., Sliwinski, M., & Aram, D. (1996). The use of spontaneous language measures as criteria for identifying children with specific language impairment: An attempt to reconcile clinical and research incongruence. *Journal of Speech, Language, and Hearing Research, 39,* 643-654.

Fenson, L., Bates, E., Reznick, J.S., Dale, P.S., & Thal, J. (2003). *MacArthur-Bates Communicative Development Inventories.* Paul H. Brookes Publishing Company.

Fenson, L., Dale, P.S., Reznick, J.S., Bates, E., Thal, .J., & Pethick, S.J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59* (Serial No. 242).

Fey, M.E. (1986). *Language intervention with young children.* Boston: Allyn & Bacon.

Fox Tree, J.E. (2001). Listeners' uses of um and uh in speech comprehension. *Memory and Cognition, 29,* 320-326.

Gallagher, T.M., & Craig, H.K. (1978). Structural characteristics of monologues in the speech of normal children: Semantic and conversational aspects. *Journal of Speech and Hearing Research, 21,* 103-117.

Gardner, M. (1985). *Receptive One-Word Picture Vocabulary Test.* Novato, CA: Academic Therapy Publications.

Gardner, M. (1990). *Expressive One-Word Picture Vocabulary Test-Revised.* Novato, CA: Academic Therapy Publications.

Gillam, R., & Johnston, J. (1992). Spoken and written language relationships in language/learning impaired and normally achieving school-age children. *Journal of Speech and Hearing Research, 35,* 1303-1315.

Gleser, G.C., Gottschalk, L.A., & Watkins, J. (1959). The relationship of sex and intelligence to choice of words: A normative study of verbal behavior. *Journal of Clinical Psychology, 15,* 183-191.

Goldfield, B.A., & Reznick, J.S. (1996). Measuring the vocabulary spurt: A reply to Mervis and Bertrand. *Journal of Child Language, 23,* 241-246.

Gordon, P., Luper, H., & Peterson, H. (1986). The effects of syntactic complexity on the occurrence of disfluencies in 5-year-old nonstutterers. *Journal of Communication Disorders, 11,* 79-83.

Gray, S., Plante, E., Vance, R., & Henrichsen, M. (1999). The diagnostic accuracy of four vocabulary tests administered to preschool-age children. *Language, Speech, and Hearing Services in Schools, 30,* 196-206.

Haas, C.W., Haug, H.T., & Landry, R.G. (1989). Sample size and type-token ratios for oral language of preschool children. *Journal of Speech and Hearing Research, 32,* 536-540.

Halliday, M.A.K. (1977). *Exploration in the functions of language.* New York: Elsevier North-Holland.

Halliday, M.A.K. (1978). *Language as a social semiotic: The social interpretation of language and meaning.* Baltimore: University Park Press.

Hart, B., & Risley, T.R. (1995). *Meaningful differences in the everyday experience of young American children.* Baltimore: Paul H. Brookes.

Hoff, E. (2005). *Language Development.* 3rd Ed. Belmont, CA: Thomson Wadsworth.

Hoff, E., Laursen, B., & Tardif, T. (2002). Socioeconomic status and parenting. In M.H. Bornstein (Ed.), *Handbook of Parenting, Volume II: Ecology and Biology of Parenting* (pp. 161-188). Mahwah, NJ: Lawrence Erlbaum Associates.

Hoff, E. & Naigles, L. (2002). How children use input to acquire lexicon. *Child Development, 73,* 418-433.

Horwitz, S.M., Irwin, J.R., Briggs-Gowan, M.J., Heenan, J.M.B., Mendoza, J., & Carter, A.S. (2003). Language delay in a community cohort of young children. *Journal of the American Academy of Child & Adolescent Psychiatry, 42,* 932-940.

Howland, P., & Kendall, L. (1991). Assessing children with language tests- which tests to use? *British Journal of Disorders of Communication, 26,* 355-367.

Hresko, W.P., Kim Reid, D., & Hammil, D.D. (1981). *The Test of Early Language Development.* Austin, TX: PRO-ED.

Hresko, W.P., Kim Reid, D., & Hammil, D.D. (1999). *The Test of Early Language Development.* (3rd ed.). Austin, TX: PRO-ED.

Hudson Kam, C.L., & Edwards, N.A. (2008). The use of uh and um by 3- and 4- year old native English-speaking children: Not quite right but not completely wrong. *First Language, 28,* 313-327.

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology, 27,* 236-248.

Irwin, J., Carter, A., & Briggs-Gowan, M. (2002). The social-emotional development of late-talking toddlers. *Journal of the American Academy of Child & Adolescent Psychiatry, 41,* 1324-1332.

Jaeger, J. J. (1992). Not by the chair of my hinny hin hin: Some general properties of slips of the tongue in young children. *Journal of Child Language, 19*, 335-366.

Kaderavek, J.N., & Sulzby, E. (2000). Narrative production by children with and without specific language impairment: Oral narratives and emergent readings. *Journal of Speech Language and Hearing Research, 43,* 34-49.

Kaplan, C. (1996). Predictive validity of the WPPSI-R: A four year follow-up study. *Psychology in the Schools, 33,* 211-220.

Kaufman, A.S. (1973). Comparison of the WPPSI, Stanford-Binet, and McCarthy Scales as predictors of first- grade achievement. *Perceptual and Motor Skills, 36,* 67-73.

Kaufman, A. S., & Kaufman, N. L. (1990). *Kaufman Brief Intelligence Test manual.* Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Kaufman, N. L. (1993). Test review: Kaufman Adolescent and Adult Intelligence Test (2nd ed.). *Journal of Psychoeducational Assessment, 12*, 190-196.

Klee, T. (1992). Developmental and diagnostic characteristics of quantitative measures of children's language production. *Topics in Language Disorders, 12,* 28-41.

Krebs, E.G. (1969). The Wechsler Preschool and Primary Scale of Intelligence and prediction of reading achievement in first grade. *Dissertation Abstracts International,* 30, 4279-A (University Microfilms, No. 70-3361).

Levelt, W. J. M. (1983). Monitoring and self repair in speech. *Cognition*, 14, 41–104.

Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation.* Cambridge, MA: MIT Press.

Lieblich, A., & Shinar, M. (1975). The predictive validity of the WPPSI with Israeli children. *Educational and Psychological Measurement, 35,* 473-475.

Liles, B.Z., & Duffy, R.J. (1995). Measurement of narrative discourse ability in children with language disorders. *Journal of Speech, Language, and Hearing Research, 38,* - 415-425.

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk. 3rd Edition.* Mahwah, NJ: Lawrence Erlbaum Associates.

MacWhinney, B., & Osser, H. (1977). Verbal planning functions in children's speech. *Child Development, 48*, 978-985.

Magnusson, D. (1967). *Test Theory* .Reading, MA: Addison-Wesley Publishing Company.

Marini, A., Tavano, A., & Fabbro, F. (2008). Assessment of linguistic abilities in Italian children with Specific Language Impairment. *Neuropsychologia, 46,* 2816-2823.

McCauley, R.J., & Swisher, L. (1984). Psychometric review of language and articulation tests for preschool children. *Journal of Speech and Hearing Disorders, 49,* 34-42.

McLaughlin, S., & Cullinan, W. (1989). Disfluencies, utterance length, and linguistic complexity in nonstuttering children. *Child Development, 48,* 978-985.

Mehl, M.R., & Pennebaker, J.W. (2003). The sounds of social life: A psychometric analysis of students' daily social environments and natural conversations. *Journal of Personality and Social Psychology, 84,* 857-870.

Merrell, A.W., & Plante, E. (1997). Norm-referenced test interpretation in the diagnostic process. *Language, Speech, and Hearing Services in Schools, 28,* 50-58.

Merritt, D.D., & Liles, B.Z. (1987). Story grammar ability in children with an without language disorder: Story generation, story retelling and story comprehension. *Journal of Speech and Hearing Research, 30,* 539-552.

Nelson, K. (1973). Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development, 38,* (1 and 2, Serial No. 149).

Osherson, D., Stob, M., & Weinstein, S. (1984). Learning theory and natural language. *Cognition, 17,* 1–28.

Paul, R., & Smith, R.L. (1993). Narrative skills in 4-year-ols with normal, impaired, and late-developing language. *Journal of Speech, Language, and Hearing Research, 36,* 592-598.

Pasewark, R., Scherr, S., & Sawyer, R. (1974). Correlations of scores on the Vane Kindergarten, Wechsler Preschool and Primary Scale of Intelligence, and Metropolitan Reading Readiness tests. *Perceptual Motor Skills, 38*, 518.

Pearl, S.Z., & Bernthal, J.E. (1980). The effect of grammatical complexity upon disfluency behavior of nonstuttering preschool children. *Journal of Fluency Disorders, 5,* 55-68.

Pence, K.L., & Justice, L.M. (2008). *Language Development from Theory to Practice.* Upper Saddle River, New Jersey: Pearson Education.

Pennebaker, J.W., & King, L.A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality and Social Psychology, 77,* 1296-1312.

Peterson, C. & McCabe, A. (1992). Parental styles of narrative elicitation: Effect on children's narrative structure and content. "First Language", 12, 299-321.

Piaget, J. (1972). *The child's conception of the world.* Towota, NJ. Littlefield Adams.

Pine, J.M. (1994). The language of primary caregivers. In C. Gallaway & B.J. Richards (Eds.), *Input and interaction in language acquisition* (pp. 15-37). Cambridge: Cambridge University Press.

Pinker, S. (1979) Formal Models of Language Learning. *Cognition, 7,* 217-283.

Plante, E., & Vance, R. (1994). Selection of preschool language tests: A data-based approach. *Language, Speech, and Hearing Services in Schools, 25,* 15-24.

Qi, C.H.., & Kaiser, A.P. (2004). Problem behaviors in low-income children with language delays: An observation study. *Journal of Speech, Language, and Hearing research, 47,* 595-609.

Redmond, S.M. (2004). Conversational profiles of children with ADHD, SLI and typical development. *Clinical Linguistics and Phonetics, 18,* 107-125.

Reid, D.K., Hresko, W., & Hammill, D. (2002). *Test of Early Reading Ability* (3rd ed.). Austin, TX: PRO-ED.

Reynolds, C.R., Wright, D., & Dappen, L., (1981). A comparison of the criterion-related validity (academic achievement) of the WPPSI and the WISC-R. *Psychology in the Schools, 18,* 20–23.

Rice, M.L., & Bode, J.V. (1993). GAPS in the verb lexicons of children with specific language impairment. *First Language, 13,* 113-131.

Rice, M.L., Buhr, J.C., & Nemeth, M. (1990). Fast mapping word-learning abilities of language-delayed preschoolers. *Journal of Speech and Hearing Disorders, 55,* 33-42.

Rice, M.L., Buhr, J.C., & Oetting, J.B. (1992). Specific-language-impaired children's quick incidental learning of words: The effect of a pause. *Journal of Speech and Hearing Research, 35,* 1040-1048.

Rice, M.L., & Wexler, K. (1996). Toward tense as a clinical marker of a specific language impairment in English-speaking children. *Journal of Speech and Hearing Research, 39,* 1239-1257.

Richards, B. J., & Malvern, D. D. (1997a). *Quantifying Lexical Diversity in the Study of Language Development*. The University of Reading New Bulmershe Papers, Reading.

Rispoli, M., & Hadley, P. (2001). The leading edge: The significance of sentence disruptions in the development of grammar. *Journal of Speech, Language, and Hearing Research, 44,* 1131-1143.

Rispoli, M., Hadley, P., & Holt, J. (2008). Stalls and revisions: A developmental perspective on sentence production. *Journal of Speech, Language, and Hearing Research, 51,* 953-966.

Rock, D.A., & Stenner, A.J. (2005). Assessment issues in the testing of children at school entry. *The Future of Children, 15,* 15-34.

Roth, F.P., Speece, D.L., & Cooper, D.H. (1997). *Longitudinal examination of the relationship of oral language and early reading development.* Paper presented at the annual meeting of the American Speech-Language-Hearing Research Association. Boston, MA, November.

Sattler, J. (1992). *Assessment of children.* (3rd ed.). San Diego, California: Jerome M. Sattler, Publisher, Inc.

Scanlon, D.M., & Vellutino, F.R. (1996). Prerequisite skills, early instruction, and success in first-grade reading: Selected results from a longitudinal study. *Mental Retardation and Developmental Disabilities Research Reviews, 2,* 54-63.

Schickedanz, J. A., Schickedanz, D. I., Forsyth, P. D., & Forsyth, G. A. (1998). *Understanding children and adolescents.* (3rd ed). Boston: Allyn and Bacon.

Scott, C.M., & Stokes, S.L. (1995). Measures of syntax in school-age children and adolescents. *Language, Speech, and Hearing Services in Schools, 26,* 309-319.

Seidenberg, M. (1997). Language acquisition and use: Learning and applying probabilistic constraints. *Science, 275,* 1599-1603.

Stromswold, K. (2001). The heritability of language: A review and metaanalysis of twin, adoption, and linkage studies. *Journal of the Linguistic Society of America, 77,* 647-723.

Stone, P.J., Dunphy, D.C., Smith, M.S., & Ogilvie, D.M. (1966). The general inquirer: *A computer approach to content analysis.* MIT Press (Eds.).

Stothard, S.E., Snowling, M.J., Bishop, D.V.M., Chipchase, B.B., & Kaplan, C.C. (1998). Language impaired preschoolers: A follow up into adolescence. *Journal of Speech, Language and Hearing Research, 41,* 407-418.

Tager-Flusberg, H. (1995). Once upon a ribbit: Stories narrated by autistic children. *British Journal of Developmental Psychology, 13,* 45-59.

Tew, B.J., & Laurence, K.M. (1983). The relationship between spina bifida's children intelligence scores on school entry and at school leaving: A preliminary report. *Child: Care, Health, and Development, 9,* 13-17.

Tomasello, M., & Todd, J. (1983). Join attention and lexical acquisition style. *First Language, 4,*197-212.

Ukrainetz, T.A., & Blomquist, C. (2002). The criterion validity of four vocabulary tests compared to a language sample. *Child Language Teaching and Therapy, 18,* 59-78.

Van der Lely, H.J.K. (1997). Narrative discourse in grammatical specific language impaired children: A modular deficit? *Journal of Child Language, 24,* 221-256.

Watkins, R.V., & Kelly, D.J. (1995). Measuring children's lexical diversity: Differentiating typical and impaired language learners. *Journal of Speech & Hearing Research, 38,* 1349-1356.

Weintraub, W. (1981). *Verbal behavior: Adaptation and Psychopathology.* New York, NY: Springer.

Wechsler, D. (2002). *Wechsler Preschool and Primary Scale of Intelligence, 3rd. ed. (WPPSI-III).* Psychological Corporation.

Wexler, K., & Culicover, P. (1980). *Formal Principles of Language Acquisition.* Cambridge, MA: MIT Press.

Wiig, E.H., Secord, W.H., & Semel, E. (2004). *Clinical Evaluation of Language Fundamentals- Preschool* (2nd ed.). San Antonio, TX: Harcourt Assessment.

Wijnen, F. (1992). Incidental word and sound errors in young speakers. *Journal of Memory and Language, 31*, 734-755.

Williams, K. (1997). *Expressive Vocabulary Test.* Circle Pines, MN: American Guidance Service.

Winsler, A., Carlton, M.P., & Barry, M.J. (2000). Age related changes in preschool children's systematic use of private speech in a natural setting. *Journal of Child Language, 27,* 665-688.

Woodcock, R.W. (1998). *Woodcock Reading Mastery Revised, Normative Update.* Circle Pines, MN: American Guidance Service.

Yaruss, J.S., Newman, R.M., & Flora, T. (1999). Language and disfluency in nonstuttering chlidren's conversational speech. *Journal of Fluency Disorders, 24,* 185-207.

Yule, W., Gold, D.R., & Busch, C. (1982). Long-term predictive validity of the WPPSI: An 11-year follow-up study. *Personality and Individual Differences, 3,* 65-71.

Zimmerman, I.L., Steiner, V.G., & Pond, R.E. (2002). *Preschool Language Scale* (4th ed.). San Antonio, TX: Psychological Corporation.

12827 09