Predicting Later Language Outcomes from the Language Use Inventory

Diane Pesco, Concordia University

Daniela K. O'Neill, University of Waterloo

now published in *Journal of Speech, Language, and Hearing Research:*

http://jslhr.asha.org/cgi/content/abstract/55/2/421

Author Note:

Diane Pesco, Department of Education, Concordia University, Montreal, QC, Canada; Daniela K. O'Neill, Department of Psychology, Waterloo, ON, Canada

Correspondence concerning this article should be addressed to Diane Pesco, Department of Education (LB 501-3), Concordia University, 1455 boul. de Maisonneuve West, Montreal, QC, H3G 1M8, Canada. E-mail: dpesco@education.concordia.ca. Phone 514-848-2424 ext. 7338. Fax 514-848-4520.

To appear in *Journal of Speech, Language, and Hearing Research*

Abstract

Purpose: To examine the predictive validity of the Language Use Inventory (LUI), a parent report of language use by children 18 to 47 months old (O'Neill, 2009). **Method:** 348 children whose parents had completed the LUI were re-assessed at 5 to 6 years old with standardized, norm-referenced language measures and parent report of developmental history. The relationship between scores on the LUI and later measures was examined through correlation, binary classification, and ROC curve analysis. **Results:** For children aged 24 to 47 months at the time of LUI completion, LUI scores correlated significantly with language measure scores. Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) was also calculated for four cut-off scores on the LUI, including -1.64 SD, a score that maximized sensitivity to 81% and specificity to 93%. For children aged 18-23 months at the time of LUI completion, specificity and NPV were high but sensitivity and PPV were lower than desirable. **Conclusions:** The results provide initial support for the LUI's predictive validity, particularly for children 24-47 months, and suggest the LUI can serve as an indicator of later language outcomes in referred populations. The results compare favourably to findings for other early child language measures.

Keywords: predictive validity, pragmatics, parent report, language assessment, language delay

Predicting Later Language Outcomes from the Language Use Inventory

Introduction

The Language Use Inventory

In the present study, we examine the predictive validity of the Language Use Inventory (LUI), a new, standardized, norm-referenced parent report of language use by children 18 to 47 months old. The LUI is premised on an understanding of language as inherently social and a view of language development as entwined with growth in social cognition, especially children's growing understanding of mind (O'Neill, 2009). It captures children's use of language in a range of everyday situations, with various interlocutors, and for a variety of purposes: to achieve instrumental goals; to interact socially with others; to comment on the immediate environment; to communicate about absent people and events; and to express the emotions, thoughts, and beliefs of themselves and others (O'Neill, 2009). The orientation of the LUI is partly revealed in the subscale names, including, for example: "your child's requests for help", "how your child uses words to get you to notice something", "your child's questions and comments about themselves or other people", and "how your child adapts conversations to other people".

The LUI's focus on children's language use in daily life and its reliance on parents' knowledge contributes to its ecological validity. It fulfills a need for pragmatic measures (Adams, 2002) and is consistent with contemporary models of health and disability that identify participation in 'natural' environments as a critical element of evaluation (Washington, 2010; World Health Organization, 2007). These characteristics are combined with strong psychometric properties. The norming sample (described further in the Method) included children from across Canada and aimed to represent the Canadian population of children under 6 years old with respect to family income; parental education; family structure (single or dual parent); and

'visible minority' status, a census category reflecting respondents' ethnic and cultural origins. The large sample (N = 3563) allowed for norms for every month from 18 to 47 months, with norms further broken down by sex given differences in total scores between boys and girls at all ages (O'Neill, 2009). Studies conducted before norming attest to the LUI's high test-retest reliability, strong content validity, internal consistency, and concurrent validity (O'Neill, 2007).

The LUI is comprised of 180 items that parents report are easy to complete in 20-30 minutes. The items are organized into 14 subscales, 10 of which contribute to a total score out of 161. Of the remaining subscales, two involve unscored open-ended questions, and two assess gestures but are excluded from the Total Score which focuses on oral language (O'Neill, 2007).

The Relationship of Pragmatics to Other Linguistic Domains

While the LUI frames children's language in terms of communicative function, it also entails some aspects of semantics and syntax. For example, some subscales include questions about specific words or word classes, ranging from names of people and common objects and actions to later-developing mental state terms, conjunctions, and modals to express uncertainty. Similarly, the measures of pragmatics administered between the ages of 5 and 6 in the present study rely to some degree on semantic and syntactic knowledge.

The perspective adopted here is that pragmatics, semantics, and syntax are related rather than fully distinct and autonomous systems. Evidence for the relatedness of pragmatics to other language domains can be found in studies of typical and atypical development. For instance, typically-developing children's pragmatic abilities at 18 months (specifically, their use of language to participate in social exchanges and initiate joint attention) predicted syntax at 31 months, accounting for 36% of the variance (Rollins & Snow, 1998). Charman et al. (2005) found that the mean rate of verbal and nonverbal communicative acts by children with autism at

24 months was associated with their expressive vocabulary at 36 months (r = .52).

In clinical populations, deficits, while perhaps more pronounced in one domain of language, are rarely fully isolated. For example, semantic and grammatical difficulties have been found in populations in which pragmatics is considered the core deficit (e.g., children with autism spectrum disorders (ASD)) (Condouris, Meyer, & Tager-Flusberg, 2003; Philofsky, Fidler, & Hepburn, 2007; Rapin & Dunn, 2003). In a similar vein, pragmatic deficits have been reported for children with specific language impairment (SLI) (Osman, Shohdi, & Aziz, 2011), for whom grammar is often described as the primary area of difficulty (Rice, Warren, & Betz, 2005), as well as for children diagnosed with language impairments on the basis of semantic and syntactic criteria (Geurts & Embrechts, 2010). Moreover, longitudinal studies demonstrate that the domain of language most affected may change over time. Some children with language impairments, for example, present with primarily semantic and pragmatic difficulties at one age, but exhibit difficulties primarily with syntax later (Conti-Ramsden & Botting, 1999).

Use of the LUI to Identify Language Impairment and Predict Later Language Outcomes

Given converging evidence for the relationship of pragmatics to other linguistic domains, we propose that the LUI will predict scores on later composite measures of language, including but not limited to pragmatics. Earlier findings from a discriminative validity study lend further support to this proposal. O'Neill (2007) studied 49 children (18 to 47 months old) who were awaiting speech-language assessment when their parents completed the LUI and who were diagnosed two months later (on average) with language impairment (43/49 children) or phonological/speech impairment (6/49). Forty-nine typically-developing children were matched by age and sex to the clinical group. Discriminant analysis showed that the two groups could be distinguished highly accurately on the basis of their LUI scores, as indicated by sensitivity and

specificity values of nearly 96%. The difference between the clinical and control groups was also evident in their mean LUI scores: respectively, 27.4 (*SD* 23.3) and 106.5 (*SD* 27.2) (O'Neill, 2007). The standardized mean difference calculated from these values is -3.12, a very large effect that exceeds effects reported for 33 other child language tests (Spaulding, Plante, & Farinella, 2006). The findings suggest that the LUI might predict later language abilities, assessed globally.

The present study examines further the LUI's potential role in screening and diagnosis by investigating its ability to predict language "outcomes", defined in terms of children's performance on language measures of syntax, semantics, and pragmatics at ages 5 to 6 years and a developmental history provided by parents at the same follow-up visit. Prior to presenting the research questions, we consider briefly the results from predictive validity studies of other parent report measures of child language and address some methodological issues.

Accuracy and Predictive Validity of Parent Reports of Early Child Language

Parents have been shown to be accurate in reporting their children's language when questions tap current behaviours and require recognition rather than recall (Fenson et al., 2007; Glascoe & Dworkin, 1995), conditions met in the LUI's design. The LUI was also subjected to readability analyses and piloted extensively to ensure its accessibility and acceptability to a wide range of parents (O'Neill, 2007). Although some studies have found that socioeconomic status (SES) or, more specifically, maternal education, affects the accuracy of maternal report, the findings are inconclusive (see Law & Roy, 2008 for a review) and outweighed by the advantages of parent reports (Bishop & McDonald, 2009; O'Neill, 2007), including levels of accuracy that approximate direct measures (e.g., Sachse & Von Suchodoletz, 2008; Volden & Phillips, 2010).

Studies of the reliability and validity of parent reports of child language include a limited

number investigating predictive validity. In these studies, and in the present one, both correlations and predictive values have been reported. Predictive values comprise *sensitivity* and *specificity*, defined in order as the proportion of positive and negative cases on the outcome measure correctly identified by the predictor(s); *positive predictive value (PPV),* the proportion of cases that are positive on the predictor and remain positive on the outcome measure; and *negative predictive value (NPV),* the proportion of negative cases that remain negative. The values thus reflect agreement (true positives and true negatives) and disagreement (false positives and false negatives) between the predictor and outcome variables.

Predictive validity studies of other norm-referenced parent reports of young children's language, including the MacArthur-Bates Communicative Development Inventory (CDI) Words and Sentences Vocabulary Production subscale (Feldman et al., 2005); a German adaptation of the CDI - Toddler Form (Sachse & Von Suchodoletz, 2008); and the Language Development Survey (LDS) (Klee, Pearce, & Carson, 2000; Rescorla, 1989, 2002), indicate substantial numbers of false positives and, when reported, lower sensitivity than specificity. High rates of false positives are also evident in large-scale outcome studies. In one, fewer than half of 2-year-olds diagnosed with language delays still tested positive at ages 3 and 4 (Dale, Price, Bishop, & Plomin, 2003) and in another, less than a quarter diagnosed with 'late language emergence' at age 2 were diagnosed as language impaired at age 7 (Rice, Taylor, & Zubrick, 2008).

False Positives

One concern with false positives is that they are likely to result in referrals of children for more complete assessment and may thus contribute to lengthening waiting lists and overtaxing services. Referrals may also cause undue worry to parents, loss of time, and financial burden should private services be obtained. Children, too, may be stressed by the assessment process.

At the same time, the impact of a language assessment on families is not necessarily negative. Moreover, overestimation of positive results entails 'erring on the side of caution'; children with difficulties are unlikely to be missed. Nevertheless, directing services to families who need them most is a legitimate concern (cf Bishop & McDonald, 2009).

Marks et al. (2007) describe predictions in any developmental domain as "thorny" and discuss several reasons for false positives, including the many factors that can intervene in the period between initial and later assessment, inexact correspondence in terms of the focus of the predictor and outcome measures, and the variability of development in early childhood. Variability in language development, in particular, is evident in the literature on typical development (Fenson et al., 2000) and transient language delays imply variability in language growth (e.g., Dale et al., 2003; Rice et al., 2008). Given these many factors, it is unrealistic to expect close to perfect predictive values for measures of child language administered in the toddler and perhaps the early preschool years. Still, the search for maximal predictions is an important one with the potential of improving both specific tests and our understanding of the language development of children with and without impairments.

Predictive Validity and the Purpose and Context of Assessment

Presence or absence of language impairment in validity studies and in clinical practice is typically partly determined based on children's performance on norm-referenced language measures relative to a cut-off score. The cut-off score can be decided arbitrarily or, as has been recommended, set at levels that maximize sensitivity and specificity (Spaulding, et al., 2006). Receiver operating characteristic (ROC), used in the present study, is a technique for examining classification accuracy and empirically determining cut-off scores that has been applied to child language measures (e.g., Klee et al., 2000; Redmond, Thompson, & Goldstein, 2011; Van Agt,

Van der Stege, De Ridder-Sluiter, & De Koning, 2007; Westerlund, Bergland, & Eriksson, 2006). Illustrated in Figure 1, the ROC procedure graphically plots true positive rate (i.e., sensitivity) and false positive rate (1-specificity) as a function of all possible cut-off scores on the predictor. The technique can thus be used to identify a cut-point that jointly maximizes sensitivity and specificity. The area under the curve of the cut-point is the probability that the predictor will distinguish the presence or absence of difficulties, with values closer to the upper bound of 1.0 indicating larger effects (Fletcher & Fletcher, 2005; Hanley & McNeil, 1982).

Sensitivity and specificity values, however, trade off; as one goes up, the other goes down. The optimal *balance* of the two must be decided in light of the goals of assessment. If the goal is to identify the maximal number of children who presently have language impairment or, in a predictive context, might be at risk for persistent impairment, sensitivity is most important. If, however, mislabelling a child as having language impairment when he or she does not is of greater concern, then specificity becomes paramount. In the present study, we investigated and report the cut-off score for which sensitivity and specificity levels were minimally .70 and also provide sensitivity and specificity levels associated with cut-off scores widely applied in practice and research. Although references to acceptable or ideal levels of sensitivity and specificity are rare in the literature, levels of .70 to .80 have been suggested as acceptable for developmental screening measures (American Academy of Pediatrics, 2006) and .80 has been occasionally suggested as a criterion level for measures used to screen for language delay (Law, Boyle, Harris, Harkness, & Nye, 2000) or diagnose language impairment (Plante & Vance, 1994).

Like sensitivity and specificity, PPV and NPV are inversely related and depend on cut-off scores, but additionally vary systematically with base rates of disorder; for any test, PPV rises with base rate and NPV falls (Meehl & Rosen, 1955; Streiner, 2003). In research samples

comprised of randomly selected children, the base rate of language impairment might approximate population prevalence rates (for example, the 7% rate of SLI found amongst kindergarteners) (Tomblin et al., 1997). In clinical samples (e.g., children referred for speech-language services) base rates are typically higher, often exceeding 50%. For example, speech-language pathologists in the U.S. reported that 70-90% of preschool age children referred for speech-language evaluation were subsequently diagnosed with language impairment (Wiig, Secord, & Semel, 2004). Base rates could also be expected to be higher than population prevalence amongst children who have not been referred but are at heightened risk for language difficulties (e.g., children with familial histories of language impairment).

Given the variability of base rates and their influence on predictive values, we examined and report PPV and NPV for the LUI as a function of several base rates. Such information plays an important role in ascertaining whether a measure has potential as a screening tool, as a diagnostic measure with predictive value for clinical populations, or both.

The Present Study

The purpose of the present study was to examine the predictive validity of the LUI for children ranging in age from 18 to 47 months at the time of LUI completion. The principal research questions follow.

1. (a) Does the LUI Total Score, expressed in standard units and indexing performance relative to age expectations, correlate with standard composite scores on norm-referenced language measures at ages 5 to 6 years? (b) Does the age of the children at the time of LUI completion influence the correlations?

2. When children in our sample are classified as ‘positive’ or ‘negative’ for language impairment based on their performance on language outcome measures and developmental

history: (a) What is the proportion of positive cases?, (b) What cut-off score on the LUI yields sensitivity and specificity values ≥ .70 according to ROC curve?, and (c) What are the effect sizes (area under the curve and risk ratio) associated with this cut-off score?

3. (a) What precisely is the LUI's sensitivity and specificity when the ROC-derived cut-off and other widely used cut-off scores are employed? (b) What are the positive and negative predictive values (PPV and NPV) for the cut-off scores at different base rates?

Method

Participants

Recruitment. Ethical approval for the research was granted by the Human Research Ethics Committee at the University of Waterloo. The participants were recruited from a database of children who participated in the LUI norming study. While the norming study involved a pan-Canadian sample, for the present study we recruited children living in an approximately 50-mile radius of the research site at the University of Waterloo in Ontario, Canada. We exhaustively recruited children with LUI scores below the 16th and in the 17th-49th percentiles in order to ensure a sample with adequate representation of relatively low scoring children. We simultaneously randomly sampled children with LUI scores ≥ 50th percentile until there were approximately equal numbers in the <50th and ≥ 50th percentile groups (see Results for the final distribution of participants). Contact with 678 families was attempted. Children from over half (52%) of these families participated. Of the remaining families, some had moved out of the area or could not be located (22%), while others either declined or did not respond to calls (26%).

Description of sample. The participants were 348 children for whom a parent had completed the LUI when their child was between 18 and 47 months old. At the time of LUI completion, participants met the norming study inclusionary criteria (O'Neill, 2009): if exposed

to languages other than English at time of LUI, exposed less than 20% of waking hours; born full-term or no more than two weeks premature or born at a weight exceeding 5 pounds 5 ounces (2.5 kg); and had not been diagnosed with hearing loss, speech/language delay, or developmental delay (suspected cases, however, were retained).

According to demographic information collected from parents when the LUI was completed, most children in the present study lived with two parents (96.6% vs. 3.4% with single parents), in households where incomes exceeded Canadian low-income thresholds (90.2% vs. 6.1% low income and 3.7% no response). Most did not belong to a visible minority (82.2% vs. 8.6% visible minority and 4.3% no response). Maternal education varied: 13.7% of mothers had a high school education or less; 29.6% had a trade or educational certification below the bachelor's level; and 56.7% had a university degree at the bachelor's level or higher. Public health data indicate that the sample represented the region in terms of mothers with university degrees, but involved fewer mothers with 'high school education or less' (Sanderson & Drew, 2009). The proportion of mothers with 'trades or educational certification' was consistent with regional census data for women 25-34 years old (Statistics Canada, 2007).

Children's mean age at follow-up was 5;8 (*M* in months 68.11, *SD* = 3.85). For some of the analyses to follow, the participants were grouped based on age at time of LUI completion (henceforth, LUI Age) into LUI Age Groups: 18-23, 24-29, 30-35, 36-41, and 42-47 months. These blocks divide the ages covered by the LUI into half-year intervals and allow for closer comparisons with findings for other measures where the age span is less large or demarcated by year. Table 1 provides the number of participants by LUI Age Group and Child Sex.

[Insert Table 1 about here]

Examiners

The first author, a certified speech-language pathologist with expertise in language assessment, conducted the first 30 of the 348 follow-up assessments (8.6%) for the purposes of training two full-time research assistants. The assistants received training in test administration, observed the first author during the training period, and were observed periodically to ensure fidelity of procedures and consistency between examiners. The assistants completed the final 318 of 348 assessments (91.4%). The three examiners were blind to the child's LUI score.

Procedures

Each child and his/her parent(s) attended a single testing session at the University of Waterloo Centre for Child Studies. Parental consent was obtained at the beginning of the session, as was verbal assent of the child. The parent was interviewed about the child's development and then independently rated the child's communicative abilities (see Outcome Measures).

For testing, the child and examiner sat at a child-sized table in one room, and the parent(s) observed the session from an adjoining room fitted with a one-way observation window and speakers. Two language tests were administered to the child in counterbalanced order. Sessions were video recorded to permit offline scoring. The testing sessions typically lasted about 70 minutes, excluding two short breaks. The children received a certificate acknowledging their participation in the study and a small gift (e.g., sticker book) once testing was completed.

Predictor (Index) Measure: The LUI

Several of the LUI's characteristics were described in the Introduction. The LUI Total Score served as predictor. This score, as defined in the user manual (O'Neill, 2009), is tallied from 161 items spread across 10 subscales. The LUI Total Scores were converted to a percentile as per the LUI manual and then to standard scores ($M = 100$, $SD = 15$).

Outcome Measures

Time between predictor and outcome measures. The time interval between the LUI and the administration of outcome measures ranged from 14.54 to 54.76 months (M = 39.26, SD = 9.03), depending on the age at which the LUI was initially completed.

Diagnostic Evaluation of Language Variation - Norm Referenced (DELV-NR). The DELV-NR (Seymour, Roeper, & de Villiers, 2005) is a standardized test of language for English-speaking children aged 4;0 to 9;11, normed in the U.S. Scaled scores on the Syntactic, Semantic, and Pragmatic Domains can be aggregated into a "total language composite score". The DELV-NR was selected based on several criteria: it has a strong theoretical and empirical basis; its psychometric properties are strong; it takes into account dialectical variation in English; and it is the only omnibus language test that offers separate norms for pragmatics for children of the age in the present study. Another advantage of this test, and the other two language tests used, is that the technical manual provides data on diagnostic accuracy that were crucial in determining appropriate cut-off scores for identifying difficulties in our sample.

Clinical Evaluation of Language Fundamentals Preschool (2nd Edition) (CELF-P2). The CELF-P2 (Wiig et al., 2004) is a standardized test for English-speaking children aged 3;0 to 6;11, also normed in the U.S. The test is comprised of several subtests that can be combined to yield various composites. One of these is a Core Language score, based on performance on three subtests: Expressive Vocabulary, Word Structure, and Sentence Structure. The Core Language score is the sole score the test developers used in establishing the CELF-P2's diagnostic accuracy (Wiig et al., 2004) and was thus the most fitting composite score for our purposes. Speech-language pathologists in the U.S. report using the test frequently in the diagnosis of SLI (Betz, Sullivan, & Eikhoff, 2010) and the word and sentence structure subtests have been shown to

correlate with spontaneous speech measures for children with autism (Condouris et al., 2003).

Children's Communication Checklist-Second Edition, U.S. Edition (CCC-2). The CCC-2 (Bishop, 2006) is a standardized, norm-referenced parent (or caregiver) report of the communication of youth aged 4;0 to 16;11, developed and validated in the United Kingdom and subsequently adapted for and normed on a U.S. population. Parents rate the frequency of their child's communication behaviours on a four-point scale. The ratings yield subscale scores for speech, syntax, semantics, coherence, and four areas of pragmatics: inappropriate initiations, stereotyped language, use of context, and nonverbal communication. The subscale scores are also summed to a General Communication Composite which has been shown to distinguish children with communication impairments from a control group (Norbury, Nash, Baird, & Bishop, 2004) and to be more effective at identifying children with ASD than the Test of Pragmatic Language (Volden & Phillips, 2010), a standardized, norm-referenced measure administered directly to the child (Phelps-Terasaki & Phelps-Gunn, 1992).

Parent interview. We developed a parent interview to systematically gather information about the child's developmental status since LUI completion. The questions covered hearing and vision, speech, language, reading, autism, behaviour, and emotional well-being. For each area, the parent was asked whether they or any other adult presently had concerns about the child and whether any difficulties had been formally identified by a professional since completion of the LUI. If responses to the questions were affirmative, additional information was requested: what the specific difficulties were; whether they had been treated, by whom, and for how long; and whether any difficulties noted had resolved or were still being treated or monitored.

Scoring of Outcome Measures

Testing sessions with the child were videorecorded. A research assistant was trained by the

first author to score the child measures and parent rating scale, following test instructions. Parent responses to interview questions regarding diagnosis and treatment of developmental disorders since LUI completion were assigned binary (yes/no) codes upon data entry.

Criteria for Positive Cases at Outcome

Based on diagnostic accuracy data for different cut-off scores and base rates provided in the test manuals of the CCC-2, CELF-P2, and DELV-NR, children with scores ≤ 7th percentile (rounded, -1.5 SD) on any composite were considered 'positive' for language difficulties on outcome (positive test criterion). If a parent reported that their child had been diagnosed with a language delay or impairment or with language difficulties associated with autism since completing the LUI, the case was also considered positive (positive history criterion).

Results

Descriptive Statistics

The total number of participants was 348. Table 1 presented the sample by LUI Age Group and Child Sex. Participants who scored below the 16th percentile on the LUI were slightly over-represented and those who scored above were slightly under-represented; the proportions for the ≤ 16th, 17th-49th, and ≥ 50th percentile groups were, respectively, 25%, 29.6%, and 45.4% (one would expect 16%, 34%, and 50% in a fully normal distribution). To take this into account, the cases were weighted for correlations and means as advised by our statistical consultant, an expert in the analysis of survey data (M. Thompson, personal communication, September 12, 2005). Weighting corrects for disproportional sample sizes by multiplying the data for every case by a weighting factor (e.g., 16% cases expected in population/25% in sample = weight of .64). Here, weights were calculated across or within LUI age groups, depending on the analysis.

Table 2 presents the means and SDs for the LUI Total Score and the composite scores for

each of the outcome measures for each LUI age group. The number of participants for each measure varies slightly due to missing data related to equipment failure (5 cases) or discontinuance of testing due to fatigue, discomfort, or marked distraction of the child (3 cases).

[Insert Table 2 about here]

The next set of results directly addresses the research questions identified in the Present Study section and are presented in the same numerical order.

Correlations of LUI Score with Language Outcome Measure Scores by LUI Age Group

Pearson correlations were conducted to determine the relationship between the LUI Total Standard Score and the standard scores on the language outcome measures for each LUI Age Group. As shown in Table 3, the correlations between LUI score and composite scores for each of the three language outcome measures were significant (r values ranged from .289 to .580, $p \leq .01$) in the middle LUI Age Groups 24-29, 30-35, 36-41. Significant correlations were also found between the LUI and DELV-NR scores for LUI Age Group 18-23 ($r = .190$), and between the LUI and CELF-P2 scores for LUI Age Group 42-47 ($r = .319$), $p \leq .02$.

[Insert Table 3 about here]

Correlations were also conducted between LUI Total Standard Score and scaled scores on the pragmatic subscales contributing to the CCC-2 and DELV-NR composites. As shown in Table 4, these were significant in the middle LUI Age Groups for particular subscales and r values ranged from .224 to .487, $p \leq .02$, depending on the subscale.

[Insert Table 4 about here]

Given the single significant correlation in LUI Age Group 18-23 for the language measure composites and their pragmatic subscales (DELV-NR and CCC-2), we thereafter analysed the data for the group independently from LUI Age Groups 24-47. As the reader will see, ROC

curve data further confirmed the decision to treat the groups separately.

Number of Positive Cases at Outcome

LUI Age Groups 24-47. Within this group, 21 of 236 children (8.9%) were identified as having language impairment at outcome: 15 based on the positive test criterion and 6 based on the positive history criterion. Of the 15 children meeting the positive test criterion, 5 also had a positive history (3 had been treated for language difficulties and 2 for ASD). Each of the 6 children meeting only the positive history criterion had been treated for language difficulties, and 4 of the 6 scored between -1 and -1.4 SD on at least one of the language outcome measures.

LUI Age Group 18-23. Within this group, 10 of 112 children, 8.9% (the same proportion as in the 24-47 group) were identified as having language impairment at outcome: 9 based on the positive test criterion and 1 based on the positive history criterion. Of the 9 children meeting the positive test criterion, 1 also had a positive history (treated for language difficulties). The single child meeting only the positive history criterion had been diagnosed with and treated for language difficulties, but scored above -1 SD on all the language outcome measures.

LUI Cut-offs Determined by ROC Curve and Effect Size

LUI Age Groups 24-47. ROC curve analysis was used to determine the LUI cut-off score that maximized sensitivity and specificity at minimum levels of .70. For LUI Age Groups 24-47 combined, the cut-off indicated by ROC coordinates for which both sensitivity and specificity exceeded .70 was -1.64 SD (5th percentile). At the -1.64 cut-off, both values exceeded .70 for LUI Age Groups 24-29, 30-35, and 36-41 considered separately, while for LUI Age Group 42-47, only specificity exceeded .70. Overall, the results confirmed our decision based on the correlations to treat LUI Age Groups 24-47 as a whole.

The coordinates for the LUI Age Groups 24-47 were plotted using the ROC procedure,

resulting in the curve provided in Figure 1. The figure shows the curve rising towards the upper left hand corner of the graph, an indicator of high sensitivity and specificity. The *area under the curve*, a measure of effect size with a value of .5 under the null hypothesis and a maximal value of 1.0, was significant: .910, $p < .001$ (CI 95% = .851 - .969).

[Insert Figure 1 about here]

The association between binary predictors and outcomes can be measured with a variety of other statistics which are related to one another in that they all express relationships between cells in the 2x2 contingency table. Amongst these is the risk ratio: the ratio of true positives to false negatives, equal to 1 under the null hypothesis. For LUI Age Group 24-47, the risk ratio was 27.09 (CI 95%, 10.72-72.49). These results show that a child with a score at the -1.64 SD cut-off on the LUI had a 27 times greater probability (risk) of exhibiting later language difficulties than a child with a score above the cut-off. Even at the lower bound of the confidence interval, the risk was tenfold.

LUI Age Group 18-23. For LUI Age Group 18-23, the ROC curve coordinates showed that no cut-off along the curve led to both sensitivity and specificity meeting our criterion level of .70. For example, at the -1.64 cut off determined for LUI Age Group 24-47, specificity was .87 but sensitivity was only .30. Given these findings, the results for PPV and NPV are restricted to LUI Age Group 24-47.

PPV and NPV at Different Cut-Offs and Base Rates

LUI Age Groups 24-47. The exact values for sensitivity and specificity at the -1.64 cut-off determined by ROC curve are provided in Table 5: .81 for sensitivity and .93 for specificity. Values for other cut-offs widely used in clinical and research contexts are also provided for comparative purposes. Table 5 also shows PPV and NPV as a function of several base rates.

[Insert Table 5 about here]

At a base rate of 10%, as one might observe in population screening, over half of children with scores -1.64 SD on the LUI might be expected to later exhibit language difficulties (PPV = .563), while the vast majority of children scoring above the cut-off would not (NPV = .978). At a higher base rate of 60%, as one might find in clinical samples, and again with a -1.64 cut-off, a far greater majority of children (PPV = .946) might be expected to later exhibit language difficulties. In contrast, over three-quarters scoring above the cut-off would not (NPV = .765). The increases in PPV were accompanied by decreases in NPV, as is always the case given calculation of these values; false positives decrease at the expense of missing some cases. Thus, the values in Table 5 must be interpreted in light of the use to which the LUI might be put, an issue returned to in the discussion.

Discussion

The aim of the study was to assess the degree to which the LUI, a parent report of young children's language use, predicts later language outcomes. Correlational data, ROC curve analysis, sensitivity and specificity values, and PPV and NPV at different base rates together provided initial support for the LUI's predictive validity. The results are discussed in turn, with limitations and strengths discussed in each section, and summarized in the conclusion.

Correlational Data

For children in the middle age groups covered by the LUI – that is, with a LUI Age of 24 to 41 months – significant correlations were found between the LUI Total Standard Score and children's scores on three standardized, norm-referenced language measures that served as outcome measures: the DELV-NR and CELF-P2, administered directly to children, and the CCC-2, a parent report. The r values ranged from low to moderate, were most consistent in the

LUI Age Group 30-35 months (r = .51 for the three language measures), and peaked at .58 for the CCC-2 for children in LUI Age Group 36-41 months. These values and the lower ones observed at other ages, which clearly leave a good deal of variance unaccounted for, might be partly explained by the time interval between predictor and outcome tests. This interval spanned over 3 years, on average, and as long as 4.5 years. During this time period, the child's language might have been influenced by a number of factors, ranging from growth in cognition to changes in linguistic input related to, for instance, preschool or school attendance. The moderate correlations found here were consistent with results from the rare studies that examine the relationship between toddlers' or preschoolers' language and their language during the early school years (Rescorla, 2002). In the oldest LUI Age Group (42-47 months), only one significant correlation was found. This result was surprising given the shorter time span between LUI completion and outcome measures but might be partly explained by the lesser variability (but no ceiling effect) in this group's LUI scores: M = 100.75, SD = 5.32. In contrast, the SDs for the LUI exceeded 13 for the other age groups in the sample, as was illustrated in Table 2.

Correlations between children's scores on the LUI and on the pragmatic subscales of the DELV-NR and CCC-2 (each contributing to the composites) were also examined. The LUI Total Score correlated significantly with scores on the Pragmatics Domain of the DELV-NR and three of the four pragmatic subscales of the CCC-2 for LUI Age Groups 24-29 and 30-35, as well as with scores on two CCC-2 subscales for LUI Age Group 36-41. Although the relationship of LUI scores to later pragmatics was not the primary focus of the present investigation, we did anticipate that LUI scores would correlate more uniformly and strongly with pragmatic measures than they did. A possible explanation for why they did not do so relates to the nature of the items on the LUI and the pragmatic measures. The DELV-NR Pragmatics Domain, one of the

few measures of pragmatics normed for children of the age we studied, has three parts (for which separate norm-referenced scores are not available): Communicative Role-taking, Question-Asking, and Short Narrative. These subparts relate to the LUI's content but only broadly. The CCC-2 subscales – in particular, inappropriate initiations and scripted language – also have no exact parallel in the LUI subscales. Thus, differences in the content and scope of pragmatic measures (see Russell & Grizzle, 2008 for an exhaustive review) coupled with the changing "face" of children's pragmatic abilities over time, are likely partially responsible for the results observed here. Still, the findings warrant further investigation. Perhaps one would find a stronger relationship between particular LUI subscales and either the pragmatic subscales or clusters of items within them, a possibility that has yet to be explored.

For the pragmatic measures discussed immediately above, analyses were restricted to correlations. This choice was deliberate. In lieu of setting arbitrary cut-offs for the outcome measures to define positive cases (presence of language difficulties), we applied cut-offs that had some empirical basis as they maximized diagnostic accuracy according to each test manual. We thus felt confident in using the composite scores as outcome measures. The DELV-NR Pragmatics Domain and CCC-2 pragmatic subscales have not been validated in terms of diagnostic accuracy and would have served poorly in this capacity.

ROC Curve and Predictive Values

ROC curve analysis was conducted using the continuous LUI Total Standard Score and a binary classification of language outcome based on a child's performance on language measures and developmental history (specifically, the presence or absence of language impairment in the interval between LUI completion and outcome measures).

The analysis showed -1.64 SD (the 5th percentile) to be the optimal cut-off score on the

LUI. Using this cut-off, sensitivity (.81) and specificity (.93) each exceeded our minimal criterion of .70 for LUI Age Groups 24-47 combined. The .80 criterion occasionally recommended in the literature for language measures was also met (Law et al., 2000; Plante & Vance, 1994). The area under the ROC curve, a measure of effect size and summary statistic of test accuracy, was .910, a value described elsewhere as "excellent" (Feldman et al., 2005; Jordan, Glutting, Ramineni, & Watkins, 2010). The risk ratio ("relative risk") was also calculated. Its value can extend from one to infinity and is thus more difficult to qualify, but at 27 in the present study, clearly demonstrated that children scoring below the ROC-determined cut-off on the LUI were far more probable to have later language difficulties than those scoring above the cut-off.

In interpreting the preceding results, the reader should keep in mind that the cut-off and the values stemming from it (sensitivity and specificity, effect size) are sample-specific. If the present study was replicated or if the sampling strategy was altered, a different cut-off might result and be further accompanied by different predictive values. There is, however, no reason to believe that applying a cut-off for the LUI based on previous experience with a different measure or on local protocols would be a better clinical choice than application of the cut-off score derived here for a single but reasonably large and diverse sample in which language impairment approximated population rates (8.9% in our sample).

The sensitivity of .81 showed that the LUI successfully detected later language difficulties in the present sample. The value is higher than the values for other parent report measures mentioned in the Introduction and administered in whole or part at age 2: the LDS (Rescorla**,** 1989, used in Klee et al., 2000); the Macarthur CDI: UK Short Form (Dionne, Dale, Boivin, & Plomin, 2003, used in Dale et al., 2003); and parts of the LDS and the Ages and Stages Questionnaire (ASQ) (Bricker & Squires, 1999, used in Rice et al., 2008). While Klee et al. used

an empirically-derived cut-off score, the basis of cut-offs in the latter two studies was unclear. Sensitivity of .81 also exceeds average values for language screening measures ranging in administration time from infancy to age 8: .65 for "high grade" studies involving typically-developing children and .72 for high-grade studies including clinical populations (Law et al., 2000). These values were for concurrently-administered index and reference measures, yet lower than in the present study where the index and reference measures were separated in time. That the LUI — a longer measure — performed better is a good indicator of its predictive validity.

The specificity value of .93 for the LUI indicated that children who did not exhibit difficulties at outcome had likewise scored negative (above the -1.64 cut-off) at an earlier age; the value thus permits a retrospective look from outcome and reflects agreement between the two sets of results. We can also track the results in a forward direction by calculating the proportion of children who scored above cut-off on the LUI and negative again on the outcome measures. This proportion – NPV – was very high at .98 for a 10% base rate (the sample rate was 8.9%; the 10% base rate is thus a minor extrapolation). It should be noted, however, that the probability that most children would later test negative for language difficulties is high *a priori*, given population rates of language impairment. Nevertheless, the results here indicate small improvements in those predictions and indicate that if the LUI were used in a screening context with the -1.64 cut-off, negative results would be highly consistent with later language status.

PPV was also reported for the -1.64 cut-off and for different base rates. Predictive values at other cut-offs were also provided to allow readers to compare the data here to other studies. For the -1.64 SD cut-off, the PPV of .56 at a 10% base rate means that more than half of children in populations with a similar base rate might exhibit later language difficulties. This value,

however, indicates a considerable number of false positives. The present findings thus do not provide adequate support for using the LUI to predict *long-term* outcomes in a sample with low base rates. However, the discriminative validity study (O'Neill, 2007) discussed in the Introduction demonstrated that the LUI can be used effectively even at low base rates (i.e., for screening purposes) in a *concurrent* context (that is, screening soon followed by assessment). Indeed, the vast majority of children in that study who screened positive on the LUI were soon after diagnosed with language impairment, and so, were "true" positives. (It can be argued that the LUI is simply too long for screening, but that issue is separate from accuracy and is outside the scope of this discussion). The higher rate of false positives in the predictive context examined here (relative to the concurrent one in O'Neill, 2007) is not surprising given that some children who exhibit early language delays do appear to "outgrow" them or eventually obtain language test scores within normal limits, even without treatment.

Returning to the present study, PPV values at higher base rates showed that in a referred or higher-risk population, the LUI would have good predictive power. It thus appears suitable for assessing the probability of later language difficulties for children who score near the -1.64 SD cut-off. Caution, however, should be exercised at base rates of 80% or 90%, as NPV declines to undesirable rates. It is difficult to compare the observed PPV and NPV values with those for other measures, given that the values are not routinely reported and that when they are reported, the base rate for which they were calculated is often either not specified (e.g., see studies reviewed in Law et al., 2000) or limited to the base rate observed in the sample (e.g., Feldman et al., 2005). More detailed reporting is needed to adequately compare the data.

Predictions for Children Below Age 2

We turn, finally, to the results for LUI Age Group 18-23 months. For this group, over 90%

of children who scored above the -1.64 SD cut-off also scored above the cut-offs on the language outcome measures and had no history of a language delay or disorder since LUI completion. Thus, a negative result on the LUI for even the youngest children can be viewed as a good indicator of later language status. The majority of children with LUI scores below the cut-off, however, no longer exhibited language difficulties at outcome, a finding consistent with results from other predictive validity studies involving children less than 24 months old. Westerlund et al. (2006), for instance, examined the relationship between 18 month-olds' scores on a screening measure derived from the Swedish MacArthur-Bates CDI (Berglund & Eriksson, 2000) and their scores on a language measure administered at age 3 years. Sensitivity was .50 and PPV was .176 (at a base rate of 4% calculated from the report) – quite low values. When children's scores on the Galician CDI (Miguel & Mariela, 2007) at 18 and 24 months were compared to subscales of the Reynell Developmental Language Scales III (Edwards, Garman, Hughes, Letts, & Sinka, 1999) at age 4, correlations were weaker for the younger group (Pérez-Pereira & Resches, 2010). Concerns about the diagnostic accuracy of communication screens (i.e., the ASQ) administered before the age of 2 years have also been raised (Rydz, et al., 2006). In summary, the numbers of false positives observed in the present study for the 18-23 group are congruent with other research, and the studies jointly suggest that younger than 24 months may be too early to screen for language impairment or to identify impairments that will persist into the school-age years.

Conclusion

The study was novel in that it examined, for the first time, the predictive validity of a norm-referenced measure that focuses explicitly on early pragmatics. One of the study's strengths was the inclusion of children in the full range of LUI Age Groups, leading to a more nuanced picture of the measure's predictive validity than would have been possible had we focused on a single

age. Other strengths were the inclusion of children scoring all along the distribution and a sample size equal to nearly 10% of the norming sample. The study relied on sources of information similar to those used in clinical evaluation to identify language difficulties: parent report of children's communicative skills; developmental history, including diagnosis and treatment related to language delay; and the child's performance on norm-referenced language measures validated in terms of their diagnostic accuracy. Data from these sources suggest that the LUI, a parent report emphasizing children's use of language, can play a significant role in predicting future language, a finding we hope shall be replicated. The results, notwithstanding the caveats raised throughout about the interpretation of predictive values, were promising and add to a growing body of evidence confirming the validity of the LUI.

Acknowledgments

This work was made possible by a grant from the Canadian Institutes of Health Research (CIHR), 2006-2009. The authors gratefully acknowledge Mare Appleby and Jillian Spratt for their research assistance, Professor Mary Thompson at the University of Waterloo for statistical consulting, and the many families and children who participated in the study.

The second author, Daniela K. O'Neill, is the author of the Language Use Inventory (LUI) and has a financial interest in the form of ownership in Knowledge in Development, Inc., the company that publishes the LUI. This financial interest is applicable as of 2009 when the company was founded. The research presented in this paper was designed and conducted prior to 2009 and was wholly funded by a 3-year grant awarded in 2006 to Daniela O'Neill (PI) and Diane Pesco (Co-investigator) by CIHR. The first author, Diane Pesco, was the lead investigator for the design and execution of the study and the statistical analysis of the results presented in this paper. She has no financial interest in the LUI or its publisher.

References

Adams, C. (2002). Practitioner review: The assessment of language pragmatics. *Journal of Child Psychology and Psychiatry, 43*(8), 973-987. doi:10.1111/1469-7610.00226

American Academy of Pediatrics (2006). Identifying infants and young children with developmental disorders in the medical home: An algorithm for developmental surveillance and screening. *Pediatrics, 118*(1), 405-420. doi:10.1542/peds.2006-1231

Berglund, E., & Eriksson, M. (2000). Communicative development in Swedish children 16–28 months old: The Swedish Early Communicative Development Inventory - Words and Sentences. *Scandinavian Journal of Psychology, 41*, 133–144. doi 10.1111/1467-9450.00181

Betz, S., Sullivan, S., & Eikhoff, J. (June, 2010). *Factors impacting the selection of standardized tests for the diagnosis of SLI*. Poster presented at the Symposium on Research in Child Language Disorders, Madison, WI.

Bishop, D. V. M. (2006). *Children's Communication Checklist 2 – U.S. edition*. San Antonio, TX: Harcourt Assessment.

Bishop, D. V. M., & McDonald, D. (2009). Identifying language impairment in children: Combining language test scores with parental report. *International Journal of Language & Communication Disorders*, *44*(5), 600-615. doi:10.1080/13682820802259662

Bricker, D., & Squires, J. (1999). *Ages and Stages Questionnaires: A Parent-Completed, Child-Monitoring System* (2nd ed.). Baltimore: Paul H. Brookes.

Charman, T., Taylor, E., Drew, A., Cockerill, H., Brown, J., & Baird, G. (2005). Outcome at 7 years of children diagnosed with autism at age 2: Predictive validity of assessments conducted at 2 and 3 years of age and pattern of symptom change over time. *Journal of*

Child Psychology and Psychiatry, *46,* 500–513. doi:10.1111/j.1469-7610.2004.00377.x

Condouris, K., Meyer, E., & Tager-Flusberg, H. (2003). The relationship between standardized measures of language and measures of spontaneous speech in children with autism. *American Journal of Speech-Language Pathology, 12*(3), 349-358. doi:10.1044/1058-0360(2003/080)

Conti-Ramsden, G., & Botting, N. (1999). Classification of children with specific language impairment: Longitudinal considerations. *Journal of Speech, Language, and Hearing Research*, *42*(5), 1195-1204.

Dale, P. S., Price, T. S., Bishop, D. V. M., & Plomin, R. (2003). Outcomes of early language delay: Predicting persistent and transient language difficulties at 3 and 4 years. *Journal of Speech, Language, and Hearing Research, 46*(3), 544-560. doi:10.1044/1092-4388(2003/044)

Dionne, G., Dale, P. S., Boivin, M., & Plomin, R. (2003). Genetic evidence for bidirectional effects of early lexical and grammatical development. *Child Development, 74*(2), 394-412. doi:10.1111/1467-8624.7402005

Edwards, S., Garman, M., Hughes, A., Letts, C., & Sinka, I. (1999). Clinical forum assessing the comprehension and production of language in young children: An account of the Reynell Developmental Language Scales III. *International Journal of Language & Communication Disorders, 34*(2), 151-171. doi: 10.1080/136828299247487

Feldman, H. M., Dale, P. S., Campbell, T. F., Colborn, D. K., Kurs-Lasky, M., Rockette, H. E., & Paradise, J. L. (2005). Concurrent and predictive validity of parent reports of child language at ages 2 and 3 years. *Child Development, 76*(4), 856-868. doi:10.1111/j.1467-8624.2005.00882.x

Fenson, L., Bates, E., Dale, P., Goodman, J., Reznick, J. S., & Thal, D. (2000). Measuring variability in early child language: Don't shoot the messenger. *Child Development, 71*(2), 323-328. doi:10.1111/1467-8624.00147

Fenson, L., Marchman, V. A., Thal, D., Dale, P. S., Reznick, J. S., Bates, E. (2007). *MacArthur-Bates Communicative Development Inventories: User's guide and technical manual* (2nd ed.). Baltimore, MA: Paul H. Brookes.

Fletcher, R. H., & Fletcher, S. W. (2005). *Clinical epidemiology: The essentials* (4th ed.). Philadelphia: Lippincott Williams & Wilkins.

Geurts, H., & Embrechts, M. (2010). Pragmatics in pre-schoolers with language impairments. *International Journal of Language & Communication Disorders*, *45*(4), 436-447. doi:10.3109/13682820903165685

Glascoe, F. P. & Dworkin, P. H. (1995). The role of parents in the detection of developmental and behavioral problems. *Pediatrics, 95,* 829-836.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, *143*, 29-36.

Jordan, N. C., Glutting, J., Ramineni, C., & Watkins, M. W. (2010). Validating a number sense screening tool for use in kindergarten and first grade: Prediction of mathematics proficiency in third grade. *School Psychology Review*, *39*(2), 181-195.

Klee, T., Pearce, K., & Carson, D. (2000). Improving the positive value of screening for developmental language disorder. *Journal of Speech, Language, and Hearing Research, 43*(4), 821-833.

Law, J., Boyle, J., Harris, F., Harkness, A., & Nye, C. (2000). The feasibility of universal screening for primary speech and language delay: Findings from a systematic review of the

literature. *Developmental Medicine & Child Neurology*, *42*, 190-200. doi:10.1111/j.1469-8749.2000.tb00069.x

Law, J., & Roy, P. (2008). Parental report of infant language skills: A review of the development and application of the Communicative Development Inventories. *Child & Adolescent Mental Health*, *13*(4), 198-206. doi:10.1111/j.1475-3588.2008.00503.x

Marks, K., Glascoe, F. P., Aylward, G.P., Shevell, M. I., Lipkin, P. H., & Squires, J. K. (2007). The thorny nature of predictive validity studies on screening tests for developmental-behavioral problems. *Pediatrics, 122*(4), 866-868. doi: 10.1542/peds.2007-3142

Meehl, P. E., & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin*, *52*(3), 194-216. doi:10.1037/h0048070

Miguel, P., & Mariela, R. (2007). Short forms of the Inventario do Desenvolvemento de Habilidades Comunicativas: Normative data and psychometric properties. *Infancia y Aprendizaje, 30*(4), 565-588. doi:10.1174/021037007782334292

Norbury, C. F., Nash, M., Baird, G., & Bishop, D. V. M. (2004). Using a parental checklist to identify diagnostic groups in children with communication impairment: A validation of the Children's Communication Checklist-2. *International Journal of Language & Communication Disorders, 39*(3), 345-364. doi:10.1080/13682820410001654883

O'Neill, D. (2007). The Language Use Inventory for young children: A parent-report measure of pragmatic language development for 18- to 47-month old children. *Journal of Speech, Language, and Hearing Research, 50*(1), 214-228. doi:10.1044/1092-4388(2007/017)

O'Neill, D. (2009). *Language Use Inventory: An assessment of young children's pragmatic language development for 18- to 47-month old children*. Manual. Waterloo, ON:

Knowledge in Development.

Osman, D., Shohdi, S., & Aziz, A. (2011). Pragmatic difficulties in children with Specific Language Impairment. *International Journal of Pediatric Otorhinolaryngology*, *75*(2), 171-176. doi:10.1016/j.ijporl.2010.10.028

Pérez-Pereira, M., & Resches, M. (2010). Concurrent and predictive validity of the Galician CDI. *Journal of Child Language*, 1-20. Available on Cambridge Journals Online 09 Mar 2010 doi:10.1017/S0305000909990262

Phelps-Terasaki, D., & Phelps-Gunn, T. (1992). *Test of Pragmatic Language*. East Moline, IL: Linguisystems.

Philofsky, A., Fidler, D. J., & Hepburn, S. (2007). Pragmatic language profiles of school-age children with autism spectrum disorders and Williams syndrome. *American Journal of Speech-Language Pathology*, *16*(4), 368-380. doi:10.1044/1058-0360(2007/040)

Plante, E., & Vance, R. (1994). Selection of preschool language tests: A data-based approach. *Language, Speech, and Hearing Services in Schools, 25*(1), 15-24.

Rapin, I., & Dunn, M. (2003). Update on the language disorders of individuals on the autistic spectrum. *Brain & Development*, *25*(3), 166-172. doi:10.1016/S0387-7604(02)00191-2

Redmond, S. M., Thompson, H. L., & Goldstein, S. (2011). Psycholinguistic profiling differentiates specific language impairment from typical development and from attention-deficit/hyperactivity disorder. *Journal of Speech, Language, and Hearing Research, 54*(1), 99-117. doi:10.1044/1092-4388(2010/10-0010)

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

Rescorla, L. (2002). Language and reading outcomes to age 9 in late-talking toddlers. *Journal of*

Speech, Language, and Hearing Research, *45*(2), 360-371. doi:10.1044/1092-4388(2002/028)

Rice, M. L., Taylor, C. L., & Zubrick, S. R. (2008). Language outcomes of 7-year-old children with or without a history of late language emergence at 24 months. *Journal of Speech, Language, and Hearing Research, 51*(2), 394-407. doi:10.1044/1092-4388(2008/029)

Rice, M. L., Warren, S. F., & Betz, S. K. (2005). Language symptoms of developmental language disorders: An overview of autism, Down syndrome, fragile X, specific language impairment, and Williams syndrome. *Applied Psycholinguistics*, *26*(1), 7-27. doi:10.1017/S0142716405050034

Rollins, P., & Snow, C. E. (1998). Shared attention and grammatical development in typical children and children with autism. *Journal of Child Language*, *25*(3), 653-673. doi:10.1017/S0305000998003596

Russell, R. L., & Grizzle, K. L. (2008). Assessing child and adolescent pragmatic language competencies: Toward evidence-based assessments. *Clinical Child and Family Psychology Review, 11*(1-2), 59-73. doi:10.1007/s10567-008-0032-1

Rydz, D., Srour, M., Oskoui, M., Marget, N., Shiller, M., Birnbaum, R., Majnemer, A., & Shevell, M. I. (2006). Screening for developmental delay in the setting of a community pediatric clinic: A prospective assessment of parent-report questionnaires. *Pediatrics, 118*(4), 1178-1186. doi:10.1542/peds.2006-0466

Sachse, S., & Von Suchodoletz, W. (2008). Early identification of language delay by direct language assessment or parent report? *Journal of Developmental and Behavioral Pediatrics, 29*, 34-41. doi:10.1097/DBP.0b013e318146902a

Sanderson, R., & Drew, S. (2009). *Maternal and Child Indicator Report for Waterloo Region:*

HBHC-ISCIS Report 2007. Waterloo, Ontario: Region of Waterloo Public Health.

Seymour, H., Roeper, T., & de Villiers, J.G., (2005). *Diagnostic Evaluation of Language Variation – Norm Referenced (DELV-NR*). San Antonio, TX: Harcourt Assessment.

Spaulding, T. J., Plante, E., & Farinella, K. A. (2006). Eligibility criteria for language impairment: Is the low end of normal always appropriate? *Language, Speech, and Hearing Services in Schools, 37*(1), 61-72. doi:10.1044/0161-1461(2006/007)

Statistics Canada (2007). *2006 Community Profiles*. Statistics Canada Catalogue no. 92-591-XWE Catalogue no. 92-591-XWE. Ottawa.

Streiner, D. L. (2003). Diagnosing tests: Using and misusing diagnostic and screening tests. *Journal of Personality Assessment*, *81*(3), 209–219. doi: 10.1207/S15327752JPA8103_03

Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech and Hearing Research, 40*(6), 1245-1260.

Van Agt, H. E., Van der Stege, H. A., De Ridder-Sluiter, J. G., & De Koning, H. J. (2007). Detecting language problems: Accuracy of five language screening instruments in preschool children. *Developmental Medicine & Child Neurology, 49*(2), 117-122. doi:10.1111/j.1469-8749.2007.00117.x

Volden, J., & Phillips, L. M. (2010). Measuring pragmatic language in speakers with autism spectrum disorder: Comparing the Children's Communication Checklist-2 and the Test of Pragmatic Language. *American Journal of Speech-Language Pathology, 19*, 204-212. doi:10.1044/1058-0360(2010/09-0011)

Washington, K. N. (2010). Focus on practical application of the International Classification of Functioning, Disability and Health (ICF): Using the ICF-CY in pediatric speech-language

pathology in day-to-day clinical practice. *Communiqué, 24*(2), 4, 8.

Westerlund, M., Berglund, E., & Eriksson, M. (2006). Can severely language delayed 3-year-olds be identified at 18 months? Evaluation of a screening version of the MacArthur–Bates Communicative Development Inventories. *Journal of Speech, Language, and Hearing Research, 49*, 237-247. doi:10.1044/1092-4388(2006/020)

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

World Health Organization (2007). *International classification of functioning, disability and health - Version for children and youth: ICF-CY*. Geneva: World Health Organization.

Table 1

Participants by LUI Age Group and Child Sex

	Numbers of Participants		
	Boys	Girls	Total
LUI Age Group (mean age in months)			
18-23 (*M* = 20.75)	60	52	112
24-29 (*M* = 26.26)	61	33	94
30-35 (*M* = 31.93)	39	28	67
36-41 (*M* = 38.34)	22	10	32
42-47 (*M* = 43.77)	17	26	43
Total	199	149	348

Table 2

Means on LUI Total (Standard) Score and Language Outcome Measures (Standard Scores) by LUI Age Group

		Language Outcome Measures		
	LUI Total Score	CCC-2 General Communication Composite	CELF-P2 Core Language	DELV-NR Language Composite
LUI Age Group	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
18-23 (*n* =112)[a]	100.73 (15.91)	103.39 (12.22)	105.85 (9.86)	101.61 (14.53)
24-29 (*n* = 94)[b]	100.24 (17.34)	105.09 (13.12)	105.64 (10.78)	101.62 (13.64)
30-35 (*n* = 67)[c]	97.42 (14.50)	104.12 (14.13)	103.84 (11.73)	100.94 (14.38)
36-41 (*n* = 32)[d]	97.12 (13.22)	100.62 (11.82)	104.14 (11.32)	101.35 (13.27)
42-47 (*n* = 43)[e]	100.75 (5.32)	107.99 (11.71)	105.89 (10.04)	100.75 (12.30)

Note. *M* = 100, *SD* = 15 for all measures.

[a]*n* = 111 for CELF-P2 and DELV-NR; [b]*n* = 93 for CELF-P2 and DELV-NR; [c]*n* = 65 for CELF-P2 and DELV-NR; [d]*n* = 31 for CELF-P2; [e]*n* = 42 for CELF-P2 and DELV-NR (see p. 17 regarding occasional missing data).

Table 3

Pearson Product-Moment Correlations, by LUI Age Group, of the LUI Total (Standard) Score with Language Outcome Measures (Standard Scores)

	Language Outcome Measures		
	CCC-2 General Communication Composite	CELF-P2 Core Language	DELV-NR Language Composite
LUI Age Group			
18-23 (n = 112)[a]	.059	.120	.190*
24-29 (n = 94)[b]	.289**	.342**	.334**
30-35 (n = 67)[c]	.506**	.519**	.519**
36-41 (n = 32)[d]	.580**	.411**	.366**
42-47 (n = 43)[e]	.174	.319*	.212

Note. [a]n = 111 for CELF-P2 and DELV-NR; [b]n = 93 for CELF-P2 and DELV-NR; [c]n = 65 for CELF-P2 and DELV-NR; [d]n = 31 for CELF-P2; [e]n = 42 for CELF-P2 and DELV-NR.
*$p \leq .02$, **$p \leq .01$, one-tailed. Given intercorrelations between the measures, the usual adjustment for multiple comparisons (.05/3) is overly conservative, but had the adjustment been made, the two values marked by a single asterisk would no longer be significant.

Table 4

Pearson Product-Moment Correlations, by LUI Age Group, of the LUI Total (Standard) Score with Pragmatic Subtests (Scaled Scores) of Language Outcome Measures.

	Pragmatic Subscales				
	CCC-2 Initiation	CCC-2 Scripted language	CCC-2 Context	CCC-2 Nonverbal	DELV-NR Pragmatics
LUI Age Group					
18-23 (n = 112)[a]	.053	.013	-.031	-.080	.126
24-29 (n = 94)[b]	.046	.268**	.254**	.230**	.224*
30-35 (n = 67)[c]	.368**	.271**	.342**	.347**	.429**
36-41 (n = 32)[d]	.070	.097	.413**	.487**	.262
42-47 (n = 43)[e]	-.011	-.047	.161	.244	.073

Note. [a]n = 111 for CELF-P2 and DELV-NR; [b]n = 93 for CELF-P2 and DELV-NR; [c]n = 65 for CELF-P2 and DELV-NR; [d]n = 31 for CELF-P2; [e]n = 42 for CELF-P2 and DELV-NR.
*$p \leq .02$, **$p \leq .01$, one-tailed. Given intercorrelations between the measures, the usual adjustment for multiple comparisons (.05/5) is overly conservative, but had the adjustment been made, the value marked by a single asterisk would no longer be significant.

Table 5

PPV and NPV at Different Base Rates and Cut-offs for Combined LUI Age Groups 24-47

				Base Rates					
				Screening		Clinical Sample (i.e., Referred)			
LUI Total Score (in SD)	Sensitivity	Specificity		10%	20%	50%	60%	70%	80%
-1.00	.81	.84	PPV	.363	.562	.837	.885	.923	.954
			NPV	.976	.947	.816	.747	.655	.526
-1.50	.81	.91	PPV	.506	.697	.902	.932	.956	.974
			NPV	.978	.950	.828	.762	.673	.545
-1.64	.81	.93	PPV	.563	.743	.921	.946	.964	.979
			NPV	.978	.951	.830	.765	.677	.550
-2.00	.62	.95	PPV	.594	.767	.929	.952	.968	.981
			NPV	.957	.909	.714	.625	.517	.385

Note. Grey shading represents the cut-off that maximized sensitivity and specificity values while maintaining each at a level ≥ .70 (criterion level). Other cut-off values and data are provided to allow comparisons with other studies.

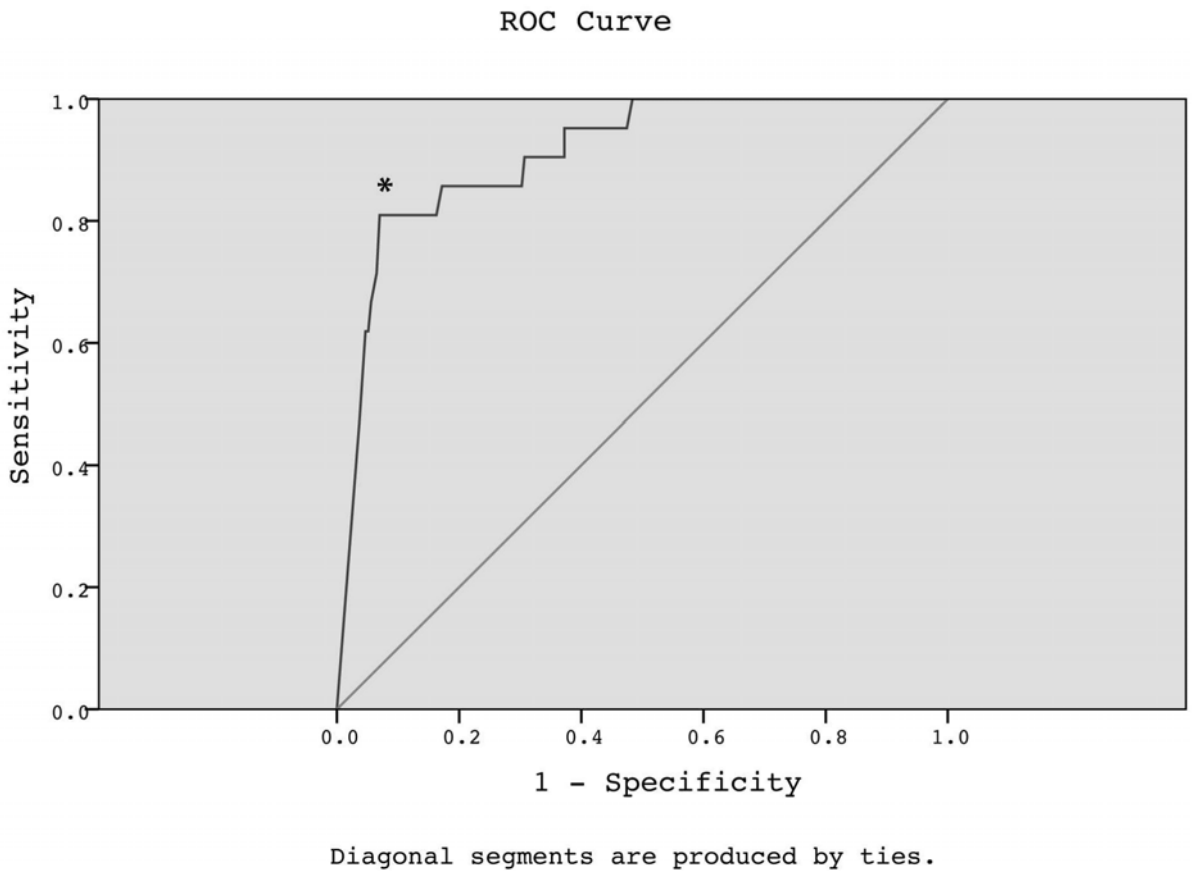

Figure 1. ROC curve illustrating classification accuracy of the LUI and .910 area under the curve.

Note. The area under the curve under the null hypothesis (.5) is indicated by the diagonal line. The asterisk indicates sensitivity at .81 and specificity at .93 (1-specificity on the x axis), observable with a LUI cut-off at -1.64 SD (5th percentile).