Typical Performance on Tests of Language Knowledge and Language Processing of French-Speaking 5-Year-Olds

Performance type lors d'examens de connaissances et de traitement du langage chez les enfants francophones de cinq ans

Elin Thordardottir
Eva Keheyia
Nicole Lessard
Ann Sutton
Natacha Trudeau

Abstract

The evaluation of the language skills of francophone children for clinical and research purposes is complicated by a lack of appropriate norm-referenced assessment tools. The purpose of this study was the collection of normative data for measures assessing major areas of language for 5-year-old monolingual speakers of Quebec French. Children in three age-groups (4;6, 5;0 and 5;6 years, n=78) were administered tests of language knowledge and linguistic processing, addressing vocabulary, morphosyntax, syntax, narrative structure, nonword repetition, sentence imitation, rapid automatized naming, following directions, and short term memory. The assessment measures were drawn from existing tools and from tools developed for this study, and included formal tests as well as spontaneous language measures. Normative data are presented for the three age groups. Results showed a systematic increase with age for most of the measures. Correlational analysis revealed relationships of varying strength between the measures, indicating some overlap between the measures, but also suggesting that the measures differ in the linguistic skills they tap into. The normative data presented will facilitate the language assessment of French-speaking 5-year-olds, permitting their performance to be compared to the normal range of typically developing monolingual French-speaking children and allowing the documentation of children's profiles of relative strengths and weaknesses within language.

Abrégé

L'évaluation des capacités langagières des enfants francophones à des fins clinique et de recherche est compliquée en raison du manque d'outils d'évaluation normalisés adéquats. Le but de cette étude était de recueillir des données normatives pour différentes mesures qui évaluent les principaux aspects du langage chez les enfants de cinq ans unilingues francophones québécois. Des enfants de trois groupes d'âge (4;6, 5;0 et 5;6, n=78) ont passé des examens sur les connaissances et le traitement du langage concernant : le vocabulaire, la morphosyntaxe, la syntaxe, la structure narrative, la répétition de non-mots, l'imitation de phrase, la dénomination rapide automatisée, l'application de consignes et la mémoire à court terme. Les mesures d'évaluation ont été élaborées à partir d'outils existants et d'outils créés pour la présente étude. Elles étaient composées de tâches formelles ainsi que de mesures du langage spontané. Les données normatives sont présentées pour les trois groupes d'âge. Les résultats de la plupart des mesures ont montré une amélioration systématique avec l'âge. L'analyse corrélationnelle a révélé des relations de forces variées entre les mesures, indiquant un certain chevauchement entre certaines d'entre elles, mais suggérant aussi que les mesures varient en fonction des capacités langagières mises à profit. Les données normatives présentées faciliteront l'évaluation du langage chez les enfants francophones de cinq ans. Elles permettront de comparer leur performance à celles des enfants unilingues francophones dont le développement est dans la norme et de documenter le profil des enfants quant à leurs forces et faiblesses relatives au langage.

Elin Thordardottir, Ph.D.
McGill University
Montreal, Quebec
Centre for Research on Language, Mind and Brain (CRLMB)
Centre de recherche interdisciplinaire en réadaptation du Montréal métropolitain (CRIR)

Eva Keheyia, Ph.D.
McGill University
Montreal, Quebec
Centre de recherche interdisciplinaire en réadaptation du Montréal métropolitain (CRIR)

Nicole Lessard, Ph.D.
Centre de réadaptation Marie-Enfant
Montreal, Quebec

Ann Sutton, Ph.D.
Université de Montréal
Montreal, Quebec

Natacha Trudeau., Ph.D.
Université de Montréal
Montreal, Quebec

Key Words: Assessment, evaluation, norms, vocabulary, morphology, syntax, narration, nonword repetition, processing, French

The language assessment of francophone children in Quebec is complicated by the lack of language measures developed and appropriately norm-referenced for this population. This problem is all too well known to Speech-Language Pathologists working with francophone populations, as well as to researchers. Several authors have written about the severity of the situation (e.g. Garcia, Paradis, Sénécal & Laroche, 2006). Given the relative paucity of French language measures, clinical decisions must frequently be made on the basis of subjective criteria, informal tests, as well as translations of formal tests, with results either interpreted informally or reported to normative data collected in another language. A few language tests have been available for French-speaking children in Quebec, such as the *Échelle de vocabulaire en images Peabody* (*EVIP*, Dunn, Thériault-Whalen, & Dunn, 1993), and an adaptation of the Test for Auditory Comprehension of Language (Carrow-Woolfolk, 1985), with norms collected for a limited age group (Groupe Coopératif en orthophonie, 1999). More tests and assessment procedures are currently being developed. For example, the Quebec version of the MacArthur-Bates CDI (Trudeau, Frank, & Poulin-Dubois, 1999; Boudreault, Cabirol, Trudeau, Poulin-Dubois, & Sutton, 2007) is an important addition to the tests available for preschool children, and a Quebec French adaptation of spontaneous language sample analysis using the clinical SALT software permits norm-referenced analysis of utterance length, as well as lexical and morphosyntactic development (Elin Thordardottir, 2005). Norms for preschool-age children have been collected for a Quebec French version of the Reynell Developmental Language Scales (Boucher, Lavoie & Bergeron, 2004; Reynell & Gruber, 1990), and normative data for phonological development of monolingual French-speaking preschool children have recently been reported (MacLeod, Sutton, Trudeau & Thordardottir (under review)). Further, tools are being developed for bilingual speakers of French (Elin Thordardottir, 2008a). However, even with these developments, there is a serious lack of norm-referenced language tests for francophone children in Quebec and more tests are urgently needed.

The lack of appropriate language measures not only makes clinical assessment of individual children difficult, but also affects the way that language impairment is defined and conceptualized. Without clear, reliable language measures, it is hard (or impossible) to define language impairment in terms of specific linguistic skills that must be found to be lacking for an impairment to be formally identified. As well, the lack of norms (documentation of the mean performance and normal variability) precludes the setting of firm cut-off criteria that separate performance within and outside the normal range, or that indicate levels of severity of the impairment. The lack of measures also makes it difficult to accurately establish a profile of areas of relative strengths and weaknesses for individual children, which is an important part of selecting and prioritizing clinical goals. Finally, without adequate measures, clinicians are limited in the extent to which they can objectively track children's performance over time to document treatment gains or to monitor the development of children considered at risk.

The present study was undertaken to provide language assessment tools for French-speaking children aged five. This is the age at which many children who experience slow language development are first seen for a formal evaluation in Quebec. A series of measures were selected for this preliminary norming effort. Our intention was to cover a range of language skills to enable a comprehensive assessment of the major domains of language which will permit the establishment of a profile of strengths and weaknesses for individual children and to examine the relationships between areas of language in the development of French. We included assessment of both receptive and productive skills in lexical, syntactic, and morphological development. We also included assessment of narrative skills and language processing measures. The selection of areas to include was guided by previous research on the clinical identification of language impairment (Conti-Ramsden, Crutchley & Botting, 1997; Dollaghan & Campbell, 1998; Ellis Weismer et al., 2000; Tomblin, Records & Zhang, 1996). Research focusing primarily on English has supported the inclusion of a range of language measures for a comprehensive language assessment, but has also identified certain measures particularly accurate clinical markers indicating the presence of language impairment. The existence of normative data on a range of measures will ultimately permit the evaluation of the usefulness of different measures in French. Tomblin et al. (1996) developed an assessment system known as epi-SLI. In that research, the initial choice of language measures was motivated by clinical practice and expert opinion and was designed to cover the major areas of language. Tomblin et al. included a narrative task because of the link between narrative skills and reading and academic achievement. Our test selection is also similar to that used by Conti-Ramsden, Crutchley & Botting (1997). In addition to the areas covered by Tomblin et al., we also use tests of verbal processing in light of the strong relationship that has now been demonstrated between such measures and language test scores, as well as recent findings demonstrating the diagnostic utility of such processing measures in English and in other languages (Bishop et al., 1999; Conti-Ramsden, 2003; Conti-Ramsden, Botting & Faragher, 2001; Dollaghan & Campbell, 1998; Ellis Weismer et al., 2000; Elin Thordardottir, 2008b; Girbau & Schwartz, 2007; Sahlén, Reuterskjöld-Wagner, Nettelbladt & Radeborg, 1999).

It was important to us to assess language skills both with formal tests, which have the advantage of targeting specific language structures, as well as with samples of spontaneous language, which give a less targeted measurement, but by the same token assess the child in a relaxed everyday setting, providing a measure of greater ecological validity. Spontaneous language samples were collected

for this purpose in a conversational context. Analysis of spontaneous language yielding the measures of mean length of utterance (MLU) in words and in morphemes is based on the work of Roger Brown (1973). MLU has been among the most widely used yardsticks of language development. Its use as such is based on the observation that progress in early language development results in lengthening of the utterance. Children progress on one hand from using predominantly one-word utterances to using two-word and three-word utterances etc. Further, if the use of grammatical morphemes is also viewed as additions to the length of the utterance, it becomes possible to track the combination of word use and use of the associated grammatical morphology as increased utterance length. French is considerably richer than English in grammatical morphology. MLU in words and morphemes has been shown to be a developmentally sensitive measure in French (Elin Thordardottir, 2005). The assessment of narrative skills targets a different aspect of spontaneous language than conversational language sampling. Narratives are more structured at the discourse level than conversations, requiring planning and organization of a larger set of utterances. Narratives are increasingly included in clinical assessments to evaluate discourse skills and because of the documented links between narrative skills and academic and reading achievement (Tomblin et al., 1996; Griffin, Hemphill, Camp, & Palmer, 2004).

The usefulness of verbal processing measures has been strongly supported by research and they are becoming increasingly common in clinical assessment protocols. The results of Dollaghan and Campbell (1998) and those of Ellis Weismer et al. (2000) and Conti-Ramsden (2003) support the utility of nonword repetition scores in the correct identification of specific language impairment, as part of a larger diagnostic protocol (see however Archibald & Joanisse (2009) who find nonword repetition to have good sensitivity but fairly low specificity in the identification of specific language impairment). Nonword repetition tests have the advantage of rapid administration and scoring and there are indications that they are relatively immune to dialectal differences (Dollaghan & Campbell, 1998). However, they do not provide accurate results for children who have significant articulation deficits. Similarly, a test as simple as Sentence Imitation has been shown to add significantly to the diagnostic protocol (Conti-Ramsden, Botting, & Faragher, 2001; Archibald & Joanisse, 2009). Additional processing tests included Following Directions, which requires children to comprehend and remember increasingly long and complex instructions, and demonstrate understanding by pointing to pictures or manipulating items. This test emphasized the processing or working memory part of this task, using easily comprehensible vocabulary. Short-term memory and working memory were assessed further by the use of tests of Forward and Backwards Digit Span. A final measure included in our protocol was Rapid Automatized Naming of animals (RAN). This task was included as an assessment of prerequisite skills for reading, given its strong relationship with subsequent word recognition skills (Catts, 1993). At age 5, most of the children in the study are too young to be tested on reading achievement.

The purpose of the present study was to collect preliminary normative data for the language measures described above for monolingual francophone children aged 5. In order to get a sense of developmental variation around age 5, children aged 6 months younger and 6 months older were included in addition to children aged 5. Our main interest here is to report the normative values for these measures, and discuss their developmental sensitivity in this age range.

Methods

Participants

Participants included 78 monolingual French-speaking children from the Montreal area ranging in age from 49 months to 71 months, with a mean and median age of 59 months. The children were divided into three age groups representing 4 ½ year-olds (49 to 56 months inclusive), 5-year-olds (57 to 63 months inclusive) and 5 ½ year-olds (64 to 71 months inclusive), yielding groups with the following mean ages: 52.41 (SD 1.78), 60.15 (SD 2.23) and 66.2 (SD 2.32). The children were recruited from daycares and preschools in the Montreal area by invitation letter sent home with the children by the daycare staff. Sixty-five of the children were recruited and tested as part of this study. Twelve additional children were tested within other studies conducted by the same team of researchers, using the same methods (Sutton et al., 2009). For this reason, however, some of the measures are not available for all the children (see Table 2). All children were from monolingual French-speaking homes and had no developmental concerns, pre-or perinatal complications, major illnesses or hospitalizations as per parent report (a history of otitis media was not a exclusionary criterion in this study given its relatively high frequency). The children had had no significant regular exposure to languages other than French (defined as less than five hours per week). Nonverbal cognitive development was tested using the Leiter International Performance Scale-Revised Brief IQ (Roid & Miller, 1997). This is a test whose stimuli and administration are entirely nonverbal. The brief IQ measure is based on four subtests: Figure Ground, Form Completion, Sequential Order, and Repeated Patterns. To be included in the study, children were required to obtain a standard score of 70 or higher on this test in order for the sample to represent the normal variation of cognitive levels while excluding children with scores consistent with mental retardation.

A hearing screening was conducted under earphones using a portable audiometer on the day of the test at octave frequencies from 500 to 4000 Hz, to ensure that no significant hearing loss was present. Children passed a screening at 10 dB HL, with some exceptions involving somewhat higher thresholds at individual frequencies. A reliable result could not be obtained at 10 dB HL in many cases due to background noise as the test was not conducted in a sound-proof booth. Children were not tested if their hearing

Table 1
Background Characteristics: Mean and (standard deviations) for age, nonverbal cognition scores and maternal education by age group.

Age group:	4 ½ years	5 years	5 ½ years
Number of girls	14	16	11
Number of boys	13	16	8
Age in months*	52.4 (1.8)	60.2 (2.2)	66.2 (2.3)
Nonverbal cognition (Leiter Brief IQ)	102.4 (14.8)	101.3 (19.1)	102.3 (19.3)
Maternal education (Number of years)	15.7 (2.8)	17.4 (2.9)	16.9 (2.4)

* A significant group difference was found for age

was considered to be impaired based on the screening. Background variables including age, nonverbal cognition and maternal education are reported in Table 1. One-way ANOVA analysis revealed a significant group difference in age ($F(2,74)=240.552$, $p=.000$), but no significant group difference in maternal education ($p=.074$) or nonverbal cognition (.978).

Procedures

Children were tested individually by trained research assistants who were native speakers of Quebec French. Each child attended one test session lasting approximately 2 to 2 1/2 hours, which was conducted either at their school, at their daycare, or in a research laboratory at McGill University or University of Montreal. Breaks were provided as needed. All testing was audio- and video-recorded. The particular measures used are described below. These include standardized tests that had previously been adapted from English, test procedures adapted from English for the purpose of this study, and procedures developed by members of the research team previously or for this study. It was beyond the scope of this study to develop new test materials in all the areas to be assessed for this preliminary look at the language profile of Montreal francophone children. Several test translations/adaptations have been in wide clinical use in Quebec, with or without normative data. Insofar as such tests 1) are based on well established English tests, 2) include careful adaptation going beyond simple translation, and 3) are supported by clinical experience with francophone populations in Quebec, they were considered potential candidates for inclusion in this study. In other cases, materials were developed for this study. A major goal of the study was to examine the usefulness of these materials. During the test session, children were administered the following measures (the order of tests varied across children):

1. The *Échelle de vocabulaire en images Peabody* (EVIP, Dunn, Thériault-Whalen, & Dunn, 1993) was used to assess receptive vocabulary (comprehension of single words in isolation). This test was originally developed in English, but was adapted for Canadian speakers of French and normed for this group. However, it has been shown that the published norms underestimate the typical skills of Quebec francophone children (Godard & Labelle, 1995), most likely because the norms included French-speaking children who are monolingual speakers and also bilingual French-speakers from across Canada. We considered the EVIP to be a well established test of receptive vocabulary, however, needing to be reevaluated for the appropriateness of its norms for Quebec. In addition to its wide clinical use, the EVIP is frequently used in research for purposes of participant selection and matching. The EVIP was administered and scored as directed in the manual.

2. The Carrow (Groupe coopératif, unpublished Quebec French adaptation of the Test for Auditory Comprehension of Language [TACL-R]; Carrow-Woolfolk, 1985) was used to assess receptive language skills (vocabulary, morphosyntax and syntax). The Carrow is a Quebec French translation and adaptation, a well-known and widely used test both in clinical practice and research. The adaptation of this test to French involved translation and reordering of items to better reflect the development of French within a previous norming study (Groupe coopératif, 1999). The Carrow has enjoyed considerable clinical popularity in Quebec and the available norms for Kindergarten (maternelle) children indicate that it is sensitive to the development of French within the age range of this study. This test has three subtests: *Classes de mots et relations* [Vocabulary], *Morphèmes grammaticaux* [Grammatical Morphemes], and *Phrases complexes* [Elaborated Phrases and Sentences]. The test was administered and scored using the test materials of the TACL-R (with items reordered as directed in the adaptation) and scored as directed in the TACL-R manual.

3. A spontaneous language sample was collected in a free-play conversational context based on the guidelines of Leadholm & Miller (1992) using age-appropriate toys (such as playmobil hospital and school, and polly pockets). The language samples were transcribed orthographically using the SALT software (Systematic Analysis of Language Transcripts, Miller & Chapman, 1984-2002). Coding of grammatical morphology followed the Quebec-French adaptation of SALT conventions developed by Elin Thordardottir (2005). MLU computed following this procedure has been shown to be developmentally sensitive in French with MLU in morphemes being higher for

French-speaking children than it is for English-speaking age mates, reflecting the greater frequency of grammatical morphemes in French. The measures reported are based on 100 utterances (excluding imitations) and include Mean Length of Utterance in words (MLUw) and in morphemes (MLUm). Those language samples that were not collected as part of this study were rechecked to ensure that the same grammatical coding procedure was applied uniformly to all the samples and that the measures were based on the same transcript cut. Reliability checks were performed by an independent scorer on a randomly selected subset of 23 language samples (29% of the samples). Interrater agreement for MLUw was 96.2 % (SD 3.05) and for MLUm 95.4 (SD 4.2).

4. The Edmonton Narrative Norms Instrument (ENNI), developed by Schneider, Dubé and Hayward (2002-2006), was used in a French adaptation developed within this study (Thordardottir & Gagné, 2006; Gagné & Thordardottir, 2006). A number of measures have been used clinically and in research to assess narratives. Just as with any assessment of spontaneous language, the elicitation context influences the child's productions. Elicitation methods for narratives vary notably in whether they use picture support and in whether the children formulate a story or retell a story. The ENNI, developed recently in Canada, comprises a set of wordless picture stories that are carefully constructed to increase in length and in the complexity of the story grammar. Story grammar refers to the main categories of information that are part of a good story and form the macrostructure of the story (Schneider, Dubé, & Hayward, 2002-2006). A measure of narrative microstructure, First Mentions, is also derived from this instrument. This measure assesses the children's use of referring expressions to introduce characters and objects as they tell the story. In the elicitation procedure, the children are presented novel stories in pictures, but are not told the stories. The procedure, therefore, targets children's ability to formulate the story rather than their ability to retell a story. Schneider et al. have demonstrated that this instrument is sensitive to development with English-speaking children aged 4 to 9 and that significant differences are found on this test between children with and without language impairment. In the ENNI norming study (Schneider et al.), children were administered two sets of stories: A and B. For this study, administration of both stories would have been too time-consuming given the larger protocol of tests. Initially, children were administered either set A or B, counterbalanced across children. However, it became apparent that children responded less well to set B, indicating that the two sets were not equivalent in the French application. In subsequent testing, story set A was administered to all children. This change in procedure accounts for missing data in the ENNI dataset, as data from those children administered stories B were not included.

Each story was administered in the following manner: the child was first shown the entire story such that the examiner does not see them. The child was then shown the pictures again and asked to tell the story to the experimenter. The child was asked/prompted to say what happened in the story rather than to describe the pictures. The stories were transcribed orthographically and scored in terms of Story Grammar Complexity and First Mentions. This scoring followed the ENNI manual (Schneider et al., 2002-2006), using a French adaptation of the scoring forms developed for this study. The Story Grammar Score gives an indication of the child's ability to include story grammar elements, while First Mentions addresses whether the child provides sufficient information on first mention, subsequently using pronouns to refer to previously presented entities. The full scoring of First Mentions as described in the ENNI manual (Schneider et al., 2002-2006) uses information from stories A and B. Our version of First Mentions is based only on the A set. In the present analysis, only these two scores are reported (additional scores that can be derived include MLU and production of grammatical morphology and complex syntax – however, MLU was obtained from spontaneous samples and the main interest in the ENNI here was in narrative assessment).

5. Nonword repetition, comprising a set of 40 words ranging in length from 1 to 5 syllables was derived from word lists developed expressly for Quebec French by Courcy (2000). The nonwords were administered from a tape and were scored online with subsequent rechecking of the recording of the responses. Scoring followed the procedure of Dollaghan & Campbell (1998), whereby children are credited for the number of correct phonemes produced. Credit is given for phonemes that are produced in a distorted form, and no points are subtracted for phonemes that are added. Children do not receive credit for phonemes that are missing or that are substituted, or for phonemes that appear in the wrong order relative to other phonemes in the nonword.

6. Sentence imitation was tested using an adaptation of the sentence imitation subtest of the CELF-P (Clinical Evaluation of Language Fundmentals-Preschool, Semel, Wiig, & Secord, 1992). This subtest involves a story book entitled *The Moving Scene*, wherein children are told a story as they look at the pictures. The story is then retold by the examiner, and the child is asked to repeat selected sentences. The French adaption of this subtest, *Le grand déménagement,* was developed by Royle & Thordardottir (2003). The adaptation involved a translation of the story and modification of the scoring guidelines. This particular sentence imitation task was selected because the sentences are embedded in a story with picture support, which is appropriate for an imitation task for children in the age range of this study in that it aids their memory and increases their motivation. The appropriateness of this particular story is supported by its long-standing clinical use. Instead of sentences receiving scores ranging from 0 to 3, the adapted scoring reflects the percentage of words repeated, which gives a more finely graded results scale of partial credit for partially repeated sentences.

7. Rapid Automatized Naming (RAN) (Catts, 1993) was tested by presenting a sheet with pictures of a horse (*cheval*), a cow (*vache*) and a pig (*cochon*) in black (*noir*), red (*rouge*) or blue (*bleu*), arranged in random order in 4 rows of 6 animals each for a total of 24 animals. The children were asked to name each picture, naming the animal and its color as fast as they could, going from left to right, and through each row from top to bottom. Before the test, the examiner ascertained that the children recognized the animals and knew their names, as well as the colors. Two scores were derived from this test: the number of errors made by the child, and the time in seconds required for completion (starting from when the child named the first animal and ending when the child had named the last animal), thus addressing both accuracy and speed. In counting the number of errors, only the child's last production for each animal was considered (thus, the child was allowed to self-correct). Unlike in English, the color term comes after the animal name (e.g. *cheval noir* [black horse]). No error was counted where the child named the color and animal in the wrong order.

8. The children's ability to follow auditory directions was assessed using an unpublished preliminary French adaptation of the Following Directions subtest of the Clinical Evaluation of Language Fundamentals-4 (CELF-4, Semel, Wiig, & Secord, 2003) developed by Boulianne & Labelle (2006). This subtest requires children to carry out directions by pointing to picture stimuli. For example, they might be asked to point to item A and then to item B, or to point to item B before they point to item A. Testing started at item 1 for each child and was discontinued after 9 consecutive errors.

9. Digit span was administered using an unpublished preliminary French adaptation of the Forward and Backwards Digit Span subtests of the CELF-4 (Clinical Evaluation of Language Fundamentals-4, Semel, Wiig, & Secord, 2004) developed by Boulianne & Labelle (2006). The two subtests were administered according to the instructions of the CELF-4 manual.

Detailed administration procedures as well as materials and score sheets for those procedures that were developed for this study may be obtained by contacting the first author.

Results

Means and standard deviations for each of the measures are reported for each age group of children in Table 2. EVIP scores are reported as raw scores and as standard scores, referenced to the published norms (Dunn et al., 1993). The scores for the Carrow/TACL-R and the CELF-4 subtests (Following Directions, Forward Digit Span and Backwards Digit Span) are raw scores. The results of Nonword Repetition and Sentence Imitation are reported as percent correct scores. RAN accuracy scores are reported in terms of the number of errors (out of 24 animals to be named) and in terms of the time required to complete the task in seconds. Mean Length of Utterance in words (MLUw) reflects the average number of words found in each child utterance in the sample, whereas MLU in morphemes (MLUm) reflects the average number of words as well as grammatical morphemes in each utterance. In addition to means and standard deviations, Table 2 reports the number of children for whom data were available for each measure. Smaller sample sizes are seen notably for the youngest age group for Nonword Repetition, Sentence Imitation, Following Directions and Digit Span. This is due to the fact that 12 children in this group came from another study in which these measures were not administered. Other variations in numbers are more minor and are due to children not finishing tasks or experimenter mistakes. Missing data for the ENNI result from the exclusion of story B data, as discussed in the Methods section.

Inspection of Table 2 reveals that the measures in general increase systematically with age, with the exception of RAN time, which decreases with age (reflecting increased speed in completing the task with age). Given that the study covers an age range of less than 2 years (49 to 71 months), some developmental change would be expected in the measures, but this change might not be large for some of the measures, and might vary between measures, reflecting differences in the rate of development of different linguistic skills within this time period. The systematic changes with age suggest that the measures are developmentally sensitive, but the numbers also indicate that some change more in this period than others. A series of one-way ANOVAs were used to examine the effect of age group on mean scores for each of the measures. Thirteen ANOVAs in total were conducted with the critical alpha level set at .05/13, or .004 as a result of a Bonferroni correction for multiple comparisons (for the EVIP, only raw scores were included given that normalized scores are expected to be more stable across age and for the Carrow, only the total score was included to limit the number of comparisons). Significant group effects were further examined with Fischer LSD post hoc comparisons. A significant effect of age group was found for the following measures: EVIP raw score ($F(2,77)=15.86$, $p<.001$, $\eta^2 = .29$), MLUw ($F(2,74)=13.12$, $p<.001$, $\eta^2 = .27$), MLUm ($F(2,74)=12.02$, $p<.001$,), $\eta^2 = .25$), ENNI First Mentions ($F(2,68)=8.13$, $p=.001$, $\eta^2 = .20$), ENNI Story Grammar ($F(2,67)=13.20$, $p<.001$, $\eta^2 = .29$), and RAN Time ($F(2,70)=6.63$, $p=.002$, $\eta^2 = .16$). Post hoc tests revealed that for MLUw, ENNI First Mentions and ENNI Story Grammar, the means of all three age groups differed from each other. For the EVIP and RAN time, the youngest group differed from each of the older groups, but the middle and oldest groups were not significantly different from each other. For MLUm, both the youngest and middle groups differed significantly from the oldest group, but the youngest and middle groups did not differ significantly from each other

The relationship between different measures of language and verbal processing was examined by correlational analysis. With the number of measures administered in this study, inclusion of all the measures in

Table 2
Results of language measures by age group, reported as means and (standard deviations). P values are reported for the age group comparison for each measure.

Age group:	4 ½ years	5 years	5 ½ years	p
EVIP raw	54.6 (16.3)	69.7 (15.4)	78.2 (13.6)	<.001
n	*27*	*32*	*19*	
EVIP ss	111.4 (18.0)	116.2 (17.7)	22.4(13.5)	.101
n	*27*	*32*	*19*	
Carrow total	74.8 (17.0)	82.9 (13.9)	90.2 (11.9)	.006
n	*25*	*29*	*16*	
1. Classes de mots	30.2 (6.1)	32.4 (2.9)	33.9 (2.7)	
2. Morphèmes	24.5 (5.5)	26.7 (6.1)	27.3 (7.4)	
3. Phrases complexes	19.6 (8.5)	23.7 (6.3)	27.4 (4.7)	
MLUw	3.97 (1.1)	4.72 (1.1)	5.89 (1.7)	<.001
n	*27*	*30*	*18*	
MLUm	5.15 (1.3)	5.90 (1.4)	7.59 (2.3)	<.001
n	*27*	*30*	*18*	
ENNI SG	13.7 (4.9)	16.9 (5.3)	21.9 (5.9)	<.001
n	*27*	*24*	*17*	
ENNI FM	13.1 (3.7)	16.9 (5.3)	21.9 (5.4)	.001
n	*27*	*25*	*17*	
Nonword Rep	89.4 (7.5)	86.9 (9.5)	91.8 (4.0)	.149
n	*13*	*27*	*16*	
Sent. Imit.	74.2 (29.2)	76.3 (16.7)	88.7 (10.0)	.057
n	*14*	*30*	*17*	
RAN error	1.0 (2.0)	0.97 (1.3)	1.2 (1.9)	,901
n	*25*	*31*	*15*	
RAN time	112.4 (44.0)	85.4 (24.4)	79.6 (21.8)	.002
n	*25*	*31*	*15*	
Foll. Dir.	24.6 (8.6)	27.3 (9.2)	32.3 (13.3)	.166
n	*11*	*25*	*14*	
F digit span	5.0 (1.6)	5.5 (1.7)	6.7 (2.3)	.053
n	*13*	*32*	*14*	
B digit span	0.7 (1.1)	1.1 (1.4)	1.9 (1.2)	.040
n	*13*	*32*	*15*	

this analysis would have resulted in a prohibitive number of correlations. Therefore, six key measures were selected for this analysis: EVIP raw score, Carrow total score, MLUm, ENNI Story Grammar, Nonword Repetition and Sentence Imitation. These were selected because they address major domains of language skills considered part of a comprehensive evaluation of language (receptive vocabulary, receptive morphosyntax and syntax, syntactic production, and production of narrative discourse structure). As well, Nonword Repetition and Sentence Imitation are increasingly part of clinical evaluation protocols. The correlations between these measures are presented in Table 3. With a total of 15 correlations, the critical alpha level was set at .0038, applying a Bonferroni correction. These correlations need to be interpreted with some caution due to the fact that they are not all based on an equal '*n*' due to missing data, notably for Nonword Repetition and Sentence Imitation. The Table reveals significant correlations ranging from weak to strong between the EVIP and all the other measures except Nonword Repetition and MLUm. The strongest correlation for the EVIP is with the Carrow (r=.714). The Carrow is moderately to strongly correlated with all measures except MLUm. ENNI Story Grammar is significantly moderately correlated with the EVIP and the Carrow. In contrast, MLUm is not significantly correlated with any of the other measures. The two processing measures, Nonword Repetition and Sentence Imitation are correlated with each other and with the Carrow. The correlation with the Carrow is stronger for Sentence Imitation than Nonword Repetition (r=.741 vs. .456). Sentence Imitation is, in addition, moderately correlated with the EVIP, whereas Nonword Repetition is not.

Table 3
Correlations between language measures: EVIP, MLUm, Carrow, Story Grammar, Nonword Repetition and Sentence Imitation

	EVIP	MLUm	Carrow	SG	NWR	SI
EVIP	-					
MLUm	.319 .005	-				
Carrow	.714 .000*	.216 .074	-			
SG	.594 .000*	.181 .139	.526 .000*	-		
NWR	.314 .018	.091 .504	.456 .001*	.260 .066	-	
SI	.547 .000*	.188 .148	.741 .000*	.370 .005	.595 .000*	-

* correlation significant at the .003 level

EVIP: raw score
MLUm: MLU in morphemes
Carrow: total scores
SG: ENNI story grammar score
NWR: Nonword repetition score
SI: Sentence Imitation

Discussion

This article presents preliminary normative data on a number of measures of language knowledge and language processing for francophone children from the Montreal area between the ages of 4;6 and 5;6. The results are preliminary in that they are based on a relatively small sample of children and cover a fairly limited age range. In addition, the children are all from the Montreal area and therefore, these normative data cannot be assumed to reflect all francophone children in the province of Quebec. For some of these measures, larger datasets targeting this age range or larger age ranges are being collected and analyzed (Elin Thordardottir et al., 2009; Sutton et al., 2009), including children from Quebec city and rural areas. In spite of their preliminary nature, however, these results do represent a useful reference base for the evaluation of language skills in francophone children within this age range for both clinical and research purposes. The age range covered here corresponds to the age at which many children receive their first formal evaluation and in which language impairment is first identified. Some of the data reported here are for tests that are already in clinical use. All of the tests are easy to use clinically and are accessible to clinicians. The existence of normative data for these measures will permit clinicians to compare the scores of individual children to the typical performance of French-speaking children in Quebec and thus to ascertain whether the performance of the child falls within the normal range or significantly below it. The tables provided in the Appendix are designed to facilitate this comparison.

To participate in this study, children had to be monolingual speakers of French and had to have had no serious developmental concerns or major illnesses. They also had to receive nonverbal cognitive scores within the normal range. The normal range was fairly broadly defined, as encompassing children who score as much as 2 SD below the mean. This was done so as to allow within the sample a typical variability of cognitive levels, rather than using an overly restrictive criterion. The resulting background characteristics indicate that the sample does indeed reflect considerable variability in this respect: The mean cognitive score is right around 100 in each age group, with a standard deviation of 14 to 20 points, which is similar to the norms for this test. Therefore, the results for the participants of this study closely resemble the normative sample for this test and can be seen as providing a good representation of the normal variability of cognitive levels in the population.

The pattern of scores for each of the language measures shows a systematic increase with age, suggesting that each is sensitive to developmental increases in skill. However, the developmental increase was larger for some measures than others, as reflected in that most, but not all of the tests showed a significant effect of age group, and by the fact that some showed significant effects between all adjacent age groups, whereas others did not. This is not unexpected. Different language skills show periods of more rapid vs. slower growth at different ages. Given that this study covers a fairly small age range, the lack of an age effect does not mean that the measure in question would not show developmental changes given a larger age range. Among the measures that had a significant age effect were the EVIP raw score, MLUw and MLUm in spontaneous language, both ENNI measures: Story

Grammar and First Mentions, and RAN time. The strong age progression seen in MLU (with significant difference between all groups for MLUw) is noteworthy given that MLU is regarded by many as having limited developmental sensitivity beyond a certain level, expressed in terms of an age level (age 5) or an MLU level (often considered to be an MLUm of 3.0). Indeed, several studies have reported MLU to increase linearly up to a certain point, defined in terms of age or MLU level, with a subsequent leveling off of the developmental increase (Miller & Chapman, 1981; Rondal, Ghiotto, Bredart, & Bachelet, 1987; Scarborough, Wyckoff, & Davidson, 1986). The present data clearly indicate that MLU in French, both as measured in words and morphemes, continues to show an important systematic increase with age, at least up to the age of 5;6 and an MLU level of 5.9 words and 7.6 morphemes. In fact, the increase between 5;0 and 5;6 is significant for both MLUw and MLUm, suggesting that the developmental curve does not yet show signs of leveling off. This suggests that MLU is a useful measure for French-speaking children in this age range. It should also be noted that in spite of previous results showing nonlinearities in MLU development fairly early in development as noted above, MLU norms for English have been collected up to age 13 (Leadholm & Miller, 1992; Miller & Chapman, 1984-2002), showing developmental increases throughout this age range, especially for narrative samples which require the use of more complex syntax than does conversation. Further, significant group differences in MLU have been documented between groups of English-speaking school-age children with and without specific language impairment (Elin Thordardottir, 2008b), suggesting that even in English, MLU should not be too easily discarded for older preschoolers and school-aged children. Given that French is more highly inflected than English, producing a considerably higher MLU within a given age group (Elin Thordardottir, 2005), and the finding that the acquisition of grammatical morphology is more protracted in French than in English, MLUm in French is likely to show developmental changes longer than it does for English. Continued increase in MLU throughout the school age range has been demonstrated also for Icelandic, another language considerably more highly inflected than English (Elin Thordardottir, 2008b).

The ENNI was another spontaneous measure showing a significant age effect. This indicates that the discourse formulation skills tapped by this instrument are developing at a relatively rapid rate within this age range. We did not expect that this measure would yield identical scores in our French version as it does in the original English version. Story Grammar addresses higher level story organization and inclusion of information, which is a skill relatively independent of language structure as such, and therefore might be expected to yield similar scores across languages. However, cultural variation is also known in story telling style, and it is possible that the picture materials interact in a different way with cultural traditions and knowledge. The ENNI norms are given for children age 4 to 9 at 1-year intervals (Schneider et al., 2002-2006). The closest comparisons are between our 5 ½ year-olds, who compare closely in age to the normative group of 5-year-olds and between our 4 ½ year-olds and the ENNI 4 year-olds. These comparisons of our data with the English ENNI data show that our 5 ½ year-olds obtain a mean Story Grammar score for story A3 of 21.9 (SD 5.9) compared to 21.25 (SD 4.97) in the ENNI norms. The corresponding comparison for the 4-year-olds is 13.7 (SD 4.9) versus 17.06 (SD 6.45). The French and English results, therefore, appear to agree closely for the 5-year olds, but somewhat less well for the 4-year-olds. This provides preliminary evidence that the tests are fairly comparable in the two languages in terms of Story Grammar scores for Story A. Further examination including exact age matches is required to ascertain whether the observed cross-linguistic differences in scores for the younger group are meaningful. Cross-linguistic comparison with the First Mentions (FM) scores is more problematic due to a difference in procedures. The full ENNI FM score is obtained by scoring elements from both the A and B series of stories. In this study, we administered only the A series. As a result, our FM scores reflect a truncated version in comparison to the FM scores reported in the ENNI norms. Indeed, comparison reveals that our scores are approximately half the ENNI scores. Further comparison is not meaningful due to the differences in administration.

It was of interest as well to compare our EVIP scores to the published norms (Dunn et al., 1993) given previous indications that these norms do not adequately represent monolingual French-speaking children in Quebec (Godard & Labelle, 1995). Table 2 reports raw scores for the EVIP, as well as standard scores, which are obtained by converting raw scores using the published norms. The EVIP standard scores have a mean of 100 and a SD of 15. Thus, the average child is expected to obtain a standard score around 100. Our results for each of the 3 age groups, show standard score means well above 100 (112, 116 and 122, respectively). Thus, the difference between our results and the published norms is on the order of a standard deviation – a non negligible difference. It is unlikely that this finding results from our sample being non-representative of its population. First, we made no effort to include only children with relatively high language abilities. The mean nonverbal cognitive scores and the range of scores both indicate that the sample represents children of various ability levels, with the means for nonverbal cognition close to 100 for each age group, and thus agreeing closely with the published norms for the Leiter test of nonverbal cognition. Secondly, this result constitutes a replication of the findings of Godard and Labelle. This finding has therefore been reported in two independent studies. The difference may result from several factors. The published EVIP norms, in their effort to represent the whole population of francophone children in Canada, included children whose exposure to French varies considerably. This may partly explain why the population that possesses the most advanced French skills – monolingual francophone native speakers, performs better than the nation-wide norm.

In addition, the typical performance may have changed since the EVIP was normed. The clear implication is that the interpretation of the EVIP, when used with this population, must draw on additional local norms. The same may be true of other francophone populations within Canada.

Significant correlations were found between many of the measures, most of which were in the weak to moderate range, but some of which were fairly strong. Strong correlations were found notably between the Carrow and Sentence Imitation and the Carrow and EVIP, and moderate ones between Sentence Imitation and the EVIP and Nonword Repetition, respectively, and between the EVIP and ENNI Story Grammar. Interestingly, MLUm was not significantly correlated with any other measure. The pattern of correlations indicates that the different language measures do overlap to a certain degree, but at the same time, they also tap different skills. In particular, MLU in spontaneous language appears to tap a different skill than the other measures, not only in comparison to formal tests of receptive language (EVIP and Carrow), but also in comparison to the production of narrative structure (ENNI). The pattern of correlations suggests that in spite of some overlap, these measures do contribute unique information regarding the children's language skills.

Normative data are an important tool for clinicians in Speech-Language Pathology. They allow an individual child´s performance to be compared to that of other children of the same age and who have had similar language experiences. Language impairment generally results in significantly low scores in one or more areas of language. Therefore, norm-referenced tests play a key role in accurate identification of language impairment and in the determination of severity levels, the documentation of the child's profile of strengths and weaknesses and the detailed documentation of the current level of functioning used to select appropriate treatment targets. Not all tests are equally well suited for each of these purposes, however. Not all areas of language may be equally severely affected by a language impairment. Further, formal tests are targeted to specific areas of language, but use a more contrived elicitation method than spontaneous measures. Therefore, clinical work is greatly facilitated by the availability of different types of language measures. Normative data are important in guiding treatment decisions in that they provide information on the normal sequence of development, thereby helping to prioritize goals based on the child's needs and level of readiness for new learning. However, norms are not sufficient by themselves to accurately identify the presence of language impairment in all cases. While very severe cases are easily identified by very low scores, decisions are more complicated in moderate and borderline cases. Language tests differ in how sensitive they are to the presence of language impairment (Conti-Ramsden, Botting & Faragher, 2001; Plante & Vance, 1995; Tomblin, Records & Zhang, 1996). Therefore, the next step in this research, aiming to develop an assessment protocol is the verification of how accurately these tests are able to rule in or rule out the presence of language impairment. Such work is currently underway.

Acknowledgements

This study was funded by a research grant from the Ministère de la santé et des services sociaux of Québec (MSSS) [Ministry of Health and Social Services], awarded to Elin Thordardottir and Eva Kehayia (PIs), and to collaborators André Courcy (deceased) Nicole Lessard, Barbara Mazer, Annette Majnehmer, Ann Sutton and Natacha Trudeau. The authors acknowledge the assistance of Julia Levy, Veronica Osborne, and Marie-Hélène Vanier. The authors are grateful to the children and parents who participated in this study.

References

Archibald, L. & Joanisse, M. (2009). On the sensitivity and specificity of nonword repetition and sentence recall to language and memory impairments in children. *Journal of Speech, Language and Hearing Research*, 52, 899-914.

Bishop, D., Bishop, S., Bright, P., James, C., Delaney, T., & Tallal, P. (1999). Different origins of auditory and phonological processing problems in children with language impairment: Evidence from a twin study. *Journal of Speech, Language and Hearing Research*, 42, 155-168.

Boudreault, M.-C., Cabirol, É.-A., Trudeau, N., Poulin-Dubois, D. & Sutton, A. (2007). Les inventaires MacArthur du développement de la communication : validité et données normatives préliminaires. *Revue canadienne d'orthophonie et d'audiologie*, 31, 27-37.

Boucher, N., Lavoie, A., & Bergeron, F. (2004). Normalisation du Reynell Developmental Language Scales, volet réceptif, chex une population d'enfants québecois âgés de 3 ans zéro mois à trois ans onze mois. Unpublished manuscript. Université Laval, Faculté de médecine, Département de réadaptation, Programme d'orthophonie.

Boulianne, L. & Labelle, M. (2006). Unpublished working version of a French adaptation of the digit span subtest of the CELF-4 and the Following Directions subtest. Montreal.

Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press.

Carrow-Woolfolk, E. (1985). Test for the Auditory Comprehension of Language-Revised. Allen,TX: DLM Teaching Resources.

Catts, H. W. (1993). The relationship between speech-language impairments

and reading disabilities. Journal of Speech and Hearing Research, 36, 948-958.

Conti-Ramsden, G. (2003). Processing and linguistic markers in young children with Specific Language Impairment. *Journal of Speech, Language, and Hearing Research*, 46, 1029-1037.

Conti-Ramsden, G., Botting, N & Faragher, B. (2001). Psycholinguistic markers for specific language impairment (SLI). *Journal of Child Psychiatry*, 42, 741-748.

Conti-Ramsden, G., Crutchley, A. & Botting, N. (1997). The extent to which psychometric tests differentiate subgroups of children with SLI. *Journal of Speech, Language and Hearing Research*, 40, 765-777.

Courcy (2000). Conscience phonologique et apprentissage de la lecture. Unpublished doctoral dissertation. Montréal: Université de Montréal.

Dollaghan, C. & Campbell, T. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language and Hearing Research*, 41, 1136-1146.

Dunn, L., Thériault-Whalen, C., & Dunn, L. (1993). *Échelle de vocabulaire en images Peabody: Adaptation française du Peabody Picture Vocabulary Test*. Toronto, ON : PsyCan.

Thordardottir, E. (2008a). L'évaluation du langage des enfants bilingues. *Fréquences*, 20, 19-22.

Elin Thordardottir (2008b). Language specific effects of task demands on the manifestation of specific language impairment: A comparison of English and Icelandic. *Journal of Speech, Language and Hearing Research*, 51, 922-937.

Elin Thordardottir. (2005). Early lexical and syntactic development in Quebec French and English: Implications for cross-linguistic and bilingual assessment. *International Journal of Language and Communication Disorders*, 40, 243-278.

Elin Thordardottir. & Gagné, A. (2006). Limites et possibilités de la narration comme outil de dépistage du trouble spécifique de la lecture. Presented at the Annual Conference of the Ordre de orthophonistes er audiologistes du Québec (OOAQ), June.

Elin Thordardottir, Kehayia, E., Lessard, N., Sutton, & Trudeau (2009). Syntactic and morphological development of French-speaking children age 2 to 6. Manuscript in preparation.

Ellis Weismer, S., Tomblin, B., Zhang, X., Buckwalter, P., Chynoweth, J., & Jones, M. (2000). Nonword repetition performance in school-age children with and without language impairment. *Journal of Speech, Language and Hearing Research*, 43, 865-878.

Gagné, A. & Thordardottir, E. (2006). La petite histoire des jeunes conteurs: étude du discours narratif chez les enfants québecois francophones âgés entre 4 et 6 ans. Presented at ACFAS, Montreal, May.

Garcia, L., Paradis, J., Sénécal, I., & Laroche, C. (2006). Utilisation et satisfaction à l'égard des outils en français évaluant les troubles de la communication. *Revue d'orthophonie et d'audiologie*, 30, 239-249.

Girbau, D. & Schwartz, R. (2007). Non-word repetition in Spanish-speaking children with specific language impairment (SLI). *International Journal of Language and Communication Disorders*, 42, 59-75.

Godard, L. & Labelle, M. (1995). Utilisation de l'ÉVIP avec une population québecoise. *Fréquences*, 7, 18-21.

Griffin, T., Hemphill, L., Camp., L. & Wolf, D.P. (2004). Oral discourse in the preschool years and later literacy skills. First Language, 24, 123-147.

Groupe coopératif en orthophonie- Région Laval, Laurentides, Lanaudière (1999). Épreuve de compréhension de Carrow-Woolfolk, adaptation du TACL-R. Ordre des orthophonistes et audiologistes du Québec.

Leadholm, L. & Miller, J. (1992). *Language sample analysis : The Wisconsin guide*. Madison, WI: Wisconsin Department of Public Instruction.

MacLoed, A., Sutton, A., Trudeau, N, & Elin Thordardottir (under review). Phonological development in quebecois French: A cross-sectional study of preschool aged children.

Miller, J. & Chapman, R. (1984-2002). *Systematic Analysis of Language Transcripts: Software for analyzing English and Spanish language transcripts*. University of Wisconsin, Madison: Language Analysis Laboratory.

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

Plante, E. & Vance, R. (1995). Diagnostic accuracy of two tests of preschool language. *American Journal of Speech-Language Pathology*, 4,70-76.

Reynell, J. & Gruber, P. (1990). Reynell Developmental Language Scales. Greenville, SC: Super Duper Publications.

Roid, G. & Miller, L. (1997). *Leiter International Performance Scale-Revised*. Wood Dale, IL: Stoelting.

Rondal, J.-A., Ghiotto, M., Bredart, S., & Bachelet, J.-F. (1987). Age-relation, reliability and grammatical validity of measures of utterance length. *Journal of Child Language*, 14, 433-446.

Royle, P. & Thordardottir, E. (2003). *Le grand déménagement*. French adaptation of the Recalling Sentences in Context subtest of teh CELF-P. Unpublished research tool, McGill University.

Sahlén, B., Reuterskjöld-Wagner, C., Nettelbladt, U., & Radeborg, K. (1999). Nonword repetition in children with language impairment: Pitfalls and possibilities. *International Journal of Language and Communication Disorders*, 34, 337-352.

Scarborough, H., Wyckoff, J., & Davidson, R. (1986). A reconsideration of the relation between age and mean utterance length. *Journal of Speech and Hearing Research*, 29, 394-399.

Schneider, P., Dubé, R. & Hayward, D. (2002-2006). The Edmonton Narrative Norms Instrument. University of Alberta Facultyof Rehabilitative Medicine website. http://www.rehabmed.ualberta.ca/spa/enni

Semel, E., Wiig, E., & Secord, W. (2006). Clinical Evaluation of Language Fundamentals-Fourth Edition. San Antonio, TX: The Psychological Corporation.

Semel, E., Wiig, E., & Secord, W. (2004). Clinical Evaluation of Language Fundamentals-Preschool Second Edition. San Antonio, TX: The Psychological Corporation.

Sutton, A., Trudeau, N., Lessard, N., Jutras, B., Elin Thordardottir, & Lefebvre, P. (2009). Language development of French-speaking preschoolers. Manuscript in preparation.

Tomblin, J.B., Records, N., & Zhang, X. (1996). A system for the diagnosis of Specific Language Impairment in Kindergarten children. *Journal of Speech and Hearing Research*, 39, 1284-1294.

Trudeau, N., Frank, I, & Poulin-Dubois, D. (1999). Une adaptation en français québecois du MacArthur Communicative Development Inventory. *Journal of Speech-Language Pathology and Audiology*, 23, 31-73.

Author Note

Correspondance concerning this article should be addressed to Elin Thordardottir, School of Communication Sciences and Disorders, McGill University, 1266 Pine Avenue West, Montreal, Quebec, H3G 1A8. Email: elin.thordardottir@mcgill.ca.

Received: July 8, 2009
Accepted: November 4, 2009

Appendix

Scores corresponding to cut-off points of -1 standard deviation, -1.5 standard deviation, and -2 standard deviations for each of the measures, for each of the three age groups. In a normal distribution, the -1 SD point corresponds to the 16th %ile. The -1.5 SD point corresponds roughly

Age 4 ½ years (range 49 to 56 months inclusive)

	- 1SD	-1.5 SD	-2SD
EVIP raw score	37.30	29.17	21.05
EVIP ss	93.42	84.45	75.47
Carrow total	57.77	49.28	40.78
1. Vocabulary	42.14	21.12	18.09
2. Morphology	19.01	16.27	13.54
3. Syntax	11.11	6.85	2.58
MLUw	2.89	2.35	1.81
MLUm	3.86	3.22	2.57
ENNI First Mention	9.45	7.62	5.79
ENNI Story Grammar	8.72	6.24	3.77
Nonword Repetition	81.95	78.20	74.45
Sentence Imitation	44.89	30.27	15.64
RAN Errors	0	0	0
RAN Time	156.41	178.44	200.36
Following Directions	15.96	11.67	7.38
Forward Digit Span	3.37	2.56	1.74

Age 5 years (range 57 to 63 months inclusive)			
	- 1SD	-1.5 SD	-2SD
EVIP raw score	54.32	46.62	38.93
EVIP ss	98.47	89.61	80.75
Carrow total	70.87	64.88	58.88
1. Vocabulary	29.53	28.10	26.66
2. Morphology	20.59	17.52	14.46
3. Syntax	17.41	14.26	11.11
MLUw	3.66	3.13	2.60
MLUm	4.46	3.74	3.03
ENNI First Mention	11.75	10.13	8.50
ENNI Story Grammar	11.57	8.90	6.23
Nonword Repetition	77.36	72.59	67.82
Sentence Imitation	59.67	51.37	43.07
RAN Errors	0	0	0
RAN Time	109.76	121.97	134.17
Following Directions	18.11	13.53	8.94
Forward Digit Span	3.87	2.92	2.06

Age 5 ½ years (range 64 to 71 months inclusive)			
	- 1SD	-1.5 SD	-2SD
EVIP raw score	64.52	57.70	50.88
EVIP ss	108.86	102.11	95.36
Carrow total	78.20	72.21	66.21
1. Vocabulary	31.21	29.85	28.48
2. Morphology	19.87	16.17	12.46
3. Syntax	22.69	20.35	18.00
MLUw	4.23	3.40	2.57
MLUm	5.27	4.11	2.95
ENNI First Mention	14.30	12.81	11.31
ENNI Story Grammar	16.55	13.86	11.16
Nonword Repetition	87.69	85.64	83.59
Sentence Imitation	78.71	73.71	68.71
RAN Errors	0	0	0
RAN Time	101.33	112.2	123.06
Following Directions	18.99	12.34	5.69
Forward Digit Span	4.36	3.21	2.05